

Causes and consequences of genetic diversity in pinnipeds

Demographic, chemical, microbial and methodological studies

Submitted in fulfillment of the requirements for the degree of

Doctor rerum naturalium / Doctor of Philosophy

at

Universität Bielefeld

Faculty of Biology, Department of Animal Behaviour

and

Liverpool John Moores University

Faculty of Science, Schools of Natural Sciences and Psychology

Martin A. Stoffel

Supervised by:

Prof. Joseph I Hoffman

Dr. Hazel J Nichols

Prof. Richard Brown

October 2018

Bielefeld, Germany

Für meine Familie

Contents

Declaration of originality (Erklärung der Urheberschaft)	vii
Statement of contribution (Beitragsnachweis)	viii
Dissertation abstract (Zusammenfassung)	xvi
Chapter 1 Introduction	1
Chapter 2 Comparative demography	21
Chapter 3 Elephant seal demography	49
Chapter 4 Chemical fingerprints	67
Chapter 5 Gut microbiota	91
Chapter 6 GCalignR: R package	129
Chapter 7 inbreedR: R package	153
Chapter 8 rptR: R package	175
Chapter 9 General discussion	187
References	197
Acknowledgements	217

Declaration of originality (Erklärung der Urheberschaft)

By presenting this thesis in fulfillment of the requirements for a doctoral degree (Dr.rer.nat and PhD), as agreed upon in a Cotutelle arrangement between Bielefeld University and Liverpool John Moores University, I affirm that this work is original and has not, in whole or part, been submitted for a higher degree to any other university or institution. I affirm that I have written this thesis by myself and that I have marked all citations and references. I affirm that I understand the doctoral regulations of the Faculty and that no third parties have received either direct or indirect monetary benefits for work connected to this dissertation.

Hiermit versichere ich, dass die vorliegende Arbeit zur Erlangung des Dokortitels (Dr.rer.nat. and PhD), wie im Cotutelle Vertrag zwischen der Universität Bielefeld und der Liverpool John Moores University festgelegt wurde, ein Originalwerk ist und dass diese Arbeit nicht an einer anderen Universität oder einem anderen Institut zur Erlangung eines höheren Abschlusses eingereicht wurde. Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig verfasst habe, und dass ich alle benutzten Hilfsmittel und Quellen kenntlich gemacht habe. Ich erkläre, dass mir die Promotionsordnung der Fakultät bekannt ist und dass Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Vermittlungstätigkeiten oder für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.



Martin Stoffel, Bielefeld, Germany, September 2018

Statement of contribution (Beitragsnachweis)

This dissertation is the cumulative work of many productive collaborations with numerous scientists worldwide. My personal contribution to each of the following chapters is detailed below. All co-authors have seen this statement and have given their consensual agreement that my contributions are correct and that I may use these manuscripts in this dissertation.

Chapter 2

Martin A Stoffel, Emily Humble, Anneke J. Paijmans, Karina Acevedo Whitehouse, Barbara L. Chilvers, Bobette Dickerson, Filippo Galimberti, Neil Gemmell, Simon Goldsworthy, Hazel J. Nichols, Oliver Krüger, Sandra Negro, Amy Osborne, Teresa Pastor, Bruce Robertson, Simona Sanvito, Jennifer Schultz, Aaron B.A. Shafer, Jochen B.W. Wolf, Joseph I. Hoffman (2018) Demographic histories and genetic diversity across pinnipeds are shaped by human exploitation, ecology and life-history. Accepted in *Nature Communications*.

Conceived the study: JIH and MAS. Generated data: KA, BLC, BD, FG, NG, SDG, HJN, OK, SN, AO, AJP, TP, BCR, SS, JS, ABAS, JBWW, JIH. Analysed data: MAS, EH and JIH. Wrote the paper: MAS and JIH. All of the authors commented upon and approved the final manuscript.

Chapter 3

Martin A Stoffel, Kanchon Dasmahapatra, Emily Humble, Stephen Gaughran, Frances Gulland, David L.J. Vendrami, Hazel J. Nichols, Joseph I. Hoffman. Genomics reveals the demographic history of the Northern elephant seal *in prep.*

MAS and JIH conceived the study, FG provided samples, KD, MAS and SG conducted the labwork, MAS, EH, DV, KD, and SG did the bioinformatics, MAS analysed the data, MAS wrote the paper with input from all co-authors.

Chapter 4

Martin A Stoffel, Barbara A. Caspers, Jaume Forcada, Athina Giannakara, Markus Baier, Luke Eberhart-Phillips, Caroline Müller, Joseph I. Hoffman. Chemical fingerprints encode mother–offspring similarity, colony membership, relatedness, and genetic quality in fur seals. *Proceedings of the National Academy of Sciences*. **112**: E5005–E5012.

Conceived the study: JIH, BAC, MAS, CM. Collected data: JE. Conducted labwork: MAS, AG, MCB. Analysed data and created visualizations: MAS, BAC, LJEP. Wrote the manuscript: MAS, JIH, BAC, CM. All authors gave feedback on the manuscript.

Chapter 5

Martin A Stoffel, Karina Acevedo-Whitehouse*, Nami Morales-Duràn, Stefanie Grosser, Nayden Chakarov, Oliver Krüger, Hazel, J. Nichols, Fernando R. Elorriaga-Verplancken, Joseph I. Hoffman*. Early sex-specific differentiation, development and genetic basis of the gut microbiota in a highly polygynous marine mammal, the Northern elephant seal. *In prep*. *Shared senior-authors

Conceived the study: MAS, KA. Conducted fieldwork: MAS, KA, NMD, FE. Conducted labwork: MAS, NC, Analysed data: MAS. Wrote the paper: MAS, KA. All co-authors gave feedback to the manuscript.

Chapter 6

Meinolf Ottensmann*, **Martin A Stoffel***, Hazel J. Nichols, Joseph I. Hoffman. GCalignR: An R package for aligning gas-chromatography data for ecological and evolutionary studies (2018). *PLoS one*. **13**: e0198311. *contributed equally

MO and MAS conceived and wrote the package. MO, JIH, MAS and HJN wrote the paper.

Chapter 7

Martin A Stoffel, Mareike Esser, Marty Kardos, Emily Humble, Hazel Nichols, Patrice David & Joseph I Hoffman (2016) *inbreedR*: an R package for the analysis of inbreeding

based on genetic markers. *Methods in Ecology & Evolution*. 7:1331–1339.

MAS conceived and wrote the package with help from ME, MK and EH. MAS, JIH and HN wrote the paper with comments from all co-authors.

Chapter 8

Martin A Stoffel, Shinichi Nakagawa, Holger Schielzeth (2017) rptR: repeatability estimation and variance decomposition by generalized linear mixed -effects models. *Methods in Ecology & Evolution*. 8:1639–1644.

SN and HS conceived the idea for the package and drafted first version. MAS has implemented the new features with help of HS. All authors have contributed to testing the package, drafting and revising the manuscript.

By signing this declaration, co-authors give their consent to allow each of the above published, submitted, or drafted manuscripts to be included in the doctoral dissertation of Martin Stoffel.

Name and Affiliation

Dr. Emily Humble

Royal School of Veterinary Studies
University of Edinburgh, UK



Anneke J. Pajmans

Department of Animal Behaviour
University of Bielefeld, Germany



Prof. Dr. Karina Acevedo-Whitehouse

School of Natural Sciences
Autonomous University of Queretaro, Mexico



Dr. Barbara L. Chilvers

Institute of Veterinary, Animal and Biomedical Science
Massey University, New Zealand



Dr. Bobette Dickerson

Former: NOAA
Washington, USA

Left academia /
No e-mail address
available

Dr. Filippo Galimberti

Elephant Seal Research Group
Falkland Islands



Prof. Dr. Neil Gemmell

Department of Anatomy
University of Otago, New Zealand



Prof. Dr. Simon Goldsworthy

South Australian Research and Development Institute
Adelaide, Australia



Dr. Hazel Nichols

Department of Biosciences
University of Swansea, UK
Department of Animal Behaviour
University of Bielefeld, Germany



Name and Affiliation

Prof. Dr. Oliver Krüger
Department of Animal Behaviour
University of Bielefeld, Germany

Dr. Sandra Negro
INRA
Université Paris-Sud, France

Dr. Amy Osborne
School of Biological Sciences
University of Canterbury, New Zealand

Dr. Teresa Pastor
EUROPARC Federation
Barcelona, Spain

Dr. Bruce Robertson
Department of Zoology
University of Otago, New Zealand

Dr. Simona Sanvito
Elephant Seal Research Group
Falkland Islands

Dr. Jennifer Schultz
National Marine Fisheries Service
National Oceanic and Atmospheric Administration, USA

Digitally signed by SCHULTZ.JENNIFER.K.1384865174
Date: 2018.08.24 10:06:13 -04'00'

Dr. Aaron B.A. Shafer
Forensic Science and Environmental Life Sciences
Trent University, Canada

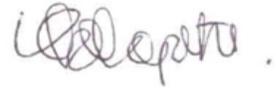
Prof. Dr. Jochen Wolf
Division of Evolutionary Biology
Ludwig-Maximilians-Universität Munich, Germany

Prof. Dr. Jochen Wolf

Prof. Dr. Joseph Hoffman
Department of Animal Behaviour
University of Bielefeld, Germany

Name and Affiliation

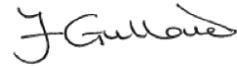
Dr. Kanchon Dasmahapatra
Department of Biology
University of York, UK



Stephen Gaughran
Department of Ecology and Evolutionary Biology
Yale University, USA



Dr. Frances Gulland
The Marine Mammal Center
Sausalitos, USA



David Vendrami
Department of Animal Behaviour
University of Bielefeld, Germany



Prof. Dr. Caroline Müller
Department of Chemical Ecology
University of Bielefeld, Germany



Dr. Barbara A. Caspers
Department of Animal Behaviour
University of Bielefeld, Germany

Signature Dr. Barbara Caspers



Dr. Jaume Forcada
British Antarctic Survey
Cambridge, UK



Athina Giannakara
Department of Animal Behaviour
University of Bielefeld, Germany



Dr. Markus Baier
Former: Department of Chemical Ecology
University of Bielefeld, Germany

Left academia /
No e-mail address
available

Dr. Luke J. Eberhart-Phillips
Max Planck Institute for Ornithology
Starnberg, Germany



Name and Affiliation

Nami Morales-Durán
School of Natural Sciences
Autonomous University of Queretaro, Mexico

Signature 
Lic. Nami del Rosario Morales-Durán

Dr. Stefanie Grosser
Division of Evolutionary Biology
Ludwig-Maximilians-Universität Munich, Germany



Dr. Nayden Chakarov
Department of Animal Behaviour
University of Bielefeld, Germany



Dr. Fernando R. Elorriaga-Verplancken
Departamento de Pesquerías y Biología Marina
CICIMAR-IPN, Mexico



Meinolf Ottensmann
Department of Animal Behaviour
University of Bielefeld, Germany



Prof. Patrice David
Centre d'Ecologie Fonctionnelle et Evolutive
Montpellier, France



Dr. Mareike Esser
Faculty of Technology, Bielefeld University
Bielefeld, Germany



Dr. Marty Kardos
Flathead Lake Biological Station
University of Montana, USA



Prof. Dr. Shinichi Nakagawa
School of Biological, Earth and Environmental Sciences
University of New South Wales, Australia



Prof. Dr. Holger Schielzeth
Population Ecology Group
University of Jena, Germany



Dissertation abstract

The rise of empirical population genetics and more recently the genomic revolution have given us the tools to explore completely new questions in ecology and evolution. Based on only a single temporal sample of individuals, genetics now allows us to glimpse into the demographic history of a species and gain insights into its population dynamics during the last glacial period or the impact of recent bottlenecks caused by human exploitation. Measuring genetic variation is also key to exploring the forces shaping phenotypic variation, which are well understood for some traits but are completely unknown for many others. Among the latter are chemical and bacterial phenotypes, which are so complex in themselves that it is not clear how they are determined by an animals' genotype. Nevertheless, a better knowledge of the interconnectivity between animal genetics, chemicals and microbiota has the potential to drastically change our understanding of ecological and evolutionary processes. However, very few studies have bridged the gap between population genetics, chemical ecology and microbiology in wild populations.

Pinnipeds are an extraordinary group of marine mammals for exploring such wide-ranging questions. From the tropical waters of the Central Pacific to the wild seas of the Southern Ocean, pinnipeds inhabit nearly every marine environment in the world. They show a remarkable variety of life-history adaptations where chemical communication and host-microbe interactions play a potentially critical role, such as the flawless mother-pup recognition in fur seals or the sex-specific feeding strategies in Northern elephant seals. Pinnipeds also differ greatly in their recent demographic histories, as large-scale commercial exploitation by 18th and 19th century sealers brought many species to the edge of extinction, while others remained largely untouched. In my dissertation, I elucidate the origins of genetic variation among pinnipeds using demographic inference, but I also explore some of the more unknown consequences of genetic variation: chemical and microbial phenotypes, and their potential functions. Lastly, inspired by the open science movement, I developed three scientific packages in R which emerged from the analyses in this dissertation. My thesis is divided into the following chapters:

In chapter 1, I give an overview of the historical context and main questions of this dissertation and describe the pinnipeds as a study system. In chapter 2, I present a comparative genetic analysis of the demographic consequences of commercial exploitation. We found that around one-third of all pinnipeds underwent severe genetic bottlenecks and that these were mediated by both ecology and life-history. Moreover, genetic diversity seems to be largely determined by contemporary population size and reduced only by very severe bottlenecks. Chapter 3 presents the first genomic investigation of the demographic history of the Northern elephant seal. Using a novel genome sequence and restriction-site associated DNA sequencing, we infer both an extreme recent bottleneck and a likely post-glacial expansion of the species. In chapter 4, we characterise the chemical basis of olfactory mother-offspring recognition in Antarctic fur seals. The skin chemical profiles were surprisingly diverse, with chemical compounds differing between populations and mother-offspring pairs and correlating with heterozygosity and genetic relatedness. Chapter 5 explores the development of gut microbiota in young Northern elephant seals. Using a diet-controlled setting, we show that gut microbiota of young seals are already highly complex and change radically within only a few weeks. Furthermore, we show that gut microbiota are highly sex-specific and linked to genotype in males but not in females. In chapter 6, we describe GCalignR, an R package for aligning gas chromatography data across many individuals for field studies in animal ecology and evolution. Chapter 7 presents inbreedR, an R package for analysing inbreeding and inbreeding depression using genetic and genomic markers. The rptR package described in chapter 8 quantifies intra-class coefficients or repeatabilities for Gaussian and non-Gaussian traits in a mixed model framework. Lastly, chapter 9 puts the findings of this thesis in a broader context, discusses its limitation with respect to the neutral theory of molecular evolution and the complexities of assigning function to chemical compounds and microbes in wild organisms and outlines future directions.

To summarise, my dissertation provides novel insights into how recent and historical demography shaped the genetic makeup of contemporary populations, thereby contributing to the old riddle of the determinants of genetic diversity. Moreover, I show that individual genetic variability shapes complex chemical signatures and gut microbial communities in the wild and explore the potential mechanisms by which these could be intertwined with pinniped ecology and evolution. Lastly, I hope that both the openly available and documented analytical pipelines as well as the R packages that have emerged from this dissertation will help to facilitate scientific progress and replicability.

Zusammenfassung der Dissertation

Die Errungenschaften der empirischen Populationsgenetik haben Ökologen und Evolutionsbiologen eine komplett neue Perspektive auf die Natur ermöglicht. Mithilfe der Genetik kann man heutzutage zum Beispiel anhand von aktuellen Blutproben einer Tierart Einsichten in deren evolutionäre Vergangenheit gewinnen, von der Populationsgröße während der letzten Eiszeit bis hin zu den Folgen der kommerziellen Ausbeutung durch den Menschen. Die Populationsgenetik ist aber ebenso wichtig um zu verstehen, warum sich Tiere und Menschen überhaupt in ihrem Erscheinungsbild oder Phänotyp unterscheiden. Obwohl für manche Merkmale, wie beispielsweise Haarfarbe, die zugrunde liegenden Gene relativ gut bekannt sind, wissen wir nahezu nichts über die Gene oder die Erbllichkeit von vielen komplexeren Merkmalen. Zwei dieser Merkmale, welche ich in dieser Doktorarbeit untersuche, könnten unser Verständnis der Biologie grundlegend erweitern. Zum einen geht es um die chemischen Moleküle und Substanzen, die sich auf der Haut eines jeden Lebewesens befinden und welche von grundlegender Wichtigkeit für die ältesten aller Sinne sind: Die chemischen Sinne, Riechen und Schmecken. Zum anderen geht es um die Millionen und Abermillionen von Mikroben, die, wie wir mittlerweile wissen, jedes Lebewesen bevölkern und von denen manche für ihren Wirt lebensnotwendig sind, und andere lebensgefährdend. Die komplexen Zusammenhänge zwischen Genen und chemischen Substanzen sowie Mikroben wurden bisher kaum bei Tieren in freier Wildbahn untersucht, obwohl sie unser Verständnis vieler biologischer Phänomene erweitern könnten.

Die Robben (Pinnipedia) sind Meeressäuger mit einer enormen Vielfalt an ökologischen und physiologischen Besonderheiten, was sie zu interessanten Studienorganismen macht, um weitreichende populationsgenetische Fragen zu erforschen. Sie sind in nahezu allen aquatischen Lebensräumen weltweit zu finden, von den tropischen Wassern Hawaiis bis hin zu den eisigen Strömungen des Südpolarmeeres, und haben sich im Laufe der Evolution dementsprechend an ihre jeweiligen Lebensräume angepasst. So brauchen zum Beispiel Antarktische Seebärenmütter einen exzellent ausgeprägten Orientierungs- Hör- und Geruchssinn, um ihre Jungen nach der Futtersuche im Meer zwischen tausenden von anderen Tieren in der Kolonie

wiederzufinden. Auf der anderen Seite des Äquators haben Männchen und Weibchen der Nördliche Seeelefanten erstaunlich unterschiedliche Lebenslaufstrategien entwickelt, wodurch sie sich zum Beispiel in ihren Jagdrevieren und ihrer Beute stark unterscheiden, und was potenziell eine Kovolution von geschlechterspezifischen Darmbakterien zur Folge haben könnte. Viele Robbenarten wurden industriell bejagt, da im 18. und 19. Jahrhundert sowohl ihr Fell als auch ihre dicke Fettschicht (für die Produktion von Ölen zur Beleuchtung oder für Kosmetik) stark nachgefragt waren. Manche Arten wurden komplett ausgerottet, andere an den Rand des Aussterbens gebracht, und wieder andere hatten das Glück, nie in die Fänge der Robbenjäger zu kommen. Im ersten Teil dieser Arbeit beschäftige ich mich mit den genetischen Konsequenzen dieser unterschiedlichen demografischen Vergangenheiten, zum Beispiel mit der Frage ob die Bejagung die genetische Diversität der verschiedenen Robbenarten reduziert hat. Im zweiten Teil untersuche ich dann die genetischen Grundlagen sowohl der chemischen Kommunikation bei Antarktischen Seebären wie auch der Interaktion zwischen Wirt und Darmbakterien in Nördlichen Seeelefanten. Inspiriert von der Open-Science Bewegung habe ich außerdem auf Basis der Analysen in diesen Studien drei wissenschaftliche Softwarepakete in der Programmiersprache R entwickelt. Die vorliegende Arbeit teilt sich in folgende Kapitel:

In Kapitel 1 gebe ich einen Überblick über die historische Entwicklung der Populationsgenetik als Wissenschaft, beschreibe die Hauptfragestellungen und stelle die Robben als Hauptstudienorganismen dieser Arbeit vor. Kapitel 2 enthält eine vergleichende, genetische Analyse der demografisch-genetischen Konsequenzen der Robbenjagd, in der wir zeigen, dass ein Drittel aller Robbenarten in Folge der Bejagung an den Rand des Aussterbens gebracht wurde, was für manche Spezies eine starke genetische Verarmung zur Folge hatte. In der Studie in Kapitel 3 benutzen wir neueste genomische Technologien um die demografische Vergangenheit einer ganz besonderen Art zu untersuchen, des Nördlichen Seeelefanten. Der Seeelefant ist mittlerweile in so gut wie jedem Biologie Lehrbuch als Musterbeispiel für einen sogenannten 'Genetischen Flaschenhals' zu finden, einer starken genetische Verarmung aufgrund des drastischen Rückgangs der Population. Die Art wurde vor gut einhundert Jahren nahezu komplett ausgerottet, hat sich aber innerhalb kürzester Zeit wieder erholt, wobei mittlerweile wieder mehr als 200,000 Exemplare an der Pazifikküste Nord- und Mittelamerikas leben. In dieser Studie untersuchen wir aber nicht nur die genetischen Folgen der Bejagung, sondern fanden auch erste Belege, dass die Art während der letzten Eiszeit nur eine relativ kleine Population hatte, welche sich nach dem Rückgang der Gletscher aber stark ausbreitete. In Kapitel 4 analysieren wir die verschiedenen chemischen Substanzen auf der Haut von Antarktischen Seebären, welche an der Geruchskommunikation beteiligt sein könnten. Erstaunlicherweise fanden wir Belege,

dass diese Substanzen überaus komplexe Informationen enthalten können, sowohl über die jeweilige Population, die familiäre Verwandtschaft als auch über die genetische Diversität der Tiere. Kapitel 5 beschreibt eine Studie zur Entwicklung der verschiedenen Darmmikroben von jungen Seeelefanten auf dem San Benitos Archipel im nordöstlichen Pazifik. Säugetiere werden nahezu ohne Darmmikroben geboren, aber innerhalb kürzester Zeit nach der Geburt bevölkern viele verschiedene Spezies den Darm. Die Ergebnisse dieser Studie zeigen, dass völlig verschiedene Bakterienarten die Därme von Männchen und Weibchen bevölkern, und dass die Gene des Wirtes zum Teil die Bakterienkolonien beeinflussen, welche sich während der Entwicklung bilden. In den letzten drei Kapiteln stelle ich verschiedene Softwarepakete vor, geschrieben in der Programmiersprache R, welche neue Analysemethoden für zukünftige Forschungsprojekte implementieren. Kapitel 6 beschreibt GCalinR, eine Software, die Wissenschaftlern ermöglicht, komplexe chemische Daten von vielen verschiedenen Tieren oder Proben zu vergleichen. In Kapitel 7 stelle ich inbreedR vor, ein Paket, das die Analyse von Inzuchtverhalten mithilfe genetischer Methoden vereinfacht. Kapitel 8 stellt schließlich rptR vor, ein Paket, das sogenannte Intra-Klassen-Koeffizienten berechnet, welche gebraucht werden, um die Stabilität von wissenschaftlichen Messungen zu berechnen aber auch um biologische Phänomene zu untersuchen, wie zum Beispiel in der Persönlichkeitsforschung. In Kapitel 9 diskutiere ich die Studien meiner Dissertation in ihrem größeren Zusammenhang und erörtere potentielle zukünftige Forschungsansätze. Alles in allem hat meine Doktorarbeit zu neuen Einsichten über den Zusammenhang von genetischer Diversität und Demografie geführt, und zeigt insbesondere wie menschliche Einflüsse die natürliche genetische Diversität der Robben bereits verringert haben. Eine weitere wichtige Erkenntnis ist jedoch, dass nur Populationen, welche auf sehr wenige Individuen reduziert wurden, tatsächlich eine starke genetische Verarmung erfahren. Das bedeutet im Rückschluss auch, dass der Schutz von nur wenigen hundert Individuen einer Art ausreichen kann um deren erfolgreiche Erholung zu garantieren. Ich habe außerdem in dieser Arbeit gezeigt, dass auch komplexe Phänotypen, wie chemische Substanzen und Darmbakterien, von der Genetik des Tieres bzw. Wirtes abhängen. Diese Ergebnisse könnten einige Fragen der Ökologie und Evolutionsbiologie in einem neuen Lichte erscheinen lassen, zum Beispiel, was denn eine biologische Einheit ausmacht, welche sich unter dem Druck der natürlichen Selektion befindet: Sind alle Lebewesen nur ihre Gene und Proteine, oder vielleicht doch ein deutlich komplexeres Geflecht aus chemischen Substanzen und Bakterienkolonien? Schlussendlich hoffe ich, dass auch die in meiner Dissertation entwickelten Methoden und Software ein wenig helfen werden, sowohl den wissenschaftlichen Fortschritt als auch die Replizierbarkeit von wissenschaftlichen Studien, ganz im Sinne der aufkommenden Open-Science Bewegung, voranzubringen.



Islas San Benito.

1

INTRODUCTION

The stuff of life and how it varies ¹

It is the year 1869, around ten years after Charles Darwin published his groundbreaking work ‘On the Origin of Species’ and only three years after Gregor Mendel discovered the principles of inheritance. A young and highly ambitious Swiss scientist named Friedrich Miescher works in a cold, stony laboratory in a castle high above Tübingen, with no less of a goal than to unravel the cellular foundations of life. He conducts a series of chemical experiments to split leucocytes from the pus in surgical bandages into proteins, which he thinks are key to understanding how cells work. Yet, during these tests, a curious side effect catches Miescher’s attention. A white substance consistently precipitates from the solution when adding acid and dissolves again when adding alkali. Who could have guessed at that time, that this seemingly inconspicuous process was the first isolation of deoxyribonucleic acid, or DNA, which would later shape our understanding of life like few other discoveries. Miescher called the substance nuclein, as it came from the nucleus of cells (Miescher, 1871). However, his finding received little attention and certainly was not connected to matters of heredity until half a century after his death (Dahm, 2005). While it took humanity 200,000 years to isolate the centerpiece of biological information, and a further 80 years to recognize that DNA actually carries genetic information (Avery et al., 1944; Hershey and Chase, 1952), and to figure out that it’s a double helix (Franklin and Gosling, 1953; Watson and Crick, 1953), the following decades saw an unparalleled scientific progress in the field of genetics. These efforts culminated in the largest collaborative biological project ever conducted, which started out in 1990 to sequence the first human genome. After eleven years of work from thousands of people, the complete human genome sequence was eventually published (Venter et al., 2001) at an overall cost of around 2.7 billion dollars.

Of course, there is no such thing as the human genome. Every human genome is different – although to only around 0.1% (Altshuler et al., 2012) – and this variation among genes is the key to understanding traits and evolution. Consequently, the rise of genetics was not just confined to single genomes or humans. In the 1960s, the first two studies appeared which described genetic variation at multiple loci in both humans and natural populations of the fruit fly *Drosophila melanogaster* (Harris, 1966; Lewontin and Hubby, 1966); the birth of empirical population genetics. Starting with studies using allozymes that simply quantified genetic variation, rapidly advancing sequencing technologies soon made it possible to sequence mitochondrial DNA (mtDNA) by the end of the 1970s, followed by microsatellites in the 1990s and single nucleotide polymorphisms (SNPs) in the 2000s (Allendorf, 2017). With these new

¹Heading inspired by Bill Bryson’s chapter on DNA in ‘A short history of nearly everything’

molecular markers, completely novel questions opened up in the fields of ecological and evolutionary genetics. In the 1960s it was merely possible to quantify genetic variation. Then, the analysis of mtDNA enabled the characterization of sex-specific gene flow and phylogeography whilst microsatellites made it possible to measure population structure, genetic relatedness and effective population sizes. Now, with access to tens to hundreds of thousands of SNPs scattered across the genome, modern population genomics is able to detect loci under selection, estimate historical demography, precisely estimate inbreeding and much more (Allendorf, 2017; Kardos et al., 2015; Lukic and Hey, 2012; Nielsen et al., 2007).

With these possibilities, modern population genetics has become a field with close ties to many other disciplines, such as microbial biology, chemical ecology and population biology, but also statistics and bioinformatics. More than merely a mathematical framework, population genetics provides the building blocks for modern ecological and evolutionary science. The studies in this dissertation reflect the opportunities emerging from the interdisciplinary and data-driven field of population genetics but also from the fascinating ecological and evolutionary questions surrounding the diverse group of pinnipeds as a study system.

Causes of genetic diversity

A short history

The degree of genetic polymorphism varies between species but also between individuals and even within genomes. We know now that understanding the causes of genetic variation is key to understanding evolution. However, before genetic polymorphism could be measured, it was entirely unclear to which degree species vary. For the ancient Greek philosophers, all variation among material bodies was only an imperfection around the ideal form (Ellegren and Galtier, 2016), rather than a functional adaptation. The first scientist who really challenged this view was probably Charles Darwin, who recognized that variation was an essential feature of natural systems (Darwin, 1859), rather than a deviation from the ideal. The genetic component was introduced by Gregor Mendel during his famous pea plant experiments. This was a major step in evolutionary genetics and biology as a whole, as it allowed him to formulate expectations for the variation among the offspring of peas (Mendel, 1886). Later on, the theory of population genetics emerged based on the work of Ronald A. Fisher, John B.S. Haldane and Sewall Wright, which, together with Darwin's and Mendel's theories, led to the modern evolutionary synthesis. Genetic diversity was now understood to emerge due to gradual changes in gene frequencies. However, as genetics could not be directly measured yet, it was still entirely unclear how much

variation there actually is.

The first theories on this question were contrasted by Theodosius Dobzhansky as the ‘classical’ and the ‘balance’ theory (Dobzhansky, 1957). The classical theory argued that most individuals are homozygous for an ‘optimal’ wild-type gene. Then, when new mutations occur, natural selection would either quickly select against deleterious alleles or quickly favor advantageous alleles, leading to their fixation. According to this theory, genetic diversity in populations would be very low. By contrast, the balance theory suggested that most loci are heterozygous and contain many alleles. Selection then maintains these high levels of diversity through a form of balancing selection, such as over-dominance or frequency dependent selection. Consequently, the balance theory predicted much higher levels of genetic variation within and among natural populations than the classical theory.

When the first technology for quantifying genetic diversity was invented in the 1960s, the high degree of variation quickly became evident. Although initially this could have been seen as support for the balance theory, a new idea emerged and changed the field drastically. Motoo Kimura formulated the neutral theory of molecular evolution, which stated that most alleles were actually selectively neutral (Kimura, 1983). This theory was further developed into the nearly neutral theory by Tomoko Ohta (Ohta, 1973), which also allowed slightly deleterious mutations to segregate in populations at frequencies inversely proportional to their population sizes. This theory largely reflects our current view on genetic diversity (Ellegren and Galtier, 2016). We know now that substantial parts of the genome are likely to be non-functional and hence neutral, and that diversity in such regions will reflect processes other than selection. In particular, genetic diversity is assumed to arise as a consequence of the demographic history of a species, which can be understood as its effective population size (N_e) over time. However, it remains unclear how exactly demography and genetic diversity are linked (Leffler et al., 2012). This is one of the objectives tackled in Chapter 2 of this thesis.

Demographic inference

With these questions in mind, characterizing population histories from genetic and more recently genomic data has been an important theme in evolutionary biology for decades. Understanding the demography of a species is interesting in its own right, but can also be used to answer fundamental questions ranging from exploring the determinants of genetic diversity (Ellegren & Galtier, 2016) to practical efforts in conservation (Shafer et al., 2015*b*) or to disentangling demographic from selective effects (Nielsen et al., 2007). The insight that contemporary neutral diversity can be understood as the consequence of N_e in the past has

motivated researchers to combine simulations with empirical genetic data for decades and has led to the development of a multitude of methods for modeling past population size changes. The key advantage of combining demographic simulations and genetic data is that it becomes possible to infer the demographic history of a species from only a single, contemporary sample. At their core, most of these methods work with the same principle: First, the observed genetic or genomic data is summarized by one or more summary statistics. These can range from simple statistics such as allelic richness, allelic range, heterozygosity-excess (Cornuet and Luikart, 1996) and M-ratio (Garza and Williamson, 2001) to more complex statistics which summarize genome-wide data such as the site frequency spectrum (Salmona et al., 2017). Second, genetic data is simulated under one or more demographic scenarios of interest. This is possible as theoretical frameworks such as the coalescent theory (Kingman, 1982) allow to explore the genetic diversity resulting from infinitely complex demographic histories and have been implemented in a variety of software (Excoffier and Foll, 2011; Hudson, 2002). Lastly, the observed and simulated genetic data are compared to infer the best fitting demographic scenario and its most likely parameters. The variety of methods developed based on this idea are now having a tremendous impact on how we understand human demography (e.g. Gravel et al. 2011; Gronau et al. 2011; Lukic and Hey 2012) but are also starting to reveal novel insights into the history of wild animal population (Foote et al., 2016; Pedersen et al., 2018; Peery et al., 2012; Trucchi et al., 2014). Depending on whether smaller panels of genetic markers or large-scale genomic data are used, two methodological frameworks have established themselves as the prevailing ways of demographic inference: Approximate Bayesian Computation for genetic data and site frequency spectrum-based methods for genomic data.

Demographic inference from genetic data

- Approximate Bayesian Computation

Arguably, the most flexible and powerful framework for demographic inference based on smaller panels of genetic markers (e.g. microsatellites) is Approximate Bayesian Computation (ABC, Beaumont et al. 2002). ABC goes back to the rejection algorithm (Tavaré et al., 1997), and the idea is as follows: First, a large number of datasets are simulated under an evolutionary scenario, based on probability distributions (Figure 1.1A). Second, these data are then reduced to genetic summary statistics and different simulations are accepted or rejected based on the distance between the simulated and the observed summary statistics (Figure 1.1B, C). Third, the sub-sample of accepted simulations can then be used to directly infer the values of the underlying model parameters and evaluate their uncertainties (histogram in Figure 1.1D). The power

of ABC is only limited by the versatility of the underlying demographic simulation software, but tools like *ms* (Hudson, 2002) or *fastsimcoal* (Excoffier and Foll, 2011) can now be used to simulate data under very complex scenarios and are even able to incorporate recombination.

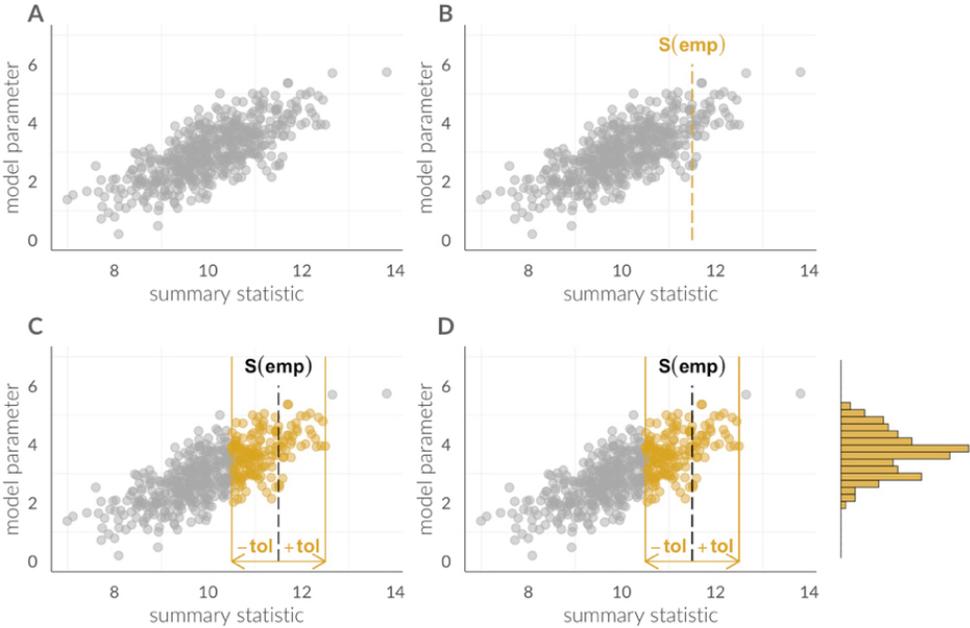


Figure 1.1: The four steps of ABC. Panel A shows simulated summary statistics generated under different values of a model parameter in a demographic model. Panel B shows the simulated summary statistics together with the value of the empirical, observed summary statistic $S(\text{emp})$. Panel C shows the rejection algorithm, which accepts all simulated summary statistics that are within a tolerance range ($-\text{tol}$, $+\text{tol}$) to the empirical value. Panel D shows how the posterior distribution of the model parameter is inferred based on the values of the accepted summary statistics. Figure adapted from Csilléry et al. 2010

Therefore, ABC makes it possible to select the best fitting among several demographic models and to estimate the underlying model parameters. Hence, it is not surprising that ABC has led to major insights into the demographic histories of species ranging from humans (Fagundes et al., 2007), chimpanzees (Wegmann and Excoffier, 2010) and *Drosophila* (Duchen et al., 2013) to pinnipeds (Hoffman et al., 2011; Shafer et al., 2015a). In Chapter 2, we use ABC for the first time in a large-scale comparative context to explore the demographic consequences of the commercial exploitation of pinnipeds.

Demographic inference from genomic data

- the site frequency spectrum

With the advent of next-generation sequencing technologies and more recently the large-scale availability of affordable sequencing, novel approaches to infer demographic histories have emerged. The main challenge with genomic data containing thousands to millions of genetic markers is to summarize these data in a way that captures the key information which reflects the demography or N_e of a species. A genomic summary statistic which is highly sensitive to demographic change and hence uniquely suited to the task of demographic inference is the site frequency spectrum (SFS), sometimes called the allele frequency spectrum (Evans et al., 2007a; Fisher, 1931). The SFS represents the distribution of alleles by classes of frequency (i.e. singletons, doubletons) and has been called the ‘most fundamental yet under-appreciated aspect of genomic data’ (Salmona et al., 2017). To illustrate how the SFS represents different demographic histories, I used coalescent-simulations to create SFS for an expanding, a stable and a bottlenecked population (Figure 1.2). The expanding population shows an excess of rare alleles (Frequency = 1 or 2), which is caused by a large proportion of new mutations due to a strong population growth. The bottlenecked population, on the other hand, shows a completely flat SFS. During the bottleneck, all rare alleles have been lost and all frequency classes are equally abundant. Finally, the SFS of the stationary population is somewhere in the middle of the two.

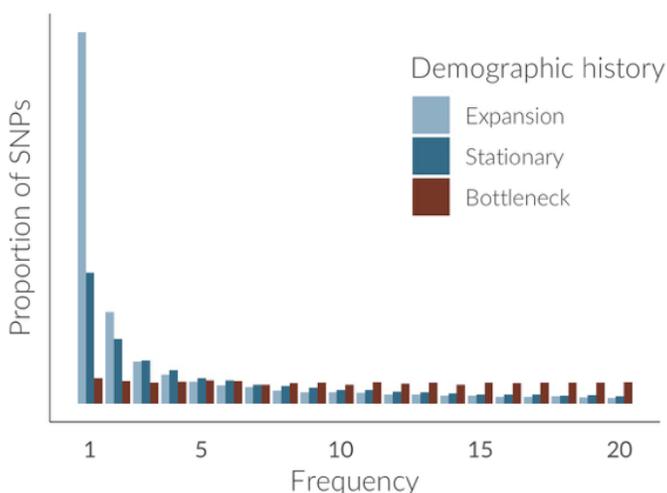


Figure 1.2: Simulated site frequency spectra for expanding, bottlenecked and stationary populations.

There are several major advantages to using the SFS for inferring demography. Its properties are well understood and can be simulated in the major population genetic frameworks, such as the coalescent (Fu, 1995; Kingman, 1982). Demographic inferences are equally computationally intensive, independent of the amount of underlying genomic data, and it has been shown that demographic models can be inferred correctly based on a small number of polymorphisms (Shafer et al., 2015a) and even from low-coverage data (Korneliussen et al., 2014).

The site frequency spectrum has become the foundation of a large body of knowledge on human historical demography (Li and Durbin, 2011; Malaspina et al., 2016), but is increasingly being used to elucidate demography, ecology and evolution of wild organisms. Recently for example, inference based on the SFS shed light on the post ice-age expansions of the the king penguin (Trucchi et al., 2014), historical European-African admixture in *Drosophila* (Duchen et al., 2013), genome-culture co-evolution in killer-whales (Foote et al., 2016) and the geographical origin of the plains zebra (Pedersen et al., 2018). However, all studies to date use genomic data to infer historical or even ancient demographic histories and it is not well understood how and whether genomic data can be used to infer contemporary demographic changes (Nunziata and Weisrock, 2018). However, drastic species declines due to anthropogenic exploitations and habitat destruction are now occurring within only years and decades (Ceballos et al., 2017; Li et al., 2016) and it is vital that we quantify these declines. The question of whether genomics can help to elucidate contemporary demography will be addressed in Chapter 3 based on the example of the iconic northern elephant seal and its drastic decline due to the 19th century sealing industry.

Consequences of genetic diversity

Bacterial and chemical phenotypes

The consequences of genetic variation among organisms can be non-existent in the case of neutral variation (Kimura, 1983), maladaptive when new mutations cause deleterious protein changes (Kryukov et al., 2007; Sunyaev et al., 2001), or adaptive when causing variation in a phenotypic trait. When bringing together a genotype and its phenotype, it is the genotype that is traditionally more complex whilst the phenotype is usually more straightforward to measure. For most traits, it is unclear whether they are caused by many loci with small effects, few loci with large effects or a mixture of both, and where in the genome the causal loci are located. This has led to the extensive use and development of genome-wide association studies (GWAS)

to elucidate the complex genetic basis of traits (Bush and Moore, 2012).

However, certain complex traits exist, for which it is still unclear whether they can even be considered a trait and as such a consequence of a genotype or not. Two particular interesting examples with the potential to greatly widen our understanding of ecology and evolution are animal-associated chemical compounds and bacteria. During the last decade, it has been increasingly recognized that animals live in both a chemical world (Wyatt, 2014a) and a bacterial world (McFall-Ngai et al., 2013), and that both domains are critical to our understanding of eco-evolutionary systems. However, the complexity of both chemical and bacterial traits has so far prevented a good understanding of them – particularly in wild organisms – due to technological and statistical problems. As in studies elucidating the genetic basis of traits, measuring chemical traits and animal-associated bacterial communities will result in hundreds, if not thousands of chemical compounds or bacterial species, of which only a proportion will be associated with a measured trait. Disentangling the chemicals or bacteria which represent the important variation in such complex phenotypes is the main challenge and requires new study designs and statistical methods. In this thesis, I elucidate the genetic basis and potential function of: (1) Skin chemical profiles (Chapter 4), which are likely an integral part of olfactory communication in mammals and (2) gut microbiota (Chapter 5), which might be critical for their host's health and life-history adaptations.

Genotype, skin chemicals and olfactory communication

The chemical senses are probably the oldest senses among animals, shared by all organisms and critical for fundamental processes such as olfactory communication. Consequently, they play a key role in important ecological processes like kin recognition and mate choice (Wyatt, 2014a). The chemical basis of olfactory communication can now be studied using metabolomics tools such as Gas Chromatography-Mass Spectrometry (GC-MS) which facilitate the quantification of chemical compounds underlying olfactory signals. Despite the importance of olfactory communication and the availability of technologies to measure its chemical basis, very little is known about the link between chemical signals and genotypes (Hurst and Beynon, 2010). Establishing such a connection is especially difficult in wild animals, as an individual's mixture of skin surface chemicals used for olfactory communication is likely a product of not just genes but also hormones, microbes or simply the environment (Hurst and Beynon, 2010). In Chapter 4, I dissect the potential factors shaping skin chemicals in Antarctic fur seals by using a study design which allowed us to partially disentangle environmental from intrinsic factors together with a novel combination of statistical methods in a system where olfactory

communication likely plays an important role in kin recognition (Dobson and Jouventin, 2003) and mate choice (Hoffman et al., 2007a).

Genotype and gut microbiota

During the last few decades, the development of large-scale sequencing technologies has revealed a bacterial world which is fundamentally changing our understanding of animal biology. It is clear now that we are not alone, but that any animal and virtually any living organism lives in a strong interdependency with hordes of associated microbes (Gilbert et al., 2012). Efforts to combine Darwinian and Mendelian principles and the modern evolutionary synthesis with novel insights on host-microbe relationships have led to the concept of a hologenome (Bordenstein and Theis, 2015). The hologenome is defined as the sum of the genes of a host and its associated microbiota, with the idea that both together can lead to variation in phenotypes. As a consequence, the hologenome can be shaped by natural selection or experience genetic drift, and can be seen as the new unit of selection (Bordenstein and Theis, 2015).

Arguably the most complex and functional of bacterial communities inhabit the gastrointestinal tract and are collectively called the ‘gut microbiota’. Gut microbiota are key to their hosts’ development, dietary function and immune system (Cheesman et al., 2011; Diaz Heijtz et al., 2011; Lathrop et al., 2011; Zhu et al., 2011) and when disturbed, can lead to severe health problems for the host (Candon et al., 2015; Cho et al., 2012; Macpherson and Harris, 2004; Russell et al., 2012). The gut microbiota are thus a highly interesting candidate to elucidate host-microbe interactions and explore the implications of the hologenome. However, a fundamental link in the chain of argumentation is still largely unexplored; most bacteria are not inherited directly, and the mammalian gut is thought to be largely sterile before birth (Perez-Muñoz et al., 2017). Most of the gut microbiota will hence only colonise the gut after an individual is born. Therefore, for a hologenome theory to make sense, there must be an association between the host genotype and its microbiota. Put another way, certain bacteria should preferably colonise individuals with a certain genotype.

Most insights on how host genetics shape microbiota to date come from studies in humans and captive mice (Kurilshikov et al., 2017). However, we cannot learn a great deal about the ecological and evolutionary function of microbiota from these studies, as human microbiota are drastically altered by different life-styles (David et al., 2014a) and captive animals have very different gut microbial communities compared to their wild relatives (Hird, 2017). Consequently, we need studies in wild animals which provide insights into the early bacterial colonization of the gut and explore associations with host genetics, but these are rare. One such study is

presented in Chapter 5, where I investigate the relationship between Northern elephant seals and their gut microbiota in a diet-controlled and longitudinal study during a critical phase in their early development. Within this study design, I explore the intrinsic factors shaping gut microbiota, such as development, genotype and sex.

Inbreeding depression

The most severe and immediate consequences of genetic variation result from its absence, and this phenomenon has been extensively studied in the framework of inbreeding depression. The detrimental consequences of inbreeding have fascinated scientists since Darwin showed that self-fertilization greatly reduces fitness in plants (Darwin, 1876). One and a half centuries later, we know that inbreeding depression is a ubiquitous feature in all diploid organisms. Furthermore, evidence for its critical impacts on a broad range of traits, individual fitness and population viability has been documented from throughout the animal and plant kingdoms (Charlesworth, 2009; Crnokrak and Roff, 1999; Joshi et al., 2015; Keller, 2002). In the course of the 21st century and beyond, inbreeding depression will become an ever more severe issue in wild animal populations, which are declining at an unprecedented pace (Ceballos et al., 2017), leading to higher and higher inbreeding rates in smaller and smaller populations. On an individual level, the decline in fitness is a function of the proportion of the genome which is identical by descent (IBD_G) or homozygous and hence reveals the effects of deleterious recessive alleles, which are considered to be the main driver of inbreeding depression (Charlesworth and Charlesworth, 1999). In principle, IBD_G can be estimated from a pedigree (Pemberton, 2004), but pedigrees are not available for most non-model organisms. Moreover, we know now that estimating an individual's IBD_G from a pedigree will be imprecise as recombination and Mendelian segregation during meiosis cause stretches of the genome to be inherited together which causes the realized IBD_G of an individual to differ substantially from the pedigree expectation (Hill and Weir, 2011).

A solution to this problem is to measure IBD_G directly using genetic markers. While smaller marker panels have been used for decades to estimate inbreeding coefficients (Coltman and Slate, 2003), next-generation sequencing now allows us to measure IBD_G using many thousands of markers resulting in estimates even more precise than pedigrees (Kardos et al., 2015). However, a major question arises when measuring inbreeding and inbreeding depression: Are the genetic markers actually reflecting inbreeding? This is critical, as variation in marker homozygosity could also reflect variation due to stochastic processes or adaptations. It is therefore important to quantify whether there is variation in inbreeding in the population and

how well homozygosity at a given panel of genetic markers reflects genome-wide homozygosity or IBD_G . To provide a solution to this problem, I implemented an extensive theoretical framework around inbreeding and inbreeding depression (Szulkin et al., 2010) in the `inbreedR` package, which is described in detail in Chapter 7.

Study system: Pinnipeds

An outstanding opportunity to study the demographic origins of genetic diversity as well as the intricate details of chemical communication, host-microbe interactions and inbreeding depression is provided by the pinnipeds (Figure 1.3), a monophyletic and highly diverse group of marine mammals comprising the three families Phocidae (true seals), Otariidae (eared seals) and Odobenidae (the walrus).

Pinnipeds inhabit nearly every marine environment across the globe, from the icy shores of the Antarctic to the rocky caves of the Mediterranean. Naturally, these habitats have led to dramatic differences in their ecological and life-history adaptations (Ferguson and Higdon, 2006). The pinnipeds show drastic differences in body sizes, from only around one meter in Baikal seals (*Pusa sibirica*) to sometimes over five meters in adult Southern elephant seal bulls (*Mirounga leonina*), which translates into a 40 times difference in weight (Wilson and Mittermeier, 2014). These enormous size differences are linked to a large variation in mating systems which range from slightly polygynous in most phocids to some of the most polygynous mating systems we know in most otariids, the walrus and the two elephant seal species (Cassini, 1999). In these polygynous systems, males guard harems of dozens of females to mate with during the breeding season (see Figure 1.4) (Lindenfors et al., 2002). Both size differences and mating systems are intertwined with a third trait upon which pinnipeds vary strongly, their sexual size dimorphism (see Figure 1.4). While in many phocids, males and females are equally large and heavy, in some otariids and the elephant seals, males can be up to four to six times heavier than females. These are some of the most extreme examples of sexual dimorphism among all vertebrates and have likely evolved due to both sexual selection and ecology (Krtüger et al., 2014).

What all pinnipeds have in common is their fur and a thick layer of blubber, which are critical for their aquatic lifestyles but which also exposed them to severe commercial exploitation during the 18th and 19th century (Wilson and Mittermeier, 2014). While seal fur was in high demand for clothing, seal blubber was processed into oils which were an important component of lighting and cosmetics at that time (Wilson and Mittermeier, 2014). This has led to the drastic decline of many species as a consequence of over-exploitation, ultimately leading the extinction

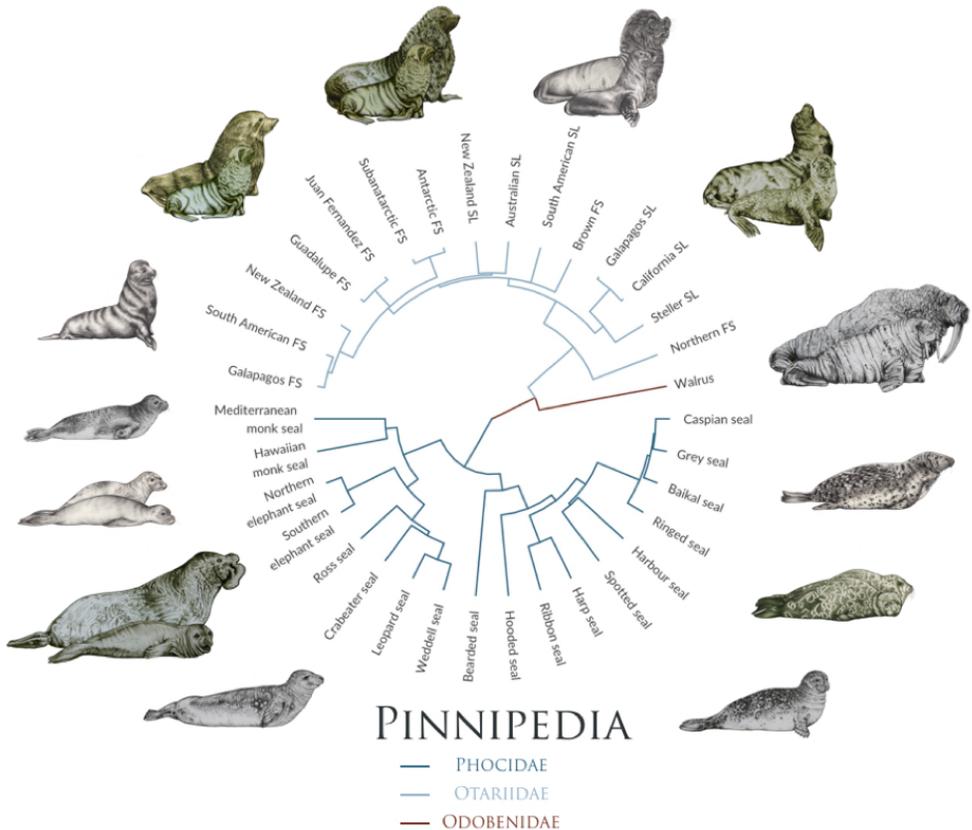


Figure 1.3: Cladogram based on a phylogeny from the 10K trees project. Shown is the phylogenetic relationship between the extant pinniped species and their families. Abbreviations: FS for fur seal and SL for sea lion. Illustrations depict (from left bottom to right bottom) the Crabeater seal, Northern elephant seal, Hawaiian monk seal, Mediterranean monk seal, Galapagos fur seal, Guadalupe fur seal, Antarctic fur seal, South American sea lion, California sea lion, Walrus, Grey seal, Ringed seal, Harbour seal. Illustrations kindly provided by Rebecca Carter for use in this thesis.

of two species, the Japanese sea lion and the Caribbean monk seal. Most extant species however, managed to recover with the introduction of international sealing bans (Wilson and Mittermeier, 2014) and are now thought to have recolonized much of their former ranges from small remaining colonies (e.g. Humble et al. 2018). Nevertheless, some pinniped species are on the decline again (Atkinson et al., 2008; Forcada and Hoffman, 2014; Robertson and Chilvers, 2011) or are still endangered due to intensive anthropogenic disturbances and limited habitats such as the Mediterranean monk seal (Karamanlidis et al., 2016) and the Hawaiian monk seal (Baker and Johanos, 2004).

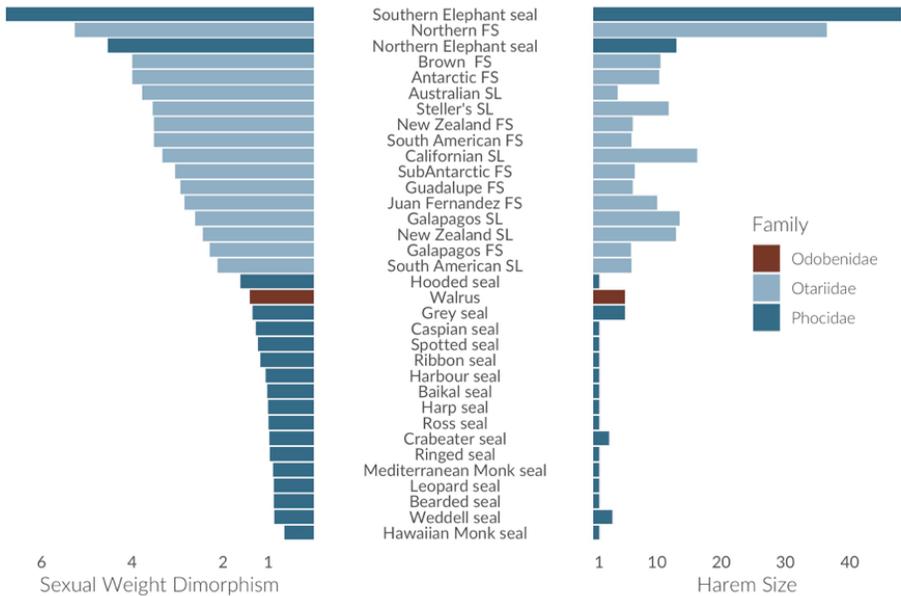


Figure 1.4: Sexual weight dimorphism (male weight / female weight) and average harem sizes of pinnipeds. A correlation between both is reflecting an increase in size and size dimorphism due to the potential of some species to monopolize large harems of females and increasing male-male competition. Data from Lindenfors et al. (2002).

The extent to which pinniped species were affected by the sealing industry differed drastically, with some species such as the Northern elephant seal having been decimated to near extinction and others having never been hunted at all, such as the crabeater seal in the Antarctic. As consequence, there is an extreme level of variation in the recent demographic histories of pinnipeds, which provides an unprecedented opportunity to evaluate the power of genetic demographic inference using techniques such as ABC, but also to explore recent demography as a determinant of genetic diversity in contemporary SL populations (Leffler et al., 2012). In Chapter 2, I explore the demographic histories of most extant pinniped species in a large comparative genetic framework to evaluate the genetic consequences of commercial exploitation and the role of both ecology and life-history.

A major life-history difference between the three families are differences in maternal care strategies. While in all pinnipeds the mother alone cares for her pup, the maternal care strategies vary drastically and can typically be categorized into ‘aquatic nursing’, ‘foraging cycle’ and ‘fasting’ strategies (Boness and Bowen, 1996). The foraging cycle strategy is reminiscent of that found in mammals and birds where the mother leaves her young to forage. Otariids such

as the Antarctic fur seal go through a cycle where mothers come ashore, give birth and then alternate between foraging at sea and nursing their young ashore. As Antarctic fur seal colonies are large and dense aggregations of individuals, well-developed mother-offspring recognition mechanisms have developed, which involve a three step mechanism using knowledge of the location, vocal distance cues and olfactory close-range recognition (Dobson and Jouventin, 2003). Consequently, it is likely that skin chemicals provide an important means of communication between mothers and pups in fur seals. Moreover, olfactory cues can also explain mate choice for genetic characteristics in this species (Hoffman et al., 2007a). In Chapter 4, I explore these possibilities through identifying the potential signals in skin chemicals of Antarctic fur seals.

A very different maternal care strategy is followed by most phocids, where mothers and pups stay together during nursing, forcing the mother to fast for up to a month in species such as the Northern elephant seal. After this nursing period, the mother abandons her offspring, and the newly weaned young stay ashore for up to two months to develop physiologically and behaviorally until they depart to the sea themselves. In Chapter 5, we used this strategy as an opportunity to explore the development of gut microbiota in young Northern elephant seals. We identified newly abandoned pups and used the critical time after weaning as a unique opportunity to study the intestinal microbiota of young elephant seals, which do not yet feed by themselves during this time. The maternal care strategy in the elephant seal hence provides a chance to study gut microbiota in a wild animal while largely controlling for major confounding factors such as differences in diet, which is a key part of the study in Chapter 5.



Looking for tagged individuals in a dense group of Northern elephant seal weaners on the San Benitos Archipelago, Mexico.

Objectives of the thesis

The broader aims of the thesis are threefold. First of all, I use genetic and genomic data to shed light on the demographic causes of genetic diversity across pinnipeds, with a special emphasis on recent demographic declines caused by commercial seal hunting and more historical post-glacial expansions. Second, I elucidate some of the intricate mechanisms by which genotypic variation underpins variation in complex chemical and microbial phenotypes using both the Antarctic fur seal and the Northern elephant seal as study species. Third and lastly, I present three R packages, which were developed during this dissertation and will hopefully facilitate both scientific progress and reproducibility. More specifically, the objectives of the chapters are as follows:

Chapter 2 - Demography and genetic diversity in pinnipeds

This chapter describes a large-scale comparative genetic study involving more 11,000 individuals from 30 pinniped species. The goal of the study was to estimate the extent to which over-exploitation by the 19th century sealing industry caused genetic bottlenecks across pinnipeds. Moreover, we explore the potentially mediating effects of life-history variation such as mating system, and ecological variation such as breeding habitat. This is probably the first study to estimate the recent demography across a whole group of species, which made it possible to gain novel insights into how severely population declines impact genetic diversity.

Chapter 3 - Genomics and the demographic history of the Northern elephant seal

This chapter explores the demographic history of the Northern elephant seal using a newly sequenced genome and restriction-site associated DNA (RAD) sequencing data from 80 individuals. The Northern elephant seal was presumably hunted to near extinction, but neither the strength nor the duration of the bottleneck have ever been estimated using genomic data. Going further back in the species' history, we also use genomic inference based on the site frequency spectrum to provide a glimpse into the unknown, pre-sealing demography of the species.

Chapter 4 - Chemical communication and genetics

This chapter aims to shed light on the chemical basis of olfactory communication in Antarctic fur seals. In particular, we use GC-MS and microsatellite genotyping to elucidate whether skin chemicals encode potential information for olfactory mother-offspring recognition and mate choice by linking chemical with genotype data. Using different statistical approaches, we further

explore in detail which specific compounds might be involved in chemical communication in fur seals.

Chapter 5 - Chemical communication and genetics

This chapter presents a longitudinal investigation of gut microbiota in Northern elephant seal pups. The study took place immediately after weaning where we started quantifying gut microbiota across a critical developmental period of up to two months. As pups do not yet feed by themselves during this period, the study naturally controlled for the impacts of diet on the gut microbiota and allowed us to explore the more subtle and intrinsic factors shaping early microbial communities, such as genotype and sex.

Chapter 6 - GCalignR: Alignment of Gas-Chromatography data.

This chapter describes GCalignR, an R package for aligning highly variable Gas-Chromatography (GC) data obtained in ecological and evolutionary studies. The package aims at streamlining and automating the alignment process, therefore avoiding manual biases and making the process transparent and reproducible. It emerged from a simpler version of the algorithm developed to align Antarctic fur seal GC profiles in Chapter 4.

Chapter 7 – inbreedR: Measuring inbreeding with genetic markers

This chapter describes inbreedR, an R package for investigating inbreeding using genetic and genomic markers. The package provides functions to measure variation in inbreeding, to plan inbreeding studies in terms of marker number and sample size using simulations, and to evaluate the power of different sets of microsatellite or SNP markers to estimate genome-wide inbreeding based on an established theoretical framework. Within this dissertation, it was used to quantify variation in inbreeding in Chapter 4 and 2 and will build the foundation for upcoming genomics projects on inbreeding depression in Northern elephant seals and Antarctic fur seals.

Chapter 8 – rptR: Repeatability estimation for Gaussian and non-Gaussian data

This chapter presents rptR, an R package for calculating intra-class coefficients or repeatabilities in a mixed model framework. Notably, the package allows one to calculate the repeatability of a trait or measurement while controlling for other variables (adjusted repeatability) for Gaussian, Binomial and Poisson data. Moreover, it implements parametric bootstrapping and permutation tests for quantifying confidence intervals and determining the statistical significance of

repeatability estimates. The package was used to compare the repeatability of genetic diversity and bottleneck measurements across different datasets in Chapter 2 and to quantify the repeatability of microbial diversity in Chapter 5.

A note on scientific software development, reproducible research and science.

Modern science has problems. While most scientific findings might even be false (Ioannidis, 2005), there is certainly a problem in the replicability of research. Although systematic problems relating to replicability were initially revealed in psychology (Aarts et al., 2015) and cancer biology (Begley and Ellis, 2012), these issues are likely to be commonplace in all empirical sciences (Baker, 2016). The replicability of scientific findings using independent investigators and methods is the standard by which all science should be evaluated but can be hindered by various aspects of the implementation of a study. These aspects start with the researchers' degrees of freedom in planning and running a study (Wicherts et al., 2016) to issues with the statistical analyses, in particular p-hacking (Benjamin et al., 2018; Halsey et al., 2015; McShane et al., 2017; Peng, 2015). One of the most critical points however, is that the analytical pipeline from raw data to published results is not reproducible, due to a lack of open data and/or documented code (Peng, 2011).

Overall, reproducibility critically relies on code-based analyses instead of point and click software and on frameworks which are widely used and provide all relevant analytical methods in a field. The R language (R Core Team, 2015) provides such a framework. Virtually all established methods in biology are available as R packages from the two major repositories CRAN and Bioconductor (Gentleman et al., 2004). Furthermore, R provides all relevant tools for reproducible research, starting from dplyr for data wrangling (Wickham et al., 2014), to ggplot2 for visualization (Wickham, 2009) and knitr for documenting code (Xie, 2014). Importantly, R provides the tools for developing scientific packages in order to make new methods readily available to the community (Wickham et al., 2013). The studies in this thesis are accompanied by both complete and documented analytical pipelines (available at <https://github.com/mastoffel>) which generally focus on effect size and confidence interval based statistical reporting (Nakagawa and Cuthill, 2007). Moreover, the three scientific R packages presented in the Chapters 6, 7 and 8 were developed with the idea to fill substantial gaps in the availability of scientific methods and are hopefully of use for the scientific community.



Only a very small number of Guadalupe fur seals survived commercial exploitation on the remote Guadalupe island. It took nearly a hundred years until the species started to expand its breeding grounds to the San Benitos Islands again.

2

Comparative demography

Demographic histories and genetic diversity across pinnipeds are shaped by human exploitation, ecology and life-history

Martin A Stoffel, Emily Humble, Anneke J Pajmans, Karina Acevedo-Whitehouse, Barbara L Chilvers, Bobette Dickerson, Fillipo Galimberti, Neil Gemmell, Simon D Goldsworthy, Hazel J Nichols, Oliver Krüger, Sandra Negro, Amy Osborne, Teresa Pastor, Bruce C Robertson, Simona Sanvito, Jennifer Schultz, Aaron BA Shafer, Jochen BW Wolf & Joseph I Hoffman

Nature Communications, in press. bioRxiv doi: <https://doi.org/10.1101/293894>.

Abstract

A central paradigm in conservation biology is that population bottlenecks reduce genetic diversity and negatively impact population viability and adaptive potential. In an era of unprecedented biodiversity loss and climate change, understanding both the determinants and consequences of bottlenecks in wild populations is therefore an increasingly important challenge. However, as most studies have focused on single species, the multitude of potential drivers and the consequences of bottlenecks remain elusive. Here, we used a comparative approach by integrating microsatellite data from over 11,000 individuals of 30 pinniped species with demographic, ecological and life history data to elucidate the consequences of large-scale commercial exploitation by 18th and 19th century sealers. We show that around one third of these species exhibit strong genetic signatures of recent population declines, with estimated bottleneck effective population sizes reflecting just a few tens of surviving individuals in the most extreme cases. Bottleneck strength was strongly associated with both breeding habitat and mating system variation, and together with global abundance explained a large proportion of the variation in genetic diversity across species. Overall, there was no relationship between bottleneck intensity and IUCN status, although three of the four most heavily bottlenecked species are currently endangered. Our study reveals an unforeseen interplay between anthropogenic exploitation, ecology, life history and demographic declines, sheds new light on the determinants of genetic diversity, and is consistent with the notion that both genetic and demographic factors influence population viability.

Introduction

Unravelling the demographic histories of species is a fundamental goal of population biology and has tremendous implications for understanding the genetic variability observed today (Salmona et al., 2017; Ellegren and Galtier, 2016). Of particular interest are sharp reductions in the effective population size (N_e) known as population bottlenecks (Nei et al., 1975; Tajima, 1989), which may negatively impact the viability and adaptive evolutionary potential of species through a variety of stochastic demographic processes and the loss of genetic diversity (Lande, 1988; Frankham et al., 1999; Spielman et al., 2004; Frankham, 2005). Specifically, small bottlenecked populations have elevated levels of inbreeding and genetic drift, which decrease genetic variability and can lead to the fixation of mildly deleterious alleles and ultimately drive a vortex of extinction (Frankham et al., 1999; Frankham, 2005; Mills and Smouse, 1994; Lande, 1994). Hence, investigating the bottleneck histories of wild populations and their determinants and

consequences is more critical than ever before, as we live in an era where global anthropogenic alteration and destruction of natural habitats are driving species declines on an unprecedented scale (Li et al., 2016; Ceballos et al., 2017).

Unfortunately, detailed information about past population declines across species is sparse because historical population size estimates are often either non-existent or highly uncertain (Wilson and Mittermeier, 2014; Lotze and Worm, 2009). A versatile solution for inferring population bottlenecks from a single sample of individuals is to compare levels of observed and expected genetic diversity, the latter of which can be simulated under virtually any demographic scenario based on the coalescent (Kingman, 1982; Hudson, 2002). A variety of approaches based on this principle have been developed, one of the most widely used being the heterozygosity-excess test, which compares the heterozygosity of a panel of neutral genetic markers to the expectation in a stable population under mutation-drift equilibrium (Cornuet and Luikart, 1996). Although theoretically well grounded, these methods are highly sensitive to the assumed mutation model, which is seldom known (Peery et al., 2012). A more sophisticated framework for inferring demographic histories is coalescent-based Approximate Bayesian Computation (ABC) (Beaumont et al., 2002). ABC has the compelling advantages of making it possible to (i) compare virtually any demographic scenario as long as it can be simulated, (ii) estimate key parameters of the model such as the bottleneck effective population size and (iii) incorporate uncertainty in the specification of models by defining priors. Due to this flexibility, ABC has become a state of the art approach for inferring population bottlenecks as well as demographic histories in general (Beaumont et al., 2002; Csilléry et al., 2010; Wegmann and Excoffier, 2010; Hoffman et al., 2011; Shafer et al., 2015a; Csilléry et al., 2012; Chan et al., 2006; Xue and Hickerson, 2015; Duchon et al., 2013; Chan et al., 2014).

Although the widespread availability of neutral molecular markers such as microsatellites has facilitated numerous genetic studies of bottlenecks in wild populations, the vast majority of studies focused exclusively on single species and were confined to testing for the presence or absence of bottlenecks. We therefore know very little about the intensity of demographic declines and how these are influenced by anthropogenic impacts as well as by factors intrinsic to a given species. For example, species occupying breeding habitats that are more accessible to humans would be expected to be at higher risk of declines, while species with highly skewed mating systems tend to have lower effective population sizes (Charlesworth, 2009) and might also experience stronger demographic declines as only a fraction of individuals contribute towards the genetic makeup of subsequent generations. Consequently, to disentangle the forces shaping population bottlenecks, we need comparative studies incorporating genetic, ecological and life history data from multiple closely related species within a consistent analytical

framework.

Another question that remains elusive due to a lack of comparative studies is to what extent recent bottlenecks have impacted the genetic diversity of wild populations. While a number of influential studies of heavily bottlenecked species have indeed found very low levels of genetic variability (Houlden et al., 1996; Hoelzel et al., 2002; Pinsky and Palumbi, 2014; O'Brien, 1994) others have reported unexpectedly high genetic variation after supposedly strong population declines (Hoffman et al., 2011; Hailer et al., 2006; Dinerstein and McCracken, 1990; Busch et al., 2007; Roman and Palumbi, 2003). Hence, it is not yet clear how population size changes contribute towards one of the most fundamental questions in evolutionary genetics - how and why genetic diversity varies across species (Lewontin, 1974; Ellegren and Galtier, 2016; Leffler et al., 2012; Romiguier et al., 2014). To tackle this question, we need to compare closely related species because deeply divergent taxa vary so profoundly in their genetic diversity due to differences in their life-history strategies that any effects caused by variation in N_e will be hard to detect and decipher (Leffler et al., 2012; Romiguier et al., 2014).

Finally, the relative contributions of genetic diversity and demographic factors towards extinction risk remain unclear. While historically there has been a debate about the immediate importance of genetic factors towards species viability (Lande, 1988; Spielman et al., 2004) there is now growing evidence that low genetic diversity increases extinction risk (Frankham, 2005; Saccheri et al., 1998) and on a broader scale that threatened species tend to show reduced diversity (Spielman et al., 2004). Nevertheless, due to a lack of studies measuring bottlenecks consistently across species, it remains an open question as to how the loss of genetic diversity caused by demographic declines ultimately translates into a species extinction risk, which can be assessed by its International Union for Conservation of Nature (IUCN) status.

An outstanding opportunity to address these questions is provided by the pinnipeds, a clade of marine carnivores inhabiting nearly all marine environments ranging from the poles to the tropics and showing remarkable variation in their ecological and life-history adaptations (Ferguson and Higdon, 2006). Pinnipeds include some of the most extreme examples of commercial exploitation known to man, with several species including the northern elephant seal having been driven to the brink of extinction for their fur and blubber by 18th to early 20th century sealers (Wilson and Mittermeier, 2014). By contrast, other pinniped species inhabiting pristine environments such as Antarctica have probably had very little contact with humans (Wilson and Mittermeier, 2014). Hence, pinnipeds show large differences in their demographic histories within the highly constrained time window of commercial sealing and thereby represent a unique 'natural experiment' for exploring the causes and consequences of recent bottlenecks.

Here, we conducted a broad-scale comparative analysis of population bottlenecks using a combination of genetic, ecological and life-history data for (Charlesworth, 2009) pinniped species. We inferred the strength of historical declines across species from the genetic data using two complimentary coalescent-based approaches, heterozygosity-excess and ABC. Heterozygosity-excess was used as a measure of the relative strength of recent population declines, while a consistent ABC framework was used to evaluate the probability of each species having experienced a severe bottleneck during the known timeframe of commercial exploitation, as well as to estimate relevant model parameters. Finally, we used Bayesian phylogenetic mixed models to investigate the potential causes and consequences of past bottlenecks while controlling for phylogenetic relatedness among species. We hypothesised that (i) extreme variation in the extent to which species were exploited by man should be reflected in their genetic bottleneck signatures; (ii) ecological and life-history traits could have an impact on the strength of bottleneck signatures across species; (iii) past bottlenecks should reduce contemporary genetic diversity; and (iv) heavily bottlenecked species with reduced genetic diversity will be more likely to be of conservation concern.

Results

Genetic data.

We analysed a combination of published and newly generated microsatellite data from 30 pinniped species, with a median of 253 individuals and 14 loci per species (see Methods and Supplementary Table 1 for details). Measures of genetic diversity, standardised across datasets as the average per ten individuals, varied considerably across the pinniped phylogeny, with observed heterozygosity (H_o) and allelic richness (A_r) varying by over two and almost five-fold respectively across species (Supplementary Table 2). Both of these measures were highly correlated ($r = 0.92$) and tended to be higher in ice breeding seals, intermediate in fur seals and sea lions, and substantially lower in a handful of species including northern elephant seals and monk seals (Figure 2.1A).

Bottleneck inference.

We used two different coalescent-based approaches to infer the extent of recent population bottlenecks. First, the amount of heterozygosity-excess at selectively neutral loci such as microsatellites is an indicator of recent bottlenecks because during a population decline the number

of alleles decreases faster than heterozygosity (Nei et al., 1975). Recent bottlenecks therefore generate a transient excess of heterozygosity relative to a population at equilibrium with an equivalent number of alleles (Cornuet and Luikart, 1996). Here, we quantified the proportion of loci in heterozygosity-excess ($prop_{het-exc}$) for each species, which was highly repeatable across a range of mutation models (see Methods and Supplementary Table 3). Consequently, we focused on a two-phase model with 80% single-step mutations (TPM80), which is broadly in line with mammalian mutation model estimates from the literature (Ellegren, 2004) as well as posterior estimates from our ABC analysis (Supplementary Table 4B, Supplementary Figure 5). Figure 2.1B shows a heatmap of $prop_{het-exc}$ across species, which is bounded between zero (all loci show heterozygosity-deficiency, an indicator of recent expansion) and one (all loci show heterozygosity-excess, an indicator of recent decline) whereby 0.5 is the expectation for a stable population. Considerable heterogeneity was found across species, with northern and southern elephant seals, grey seals, Guadalupe fur seals and Antarctic fur seals showing the strongest bottlenecks signals. By contrast, the majority of ice-breeding seals exhibited heterozygosity-deficiency, consistent with historical population expansions.

Second, we used ABC to select between a bottleneck and a non-bottleneck model as well as to estimate posterior distributions of relevant parameters. To optimally capture recent population size changes across species, we allowed N_e to vary from pre- to post-bottleneck in both models within realistic priors (see Methods for details) while the bottleneck model also included a severe decrease in N_e to below 500 during the time of peak sealing. Therefore, both models incorporate longer-term declines or expansions within realistic bounds for all species but only the bottleneck model captures a recent and severe decrease in N_e due to anthropogenic exploitation. ABC was clearly able to distinguish between the two models, with simulations under the bottleneck model being correctly classified 85% of the time and simulations under the non-bottleneck model being correctly classified 89% of the time (Supplementary Figure 1). A small amount of overlap between the models and therefore misclassification is unavoidable because both models were specified using broad priors to optimally fit a variety of species with vastly different population sizes. For each species, however, the preferred model showed a good fit to the observed data (all p -values > 0.05 , Supplementary Table 5) (Lemaire et al., 2016). As another indicator of model quality, posterior predictive checks (Csilléry et al., 2010; Gelman et al., 1995) showed that the preferred models across all species were largely able to reproduce the relevant observed summary statistics (Supplementary Figure 2). The posterior bottleneck model probability (p_{bot}) varied substantially across species and was strongly but imperfectly correlated with $prop_{het-exc}$ (posterior median and 95% credible intervals; $\beta = 0.17$ [0.04, 0.28], $R^2_{\text{marginal}} = 0.32$ [0.03, 0.59], see Supplementary Figure 3). For eleven species, the

bottleneck model was supported with a higher probability than the non-bottleneck model (i.e. $p_{bot} > 0.5$, see Supplementary Table 3). Subsequent parameter estimation was therefore based on the bottleneck model for eleven species and on the non-bottleneck model for the other 19 species.

Under the bottleneck model, prediction errors from the cross-validation were well below one for the bottleneck effective population size (N_{ebot} , Supplementary Table 4A and Supplementary Figure 4) and mutation rate (μ , Supplementary Table 4A) indicating that posterior estimates contain information about the underlying true parameter values. Similarly, under the non-bottleneck model, μ (Supplementary Table 4B) and the parameter describing the proportion of multi-step mutations (GSM_{par} , Supplementary Table 4B) were informative. By contrast, although the pre-bottleneck effective population size (N_{ehist}) also had a prediction error below one in both models, visual inspection of the cross-validation results revealed high variation in the estimates and a systematic underestimation of larger N_{ehist} values, so this parameter was not considered further. Figure 2.2 shows the eleven bottlenecked species ranked in descending order of estimated posterior modal N_{ebot} (see also Supplementary Table 4A). The parameter estimates were indicative of strong bottlenecks (i.e. $200 < N_{ebot} < 500$) in seven species including both phocids and otariids, while even smaller N_{ebot} values (i.e. $N_{ebot} < 50$) were estimated for four phocids comprising the landlocked Saimaa ringed seal, both monk seal species and the northern elephant seal. Mutation rate estimates were remarkably consistent across species, with modes of the posterior distributions typically varying around 1×10^{-4} (Supplementary Figure 5 and Supplementary Table 4), while GSM_{par} across species typically varied between around 0.2 and 0.3 (See Supplementary Figure 6 and Supplementary Table 4B). Therefore, although studies of individual species are usually limited by uncertainty over the underlying mutation characteristics, our ABC analyses converged on similar estimates of mutation model and rate across species, allowing us to appropriately parameterise our bottleneck analyses.

To explore whether our results could be affected by population structure, we used STRUCTURE (Pritchard et al., 2000) to infer the most likely number of genetic clusters (K) across all datasets (see Supplementary Table 6). For all of the species for which the best supported value of K was more than one ($n = 12$), we recalculated genetic summary statistics and repeated the bottleneck analyses based on individuals comprising the largest cluster. Using the largest genetic clusters did not appreciably affect our results, with repeatabilities for the genetic summary statistics and bottleneck signatures all being greater than 0.9 (see Supplementary Table 7 for repeatabilities and Supplementary Figure 7, which is virtually identical to Figure 2.1).

Furthermore, we tested all loci from each dataset for deviations from Hardy-Weinberg

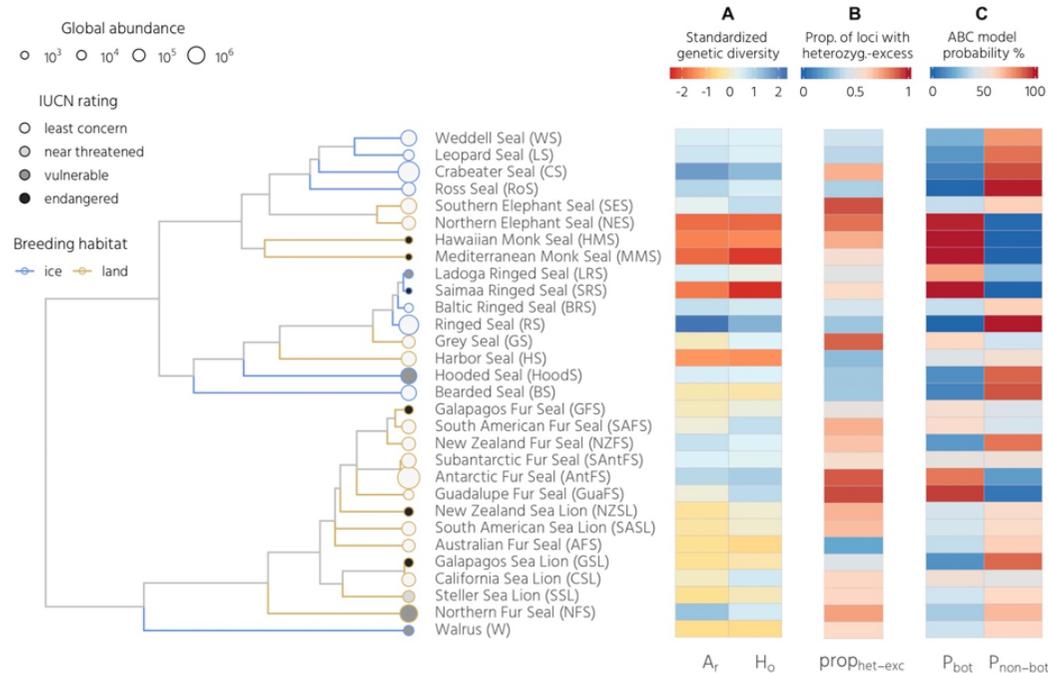


Figure 2.1: Patterns of genetic diversity and bottleneck signatures across the pinnipeds. The phylogeny shows 30 species with branches colour coded according to breeding habitat and tip points coloured and sized according to their IUCN status and global abundance respectively. Panel A shows two genetic diversity measures, allelic richness (A_r) and observed heterozygosity (H_o), which have been standardised by randomly subsampling ten individuals from each dataset 1000 times with replacement and calculating the corresponding mean. Panel B shows the proportion of loci in heterozygosity-excess ($prop_{het-exc}$) calculated for the TPM80 model (see Methods for details). Panel C summarises the ABC model selection results, with posterior probabilities corresponding to the bottleneck versus non-bottleneck model. The raw data are provided in Supplementary Tables 2 and 3.

equilibrium (HWE, see the Methods for details). Overall, 6% of loci were found to deviate from HWE in both χ^2 and exact tests after table-wide Bonferroni correction for multiple testing. To investigate whether including these loci could have affected our results, we recalculated the genetic summary statistics and repeated our bottleneck analyses after excluding them. The results remained largely unaltered, with repeatabilities all being greater than 0.97 (see Supplementary Table 8 and Supplementary Figure 8).

Finally, we considered the possibility that our inference of recent bottlenecks could have been confounded by events further back in a species' history. In particular, increased ice cover during the last glacial maximum (LGM) could have reduced habitat availability and consequently population sizes (Coyer et al., 2003; Burbrink et al., 2008; Liu et al., 2006; Burbrink et al., 2016; Gehara et al., 2017). We therefore tested whether small population sizes during the LGM followed by expansions could result in similar genetic patterns across pinnipeds to recent bottlenecks caused by anthropogenic exploitation (for details, see Supplementary Information). Specifically, we used ABC to simulate two additional demographic scenarios that were identical to the bottleneck and non-bottleneck models but which also incorporated a small population size during the LGM and subsequent expansion. ABC was not able to reliably distinguish between the alternative bottleneck models: correct classification rates were substantially lower at 64% for the bottleneck model and 60% for the bottleneck model incorporating post-glacial expansion. Similarly, the two non-bottleneck models had relatively poor classification rates (60% for the non-bottleneck model and 66% for the non-bottleneck model incorporating expansion). These rates are much lower than in our main analysis based on two models, indicating that ABC cannot reliably distinguish on the basis of our data between broadly equivalent models that do and do not include ice age effects. Regardless, all eleven of the species that supported the bottleneck model in the main analysis again showed the highest probability for one of the two models that incorporated a recent bottleneck (Supplementary Table 11). The fact that none of these species supported the non-bottleneck model with post-glacial expansion indicates that the reduction in genetic diversity produced by a recent bottleneck can be clearly distinguished from the reduction in diversity due to a small population size at the end of the last ice age. This is to be expected as many of our summary statistics such as the *M*-ratio are sensitive towards recent population size changes (Garza and Williamson, 2001).

Factors affecting bottleneck history.

Conceivably both ecological and life-history variables could have impacted the extent to which commercial exploitation affected different pinniped species. We therefore investigated the effects

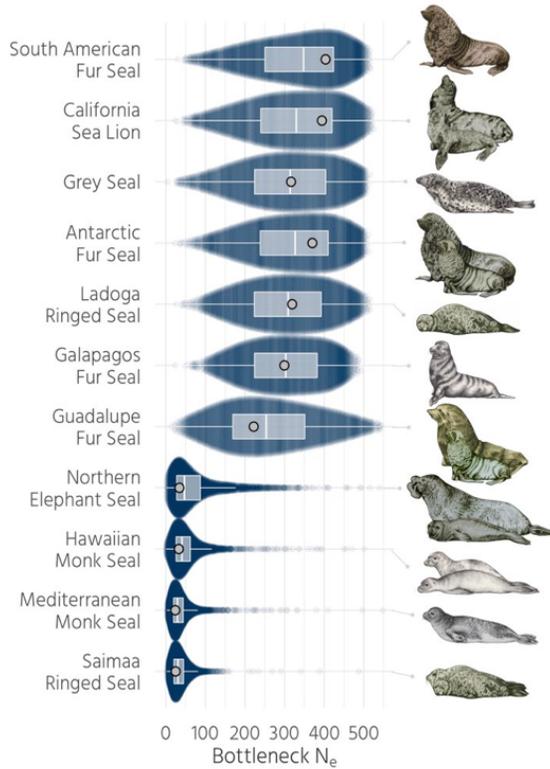


Figure 2.2: Estimated bottleneck effective population sizes. Posterior distributions of N_{e}^{bot} are shown for eleven species for which the bottleneck model was supported in the ABC analysis, ranked according to the modes of their density distributions which reflect the estimated most likely N_{e}^{bot} . Prior distributions are not shown as N_{e}^{bot} was drawn from a uniform distribution with $U[1, 500]$. For each species, parameter values for 5,000 accepted simulations are presented as a sinaplot, which arranges the data points to reflect the estimated posterior distribution. Superimposed are Tukey boxplots with light grey points representing maximum densities.

of four different variables on bottleneck signatures. First, we hypothesised that breeding habitat would be important as ice-breeding species are less accessible and more widely dispersed than their land breeding counterparts. Second, we considered sexual size dimorphism (SSD) an important life history variable as species with a high SSD aggregate in denser breeding colonies, making them more valuable to hunters, and polygyny reduces effective population size. Third, the length of the breeding season may have impacted the vulnerability of a given species to exploitation and finally, generation time could potentially mediate population recovery. We found clear differences between ice- and land-breeding seals in both $prop_{het-exc}$ and p_{bot} ,

with land-breeders on average showing stronger bottleneck signatures (Figure 2.3A, B). In addition, $prop_{het-exc}$ was positively associated with SSD but not with p_{bot} (Figure 2.3C) and the former relationship was robust to the exclusion of the southern elephant seal (Supplementary Figure 9). However, we did not find the expected positive relationships with either breeding season length or generation time (see below).

To investigate this further, we constructed two Bayesian phylogenetic mixed models with $prop_{het-exc}$ and p_{bot} as response variables respectively and breeding habitat, SSD, breeding season length and generation time fitted as predictors (see Methods for details). Both models explained an appreciable amount of variation ($prop_{het-exc}$ $R^2_{\text{marginal}} = 0.58$, CI [0.22, 0.92]; p_{bot} $R^2_{\text{marginal}} = 0.38$, CI [0.08, 0.62], Figure 2.3D). As the four predictor variables show some level of multicollinearity (Supplementary Table 9), we reported both standardised model estimates (β) and structure coefficients ($r(\hat{Y}, x)$), which represent the correlation between each predictor and the fitted response independent of the other predictors. Breeding habitat showed the largest overall effect size in both models (Figure 2.3E, Supplementary Table 9). By contrast, structure coefficients showed that breeding habitat and SSD were both strongly correlated to the fitted response in the $prop_{het-exc}$ model, while SSD indeed had a much weaker effect in the p_{bot} model (Figure 2.3F, Supplementary Table 10). Thus, breeding habitat and SSD explain variation in $prop_{het-exc}$ whereas only breeding habitat explains variation in p_{bot} . We did not find a relationship between breeding season length and bottleneck signatures, with R^2 , β and structure coefficients all being low with broad CIs overlapping zero (Figure 2.3D, E and F). While the structure coefficient of generation time in the $prop_{het-exc}$ model did not have CIs overlapping zero, a negative relationship is contrary to expectations and probably reflects the longer generation times of ice-breeding seals (Supplementary Figure 10) rather than a genuine relationship.

Determinants of genetic diversity.

To investigate the determinants of contemporary genetic diversity across pinnipeds, we constructed a phylogenetic mixed model of allelic richness (A_r) with log transformed global abundance, breeding habitat and SSD fitted as predictor variables together with the two bottleneck measures $prop_{het-exc}$ and p_{bot} (Figure 2.4). In order to avoid over-fitting the model, we did not include breeding season length and generation time, as these variables were not individually associated with A_r (breeding season: $\beta = 0.01$ CI [-0.03, 0.01], generation time: $\beta = 0.00$ CI [0.00, 0.01]). A substantial 75% of the total variation in A_r was explained (Figure 2.4C, $R^2_{\text{marginal}} = 0.75$, CI [0.52, 0.91]). Specifically, A_r decreased nearly five-fold from the species

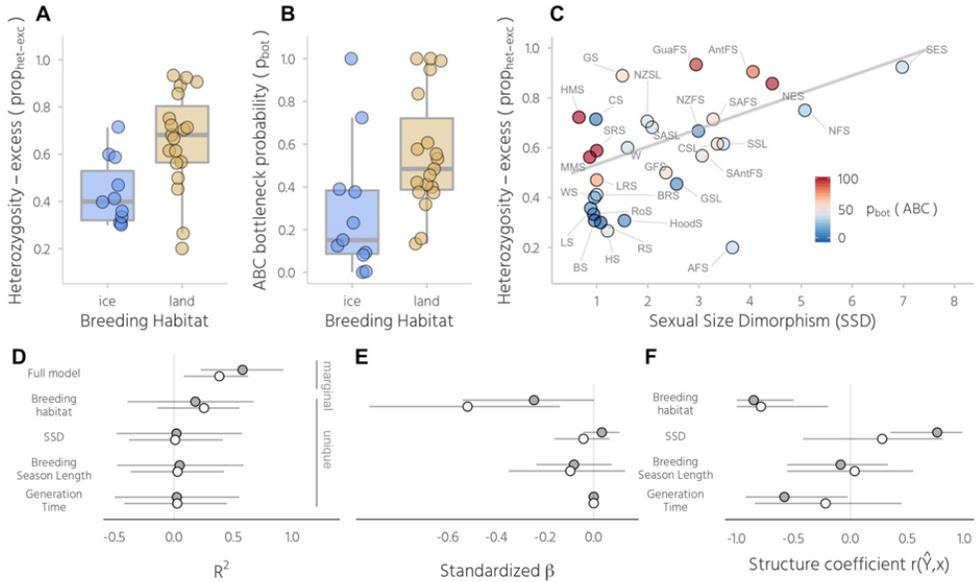


Figure 2.3: Ecological and life-history effects on bottleneck signatures. Shown are the results of phylogenetic mixed models of $prop_{het-exc}$ and p_{bot} with breeding habitat and SSD fitted as fixed effects. Panels A and B show differences between ice- and land-breeding species in $prop_{het-exc}$ and p_{bot} respectively. Raw data points are shown together with standard Tukey box plots. Panel C shows the relationship between sexual size dimorphism (SSD) and $prop_{het-exc}$, with individual points colour coded according to the ABC bottleneck probability (p_{bot}) and the line representing the predicted response from the $prop_{het-exc}$ model. Marginal and unique R^2 values, standardized β coefficients and structure coefficients are shown for models of $prop_{het-exc}$ (filled points) and p_{bot} (open points) in panels D–F, where they are presented as posterior medians with 95% credible intervals. Species abbreviations are given in Figure 2.1 and Supplementary Table 1.

with the lowest p_{bot} to the species with the highest p_{bot} ($\beta = -1.80$, CI [-3.10, -0.42] Figure 2.4A), increased by nearly five-fold from the least to the most abundant species ($\beta = 1.38$, CI [0.21, 2.47], Figure 2.4B), and was on average 27% higher in ice than in land-breeding seals ($\beta = 1.76$, CI [0.10, 3.14], Figure 2.4B). Due to multicollinearity among the five predictor variables (Supplementary Table 9), standardized β estimates (Figure 2.4D) can be hard to interpret because of potential suppression effects (Ray-Mukherjee et al., 2014). This is reflected by the low unique R^2 values of the predictors relative to the marginal R^2 of the full model (Figure 2.4C). However, the structure coefficients (Figure 2.4E) also revealed strong associations between the fitted model response and breeding habitat ($(r(\hat{Y}, x) = 0.54$, CI [0.20, 0.76]), abundance ($r(\hat{Y}, x) = 0.73$, CI [0.54, 0.91]) and p_{bot} ($r(\hat{Y}, x) = -0.78$, CI [-0.91, -0.62]) indicating that

all three variables are associated with the response.

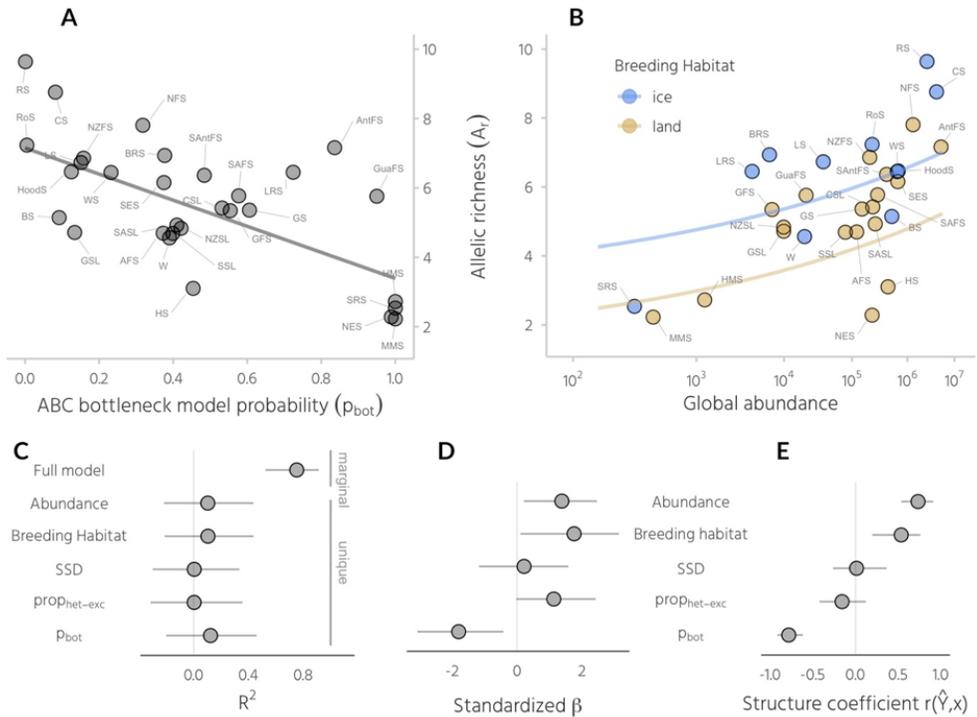


Figure 2.4: Determinants of contemporary genetic diversity across pinnipeds. Panel A shows a scatterplot of A_r versus p_{bot} with the grey line representing the model prediction. Panel B shows the relationship between global abundance and allelic richness (A_r) with the lines representing model predictions for ice- and land-breeding seals respectively. Marginal and unique R^2 values, standardised β estimates and structure coefficients for the model are shown respectively in panels C–E, where they are presented as posterior medians with 95% credible intervals. Species abbreviations are given in Figure 2.1 and Supplementary Table 1.

Conservation status, bottleneck signatures and genetic diversity.

To investigate whether population bottlenecks and low genetic diversity are detrimental to species viability, we asked whether contemporary conservation status is related to the strength of past bottlenecks and A_r . Based on data from the IUCN red list (IUCN, 2018), we classified species into two categories; the first of these, which we termed ‘*low concern*’ comprised species listed as ‘*least concern*’ and ‘*near threatened*’, while the second combined species listed as ‘*vulnerable*’ or ‘*endangered*’ into a ‘*high concern*’ category. Using a phylogenetic mixed model,

we did not find any clear differences in either heterozygosity-excess or p_{bot} with respect to conservation status (Figure 2.5A, B). By contrast, average A_r was around 1.2 alleles lower in the *high concern* category, although there was considerable uncertainty with the 95% credible interval of β ranging from -0.08 to 2.56 (Figure 2.5C).

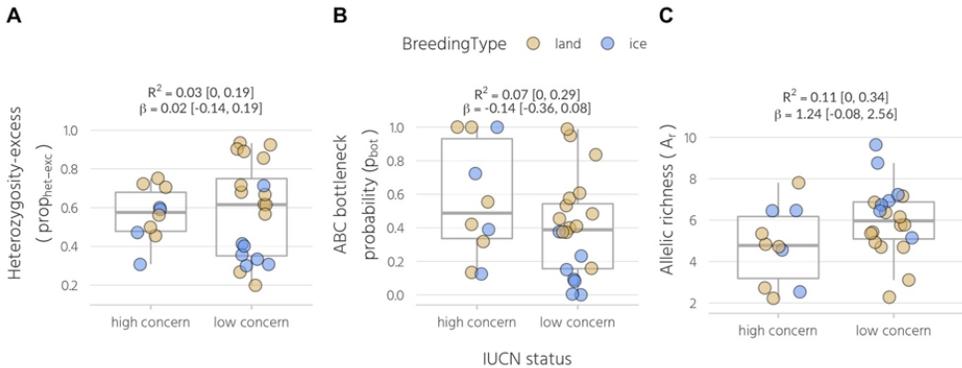


Figure 2.5: Conservation implications of bottlenecks and genetic diversity. All pinniped species were classified into either a low concern or a high concern category depending on their current IUCN status as described in the main text. Shown are the raw data for each category together with standard Tukey box plots for (A) $prop_{het-exc}$, (B) p_{bot} , and (C) A_r . Marginal R^2 and standardised β estimates are shown for Bayesian phylogenetic mixed models with standardized predictors (see Methods for details).

Discussion

To explore the interplay between historical demography, ecological and life-history variation, genetic diversity and conservation status, we used a comparative approach based on genetic data from over 80% of all extant pinniped species. To model bottleneck strength, we used two approaches that capture different but complementary facets of genetic diversity resulting from population bottlenecks. Using ABC, we contrasted a bottleneck model incorporating a severe decrease in N_e during the time of peak sealing in the 18th and 19th centuries with a non-bottleneck model. The resulting bottleneck measure, p_{bot} is the probability (relative to the non-bottleneck model) that a species' observed genetic diversity is similar to the diversity of a population that experienced a severe reduction in N_e below 500, and therefore provides an *absolute* bottleneck measure. By contrast, heterozygosity excess ($prop_{het-exc}$) theoretically captures sudden recent reductions in N_e even in fairly large populations (Cornuet and Luikart, 1996) and therefore provides a *relative* bottleneck measure. Concretely, given the average sample size of individuals

and loci used in this study, we would expect to detect an excess of heterozygosity at the majority of loci (i.e. $\text{prophet-exc} > 0.5$) when a 100- to 1000-fold reduction in N_e occurred, regardless of the magnitude of N_e (see simulations in Cornuet and Luikart 1996).

We specifically focused on two simple ABC models reflecting only recent demographic histories to test a clear hypothesis—large scale commercial exploitation caused severe bottlenecks and reduced the genetic diversity of many pinnipeds. This focus on a short time-frame and well known sealing history allowed us to clearly define our models around reasonable priors. Furthermore, although the genetic diversity simulated based on models of recent demographic history could in principle also be generated by more ancient bottlenecks, these are unlikely to be detected reliably using microsatellite data when a subsequent recovery occurred (Hoban et al., 2013).

ABC analysis supported the bottleneck model for more than a third of the species. The strongest bottlenecks ($N_{e\text{bot}} < 50$) were inferred for the northern elephant seal, a textbook example of a species that bounced back from the brink of extinction (Hoelzel, 1999), as well as for the two monk seals and the Saimaa ringed seal, species with very small geographic ranges and a long history of anthropogenic interaction (Wilson and Mittermeier, 2014). Slightly weaker bottlenecks were estimated for seven further species including Antarctic and Guadalupe fur seals, both of which share a known history of commercial exploitation for their fur (Wilson and Mittermeier, 2014). At the other end of the continuum, several Antarctic species that have not been commercially hunted such as crabeater and Weddell seals showed unequivocal support for the non-bottleneck model in line with expectations. Surprisingly, several otariid species known to have been hunted in the hundreds of thousands (e.g. South American sea lions) to millions (e.g. northern fur seals) did not show support for a bottleneck as strong as simulated in our analyses. This suggests that sufficiently large numbers of individuals must have survived despite extensive sealing, possibly on inaccessible shores or remote islands (Bonin et al., 2013).

A number of factors could potentially impact our inference of the strength of recent bottlenecks across pinnipeds. First of all, population structure and deviations from HWE can affect population genetic inference. However, we found that our measures of genetic diversity as well as bottleneck signatures were highly consistent when we repeated our analyses using the largest genetic clusters or after removing loci that were out of HWE. Second, demographic events deeper in a species' history could potentially confound our inference of recent bottlenecks. However, we believe this is unlikely given the results of our supplementary analysis of post-glacial expansion models and the fact that we chose our summary statistics including the M -ratio to be informative about recent population size changes. Importantly, all 11 species showing

strong signatures of recent bottlenecks in our main analysis did so regardless of whether these bottlenecks were preceded by reduced population sizes followed by expansions towards the end of the late Pleistocene. Moreover, for these species, models incorporating small population sizes during the LGM did not explain the observed genetic variation better than a recent bottleneck model. A third possibility, which will affect any demographic reconstruction from genetic data, is that some of the genetic markers could be linked to loci under selection. In this case, selection would have to operate in the same direction across multiple loci within species and across species to explain our comparative patterns. However, it is not necessary to invoke selection to explain the broad-scale patterns we found across pinnipeds.

We hypothesised that not all pinniped species were equally affected by commercial exploitation partly due to intrinsic differences relating to a species' ecology and life-history. In line with this, we found a strong influence of breeding habitat on bottleneck signatures, with both $prop_{het-exc}$ and p_{bot} being higher in species that breed on land relative to those breeding on ice. A likely reason for this is that terrestrially breeding pinniped species were more profitable due to their generally higher population densities and accessibility, and therefore probably experienced more intense hunting. We also found that heterozygosity-excess was strongly linked to sexual size dimorphism (SSD), with highly polygynous species like elephant seals and some fur seals showing the strongest footprints of recent decline. While this could reflect the increased ease of exploitation and thus higher commercial value of species that predictably aggregate in very large numbers to breed, species with higher SSD also have highly skewed mating systems making them potentially more vulnerable to severe decreases in N_e when key males are taken out of the system. By contrast, we did not find an effect of SSD on the ABC bottleneck probability p_{bot} , suggesting that although sexually dimorphic species experienced the greatest declines, these were not necessarily as severe as simulated in the ABC analysis ($N_e < 500$). This is probably because many species reached 'economic extinction' well above this threshold, when populations became too small to sustain the sealing industry.

Although vast numbers of species are declining globally at unprecedented rates (Ceballos et al., 2017) we still lack a clear understanding of how recent declines in N_e affect contemporary genetic diversity in wild populations (Ellegren and Galtier, 2016; Leffler et al., 2012). Here, we explained a large proportion of the five-fold variation in allelic richness (A_r) observed from the most to the least diverse pinniped species. First, A_r was strongly associated with p_{bot} but not with $prop_{het-exc}$, in agreement with the theoretical expectation that populations have to decline to a very small N_e (Nei et al., 1975), as was simulated in our ABC analysis, to lose a substantial proportion of their diversity. Second, we showed that global abundance across

species was tightly linked to A_r , despite the likely impact of bottlenecks and the limited time-window for the recovery of genetic diversity. As differences in genetic diversity across species are largely determined by long-term N_e (Ellegren and Galtier, 2016), this implies that contemporary population sizes across pinnipeds must to some extent resemble patterns of historical abundance, and hence that many bottlenecked species have to a large extent rebounded to occupy their original niches. Third, A_r was higher in ice-breeding relative to land-breeding seals. However, a low unique R^2 of breeding habitat in our model suggests that this probably reflects the more intense bottleneck histories of land-breeding seals rather than a true ecological effect.

Finally, we compared genetic diversity and bottleneck strength between species that are currently classified by the IUCN as being of conservation concern versus those that are not. We found that A_r was on average around 21% lower in species within the *high concern* category, consistent with previous evidence from a broad range of species (Spielman et al., 2004). While three out of the four pinniped species with the strongest estimated bottlenecks are currently listed as endangered, species from both categories did not overall differ in their bottleneck signatures. Our comparative study of population bottlenecks is therefore encouraging: population bottlenecks do not necessarily result in reduced genetic diversity, population viability and adaptive potential. As shown here, global bans on commercial sealing at the beginning of the 20th century allowed many surviving pinniped populations to recover in abundance. Those that have not sufficiently rebounded illustrate the two fundamental conservation challenges, especially as biodiversity loss and climate change continue at unprecedented rates: halting population declines and promoting population recovery.

Methods

Genetic data.

We obtained microsatellite data for a total of 30 pinniped species including three subspecies of ringed seal (summarised in Supplementary Table 1). First, we conducted systematic literature searches to identify previously published microsatellite datasets for 25 species (see Supplementary Information for details). Second, we generated new data for five species (see Supplementary Information for details). Sample sizes of individuals ranged between 16 for the Ladoga ringed seal to 2386 for the Hawaiian monk seal, with a median of 253 individuals. The number of loci genotyped varied between five and 35 with a median of 14.

Phylogenetic, demographic, life history and conservation status data.

Phylogenetic data were downloaded from the 10K trees website (Arnold et al., 2010) and plotted using *ggtree* (Yu et al., 2017). The three ringed seal subspecies were added according to their separation after the last ice age (Sipilä and Hyvärinen, 2014). Demographic and life-history data for each species were obtained from (Krüger et al., 2014). While most data stayed untransformed, we calculated sexual size dimorphism (SSD) as the ratio of male to female body mass, and log-transformed abundance across species to account for the several orders of magnitude differences in population sizes. Data on conservation status were retrieved from the IUCN website (<http://www.iucnredlist.org/>, 2017) (IUCN, 2018).

Data cleaning and preliminary population genetic analyses.

In order to maximise data quality, we checked all datasets by eye and generated summary statistics and tables of allele counts to identify potentially erroneous genotypes including typographical or formatting errors. In ambiguous cases, we contacted the authors to verify the correct genotypes. As several of the datasets included samples from more than one geographical location, we used a Bayesian approach implemented in *STRUCTURE* version 2.3.4 (Pritchard et al., 2000) to infer the most likely number of genetic clusters (K) across all datasets. For computational and practical reasons, we used the *ParallelStructure* package in R (Besnier and Glover, 2013) to run these analyses on a computer cluster. For all of the species for which the best supported value of K was more than one, we recalculated genetic summary statistics and repeated the bottleneck analyses based on individuals comprising the largest cluster and calculated repeatabilities including 95% confidence intervals (CIs) for all variables using the *rptGaussian* function in the *rptR* package (Stoffel et al., 2017). We also tested all loci from each dataset for deviations from Hardy-Weinberg equilibrium (HWE) using χ^2 and exact tests implemented in *pegas* (Paradis, 2010) and applied Bonferroni correction to the resulting p -values.

Genetic diversity statistics.

In order to examine patterns of genetic diversity across species, we calculated observed heterozygosity (H_o) and allelic richness (A_r) with *strataG* (Archer et al., 2016) as well as the proportion of low frequency alleles (*LEA*), defined as alleles with a frequency of <5%, using self-written code. For maximal comparability across species with different sample sizes, we

randomly sampled ten individuals from each dataset 1000 times with replacement and calculated the corresponding mean and 95% CI for each summary statistic. We did not attempt to standardise our genetic diversity measures by the number of microsatellites, as differences in the number of loci are not expected to systematically bias the mean of any summary statistic across loci.

Heterozygosity-excess.

We quantified heterozygosity-excess using the approach of Cornuet and Luikart (Cornuet and Luikart, 1996) implemented in the program BOTTLENECK version 1.2.02 (Piry et al., 1999). BOTTLENECK compares the heterozygosity of a locus in an empirical sample to the heterozygosity expected in a population under mutation-drift equilibrium with the same number of alleles as simulated under the coalescent (Kingman, 1982; Hudson, 2002). Microsatellites evolve mainly by gaining or losing a single repeat unit (Bonin et al., 2013) (the Stepwise Mutation Model, SMM), but occasional larger jump mutations of several repeat units also occur (Di Rienzo et al., 1994). Consequently, BOTTLENECK allows the user to specify a range of mutation models, from the strict SMM through two-phase models (TPMs) with varying proportions of multi-step mutations to the infinite alleles model (IAM) where every new mutation is novel. We therefore evaluated the SMM plus three TPM models with 70%, 80% and 90% single-step mutations respectively and the default variance of the geometric distribution (0.30). For each of the mutational models, the heterozygosity of each locus expected under mutation-drift equilibrium given the observed number of alleles (H_{eq}) was determined using 10000 coalescent simulations. The proportion of loci for which H_e was greater than H_{eq} ($prop_{het-exc}$) was then quantified for all of the mutation models. To quantify consistency of the measure across mutation models, we calculated the repeatability of $prop_{het-exc}$ using the rptR package (Stoffel et al., 2017) in R with 1000 bootstraps while adjusting for the mutation model as a fixed effect. Although the relative pattern across species was very consistent across mutation models (repeatability = 0.81, CI = [0.71, 0.89]), absolute values of $prop_{het-exc}$ within species decreased with lower proportions of multistep mutations (means for the TPM70, 80, 90 and SMM were 0.63, 0.58, 0.49 and 0.27, respectively). Given our posterior estimates (Supplementary Figure 6) and in line with previous studies, we therefore based our subsequent analyses on $prop_{het-exc}$ from the intermediate TPM80 model.

Demographic models.

As a second route to inferring historical population declines, we contrasted two alternative demographic scenarios (Figure 2.6) using a coalescent-based approximate Bayesian computation (ABC) framework (Kingman, 1982; Beaumont et al., 2002; Tavaré et al., 1997; Pritchard et al., 1999). To address the hypothesis that commercial exploitation from the 18th to the beginning of the 20th century led to population bottlenecks, we first defined a bottleneck model, which incorporated a severe reduction in population size within strictly bound time priors reflecting the respective time period. This model also allowed us to capture realistic changes from the pre- to post-bottleneck effective population size as both priors were drawn independently from the same distribution. Therefore, the model incorporates not only the bottleneck, but also longer term declines or expansions within realistic bounds as described below. For comparison, we defined a model that did not contain a bottleneck but which was identical in all other respects, which we called the non-bottleneck model. This model still allowed the population size to vary over time within a defined set of priors and thus captures realistic longer term variation in population size, but it does not include a severe recent bottleneck due to human exploitation.

Genetic data under both models were simulated from broad enough prior distributions to fit all 30 species while keeping the priors as tightly bound as possible around plausible values. The bottleneck model was defined with seven different parameters (Figure 2.6A). The current effective population size N_e and the historical (i.e. pre-bottleneck) effective population size N_{ehist} were drawn from a log-normal distribution with $N_e \sim \text{lognorm}[\text{logmean} = 10.5, \text{logsd} = 1]$ and $N_{ehist} \sim \text{lognorm}[\text{logmean} = 10.5, \text{logsd} = 1]$. This concentrated sampling within plausible ranges that fitted most species (i.e. with effective population sizes ranging from thousands to tens of thousands of individuals) while also occasionally drawing samples in the hundreds of thousands to fit the few species with very large populations. The bottleneck effective population size N_{ebot} was drawn from a uniform distribution between 1 and 500 ($N_{ebot} \sim U[1, 500]$) while the bottleneck start and end times $t_{botstart}$ and $t + botend$ were drawn from uniform distributions ranging between ten and 70 ($t_{botstart} \sim U[10, 70]$) and one and 30 ($t_{botend} \sim U[1, 30]$) generations ago respectively. Hence, the bottleneck time priors encompassed the last four centuries for all species, as their estimated generation times vary between approximately 7 and 19 years (Supplementary Table 1). The microsatellite mutation rate μ was refined after initial exploration and drawn from a uniform prior with $\mu \sim U[10^{-5}, 10^{-4}]$ which lies within the range of current empirical estimates (Ellegren, 2004; Selkoe and Toonen, 2006). The mutation model was defined as a generalized stepwise mutation model with the geometric parameter GSM_{par} reflecting the proportion of multistep mutations,

uniformly distributed from $GSM_{par} \sim U[0, 0.3]$. The non-bottleneck model was defined with five parameters (Figure 2.6B). N_e , N_{ehist} , μ and GSM_{par} were specified with the same priors as previously defined for the bottleneck model and the time parameter corresponding to the historical population size t_{hist} was drawn from a uniform distribution ranging between 10 and 70 generations ago ($t_{hist} \sim U[10, 70]$). All population size changes are therefore modeled as instantaneous changes at times $t_{botstart}$, t_{botend} or t_{hist} .

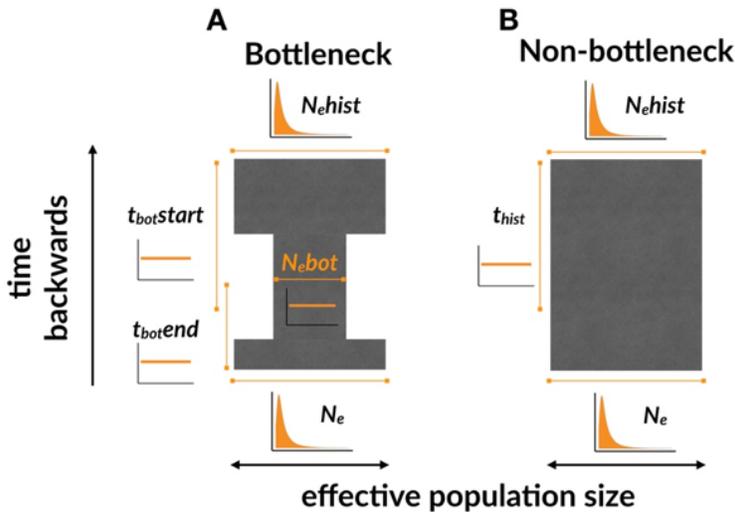


Figure 2.6: Schematic representation of two contrasting demographic scenarios and the parameter priors defining the models. All priors were drawn independently from each other, so the current N_e can be smaller or larger than N_{ehist} for a given species. This allowed both models to capture pre- to post-bottleneck variation in population size. While N_e and N_{ehist} were drawn from lognormal priors, all other parameters were specified using uniform priors. All prior distributions are also shown as small figures next to the respective parameter. The exact priors and the mutation model are given in the Methods.

ABC analysis.

ABC analysis. We simulated a total of 2×10^7 datasets of 40 individuals and 10 microsatellite loci each under the two demographic scenarios using the `fastsimcoal` function in `strataG` (Archer et al., 2016) as an R interface to `fastsimcoal2` (Excoffier et al., 2013), a continuous-time coalescent simulator. For both the simulated and empirical data, we used five different summary statistics for the ABC inference, all calculated as the mean across loci. Allelic richness (number of alleles), allelic size range, expected heterozygosity (i.e. Nei's gene diversity Nei

(1973)), the M-ratio (Garza and Williamson, 2001) and the proportion of low frequency alleles (i.e. with frequencies $< 5\%$). The summary statistics for the empirical datasets were computed by repeatedly re-sampling 40 individuals with replacement from the full datasets and calculating the mean across 1000 subsamples (for the Ladoga ringed seal and the Baltic ringed seal which had sample sizes smaller than 40, the full datasets were taken). As a small number of loci in the empirical data exhibited slight deviations from constant repeat patterns (i.e. not all of the alleles within a locus conformed to a perfect two, three or four bp periodicity), we calculated the M-ratio as an approximation using the most common repeat pattern of a locus to calculate the range of the allele size r and subsequently the M-ratio with $M = k/(r + 1)$ where k is the number of alleles. All statistics were calculated using a combination of functions from the strataG package and self-written code. For the ABC analysis, we used a tolerance threshold of 5×10^{-4} , thereby retaining 5000 simulations with summary statistics closest to those of each empirical dataset. For estimating the posterior probability for each scenario and each species, we used the multinomial regression method (Beaumont et al., 2002; Fagundes et al., 2007) as implemented in the function `postpr` in the `abc` package (Csilléry et al., 2012) where the model indicator is the response variable of a polychotomous regression and the accepted summary statistics are the predictors. To construct posterior distributions from the accepted summary statistics for the model parameters, we used a local linear regression approach (Beaumont et al., 2002) implemented in the `abc` function of the `abc` package.

Evaluation of model specification and model fit via cross-validation.

We evaluated whether ABC could distinguish between the two models by performing a leave-one-out cross validation implemented by the `cv4postpr` function of the `abc` package. Here, the summary statistics of one of the existing 10^7 simulations were considered as pseudo-observed data and classified into either the bottleneck or the non-bottleneck model using all of the remaining simulations. If the summary statistics are able to discriminate between the models, a large posterior probability should be assigned to the model that generated the pseudo-observed dataset. This was repeated 100 times and the resulting posterior probabilities for a given model were averaged to derive the rate of misclassification. We furthermore used a hypothesis test based on the prior predictive distribution (Lemaire et al., 2016) implemented in the `gfit` function in the `abc` package to check for each species that the preferred model provided a good fit to the observed data. Specifically, we used the median distance between the accepted and observed summary statistics as a test statistic, whereby the null distribution was generated using summary statistics from the pseudo-observed datasets. Hence, a non-significant p -value indicates that the

distance between the observed summary statistics and the accepted summary statistics is not larger than the expectation based on pseudo-observed data sets, i.e. the assigned model provides a good fit to the observed data.

Evaluation of the accuracy of parameter estimates via cross-validation.

In order to determine which parameters (i.e. population sizes, times and mutation rates and models) could be reliably estimated, we used leave-one-out cross validation implemented in the `cv4abc` function from the `abc` package to determine the accuracy of our ABC parameter estimates. For a randomly selected pseudo-observed dataset, parameters were estimated via ABC based on the remaining simulations using the rejection algorithm and a prediction error was calculated. This is possible because we know the ‘true’ parameter values from which a given pseudo-observed dataset was simulated. This procedure was repeated 1000 times and a mean prediction error ranging between 0 and 1 was calculated, where 0 reflects perfect estimation and 1 means that the posterior estimate does not contain any information about the true parameter value (Csilléry et al., 2012).

Posterior predictive checks.

To further confirm the fit of the preferred models, we conducted posterior predictive checks (Csilléry et al., 2010; Gelman et al., 1995) for each species. First, we estimated the posterior distribution of each parameter using ABC. Second, we sampled 1000 multivariate parameters from their respective posterior distributions and used those to simulate summary statistics *a posteriori* based on the preferred model. Last, we plotted those summary statistics as histograms and superimposed the observed summary statistics across all species (Csilléry et al., 2010).

Bayesian phylogenetic mixed models.

Finally, we used Bayesian phylogenetic mixed models in `MCMCglmm` (Hadfield, 2010) to evaluate the ecological and life-history variables affecting bottleneck strength and genetic diversity, and to test whether bottleneck history and genetic diversity are predictive of contemporary conservation status. Details of all the models are given in the Supplementary Table 10. All of the response variables were modelled with Gaussian distributions, while the predictors were fitted as fixed effects and the phylogenetic covariance matrix as a random effect. Predictors in models containing binary fixed effects were standardised by two standard deviations to allow a direct comparison between the effect sizes (Gelman, 2008; Schielzeth, 2010). In models

without binary fixed effects, the predictor variables were standardised by one standard deviation. For all models, we report the marginal R^2 as in (Nakagawa and Schielzeth, 2013). Some of the predictors in our models were correlated and multicollinearity might lead to suppression effects and make the interpretation of regression coefficients difficult (Ray-Mukherjee et al., 2014). We therefore reported standardized β estimates, structure coefficients, $r(\hat{Y}, x)$ and unique R^2 values for all variables in all models. The structure coefficients represent the correlation between a predictor and the fitted response of a model independent of the other predictors, and therefore reflect the direct contribution of a variable to that model. On the other hand, the unique R^2 is the difference between the marginal R^2 of a model including and a model excluding a predictor, which will be small when another predictor explains much of the same variation in the response (Ray-Mukherjee et al., 2014). All model estimates were presented as the posterior median and 95% credible intervals (CIs). We used uninformative priors with a belief (shape) parameter $v = 1$ for the variance-covariance matrices of the random effects and inverse-Wishart priors with $v = 0.002$ for residual variances. For each model, three independent MCMC chains were run for 110,000 iterations, with a burn-in of 10,000 iterations and a thinning interval of 100 iterations. Convergence was checked visually and by applying the Gelman-Rubin criterion to three independent chains. All of the upper 95% confidence limits of the potential scale inflation factors were below 1.05.

Data and code availability.

All data wrangling steps and statistical analyses except for the heterozygosity-excess tests (Piry et al., 1999) were implemented in R (R Core Team, 2015). The documented analysis pipeline along with the raw data can be accessed via GitHub (https://github.com/mastoffel/pinniped_bottlenecks) and is fully reproducible.

Acknowledgements

This research was supported by standard grants (HO 5122/3-1 and HO 5122/5-1) from the German Research Foundation (DFG) and as part of the SFB TRR 212 (NC³) together with a dual PhD studentship from Liverpool John Moores University. We are grateful to John Arnould, David Coltman, Corey Davis, Larissa Rosa de Oliveira, Tom Gelatt and NOAA, Melissa Gladstone, Roger Kirkwood, Melanie Lancaster, Tim Malloy, Rolf Ream, Garry Sten-son and Rob Stewart for providing access to published microsatellite datasets. We also thank

Matthias Galipaud for advice on statistical analysis, Kevin Arbuckle for advice on Bayesian phylogenetic mixed models and Luke Eberhart-Phillips for helpful comments on the figures. We are very grateful to Rebecca Carter (www.rebeccacarterart.co.uk) for the time and effort she dedicated to producing the pinniped illustrations. Finally, we want to thank two anonymous referees as well as the Okinawa Institute of Science and Technology Preprint Journal Club for their very helpful comments on the manuscript.

Supplementary Information

Demographic histories and genetic diversity across pinnipeds are shaped by human exploitation, ecology and life-history

Martin A Stoffel, Emily Humble, Anneke J Pajjmans, Karina Acevedo-Whitehouse, Barbara L Chilvers, Bobette Dickerson, Filippo Galimberti, Neil Gemmill, Simon D Goldsworthy, Hazel J Nichols, Oliver Krüger, S Negro, Amy Osborne, Teresa Pastor, Bruce C Robertson, Simona Sanvito, Jennifer Schultz, Aaron BA Shafer, Jochen BW Wolf & Joseph I Hoffman

Nature Communications, in press. bioRxiv doi: <https://doi.org/10.1101/293894>

This Supplementary Information can soon be found **online** at <https://www.nature.com/ncomms/> or in the meantime at <https://doi.org/10.1101/293894>



An dominant Northern elephant seal bull has spotted a potential competitor.

3

Elephant seal demography

Genomics reveals the demographic history of the Northern elephant seal

Martin A Stoffel, Kanchon K Dasmahapatra, Emily Humble, Stephen Gaughran, Frances Gulland, David Vendrami, Hazel Nichols & Joseph I Hoffman

in preparation

Abstract

Understanding the demographic histories of species is key to unravelling the patterns shaping genetic and phenotypic diversity and for the management of endangered species. While researchers have combined simulations and genetic data to explore simple demographic scenarios for decades, the genomic revolution has now given us the tools to infer far more complex demographic histories for virtually any species. However, as most genetic diversity is shaped by long-term demographic processes, the possibility of using genomic data to infer recent population size changes such as bottlenecks has rarely been studied, although recent demography is arguably more important for population viability and conservation. One of the most iconic examples of a recent population bottleneck is the case of the Northern elephant seal, which was nearly entirely wiped out by the 19th century sealing industry, but then underwent an unparalleled recovery to over 200,000 individuals nowadays. However, all knowledge about the species' recent history comes from either historical sealing records or from genetic studies using only handfuls of genetic markers. Moreover, close to nothing is known about the Northern elephant seal's history prior to the 19th century. Here, we used a newly sequenced Northern elephant seal genome and restriction-site associated DNA (RAD) sequencing of 80 individuals together with extensive coalescent simulations based on the site frequency spectrum to reveal the details of the species' demographic history. We show an extreme distortion of the site frequency spectrum which is likely a consequence of a severe genetic bottleneck with an effective population size of a few individuals and a duration of more than five generations. Moreover, the elephant seal exhibits an excess of low-frequency alleles, which is indicative of a strong post-glacial expansion preceding the bottleneck. Our results provide a detailed explanation for the near absence of genetic variation previously found with a variety of genetic markers and give the first genomic insights into the species' pre-sealing demography. Our results are among the first to show that genomic data can be used to shed light on recent bottlenecks, which has important implications for demographic inferences on contemporary timescales.

Introduction

A thorough understanding of a species' demographic history is necessary to distinguish demographic from selective effects across the genome (Nielsen et al., 2007) and eventually to understand the evolutionary forces shaping genetic and phenotypic diversity (Mitchell-Olds et al., 2007; Nielsen et al., 2009). As the detailed demographic histories of a species are seldom known, a variety of methods have been developed to infer demography from genomic diversity (e.g. Excoffier et al. 2013; Gutenkunst et al. 2010; MacLeod et al. 2013). While traditionally restricted to human populations where high density genomic data was available, the decreasing costs of next-generation sequencing have now made it possible to understand the demographic history of virtually any organism. For example, genomic data have been used to infer the the origins of ecological speciation in killer whales (Foote et al., 2016), post-ice age expansion of the king penguin (Trucchi et al., 2014), historical population sizes of flycatchers (Kardos et al., 2017) and the geographical origin of the plains zebra (Pedersen et al., 2018). However, most studies to date infer demography on evolutionary timescales, while recent events are arguably more important for a species' viability and adaptive potential. In particular, human impacts and climate change are causing global population declines on an unparalleled scale (Ceballos et al., 2017), but it nevertheless remains unclear whether genomic data can be used to infer recent declines or population bottlenecks Nunziata and Weisrock (2018).

The generation of vast amounts of genomic data has facilitated the development of a range of methods for inferring demographic histories. Among these, some of the most flexible and powerful methods use either coalescent simulations (Excoffier et al., 2013; Kingman, 1982) or diffusion approximation (Gutenkunst et al., 2010) and are based on a comparison between simulated and observed site frequency spectra (SFS). The SFS is the distribution of allele frequencies of a given set of loci (such as SNPs) in a sample (Evans et al., 2007*b*; Fisher, 1931) and has been termed the most fundamental yet unappreciated feature of genomic data (Salmona et al., 2017). Indeed, demographic inference based on the SFS has several advantages. For example, a reliable estimate of the SFS can be gained from a genomic sample containing a low number of polymorphisms (Shafer et al., 2015*a*) and is reliable even when inferred from a small subset of the genome (Excoffier et al., 2013) or when based on low coverage data (Korneliussen et al., 2014). Moreover, recent methodological developments such as the composite-likelihood approach implemented in fastsimcoal2 Excoffier et al. (2013) facilitate the reliable estimation of even very complex demographic models based on the SFS.

Very few species' demography's have been of greater interest than the history of the Northern elephant seal (NES). The species was severely hunted for its blubber by 19th century sealers and was assumed to be extinct for decades. However, some individuals survived and within just a century the population rebounded from presumably only a handful of animals to over 200,000 individuals in the north-east Pacific nowadays (Lowry, 2014). Not many species have been driven so close to extinction and shown such an unparalleled and exponential recovery. This is why the NES has become the focus of many studies trying to elucidate the strength of the genetic bottleneck and its consequences, particularly the potential loss of genetic variation. The evidence so far is unambiguous whereby genetic diversity has been shown to be low at mtDNA (Hoelzel et al., 1993), allozymes (Bonnell and Selander, 1974; Hoelzel et al., 1993) MHC (Weber et al., 2004) and microsatellites (Stoffel et al., 2018). Furthermore, a recent study comparing 30 pinniped species revealed that the NES exhibits some of the lowest genetic diversity of the whole clade, and that the species likely underwent an extremely severe genetic bottleneck (Stoffel et al., 2018). However, all studies to date have been conducted based on just a handful of genetic markers, but genomic data is entirely lacking. Critically, this has prevented insights into more detailed and complex aspects of the species' demography. For example, how large were the effective population sizes during and before the bottleneck and how long did the bottleneck last?

Moreover, due to decades of intensive slaughter, very little is known about NES abundance or distribution prior to the mid-nineteenth century (Le Boeuf and Laws, 1994). Curiously, the only source of knowledge are archaeological studies reporting a dearth of elephant seal remains relative to other pinnipeds of the north-east Pacific (Jones et al., 2002; Lyman, 2011; Rick et al., 2011). A testable hypothesis could be that the Northern elephant seal lacked appropriate breeding habitat due to glaciation during the last glacial maximum (LGM), but subsequently expanded with the retreat of the ice cover until it reached its pre-sealing population size. Such post-glacial expansions have indeed been commonly reported in species of the Nearctic realm (Burbrink et al., 2016), while testing these historical demographic changes was not possible for the Northern elephant seal with the genetic data available so far.

Here, we developed a novel genomic toolkit for the Northern elephant seal, which includes a high-quality genome assembly developed from 10x chromium sequencing and large-scale restriction-site association DNA sequencing (RAD) data for 80 individuals. Using this toolkit, we combine extensive coalescent simulations and a composite maximum likelihood method to shed new light on both the recent and ancient demography of the Northern elephant seal. In particular, we hypothesized that (1) the severe genetic depletion found in the Northern elephant seal is a result of an extreme and long-lasting bottleneck due to commercial overexploitation

and (2) that the elephant seal, similar to many other species in the Nearctic, experienced a strong population expansion after the LGM.

Methods

Genome sequencing and assembly

To generate the Northern elephant seal genome, a muscle sample was collected from a male Northern elephant seal that had stranded in Mendocino, California, USA (Lat: 3900938, Long: -123.695275), on 4 April 2015. High molecular weight DNA was extracted from the muscle using a Qiagen MagAttract HMW DNA Kit, and large fragment size was confirmed through gel electrophoresis. The DNA was prepared for sequencing using the 10X Genomics Chromium Genome Kit, which isolates DNA molecules and attaches identical barcodes to fragments derived from the same DNA molecule. Sequencing was then performed on two lanes of an Illumina HiSeq X, using paired-end sequencing of 150bp reads. The Northern elephant seal genome was assembled *de novo* using the 10X Genomics Supernova (v2) software. This software uses the molecule-specific barcodes on each sequencing read to infer the proximity of two given reads on the same original DNA molecule. This method of linked-read sequencing allows for the high-confidence assembly of large genomic scaffolds from short read data. The Supernova *de novo* assembly produced a genome of approximately 2.25 Gb, slightly smaller than the current assembly for the closely related Hawaiian monk seal (240 Gb). The median effective coverage for the assembly was $\sim 45\times$. The size of the scaffold N50 was 1.06 Mb.

Sample collection for RAD sequencing

Tissue samples for 80 Northern elephant seals were generously provided by the Marine Mammal Center (MMC) in Sausalito, California. These samples were collected from both male and female pups between March 2006 and March 2012. Samples were stored in 90% ethanol and at -20°C . Genomic DNA was then extracted using a standard phenol-chloroform protocol.

RAD sequencing

The preparation of RAD libraries followed the protocol used in (Etter et al., 2011) with modifications as described in (Humble et al., 2018). After quality control using FastQC v0.112 all reads were trimmed to 225 bp and demultiplexed using `process_radtags` in STACKS v2.0

(Catchen et al., 2013). The reads were then mapped to the newly sequenced Northern elephant seal genome with BWA MEM v0.7.10 (Li, 2013) using the default parameters. Reads which could not be mapped to the genome were removed using SAMtools v1.1 (Li, 2011). Subsequently, we used Picard Tools to sort the SAM files, add read groups and remove PCR duplicates. As a last step in the bioinformatics pipeline, we performed indel realignment to minimize errors due to mismatching bases using RealignerTargetCreator and IndelRealigner in GATK 3.6.

Genotype likelihood inference and site frequency spectrum

Based on the bam files from GATK, we calculated genotype likelihoods using ANGSD v0.918 (Korneliussen et al., 2014). Within ANGSD, we used several filters: (1) we kept only sites present in all 80 individuals, (2) a minimum depth of coverage at a site of 240 across all individuals, (3) a maximum depth of coverage at a site of 1600 across all individuals, (4) minimum mapping quality of 20 (5) a minimum base quality of 30, and (6) a polymorphism p -value threshold of 10^{-4} . We then used the obtained genotype likelihoods as input for the realSFS method in ANGSD to estimate the site frequency spectrum (SFS). As we did not have ancestral state information, we calculated the folded SFS.

Demographic simulations

As it is unclear how recent demographic changes affect the SFS, we used coalescent simulations in fastsimcoal2 (Excoffier et al., 2013) to simulate expected SFS under two different demographic models: (1) a model including only a recent bottleneck and (2) a model including recent bottleneck and a population expansion after the last glacial maximum (LGM) (see Figure 3.1). For both models, we simulated SFS for varying bottleneck durations, ranging from one to ten generations and for different N_eBot ranging from five to 50 individuals. As we were mainly interested in the shape of the simulated SFS in relation to the empirical SFS rather than the overall rate of polymorphism across the genome, we assumed the following parameters to be fixed: (a) current effective population size $N_eCur = 10,000$, (b) pre-bottleneck effective population size $N_eHist = 20,000$ (c) mutation rate = 2.5×10^{-8} . For the ice age expansion model, we fixed (d) the start of post-glacial expansion at 1,500 generations in the past, and (e) the effective population size during the last glacial maximum N_eLGM at 2000 individuals.

Demographic modeling

Subsequently, we used the composite maximum likelihood method in *fastsimcoal2* for demographic modeling. In a nutshell, the program simulates SFS under a given demographic model and optimizes the parameters of the model (such as population sizes or timings of historical events) to reach the highest possible agreement between the simulated and the observed SFS. The results of the program include both the estimated model parameters and a likelihood for the model itself which can be used to compare different demographic scenarios (Excoffier et al., 2013). To select a best model according to our main hypotheses, we compared the likelihood of three different demographic models which are depicted in Figure 3.1. First, we defined a model including a recent bottleneck, which broadly spans the peak of commercial exploitation of elephant seals in the 19th century. Given a generation time of 8.7 years (IUCN, 2018), the bottleneck was hence fixed between 20 and 30 generations ago. In addition, we specified initial search ranges for the current effective population size N_eCur , the pre-bottleneck effective population size N_eHist and the bottleneck effective population size N_eBot (see Figure 3.1A for the detailed parameter ranges). Second, we defined a model including both a recent bottleneck with the same parameters and in addition a population expansion after the last glacial maximum (LGM). Therefore, we specified two additional parameters, the start of the expansion after the LGM and the effective population size during the LGM, termed N_eLGM (see Figure 3.1B for the detailed parameter ranges). Third, we specified a neutral model which neither included a recent bottleneck nor an ice-age expansion as a control scenario. We defined the neutral model based on two parameters, N_eHist and N_eCur which were specified with the same ranges as in the other models (see Figure 3.1C for the detailed parameter ranges). For all models, we used a fixed mutation rate of 2.5×10^{-8} which is consistent with marine mammal mutation rate estimates from the literature (Dornburg et al., 2012). Importantly, all parameters are defined as initial search ranges, and the final parameter estimates can exceed the upper ranges (Excoffier et al., 2013).

Model selection

For all models, we used a total of 50 replicate runs of the maximum likelihood optimization, each including 40 estimation loops with 100,000 coalescent simulations, according to the recommendation of the *fastsimcoal2* authors (Excoffier et al., 2013). Of the 50 replicate runs for each model, the run with the highest maximum likelihood was used for selecting the best of the three models. To do so, we calculated Akaike's information criterion (AIC) with the following formula:

$$AIC = 2 * \ln(\text{likelihood}) + 2 * K \quad (\text{Eq. 71})$$

Where \ln is the natural logarithm, likelihood is the maximum likelihood of the model and K equals the number of estimated parameters in the model.

Parameter estimation and uncertainty

After determining the best fitting model, we used non-parametric bootstrapping to estimate the model parameters and quantify their uncertainty. Specifically, we used ANGSD to generate 100 bootstrapped SFS by subsampling loci with replacement from the SFS. We then used fastsimcoal2 to re-estimate parameters for the best model based on 10 replicate runs for each of the 100 bootstrapped SFS. For each SFS, the two runs with the highest maximum likelihood were retained. Finally, we quantified the median of the distribution and its 2.5th and 97.5th percentile as a 95% confidence interval.

Results

Genomic data reveal an extreme site frequency spectrum

Paired-end 250 bp RAD data mapped to the new elephant seal reference genome resulted in an average of 2,361,524 mapped reads per individual (range = 1,447,403 - 4,716,551). Genotype likelihood inference using ANGSD yielded information for 22,394,982 sites present in all 80 individuals, with only 0.1% variable positions. After applying a polymorphism p-value threshold in addition to several other filters (see Methods) in ANGSD, we retained 13,080 high quality polymorphic positions present in all 80 individuals which were subsequently used to infer the observed site frequency spectrum (SFS, Figure 3.2). The SFS of a population with a stable sample size is expected to show a decline in the number of SNPs with increasing frequencies (Hartl and Clark, 1997). However, the SFS of the Northern elephant seal shows a pattern which differs drastically from such an expectation. SNPs with only one individual that carries the alternative allele (singletons) are the most frequent form of polymorphism in our sample, followed by doubletons. All other frequencies, however, are nearly equally abundant. Moreover, the SFS reveals an excess of rare variants, so-called singletons and doubletons.

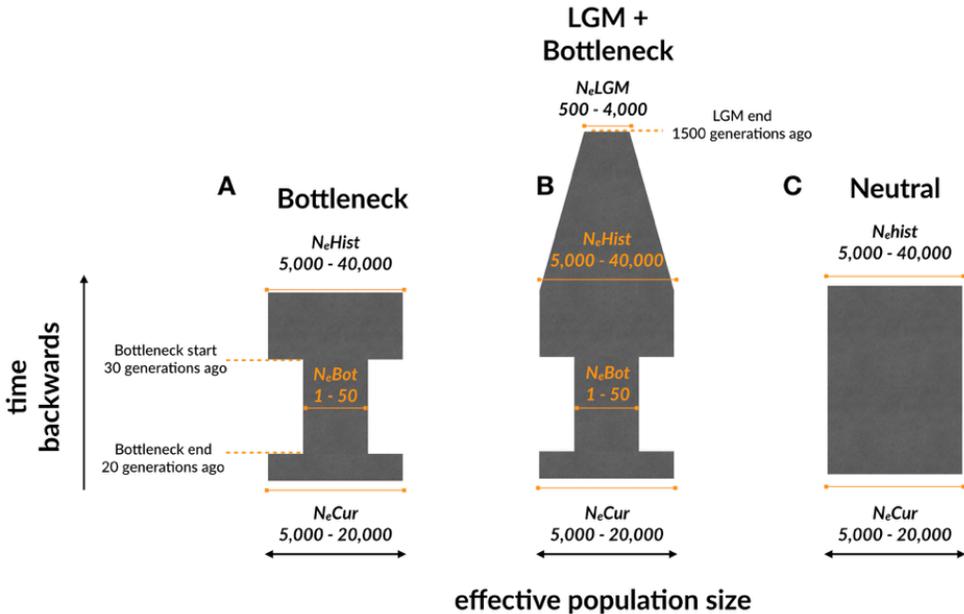


Figure 3.1: Schematic representation of the three demographic scenarios. For all models, the mutation rate was fixed to 2.5×10^{-8} . Furthermore, all time points were fixed, with the recent bottleneck ranging from 30 to 20 generations ago (see Panel A), and the last glacial maximum (LGM) ending 1,500 generations ago (see Panel B). All N_e 's define initial search ranges for the maximum composite likelihood method in fastsimcoal2. Consequently, the program is able to estimate parameter values which exceed these initial search ranges, as they only define the starting point of the optimization.

SFS simulations hint towards an ice-age expansion followed by an extreme recent bottleneck

We conducted a series of coalescent simulation in fastsimcoal2 to explore the parameters which potentially result in an SFS as extreme as in the Northern elephant seal. First of all, we simulated SFS under a recent bottleneck model. We varied both the bottleneck effective population size (N_{eBot}) and the duration of the bottleneck (T_{Bot}) in generations, while keeping all other parameters constant (see Methods). The resulting SFS based on all 16 combinations of N_{eBot} and T_{Bot} are shown in Figure 3.3. Most importantly, we looked for a pattern similar to the observed SFS, where all SNPs except for singletons and doubletons show the same abundance. Based on Figure 3.3, only the most extreme bottlenecks (right upper corner in Figure 3.3), where the population is reduced to as little as $N_{eBot} = 5$ or 10 and the bottleneck spans at least 5 or 10 generations ($T_{Bot} = 5$ or 10) can disturb the SFS strongly enough to

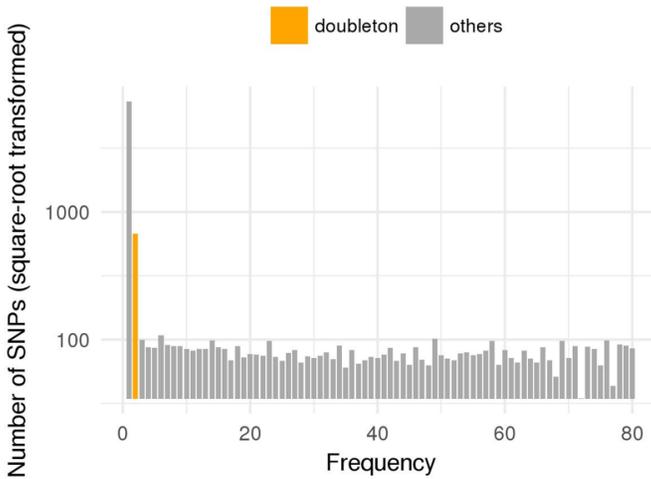


Figure 3.2: Observed site-frequency spectrum of the Northern elephant seal estimated from genotype-likelihoods.

show a pattern as flat as the observed SFS. However, an intricate detail does not yet fit. Doubletons, marked in orange across all plots, are much more abundant than SNPs with higher alternative allele frequencies in the observed SFS (Figure 3.2). We hypothesized that this could be the result of a population expansion in the past which caused an excess of rare alleles such as singletons and doubletons, a pattern previously observed in human populations (Coventry et al., 2010; Keinan and Clark, 2012).

Consequently, we repeated the simulations based on a more complex model, which included a recent bottleneck and in addition a population expansion starting from the end of the last glacial maximum (LGM, Figure 3A). Here, we see a much larger proportion of doubletons relative to higher frequency variants (right upper corner), a pattern very similar to the observed SFS. However, the flat tail of the SFS can still only be observed when the expansion is followed by an extreme recent bottleneck with N_eBot at around 5 or 10 individuals and a duration $TBot$ longer than five generations.

The best model includes a post-LGM expansion and a severe bottleneck

Based on the simulations, we used the composite maximum likelihood method in *fastsimcoal2* for a formal comparison of three alternative models, (1) a recent bottleneck model, (2) a model

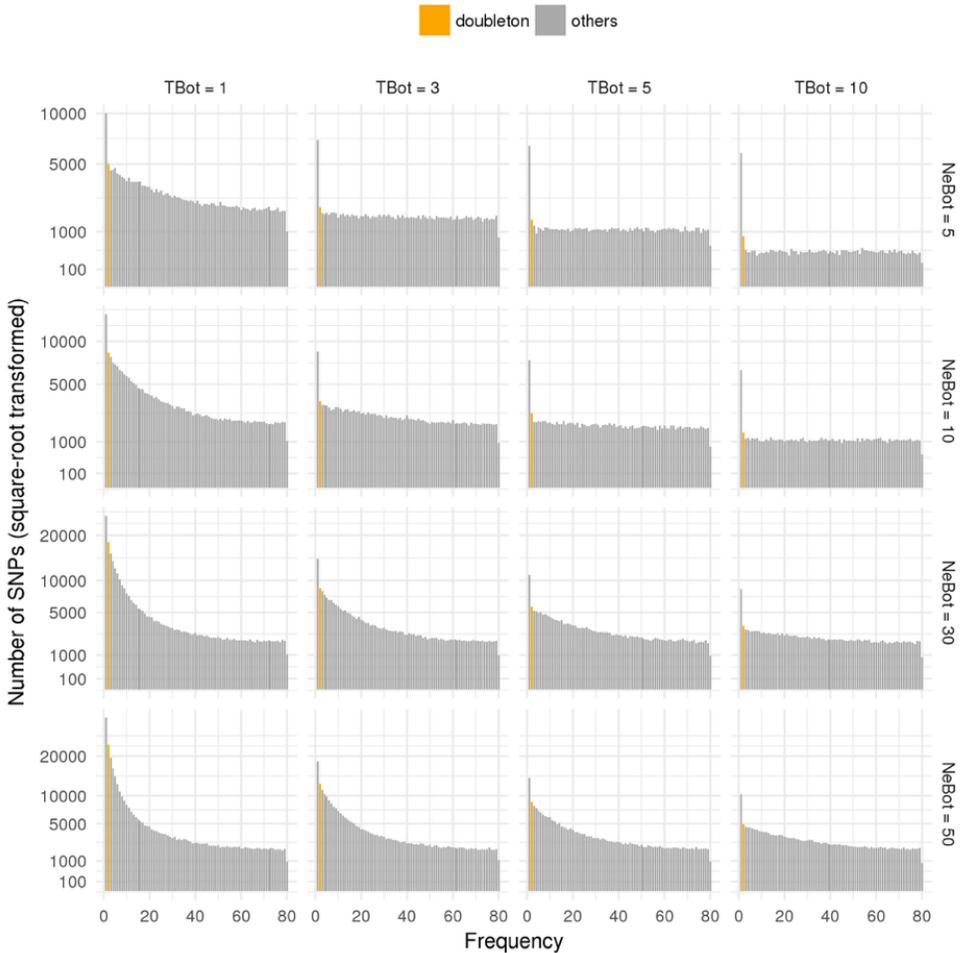


Figure 3.3: Simulated site-frequency spectra based on a model which only includes a recent bottleneck. The plot shows SFS for different bottleneck intensities. N_eBot specifies the simulated effective population size during the bottleneck and $TBot$ specifies the duration of the bottleneck in generations.

including a recent bottleneck and an expansion after the last glacial maximum (LGM) and (3) a neutral model (see Methods for details). As expected, both scenarios which included a recent bottlenecked showed a higher likelihood compared to the neutral model, as reflected in a much lower AIC value (see Table 3.1). Moreover, as indicated by the initial simulations (Figure 34), the model including a post-LGM expansion and a severe recent bottleneck showed a higher likelihood than the model only incorporating a recent bottleneck, which is also reflected in a substantially lower AIC (Table 3.1). Consequently, the Northern elephant seal SFS is likely a

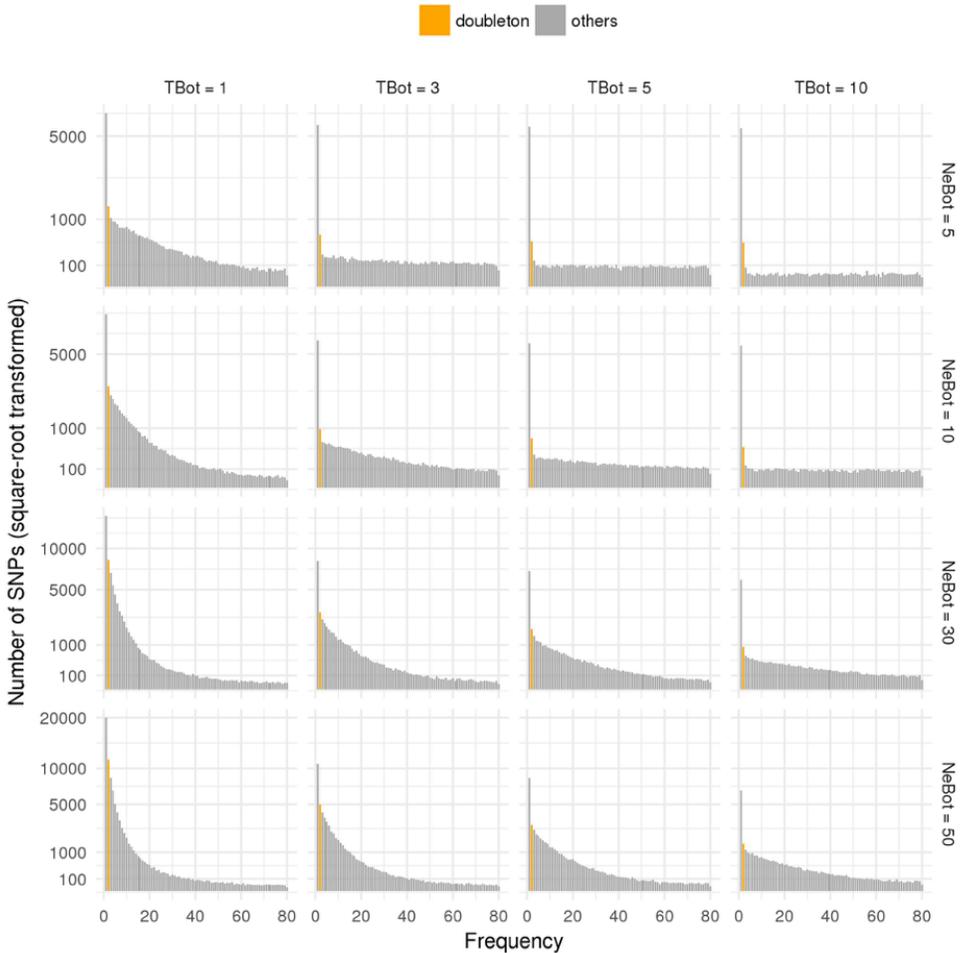


Figure 3.4: Simulated site-frequency spectra based on a model incorporating an ice age expansion and a recent bottleneck. The plot shows SFS for different bottleneck intensities. N_eBot specifies the simulated effective population size during the bottleneck and $TBot$ specifies the duration of the bottleneck in generations.

consequence of both a post-LGM expansion and a recent bottleneck.

Parameter inference reveals an N_e at the very brink of extinction

We then used non-parametric bootstrapping of the SFS to estimate the underlying model parameters of the LGM + Bottleneck model and their uncertainties Figure 3.5. Consistent with

Table 3.1: Likelihood and AIC values based on the composite maximum likelihood estimation in fastsimcoal2 for all three demographic models.

Model	Max(log ₁₀ (likelihood))*	No. of parameters	AIC
LGM & Bottleneck	-10134.46	4	46678.91
Bottleneck	-10155.12	3	46772.06
Neutral	-10565.83	2	48661.45

* Based on the best likelihood among the 50 independent runs

knowledge about the hunting history of the NES and previous genetic studies, the bottleneck was estimated extremely severely (median $N_{eBot} = 2$, 95% CI [1.5, 2]). The estimate for the current effective size had a considerable higher uncertainty (median $N_{eCur} = 4582.5$, 95% CI [3640.5, 10402]) and was estimated substantially lower than the pre-bottleneck size (median $N_{eHist} = 8041$, 95% CI = [5277, 15168.5]). Lastly, the effective size during the last glacial period was estimated to be around 10 to 20 times lower than the pre-sealing population size (median $N_{eLGM} = 355$, 95% CI [2675, 851]).

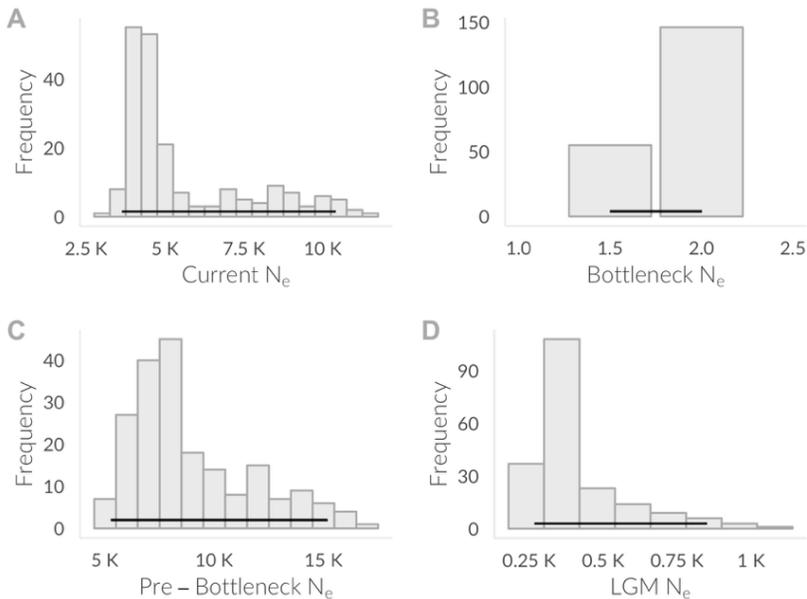


Figure 3.5: Parameter estimates from non-parametric bootstrapping of the SFS. Shown is the distribution of 200 non-parametric bootstrap estimates using the composite-maximum likelihood method in fastsimcoal2 including 95% confidence intervals indicated as black horizontal lines for (A) Current effective population size (B) Bottleneck effective population size (C) Pre-bottleneck effective population size and (D) Last glacial maximum effective population size. All estimates are presented as diploid N_e 's.

Discussion

Here, we developed a novel genomic toolkit which we used in combination with extensive coalescent simulations based on the SFS to explore the demographic history of the NES. We showed that the observed pattern of genomic variation in the NES can only be explained by a severe and long lasting recent bottleneck with an N_e of as little as two individuals and a duration of 10 generations.

There are two important implications of this result. First, from a methodological point of view, the applicability of genomic demographic inference over contemporary timescales is still questioned (Nunziata and Weisrock, 2018), as it is unclear to which degree declines in census population sizes are immediately followed by changes in N_e which are reflected by genomic variation. A recent simulation study showed that genomic data such as RADseq can indeed lead to reliable demographic inference for population size changes as close as 10 generations in the past (Nunziata and Weisrock, 2018). In line with these simulations, we showed that the NES experienced a drastic decline in N_e only some tens of generations ago, which can be reliably inferred from the SFS.

Second, our results shed novel light on a large body of research on NES demography by providing reliable estimates of both bottleneck intensity and duration. Based on the sealing literature, it was guessed that only 10 or 20 seals survived the commercial exploitation towards the end of the 19th century (Hoelzel, 1999). In particular, over the course of half a decade, from around 1880 to 1922, only a handful of seals have been observed and subsequently killed, although frequent expeditions set out for extensive searches (Hoelzel, 1999). Consequently, our estimates for both bottleneck N_e of less than five individuals and a bottleneck duration of around ten generations fit to the expectations based on sealing records. Moreover, severe depletions of genetic diversity in the NES have been reported based on nearly all types of genetic markers (Bonnell and Selander, 1974; Hoelzel et al., 1993; Hoelzel, 1999; Stoffel et al., 2018; Weber et al., 2004) and a reduction in diversity was also shown using a comparison of pre- to post bottleneck samples (Weber et al., 2000). From a theoretical point of view, a severe loss of diversity as observed in the NES is only expected from strong N_e declines which are not directly followed by a recovery (Nei et al., 1975). This supports our inference of a severe and long-lasting recent genetic bottleneck in the NES.

Despite a close fit of the empirical SFS to recent bottleneck models, our initial simulations revealed an intricate detail which required further exploration: The excess of rare variants, so-called singletons and doubletons, relative to all other frequency classes. A pattern of excessive rare alleles is usually a consequence of recent and strong population growth (Coventry et al.,

2010; Keinan and Clark, 2012), but could conceivably also be caused by immigrants into the population which carry novel alleles. However, NES has very likely expanded from just a single breeding colony on Guadalupe island after the recent bottleneck (Le Boeuf and Laws, 1994), which is why a pre-bottleneck expansion seems the more likely explanation. Both coalescent simulations and the maximum likelihood maximization method showed that SFS patterns very similar to the observed SFS can be a consequence of a long-term post-glacial expansion of the species, starting from a population size around ten to twenty times as small as the current census size. We conclude that it is likely that the NES had a smaller population size during the LGM, and expanded with the retreat of glaciation in the Nearctic realm (Hewitt, 2000), a pattern which has previously been found in many species both across Europe (Hewitt, 1999) and America (Burbrink et al., 2016). The population size of the Northern elephant seal might have been small during the LGM for two main reasons. First, massive ice-sheets during the Quaternary glaciation can have simply decreased potential breeding and molting habitats for the elephant seal, preventing a larger population. Second, hauling out on sand and gravel beaches could have made the species vulnerable to both prehistoric Native American hunters and non-human predators such as grizzly bears and mountain lions, and potentially restrained their habitats to remote offshore islands (Erlandson et al., 1998). The latter explanation is also consistent with a more in-depth archaeological study of the NES which found that most of the species' Holocene abundance might have been scattered across remote islands, further away from human settlements and predation (Rick et al., 2011). Moreover, the authors also reported that most elephant seal specimens come from the late-Holocene (3500 yrs ago-present) which again either suggests a more recent population growth consistent with post-glacial expansion (Rick et al., 2011).

Limitations

Whilst we are providing strong evidence that genomic data has the power to estimate both recent and historical population size changes, there are still a number of important limitations. First of all, while RADseq can theoretically provide an unbiased estimate of the SFS, genotype calls might be inaccurate when the coverage or the allelic frequencies are low (Fountain et al., 2016; Han et al., 2014). To circumvent this problem, we avoided calling genotypes before estimating the site frequency spectrum and instead relied on genotype likelihoods, which incorporate uncertainty inherent to sequencing errors, coverage and alignment quality and result in an accurate SFS even with very low coverage data (Korneliussen et al., 2014). However, as lower

coverage can lead to strong biases in singletons (Han et al., 2014), we excluded this SFS class from the analysis. Moreover, the non-parametric bootstrapping results suggest that most model parameters could be estimated reliably.

Second, different demographic histories can lead to the same site frequency spectrum (Myers et al., 2008), a problem inherent to all genetic demographic inference, which is why a clear theoretical expectation of the underlying demography is important. In the case of the NES, the recent bottleneck is an unquestioned part of its recent history (Le Boeuf and Laws, 1994) and hence provides a theoretically sound model to test. While the post-glacial expansion is only one of several potential scenarios causing an excess of rare alleles in the SFS of the NES, such a scenario is probably the most parsimonious and well-grounded, given the large body of research which provide similar evidence on post-glacial expansions (Burbrink et al., 2016; Hewitt, 2000) and the archaeological evidence of small population sizes during the Holocene (Rick et al., 2011).

Finally, demographic modeling based on optimization procedures such as the maximum composite likelihood method used in this study, optimizes the fit of simulated and observed SFS without evaluating an overall goodness-of-fit of a model. Such a goodness-of-fit can for example be evaluated based on a visual comparison between the optimized simulated SFS and the observed SFS. Here, we show that the simulated SFS based on our own simulations and the optimized SFS resulting from the estimated model parameters are highly similar to the NES SFS, therefore providing visual evidence for a good fit of these models to our data.

Conclusions and future studies

The NES has experienced an extremely severe bottleneck, the genetic consequences of which have now been confirmed with a range of genetic and genomic data. However, the degree to which the bottleneck impacted neutral or functional genomic variation is still to be discovered with higher-density genomic data than used in the present study. This might shed new light on the mechanisms which allowed the NES to recover so rapidly despite the apparent genomic depletion. Moreover, deeper sequencing and higher coverage will make it possible to increase the reliability in which rare variants are detected and might uncover further details of the post-glacial demography which have only been unraveled on the surface in the present study.



An Antarctic fur seal pup waiting for its mother to return her feeding trip at sea. Photo: Oliver Krüger

4

Chemical fingerprints

Chemical fingerprints encode mother-offspring similarity, colony membership, relatedness, and genetic quality in fur seals

Martin A Stoffel, Barbara A. Caspers, Jaume Forcada, Athina Giannakara, Markus Baier, Luke Eberhart-Phillips, Caroline Müller & Joseph I Hoffman

Proceedings of the National Academy of Sciences. 112(36):E5005-E5012. 2015

Significance

Understanding olfactory communication in natural vertebrate populations requires knowledge of how genes and the environment influence highly complex individual chemical fingerprints. To understand how relevant information is chemically encoded and may feed into mother-offspring recognition, we therefore generated chemical and genetic data for Antarctic fur seal mother-pup pairs. We show that pups are chemically highly similar to their mothers, reflecting a combination of genetic and environmental influences. We also reveal associations between chemical fingerprints and both genetic quality and relatedness, the former correlating positively with substance diversity and the latter encoded mainly by a small subset of substances. Dissecting apart chemical fingerprints to reveal subsets of potential biological relevance has broad implications for understanding vertebrate chemical communication.

Abstract

Chemical communication underpins virtually all aspects of vertebrate social life, yet remains poorly understood because of its highly complex mechanistic basis. We therefore used chemical fingerprinting of skin swabs and genetic analysis to explore the chemical cues that may underlie mother-offspring recognition in colonially breeding Antarctic fur seals. By sampling mother-offspring pairs from two different colonies, using a variety of statistical approaches and genotyping a large panel of microsatellite loci, we show that colony membership, mother-offspring similarity, heterozygosity, and genetic relatedness are all chemically encoded. Moreover, chemical similarity between mothers and offspring reflects a combination of genetic and environmental influences, the former partly encoded by substances resembling known pheromones. Our findings reveal the diversity of information contained within chemical fingerprints and have implications for understanding mother-offspring communication, kin recognition, and mate choice.

Introduction

The chemical senses are the evolutionarily oldest and arguably most widespread means of interacting with the outside world. Olfaction in particular is fundamental to animal communication, mediating social interactions as varied as territorial behavior, kin recognition, and mate choice (Wyatt, 2014*b*). Metabolomic tools, such as gas chromatography-mass spectrometry (GC-MS) have made it possible to generate individual-specific chemical ‘fingerprints.’ By separating compounds and quantifying their relative abundances, these fingerprints provide a wealth of information, even though not all compounds can necessarily be identified. Both volatile and contact cues are potentially hidden within the extreme complexity of chemical profiles, which is why a mechanistic understanding of chemical communication is still lacking in natural vertebrate populations (Hurst and Beynon, 2010).

In particular, ‘surprisingly little progress’ has been made in understanding the link between vertebrate chemical fingerprints and genotype (Hurst and Beynon, 2010). Experimental studies have shown that females of several species are capable of discriminating potential partners based on olfactory cues (Wedekind and Füre, 1997; Radwan et al., 2008; Olsson et al., 2003). However, very few studies have demonstrated a convincing link between the molecular composition of chemical fingerprints and genetic traits, such as heterozygosity (a measure of genetic quality) and relatedness (Boulet et al., 2009; Charpentier et al., 2008; Crawford et al., 2010; Leclaire et al., 2012). These studies were almost exclusively conducted on a captive population of lemurs, a species known for its conspicuous use of scent marking.

A functional understanding of how genotype is chemically encoded also requires knowledge of how many and which types of substances are involved. This is challenging because, especially in natural populations, an individual’s mixture of surface chemicals is not only the product of its genotype but may also be mediated by hormones, the microbial flora, body condition, and environmental factors (Hurst and Beynon, 2010). Thus, analyses based on overall chemical fingerprints may overlook subtle genetic signatures and make little if any headway toward identifying the specific substances involved. A second less-appreciated problem is that the modest panels of around 10-15 microsatellite loci typical of most studies may be underpowered to detect genetic associations because they provide relatively imprecise estimates of both heterozygosity and relatedness (Balloux et al., 2004; Hoffman et al., 2004).

In arguably the only study to report a convincing link between chemical fingerprints and genotype in a natural vertebrate population, Leclaire et al. Leclaire et al. 2012 used principle component analysis (PCA) to reduce chemical complexity. The authors identified a principle component in kittiwakes that correlated significantly with heterozygosity in both sexes and

another that correlated with relatedness, but only in adult males. However, PCA iteratively maximizes the explained variance per component instead of seeking to capture the underlying structure and dimensionality of the data, which makes the resulting components hard to interpret (Fabrigar et al., 1999). A better approach could be factor analysis (FA), a method from the field of psychology that estimates the latent variable structure of a dataset by dividing the total variability into that common to variables and a residual value unique to each variable (Eid et al., 2010). Statistical developments that allow FA to be applied to data with more variables than observations (McFerrin, 2015) have only recently made this approach amenable to studying chemical fingerprints.

Pinnipeds are an important group of marine mammals that provide an unusual opportunity to reveal insights into the basis of chemical communication. Studies of Steller's sea lions and harbor seals have revealed a large repertoire of functional olfactory receptor genes (Kishida et al., 2007) and remarkably high olfactory sensitivity (Kowalewsky et al., 2006), respectively. Individuals of many pinniped species also have a strong musky smell that has been attributed to secretions of facial sebaceous and apocrine glands (Ling, 1972). These glands are known to hypertrophy during the mating season in at least two species (Hardy et al., 1991), suggesting that olfactory cues may be particularly important during the reproductive phase of the life cycle.

Females of many otariid species breed in dense colonies and alternate lactation ashore with foraging trips at sea, necessitating accurate mechanisms for offspring localization and recognition (Insley et al., 2003). Although otariids use a combination of geographical, visual, auditory, and olfactory cues to find and recognize their pups (Insley et al., 2003), olfactory recognition is particularly important because females of many species accept or reject pups based on naso-nasal inspection (Dobson and Jouventin, 2003; Phillips, 2003). Furthermore, a recent experiment on Australian sea lions (Pitcher et al., 2011b) suggests that female pinnipeds are capable of discriminating filial from nonfilial pups using olfaction in the absence of other cues.

Antarctic fur seals (*Arctocephalus gazella*) provide a highly tractable model system for studying the importance of chemical cues in a free-ranging marine mammal. On Bird Island, South Georgia (Southwest Atlantic), a colony of fur seals has been studied intensively for over two decades (Doidge et al., 1984). In this species, olfaction is known to be important for the close-range recognition of pups (Dobson and Jouventin, 2003). However, females also show active mate choice for males who are both heterozygous and unrelated to themselves (Hoffman et al., 2007b), raising the possibility that chemical cues might be involved not only in mother-offspring recognition, but also in mate choice.

Here, we combined GC-MS fingerprinting of skin swabs and genetic analysis to explore the

chemical basis by which Antarctic fur seal mothers may recognize their pups. Because females of this species appear capable of choosing males based on heterozygosity and relatedness, we hypothesized that genotype should be chemically encoded and that this could provide a mechanism by which females could identify their pups. We therefore sampled mother-offspring pairs from two discrete but genetically indistinguishable colonies (see ‘colony differences’ in Results), which in principle allows genetically encoded substances to be disentangled from those influenced by environmental differences between colonies. We also deployed over 40 microsatellite loci to enhance the power to detect associations between chemical fingerprints and genotype. Finally, we used FA together with a variety of nonparametric approaches to explore the structure of the chemical data and to uncover specific subsets of compounds associated with chemical differences between the colonies, mother-offspring similarity, and genetic relatedness.

Results

Chemical and Genetic Data.

Chemical fingerprints and multilocus microsatellite genotypes were obtained for 41 mother-offspring pairs from two breeding colonies at Bird Island, South Georgia (Figure 4.1). After removing compounds present in the control sample or only in a single individual, the total number of substances in each individual’s chemical fingerprint averaged 359 and did not differ significantly between mothers and offspring (paired t test, $t = -0.05$, $P = 0.96$). All of the animals were genotyped at 43 highly polymorphic microsatellite loci, 41 of which did not deviate significantly from Hardy-Weinberg equilibrium (HWE) in either mothers or offspring after table-wide false-discovery rate (FDR) correction, and were therefore retained for subsequent analyses (Table S1). The mother-offspring pairs all had match probabilities of 100% (Table S2).

Colony Differences.

Multivariate statistical analysis of the relative proportions of each substance revealed highly significant differences between animals sampled from the two colonies, both overall (Figure 4.2A) [analyses of similarities (ANOSIM), global $R = 0.57$, $P < 0.0001$] and separately for mothers (ANOSIM, global $R = 0.58$, $P < 0.0001$) and offspring (ANOSIM, global $R = 0.56$, $P < 0.0001$). Bayesian structure analyses of the genetic data yielded the highest average log-likelihood value for $K = 1$ in both mothers and pups (Figure S1), indicating a lack of

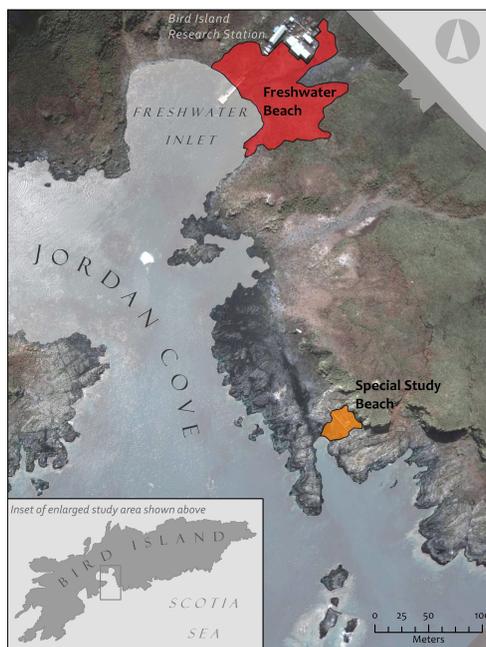


Figure 4.1: Map of the study area showing the two breeding colonies from which Antarctic fur seal mother-offspring pairs were sampled. The red and blue areas demarcate freshwater beach and the special study beach, respectively.

population structure. By implication, chemical differences between the colonies appear to reflect environmental influences (see Discussion).

Mother-Offspring Similarity.

Pups were significantly more similar to their mothers in their chemical fingerprints than expected by chance (Figure 4.2B), both overall (ANOSIM, global $R = 0.67$, $P < 0.0001$) and within each of the colonies (special study beach: ANOSIM, global $R = 0.53$, $P < 0.0001$; freshwater beach: ANOSIM, global $R = 0.45$, $P < 0.0001$). Chemical similarities between mothers and offspring could be encoded by shared genes or might simply reflect their spatial proximity. However, we found no relationship between chemical similarity and geographic distance within the special study beach, where pupping locations are recorded to the nearest square meter, either for mothers (Mantel's $r = 0.008$, $n = 20$, $P = 0.44$) or offspring (Mantel's $r = 0.06$, $n = 20$, $P = 0.31$). This finding suggests that chemical similarity is not associated with geographic proximity per se.

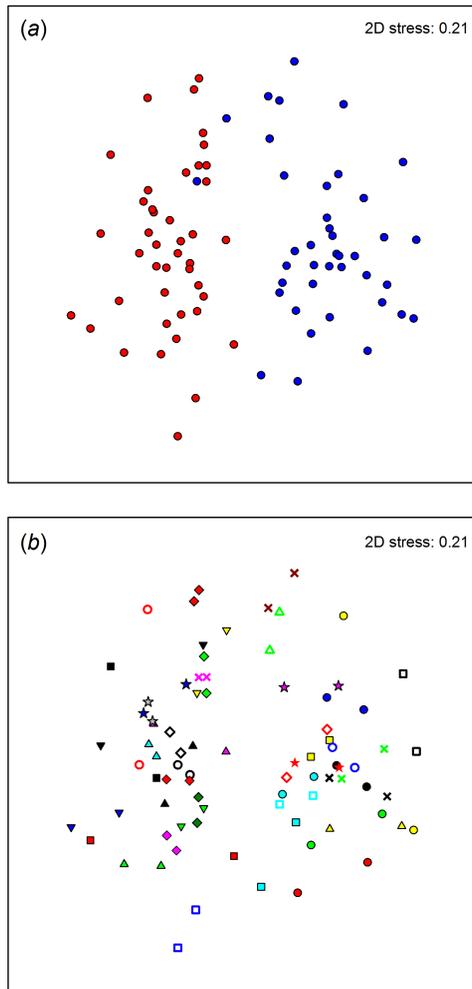


Figure 4.2: Two-dimensional nonmetric multidimensional scaling plots of chemical fingerprints of 41 Antarctic fur seal mother-offspring pairs. Bray-Curtis similarity values were calculated from standardized and $\log(x+1)$ -transformed abundance data; (a) color-coded by colony (red points: freshwater beach; blue points: special study beach); (b) plotted by mother-offspring pair, with each pair being denoted by a different symbol/color combination. The scales of the two axes are arbitrary. The closer the symbols appear on the plot, the more similar the two chemical fingerprints are.

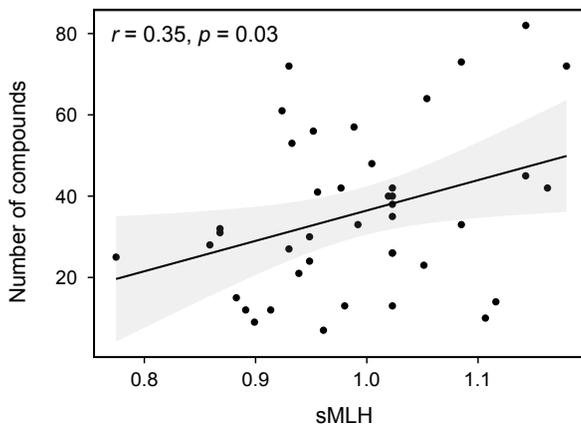


Figure 4.3: Relationship in mothers between sMLH and the number of compounds in an individual's chemical fingerprint.

Genotype and Overall Chemical Fingerprints.

To determine whether genetic relatedness is reflected in chemical similarity, we tested for an association between pairwise r (see Table S3 for summary statistics) and Bray-Curtis similarity. A highly significant relationship was obtained when all of the animals were analyzed together (Mantel's $r = 0.07$ $n = 82$, $P = 0.005$) but nonindependence of both chemical and genetic data for mothers and offspring may introduce pseudoreplication. We therefore repeated the analysis separately for mothers and offspring, finding no significant relationships (mothers, Mantel's $r = 0.06$ $n = 41$, $P = 0.10$; offspring, Mantel's $r = 0.030$ $n = 41$, $P = 0.25$).

To test for a chemical signal of genetic quality, we regressed the number of compounds in an individual's chemical fingerprint, a measure of chemical complexity, on standardized multilocus heterozygosity (sMLH). A significant positive correlation was found in mothers (Figure 4.3) ($F_{1,40} = 5.26$, $P = 0.026$) but not in offspring ($F_{1,40} = 0.50$, $P = 0.483$). The strength of correlation also increased steadily with the number of microsatellites deployed in mothers and to a lesser extent in offspring (Figure 4.4A). Conversely, the estimation error of the parameter g_2 , which quantifies the extent to which heterozygosities are correlated across loci, decreased with increasing marker number (Figure S2). Overall, g_2 was significantly positive (0.0022, $P = 0.032$ based on 1,000 iterations of the dataset), indicating that heterozygosity is correlated across the genome.

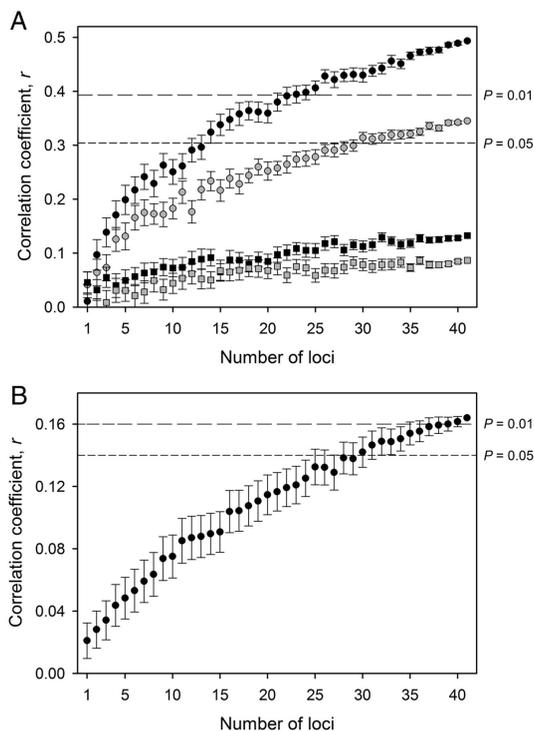


Figure 4.4: Dependency of the strength of genetic associations on the number of randomly sampled microsatellite loci. Strength of association was quantified as the correlation coefficient (r) between (A) sMLH and the number of compounds in an individual's chemical fingerprint (gray symbols) and the sum of an individual's factor 1 and factor 2 values (black symbols), plotted separately for mothers (circles) and offspring (squares); (B) relatedness and Bray-Curtis similarity at the 10 best substances in mothers (see Methods for details). Mean \pm SE of five resamplings of the data are shown for each point. The dashed lines represent significance thresholds.

Factor Analysis.

Chemical fingerprints are highly complex and may contain numerous compounds influenced by nongenetic factors. We therefore used principal axis FA to decompose the multidimensional chemical data into four factors (see Methods for details). Fitting the scores of all four factors together in a generalized linear model (GLM) of maternal heterozygosity, factors 1 and 2 were retained as significant predictor variables (Table 4.1, mother's sMLH) and together explained almost twice as much deviance as the number of compounds in an individual's chemical fingerprint (234 vs. 11.9%, respectively). A simple GLM of sMLH fitting the sum of the two factors as a single explanatory variable explained roughly the same amount of deviance (234%,

Table 4.1: Generalized linear models of sMLH in mothers and the colony from which an animal was sampled.

Term	Slope	F	df	P
Mother's sMLH (n = 41, total explained devians = 23.40%				
Factor 1	0.028	5.69	1	0.022
Factor 2	0.028	5.83	1	0.021
Colony (n = 82, total explained deviance = 56.26%				
Factor 4	0.38	102.88	1	<0.0001

$F_{1,39} = 11.91$, $P = 0.001$). In contrast, none of the factors were significantly associated with offspring heterozygosity.

To test whether any of the factors are also associated with genetic relatedness, we used partial Mantel tests to derive the statistical significance of each factor while controlling for the others (see Methods for details). Factor 1 was significantly correlated with relatedness in mothers (Mantel's $r = -0.123$, $n = 41$, $P = 0.028$) (Figure S3) but not in offspring (Mantel's $r = 0.024$, $n = 41$, $P = 0.65$). None of the other factors correlated significantly with relatedness in either mothers or offspring. As with the signal of heterozygosity, the strength of association between factor 1 and relatedness increased steadily with marker number (Figure 44B).

We next constructed a GLM to test for differences in the values of each of the four factors between the two colonies (Table 4.1, colony). Factors 1, 2, and 3 did not differ significantly, whereas factor 4 exhibited a highly significant difference between the colonies (Figure 4.5). Thus, factors 1 and 2 both show correlations with genetic traits as well as overlapping distributions between colonies, factor 3 is not significantly associated with any of the variables we measured, and factor 4 represents substances that discriminate the two colonies and must therefore be environmentally influenced.

Identification of Important Substances.

To identify substances that contribute most strongly toward chemical similarity within mother-offspring pairs, we used the 'similarity percentages' routine (SIMPER, see Methods). Selecting the two most important compounds for each of the 41 mother-offspring pairs, we identified a total of 12 substances (Table S4, mother-offspring similarity). These substances yield a much stronger pattern of within-pair mother-offspring similarity (ANOSIM, global $R = 0.68$, $P < 0.0001$) than was obtained for the full dataset. Similarly strong patterns were obtained separately for each of the colonies (special study beach: ANOSIM, global $R = 0.53$, $P < 0.0001$; freshwater beach: ANOSIM, global $R = 0.31$, $P = 0.001$).

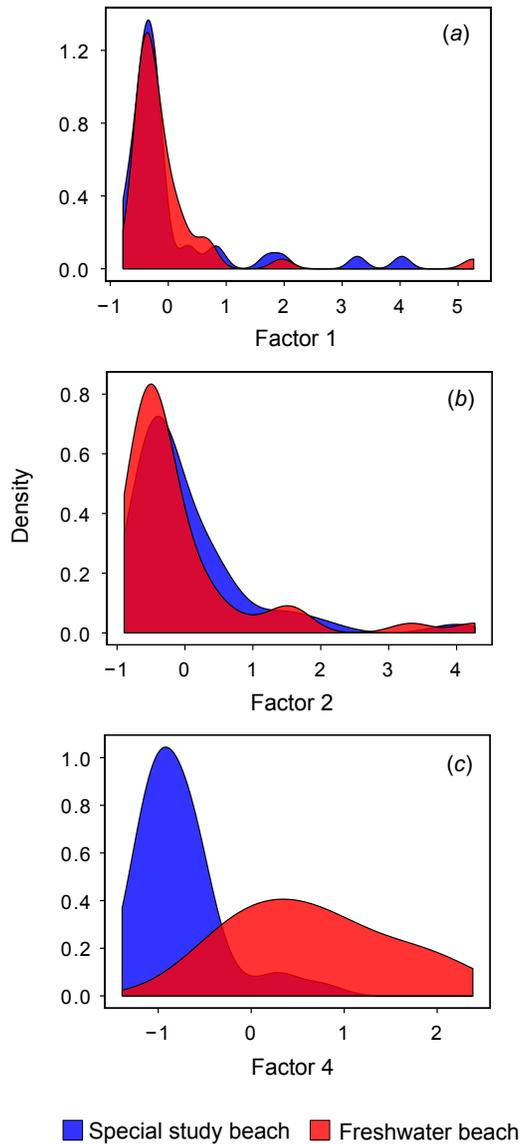


Figure 4.5: Distribution of factor scores of individuals sampled from the two seal colonies. Factors 1, 2, and 4 are shown in A, B, and C, respectively. Freshwater beach is shown in red and the special study beach is shown in blue

We also used SIMPER to search for substances accounting for most of the chemical dissimilarity between the two colonies. This approach identified a total of 15 substances (Table S4, colony dissimilarity) that collectively yield a much higher global R value (ANOSIM, global $R = 0.77$, $P < 0.0001$) than was obtained for all of the chemicals. To identify substances associated with genetic relatedness, we used the BIO-ENV procedure embedded in a bootstrap framework (see Methods for details). We obtained a subset of 10 substances (Table S4, relatedness) that consistently occurred within the ‘best’ subsets (i.e., maximizing the relationship between chemical similarity and relatedness) over all 10 x 106 bootstrap samples and collectively maximized the relationship between chemical distance and relatedness (Figure S4). Chemical similarity based on these 10 substances was significantly associated with genetic relatedness (Mantel’s $r = 0.164$, $n = 41$, $P = 0.001$).

Finally, we cross-referenced the three lists of substances to evaluate any potential overlap. Of the 12 compounds carrying the strongest signal of mother-offspring similarity, 9 also occurred in the subset of chemicals that differ between the two colonies, implying that they may be influenced by environmental conditions ashore. Remarkably, an additional two compounds overlapped with the best subset of chemicals associated with genetic relatedness. The mass spectra and Kovats indices of these substances indicate close resemblance to the known pheromones ethyl-9-hexadecenoate and heptadecanoic acid (see Discussion and Table S4, relatedness).

Discussion

Although mother-offspring recognition is under strong selection in many species, little is known about its chemical basis, particularly in natural populations of nonmodel organisms. We show that fur seal pups are highly similar to their mothers in their chemical fingerprints and that this similarity is largely encoded by a handful of substances that also carry information about either colony or genotype. Our findings provide intriguing insights into how females could use chemical information to recognize their offspring and may also help to explain how fur seals appear capable of exercising mate choice for heterozygous and unrelated partners (Hoffman et al., 2007b).

Our study was partly motivated by the discovery that female Australian sea lions can identify their pups using only olfactory cues (Pitcher et al., 2011b). In most vertebrate species, chemical fingerprints show marked differences by sex, age, and reproductive status (Caspers et al., 2011), a pattern that is partly reflected in our data because it is only the mother’s chemical fingerprints that encode genotype. However, the overall chemical fingerprints of mothers and offspring

are still very similar, raising the possibility that self-referent phenotype matching (Blaustein, 1983) could be used in mother-pup recognition. This is a conceptually simple mechanism by which the own phenotype is a representation or template used for the recognition of relatives. Self-referent phenotype matching has been demonstrated in a variety of mammalian, bird, and fish species (Hauber and Sherman, 2001). However, further experimental evidence would be needed to show that mother-offspring recognition in fur seals relies on self-matching rather than social learning. Interestingly, allosuckling rates vary considerably among pinniped species, from 6% in New Zealand sea lions to up to 90% in Hawaiian monk seals, suggesting that mother-pup recognition abilities may vary among species (Pitcher et al., 2011a). The Antarctic fur seal has one of the lowest observed rates of allosuckling (Hoffman et al., 2005), which is consistent with the strong pattern of chemical similarity we find between mothers and their pups.

Although chemical fingerprints are widely assumed to encode genetic traits, such as relatedness and individual heterozygosity, only a handful of studies have reported the expected associations. Moreover, chemical profiles typically change with age and reproductive status (Caspers et al., 2011) and genetic correlations have, to our knowledge, only been detected in breeding adults (Boulet et al., 2009; Charpentier et al., 2008). Analyzing the relationship between heterozygosity and chemical complexity separately for mothers and pups shows a clear correlation that increases with the number of loci for mothers, a pattern that is weak or lacking in pups (Figure 44A). Because of the consistency of our results with the literature, we believe this reflects a genuine functional difference between the chemical fingerprints of mothers and pups.

We also find a marked difference in the way that heterozygosity and relatedness are encoded in chemical fingerprints. Heterozygosity is detectable in the overall fingerprint, as it is correlated with the number of chemicals, whereas relatedness is encoded by a small subset of chemicals, whose signal is diluted by analyzing the overall chemical fingerprint. The diversity of chemicals reflected in heterozygosity could be the result of genetic polymorphisms in the enzymes involved in the synthesis of semiochemicals (Boulet et al., 2009) but may also be influenced by condition dependent factors (see below). In contrast, it makes sense that genetic relatedness could be encoded by a small subset of chemicals that potentially reflect certain genes, such as the MHC, a highly polymorphic cluster of immune genes detectable through scent (Yamazaki et al., 1979; Wedekind et al., 1995).

In natural populations, environmental effects on chemical fingerprints are likely to be particularly strong. The only study of a free-ranging, natural population to have detected an association with genotype used PCA to reduce the dimensionality of the chemical data (Leclaire

et al., 2012). However, this approach is not ideally suited to detecting such signals because a principal component that explains maximal variance may not necessarily provide an optimal representation of the underlying genotype. We applied PCA to our dataset but obtained no significant correlations between any of the resulting principal components and relatedness, and a weaker signal of heterozygosity than was obtained using FA. This result could be because of the so called 'simple structure' that is obtained by rotation of the factors within FA (Preacher et al., 2013). This results in each substance loading primarily on a single factor and not on the others, meaning that the factors represent subsets of variables that covary and are therefore likely to have a shared basis, such as genes or the environment.

FA was considerably more successful than PCA at detecting patterns relating to genotype within our chemical dataset. Factors 1 and 2 together explained almost twice as much of the deviance in heterozygosity as a simple regression on the number of substances, and relatedness was significantly associated with factor 1 but not with Bray-Curtis similarity based on the overall fingerprints. Because each factor mostly represents a subset of the total pool of chemicals, this finding is consistent with Hurst and Beynon's suggestion that the selective assessment of specific semiochemicals may allow individuals to assess genotype more accurately than from entire chemical fingerprints (Hurst and Beynon, 2010).

It is unclear why factor 1 carries information about both heterozygosity and relatedness, whereas factor 2 correlates only with heterozygosity. One possibility is that heterozygosity and relatedness are to some extent signaled by the same substances, potentially deriving from the MHC. As the substances loading on factor 2 are essentially uncorrelated with those loading on factor 1, we speculate that heterozygosity may influence the chemical fingerprint through two or more different pathways. Factor 1 could thus represent a direct pathway from genes to the chemical fingerprint, whereas factor 2 may represent an indirect pathway where body condition or the microbiome could be possible mediators. Future work will aim to explore these possibilities.

An important strength of our study was a sampling design that facilitated disentangling genetically encoded substances from those influenced by the environment. We found that factors 1 and 2, which both encode some aspect of genotype, did not differ significantly in the distribution of factor scores between the colonies, whereas factor 4, which carried no discernible genetic information, showed a highly significant difference. These differences could either be a result of environmental chemicals that directly contribute toward the profile, or could reflect alterations to the chemical fingerprint caused by different conditions on the beaches (e.g., temperature, wind, solar radiation). We would need to sample more colonies to determine the concrete causes.

Another important aspect of our study design was the unusually high genetic resolution provided by 41 microsatellites. Most studies use around 10-15 loci, which for our dataset was insufficient to detect a significant correlation between maternal heterozygosity and compound richness (Figure 44A). However, the strength of correlation increased steadily as more microsatellites were deployed until a highly significant relationship was obtained with the full marker panel. Similarly, the error with which the parameter g_2 was estimated from the genetic data decreased steadily with increasing marker number. This finding is consistent with the suggestion that, as long as heterozygosity is correlated across the genome (as is the case where appreciably inbred individuals are present), increasing the number of markers should improve the estimation accuracy of genome-wide heterozygosity, leading to a strengthening of effect size (Balloux et al., 2004; Hoffman et al., 2010). A similar pattern was also obtained for genetic relatedness, suggesting that, if many thousands of genetic markers could be deployed, an even greater proportion of the chemical variance should be explicable by genotype (Hoffman et al., 2004).

In many species, heterozygosity is associated with fitness (Hansson and Westerberg, 2002). In Antarctic fur seals, multilocus heterozygosity at nine microsatellites correlates with early survivorship and breeding success in females (Forcada and Hoffman, 2014), as well as reproductive success in males (Hoffman et al., 2004). Females of this species also appear to exert mate choice based on their partner's genotype (Hoffman et al., 2007a) but it is unclear how this could be achieved. The finding that heterozygosity and relatedness are both encoded in mother's chemical fingerprints lends support to the hypothesis that chemical cues could be involved, although unfortunately we were not able to include adult males in this study because they are challenging to capture and sedate. Nevertheless, as male fur seals emit a strong musky odor (Ling, 1972), which has been proposed to attract females during the mating season (Hamilton, 1956), it seems plausible that genotype could also be encoded in adult male chemical fingerprints.

To explore the extent to which genes and the environment influence mother-offspring similarity, we first attempted to identify the most important substances associated with mother-offspring similarity, colony dissimilarity, and genetic relatedness. We obtained relatively small subsets of 12, 15, and 10 chemicals, respectively. In the case of mother-offspring similarity and relatedness, these subsets yielded much stronger associations than were obtained for the overall fingerprints. This result suggests that SIMPER and BIO-ENV were successful in identifying important chemicals within the total set of 213 substances, although this does not preclude additional chemicals playing a lesser role. It is also noteworthy that as many as 10 or more chemicals appear to encode relatedness, given that a single locus is expected to provide little

power to distinguish anything other than close relatives (Hurst and Beynon, 2010).

Evaluating the overlap between the subsets of chemicals associated with mother-offspring similarity, colony dissimilarity, and genetic relatedness revealed an interesting pattern. Of the top 12 substances accounting for the similarity between mothers and their pups, 9 also occurred in the subset of chemicals that showed the greatest differences between the two colonies. Although our analysis is not exhaustive, as we focused only on the most important substances, this nevertheless suggests that chemical similarity within mother-offspring pairs is strongly influenced by the local environment. A further two substances also overlapped with the subset of chemicals associated with genetic relatedness, implying that mother-offspring similarity also has a genetic basis. Both of these substances reveal similarity to known pheromones, consistent with the previous suggestion that pheromone-like chemical signals may play an important role in mother-offspring recognition across a variety of taxa (Vaglio, 2009).

Little is currently known about the specific chemicals that signal genetic relatedness in vertebrates (Hurst and Beynon, 2010). Although we were only able to putatively identify three of the top 10 substances encoding relatedness using the National Institute of Standards and Technology (NIST) database, the mass spectra and Kovats indices of these compounds reveal close resemblance to the known pheromones ethyl 9-hexadecenoate, heptadecanoic acid and ethyl stearate (Table S4, relatedness). According to the pherobase database, all three of these substances are part of the chemical communication system of a variety of different taxa, ranging from bumblebees to badgers. Heptadecanoic acid, for example, is a known pheromone of 33 different species, including 26 vertebrate taxa. However, to act as a pheromone in a given species, a chemical must meet a number of strict criteria (Wyatt, 2015), which would require experimental evidence (see below).

Although we captured a large number of substances of varying volatility, we only recovered compounds soluble in ethanol and which could be detected by GC-MS. Extraction with other solvents was not possible because of logistic reasons. Nevertheless, even though our sampling of chemicals is likely to be incomplete, our analyses revealed a number of statistically significant and potentially biologically relevant patterns. In addition, we detected a number of chemicals that may carry important information. However, because some of these substances may have been further metabolized after extracting them from the skin (Theis et al., 2013), we cannot exclude the possibility that some of the putatively identified compounds could be breakdown products.

Finally, biologically relevant chemical cues can be transferred in a variety of ways, from volatile substances recognized by olfaction to chemicals that act when two individuals are in

physical contact (Wyatt, 2014b). Because adult female fur seals and their pups conduct naso-nasal inspections during the recognition procedure (Dobson and Jouventin, 2003), it is possible that some of the chemicals may act through contact. To unequivocally determine the biological relevance of the chemicals we have identified, as well as their precise mode of action, would require behavioral assays in the field. This will be challenging, but our results provide the basis for testable hypotheses on potential chemical signals and the substances involved.

Methods

Study Site and Field Methods.

Forty-four mother-offspring pairs were sampled from two breeding colonies: freshwater beach and special study beach, separated by ~200 m (Figure 4.1) on Bird Island, South Georgia (54°00' S, 38°02' W). Breeding females and their pups were captured and restrained on land using standard methodology (Gentry et al., 1982). Seal capture and restraint were part of annual routine procedures of the Long Term Monitoring and Survey program of the British Antarctic Survey. We obtained chemical samples by rubbing the cheek, underneath the eye, and behind the snout with a sterile cotton wool swab. Each swab was individually preserved in a glass vial in 60% (vol/vol) ethanol stored at -20°C. All of the samples were obtained immediately after capture by the same team of two seal scientists at both colonies. Tissue samples for genetic analysis were collected as described by Hoffman et al. (Hoffman et al., 2003) and stored individually at -20°C in the preservative buffer 20% (vol/vol) DMSO saturated with salt. Fieldwork was approved by the British Antarctic Survey Ethics Review Committee. Samples were collected and retained under permits issued by the Department for Environment, Food and Rural Affairs (DEFRA), and in accordance with the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES).

Chemical Analyses.

We first took 1 mL of each sample and allowed the ethanol to evaporate at room temperature under a fume hood for a maximum of 12 h before resuspending in 50 μ L dichloromethane for subsequent processing. The samples were then analyzed on a GC equipped with a VF-5 ms capillary column (30 m size x 0.25 mm inner diameter, DF 0.25, 10-m guard column; Varian) and coupled to a quadrupole mass spectrometer (Focus GC-DSQ MS system, Thermo

Electron). A blank sample (control with cotton wool and ethanol) and an alkane mix (C8-C28) were analyzed as well. One microliter of each sample was injected into a deactivated glass wool-packed liner at an inlet temperature of 225°C and processed in a splitless mode. Carrier gas (He) flow rate was held at 1.2 mL/min. The GC run was initiated at 60°C for 3 min then ramped at 10°C/min to 280°C, where it remained for 20 min. The transfer line temperature was set to 280°C and mass spectra were taken in electron ionization mode at 70 eV with five scans per second in full-scan mode (50-500 m/z). GC-MS data were processed using the program Xcalibur (Thermo Scientific). To ensure that the scoring of compounds was as objective as possible, we wrote a custom R script (available on request) that compensated for minor shifts in retention times among chromatograms by maximizing the number of shared components between samples through very small (≤ 0.03 ms) shifts in the retention time. To double-check the reliability of the scoring, $\sim 10\%$ of compounds were selected at random and scored by eye.

Genetic Analysis.

Total genomic DNA was extracted from each sample using a standard phenol-chloroform protocol and genotyped at 43 highly polymorphic microsatellite loci (see Table S1 for details). These were PCR-amplified in eight separate multiplexed reactions using a Type It Kit (Qiagen) as described in Table S1. The following PCR profile was used: one cycle of 5 min at 94°C; 24 cycles of 30 s at 94°C, 90 s at T_a °C and 30 s at 72°C; and one final cycle of 15 min at 72°C (see Table S1 for T_a). Fluorescently labeled PCR products were then resolved by electrophoresis on an ABI 3730xl capillary sequencer and allele sizes were scored automatically using GeneMarker v1.95. To ensure high genotype quality, all traces were manually inspected and any incorrect calls were adjusted accordingly.

Genepop (Raymond and Rousset, 1995) was used to calculate observed and expected heterozygosities and to test for deviations from HWE, separately for mothers and pups, specifying 10,000 dememorizations, 1,000 batches, and 10,000 iterations per batch. Two loci that deviated from HWE in either mothers or pups after table-wide correction for the FDR using Q-value (Storey, 2002) were excluded from subsequent analyses, leaving a total of 41 loci (Table S1). Because milk stealing is common in fur seals and can lead to errors in the assignment of mother-offspring pairs in the field (Hoffman et al., 2005), we used the program Colony v2.0.5.0 (Jones and Wang, 2010) to verify that all of our mother-offspring pairs were genuine. Coancestry v1.0.1.2 (Wang, 2011) was then used to generate a pairwise relatedness matrix based on Queller and Goodnight's statistic, r (Queller and Goodnight, 1989). Each

individual's heterozygosity was expressed as $sMLH$, which is defined as the total number of heterozygous loci in an individual divided by the sum of average observed heterozygosities in the population over the subset of loci successfully typed in the focal individual (Coltman et al., 1999). The two-locus heterozygosity disequilibrium g_2 , which measures the extent to which heterozygosities are correlated across loci, was then computed using the method of David et al. 2007. Sensitivity of this estimate to the number of loci was explored by randomly selecting different sized subsets of loci and recalculating g_2 1,000 times.

To test for population structure, Bayesian cluster analysis of the microsatellite dataset was implemented using Structure v2.3.3 (Pritchard et al., 2000). Structure uses a maximum-likelihood approach to determine the most likely number of genetically distinct clusters in a sample (K) by subdividing the dataset in a way that maximizes HWE and minimizes LD within the resulting clusters. Separately for mothers and pups, we ran five independent runs for each value of K ranging from 1 to 10 using 1×10^6 Markov chain Monte Carlo iterations after a burn-in of 1×10^5 , specifying the correlated allele frequencies model and assuming admixture. The most likely K was then evaluated using the maximal average value of $\ln P(D)$, a model-choice criterion that estimates the posterior probability of the data.

Statistical Analysis Framework.

Any chemicals appearing in the control sample or present in only one sample were excluded from further analyses, leaving a total of 213 substances. To explore the completeness of our sampling, we estimated the maximum number of substances present in the population using the Michaelis-Menten Function, based on a permutation procedure (9,999 iterations). Up to 229 substances might be expected in a larger sample of individuals, suggesting that we have sampled around 95% of all potential substances. Analyses were conducted on the relative proportion of each substance (%) to the total amount of substances (Sun and Müller-Schwarze, 1998). We then used a three-step analytical framework to: (i) visualize and statistically analyze overall patterns of chemical fingerprint similarity in relation to breeding colony, mother-offspring pair, relatedness, and heterozygosity; (ii) tease out subsets of chemicals containing genotypic and environmental information; and (iii) identify specific compounds involved. Computer code and documentation are provided as a PDF file written in Rmarkdown (Dataset S1) together with the data (Dataset S2).

Overall Patterns of Chemical Similarity.

The chemical fingerprint data were visualized using nonmetric multidimensional scaling (Clarke, 1999) based on a matrix of pairwise Bray-Curtis similarity values calculated from the $\log(x+1)$ -transformed data. This approach allows visualization of a high-dimensional chemical similarity space by placing each individual in a 2D scatterplot such that ranked between-individual distances are preserved, points close together representing individuals with relatively high chemical similarity. Differences between a priori defined groups (i.e., the breeding colonies and mother-offspring pairs) were then analyzed through nonparametric ANOSIM (Clarke, 1999) using 99,999 iterations of the dataset. ANOSIM is a permutation test that provides a way to evaluate whether there is a significant difference between two or more groups of sampling units without the need for assumptions concerning data distribution or homoscedasticity. These analyses were implemented in R using the *vegan* package (Oksanen et al., 2017).

Factor Analysis.

To dissect apart genetic from environmental components, we performed a principal axis FA on the chemical data. We used an oblique rotation technique (*promax*), which allows the factors to be correlated. This type of rotation was used because it is possible that certain compounds within the chemical fingerprint may encode more than one genetic characteristic (e.g., heterozygosity and relatedness) and could thus be correlated with more than one factor. FA cannot be applied when a dataset has more variables than observations ($D \gg N$) because the covariance matrix is singular and an inverse cannot be computed. We therefore used the function `factor.pa.ginv()` from the R package *HDMD*, which uses a generalized inverse matrix (McFerrin, 2015). An important step in factor analysis is choosing a reasonable number of factors to represent the data (Preacher et al., 2013). As our dataset is complex and contains many zero entries, some common methods like parallel analysis may lead to an impracticably large number of factors. Consequently, we applied two methods for determining the optimal number of factors. First, we used the Bayesian Information Criterion, which optimizes the trade-off between model complexity and model fit, and second we used a scree plot, which visually depicts the drop in the factor eigenvalue course (Preacher et al., 2013; Cattell, 1966). Both methods suggested four factors.

Generalized Linear Models.

To explore the contributions of each of the four factors toward the signal of heterozygosity, we constructed separate GLMs of mother and offspring sMLH, in which we fitted all four factors together and specified a Gaussian error structure. We then tested for factors that differ significantly between the two colonies by constructing a GLM with colony as the response variable (modeled using a binomial error structure) and the values of the four factors fitted as predictors. For each GLM, we initially implemented a full model containing all of the predictor variables and then used standard deletion testing procedures based on F tests (Crawley, 2002) to sequentially remove each term unless doing so significantly reduced the amount of deviance explained.

Partial Mantel Tests.

To test for associations between each of the factors and genetic relatedness, we used the relatedness matrix based on all 41 loci as the response variable and fitted as predictor variables matrices of pairwise similarity at each of the four factors using a Partial Mantel test implemented in the *ecodist* package (Goslee and Urban, 2007). This randomizes the rows and columns of one dissimilarity matrix but leaves the others unpermuted. Separate models were constructed for mothers and offspring, each using 10,000 permutations of the dataset. Finally, we computed the Spearman rank correlation (Mantel's r) and two-tailed P value for the association between relatedness and a factor matrix given the other factors as covariates.

Identification of Chemicals.

We next attempted to identify specific chemicals associated with breeding colony, mother-offspring similarity, and genetic relatedness. First, we assessed the contributions of specific substances to the similarity within groups, using the 'similarity percentages' routine (SIMPER) (Clarke and Warwick, 2001). This process decomposes all Bray-Curtis similarities within a group into percentage contributions from each compound, listing the compounds in decreasing order of importance. As groups, we specified (i) the two breeding colonies and (ii) the 41 different mother-offspring pairs.

Second, to explore the contributions of individual chemicals to the signal of genetic relatedness, a continuously distributed variable, we used the BIO-ENV procedure (Clarke and Warwick, 2001) to identify the 'best' subset of compounds within the chemical abundance

matrix that maximizes the rank correlation between pairwise Bray-Curtis similarities and relatedness. However, with over 200 different chemicals being present in the chemical data matrix, it seems likely that this approach could yield spurious associations, especially given that some of the chemicals were present only in a few individuals. For this reason, we embedded the BIO-ENV procedure in a bootstrap analysis as follows: (i) we randomly subsampled 20 of the 41 mothers 20,000 times; (ii) for each subsample, we randomly selected 10 chemicals, each 500 times; (iii) for each of the resulting 10 x 106 subsamples, comprising 20 individuals and 10 compounds, we applied the BIO-ENV procedure and saved the compounds present in the best subset. We then summed up the occurrences of every chemical throughout all of the subsets and sorted them in decreasing order to represent their relative importance. The basic assumption of our approach is that random correlations will not be consistent over the different subsamples of individuals and compounds, whereas compounds that genuinely encode relatedness should be recovered consistently across many subsets. This procedure was conducted in R using the `bio.env()` function in the `sinkr` package (Taylor, 2015).

Identification of putative substances encoding mother-pup similarity, colony differences, and relatedness were based on two steps: (i) comparing the mass spectrum of a specific substance with the best match of the NIST library (NIST 2005 and 2008) and (ii) calculating the Kovats Retention Index and comparing this to the literature value (obtained from [urlwww.Pherobase.com](http://www.Pherobase.com) and urlwww.chemspider.com). Kovats Indices (Kováts, 1958) were calculated by running a sample of linear alkanes (C8-C28) under the identical GC-MS conditions as described above.

Acknowledgements

We thank Mareike Esser for guidance on the implementation of `g2` in R. This research was supported by Marie Curie FP7-Reintegration-Grant within the 7th European Community Framework Programme PCIG-GA-2011-303618 (to J.I.H.) and Deutsche Forschungsgemeinschaft standard Grant HO 5122/3-1 (to J.I.H.). This article contributes to the Polar Science for Planet Earth Ecosystems and Long Term Monitoring and Survey Projects of the British Antarctic Survey, Natural Environment Research Council core funded science programme.

Supplementary Information

Chemical fingerprints encode mother-offspring similarity, colony membership, relatedness, and genetic quality in fur seals

Martin A Stoffel, Barbara A. Caspers, Jaume Forcada, Athina Giannakara, Markus Baier, Luke Eberhart-Phillips, Caroline Müller & Joseph I Hoffman

Proceedings of the National Academy of Sciences. 112(36):E5005-E5012. 2015

This Supplementary Information can be found **online** at
<https://doi.org/10.1073/pnas.1506076112>



Young elephant seals are making friends and start gathering in groups after being abandoned by their mothers.

5

Gut microbiota

Early sex-specific differentiation, development and genetic basis of the gut microbiota in a highly polygynous marine mammal, the northern elephant seal

Martin A Stoffel, Karina Acevedo-Whitehouse, Nami Morales-Duran, Stefanie Grosser, Nayden Chakaroy, Oliver Krtiger, Hazel J Nichols, Fernando R Elorriaga-Verplancken & Joseph I Hoffman

in preparation

Abstract

Throughout their life-histories, sexually dimorphic males and females often face extremely different challenges, which manifest themselves in a variety of sex-specific adaptations. Most attention to date has been given to visibly dimorphic features such as body size or colour patterns, and little is known about sexual dimorphism of more complex or subtle phenotypes. A promising new avenue for a broader understanding of life-history adaptations is the study of the microbial communities which inhabit all organisms. Despite the variety of potential adaptive functions of microbiota, studying host-microbe interactions in the wild is difficult, as environmental factors such as diet strongly shape individual microbial communities and potentially mask more subtle intrinsic influences. Here, we studied the gut microbiota of young northern elephant seals in the post-weaning period, a critical developmental time frame of several weeks where elephant seals are abandoned by their mothers and are not yet able to feed by themselves. We show that the gut microbiota changes substantially after weaning despite the absence of food, but that each individual has a core microbiota. Notably, the change in microbiota includes a substantial decrease in *Bacteriodes* and an increase in *Prevotella*, a pattern that has previously been observed in young pigs and reflects a major dietary change from milk feeding to solid food. Unlike other studies in wild organisms, we found that both alpha- and beta-diversity are highly sex-specific, with males harbouring different and more diverse microbial communities. Moreover, genetic relatedness correlated with microbial similarity in males, where a large number of taxa contribute to the signal, but not in females. Importantly, our study design controlled for dietary effects, suggesting that bacterial dynamics over time, early sex-differences and host-genetic effects could be co-evolved adaptive mechanisms rather than caused by differences in diet. Our study provides a careful examination of the early host-microbe relationship in the northern elephant seal and suggests that gut microbiota can reflect an early adaptation to markedly different sex-specific life-histories in one of the most sexually dimorphic animals on earth.

Introduction

Vertebrates are inhabited by vast amounts of microbes that increasingly emerge as key players in their host's biology and evolution (Bik et al., 2016; Ley et al., 2008; McFall-Ngai et al., 2013; Moeller et al., 2014). The richest and arguably most complex microbial communities are those that populate the gastrointestinal tract and are collectively termed the 'gut microbiota'. Gut microbes benefit their host in many ways, such as promoting the development of organs,

helping with nutrient uptake and supporting the immune system (Cheesman et al., 2011; Diaz Heijtz et al., 2011; Lathrop et al., 2011; Zhu et al., 2011). A disturbed gut microbiota can result in a series of severe consequences for the host, ranging from autoimmune diseases and infections to obesity (Giongo et al., 2011; Round and Mazmanian, 2009; Turnbaugh et al., 2008). Moreover, the gut microbiota is highly dynamic across time and space and is influenced by many factors. At a broader scale, the strongest determinants of the gut microbiota seem to be phylogeny and diet, both of which lead to remarkably different bacterial communities across host species (Bik et al., 2016; Ley et al., 2008; Muegge et al., 2011). At a finer scale, differences in the gut microbiota within species are shaped by a combination of environmental factors such as diet, location and season, behavioural factors such as social networks, and heritable factors such as host genetics (Benson et al., 2010; Kurilshikov et al., 2017; Moeller et al., 2014; Ren et al., 2017; Tung et al., 2015). Despite a growing body of studies investigating gut microbiota, a big gap in the literature is marked by the scarcity of studies on wild organisms (Hird, 2017). Most research to date focuses on humans and captive animals. Captivity, however, drastically alters microbial communities due to controlled and less diverse diets (Hird, 2017) which makes it difficult to make ecological and evolutionary inferences. Consequently, little is known about the composition, development and function of the gut microbiota in the wild, despite its potential to contribute to our fundamental understanding of ecology and evolutionary biology of mutualistic symbiotic relationships (Hird, 2017; Zilber-Rosenberg and Rosenberg, 2008).

The mammalian gut is considered to be largely sterile in the womb (Perez-Muñoz et al., 2017), but gets rapidly colonised by various microbes during and after birth. During these early stages of life, the gut microbiota is of tremendous importance, and disturbances can impact host development and impair metabolism, health and immunity (Candon et al., 2015; Cho et al., 2012; Cox et al., 2014; Macpherson and Harris, 2004; Russell et al., 2012). It is therefore of interest to identify and quantify which microorganisms populate the gut during an individual's development. Across the life-span of an organism, ontogeny appears to affect the composition of the gut microbiota of a number of species (Clark et al., 2015; Langille et al., 2014; O'Toole and Jeffery, 2015). However, the patterns of change can differ drastically depending on the host. Bacterial diversity, for example, increases during development in humans, chickens, pigs and ostriches (Ballou et al., 2016; Frese et al., 2015; Kundu et al., 2017; Videvall et al., 2018), but decreases during maturation in zebrafish and African turquoise killifish (Smith et al., 2017; Stephens et al., 2016). A mixed pattern has been observed in mice, where an early drop of diversity after the initial transmission of maternal microbiota is followed by an increase after the introduction of solid food (Pantoja-Feliciano et al., 2013). To our knowledge, patterns of microbial colonisation during early development in wild animals are as yet unknown (Ren

et al., 2017).

Every species' life-history is determined by a series of challenges to which it must adapt, both through physiological development and behaviour. A key element facilitating these adaptations might be mutualistic or commensal microbiota. A particularly strong factor that might drive the variation of microbial communities is sex, as males and females often experience very different selection pressures due to differences in behaviour and physiology (Tarka et al., 2018). Several of these differences could be directly and indirectly associated with the gut microbiota, such as sex-specific immune responses (Klein and Flanagan, 2016) or sex-specific foraging behaviour (Boeuf et al., 2000; Boinski, 1988; Lewis et al., 2002). Curiously, despite the important role of sex-specific microbiota in humans (Markle et al., 2013), the impact of sex on the gut microbiota of wild vertebrates seems to be non-existent or very small (Bennett et al., 2016; Bobbie et al., 2017; Maurice et al., 2015; Ren et al., 2017; Tung et al., 2015; Videvall et al., 2018). However, microbiota studies in wild populations are likely to be impacted by environmental factors, such as diet or social group, which might mask the effect of intrinsic factors, such as sex.

Another largely open question for natural populations is how host genetics impacts the microbiota. Most insights to date come from twin studies in humans or from different strains of laboratory mice and suggest that the influence of host genetics is modest compared to environmental effects (Kurilshikov et al., 2017). In contrast to inbred laboratory animals, wild animals are likely to exhibit greater genetic variation and more complex microbiota, potentially leading to stronger covariation. However, quantifying the impact of host genetics on gut microbiota in the wild is even more difficult due to strong environmental effects (Bik et al., 2016; Perofsky et al., 2017; Tung et al., 2015) which might blur the genetic signal. Consequently, whether host genetics influences gut microbiota in the wild is unknown, despite the importance of the question in the light of host-microbe evolution.

A particularly suitable species to investigate the intrinsic factors than shape the gut microbiota is the northern elephant seal (*Mirounga angustirostris*). This species combines an extreme lifestyle with a predictable breeding season during which the animals stay ashore and can be individually traced. Northern elephant seals are among the most sexually dimorphic mammals on earth, with males being up to 3-4 times as heavy as females (Wilson and Mittermeier, 2014). This adaptation goes hand in hand with a highly polygynous mating system, in which a single male can mate with dozens of females in a given season (Le Boeuf and Laws, 1994). Consequently, males and females face very different challenges: during the breeding season, males must continuously defend their harems against competitors, while females need to invest substantial amounts of energy into nursing their pups. Neither males nor females feed during the breeding seasons, with some males fasting for up to three months and females up to one

month, despite the high energetic investment to provide high-fat milk to their young (LeBoeuf and Ortiz, 1977). Outside the breeding season, elephant seals spend most of their lives at sea, and even there, sex differences are striking. Males and females have very different foraging strategies, with males feeding on benthic prey along the continental margin of North America, and females feeding widely on pelagic prey in deeper waters (Boeuf et al., 2000). Therefore, elephant seals have developed a series of sex-specific adaptations to these diverging life-histories, but we do not know yet whether the gut microbiota is involved in these adaptations.

Here, we studied the gut microbiota of young elephant seals immediately from weaning and during a 35-day period. This is an exceptional period to study gut microbiota, as all young seals share one month of nursing and will spend the remaining seven weeks without feeding until they leave the rookery (Reiter et al., 1978). Consequently, variation in gut microbiota will be largely intrinsic rather than caused by different diets, behaviours or habitats. Within this setting, we used repeated sampling of rectal swabs in combination with host genetic markers to investigate (1) the core gut microbiota of the northern elephant seal, (2) the development of microbiota across time, (3) sex-specific effects on the microbiota which could reflect early life-history adaptations, and (4) whether host genetics shapes the early gut microbiota. To our knowledge, this study is the first to investigate changes in gut microbiota during early development in a truly wild mammalian population and to have a diet-controlled setting to shed light on intrinsic impacts on microbiota in the wild.

Results

We investigated the development of the gut microbiota in young northern elephant seals throughout their weaning period. Specifically, we sampled rectal swabs from 40 animals, starting immediately after their mothers stopped nursing and returned to the sea (time point T1) and repeatedly sampling each individual after two (T2) and four weeks (T3). The different time points therefore reflect different age classes. As a few animals were lost or found dead during the study period, we ended up with 112 rectal swabs across three time points for which we quantified bacterial communities using 16s rRNA sequencing. After assembling the raw reads into amplicon sequence variants (ASVs) with DADA2 (Benjamin J. Callahan et al., 2016) we retained 1063 ASVs with an average of 286 ± 67 ASVs (mean \pm sd) per sample. Furthermore, to quantify host genotypes, we took a small skin sample from the flipper of each pup for microsatellite genotyping.

Broad characterization of the gut microbiota

Overall, the main phyla were typical mammalian gut microbiota (Figure 5.1), with the majority of ASVs belonging to the phyla Bacteroidetes (mean \pm sd = 34% 2%), Firmicutes (mean \pm sd = 29% 1%), Fusobacteria (mean \pm sd = 19% 3%), and Proteobacteria (mean \pm sd = 13% 1%). Across time, the relative abundances of these four phyla remained relatively stable, except for Fusobacteria, which drastically reduced during weaning (Figure 5.1). However, on a finer taxonomic scale, we found drastic changes across the three time points (see below).

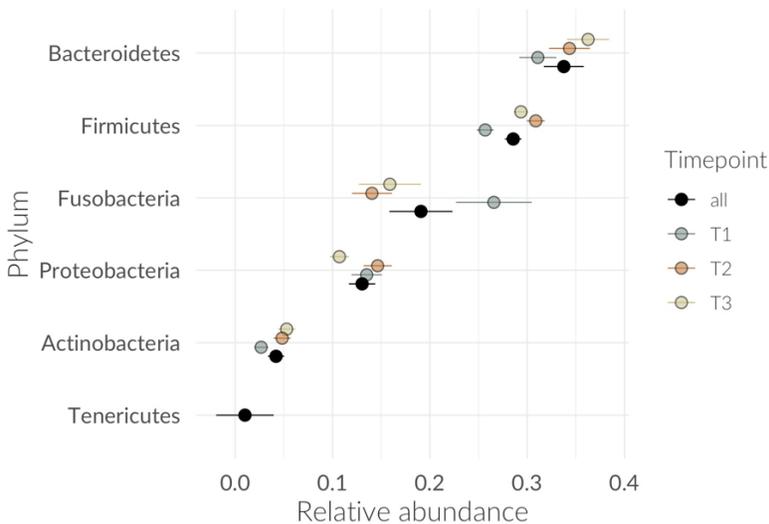


Figure 5.1: Relative abundance of the six most abundant bacterial phyla of the Northern elephant seal pup gut (mean and SD across individuals) within three sampling points and across all sampling points. ASVs with less than 50 reads in each category were discarded for this plot, reducing the dataset to six Phyla and the Tenericutes Phylum to a single estimate across all time points.

The core microbiota across individuals at different ages

We characterised the core microbiota at different developmental stages during the weaning period by extracting ASVs that appeared in at least 95% of samples at each time point (Supplementary Table 1 - 3). Directly after weaning (T1), we identified 21 core ASVs, with only two ASVs from the genera *Fusobacterium* and *Bacteroides* making up more than 25% of the average relative abundance across individuals. This pattern changed drastically at T2 and T3. Here, we identified 15 and 35 core ASVs respectively, but the dominance of the two ASVs

from T1 disappeared. Instead, a taxon not present during T1 became the most dominant ASV during T2 and T3 (with an average of 4% relative abundance) and is part of the genus *Ezakiella*. This is a recently discovered genus, of which only two species have been described; one from fecal samples of a coastal human indigenous Peruvian population (Patel et al., 2015) and one from the human female genital tract (Diop, Raoult, Bretelle, & Fenollar, 2017). Closer to the time of nutritional independence (T3), a peculiar shift happened. One taxon from the genus *Prevotella* was the most successful colonizer and became the second most abundant genus. On the other hand, the second most abundant taxon from the genus *Bacteroides* decreased substantially (Figure 5.2). Both of these genera have previously been linked to diet-specific functions, with many *Bacteroides* species being important for breaking down milk oligosaccharides and *Prevotella* being associated with plant polysaccharide consumption (Gorvitovskaia, Holmes, & Huse, 2016; Wu et al., 2011).

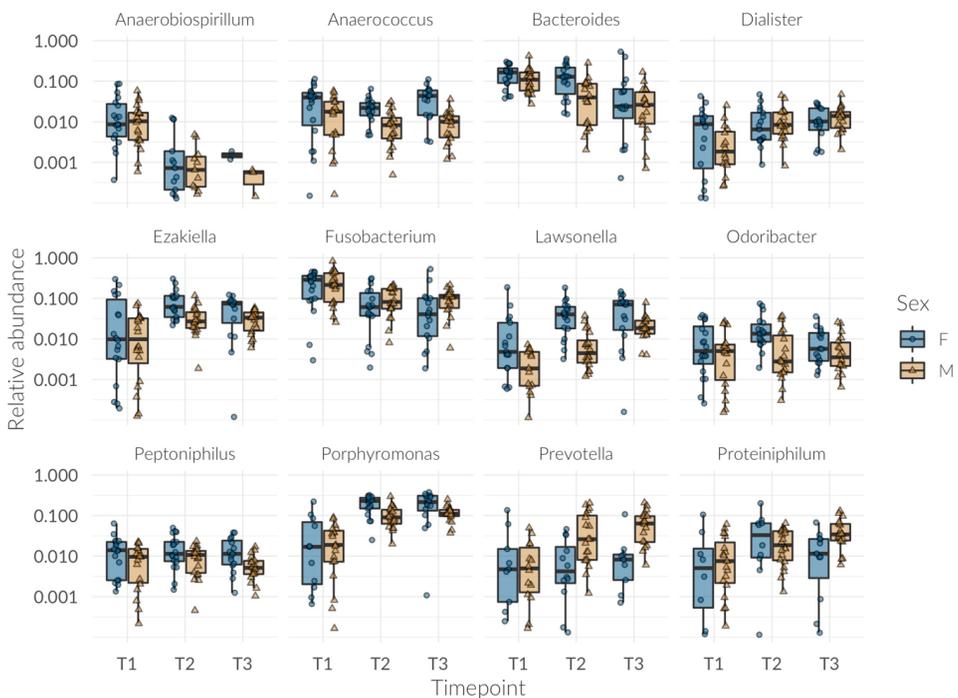


Figure 5.2: Relative abundance of northern elephant seal pup gut core microbiota across age and sex, plotted on the genus level. Shown are all taxa that appear within the top 10 core microbiota (See Supplementary table 1-3) within any of the three sampling points. Before visualization on the log scale, taxa with relative abundances below 0.01% within a sample were discarded. One genus, which could not be assigned, was excluded from the plot.

Microbiota composition changes with age and differs between sexes and individuals

To quantify the major determinants of gut microbiota similarity across samples (beta diversity), we used a multidimensional scaling plot (MDS) of the Bray-Curtis similarities between bacterial samples for visualisation and PERMANOVA (Anderson, 2001) for the statistical analysis. The MDS revealed three factors that had a strong impact on variation in the gut microbiota (Figure 5.3 and Figure 5.4). Figure 5.3 reveals strong effects of sex and age of individuals (time points). Along the first axis, which accounted for 28.5% of the multidimensional spread in the data, there is a clear transition in microbial composition from the moment of weaning (T1) to the last sampling point (T3) shortly before the young seals depart to the sea, with samples from T2 being intermediate. A strong separation is also visible along axis 2, which accounted for 13.4% of the variation and reveals differences between the two sexes across the three age stages. The third determinant of microbiota similarity across samples was the host itself, which is reflected in the close clustering of samples taken from the same individual (Figure 5.4).

To statistically analyse the observed group differences, we used PERMANOVA as a model based, non-parametric method to fit age, sex, and individual in a single model. Overall, age and sex each explained 15% of the variation in microbial similarities (age: $R^2 = 0.15$, $p < 0.001$, sex: $R^2 = 0.15$, $p < 0.001$), while between-individual differences accounted for 40% of the variation ($R^2 = 0.40$, $p < 0.001$). After fitting the model with all samples, we compared specific time points post-hoc using PERMANOVA, while still controlling for sex and individual in the model. The transition from T1 to T2 explained 10.3% of the variation ($R^2 = 0.10$, $p < 0.001$) while 4.1% was explained due to microbial differences between T2 and T3 ($R^2 = 0.04$, $p < 0.001$). The PERMANOVA assumption of multivariate homogeneity of group variances was met across all tests, as none of the contrasted groups differed in their dispersions (all $p > 0.05$). Consequently, all PERMANOVA results reflect differences in mean values across groups rather than differences in group dispersions (see Materials and methods).

Differential abundance of specific taxa across age and between sexes

At a finer scale, we used boxplots and raw data to visualize trends across time and sex for different hierarchical taxonomic ranks, from genus (Figure 5.2) to phylum, class, and order (Supplementary Figures 1-3). Here, it becomes apparent that the dynamics are quite complex, with multiple colonization and extinction events and different patterns depending on the rank. To resolve this problem, we quantified how many and which ASVs or taxa are differentially

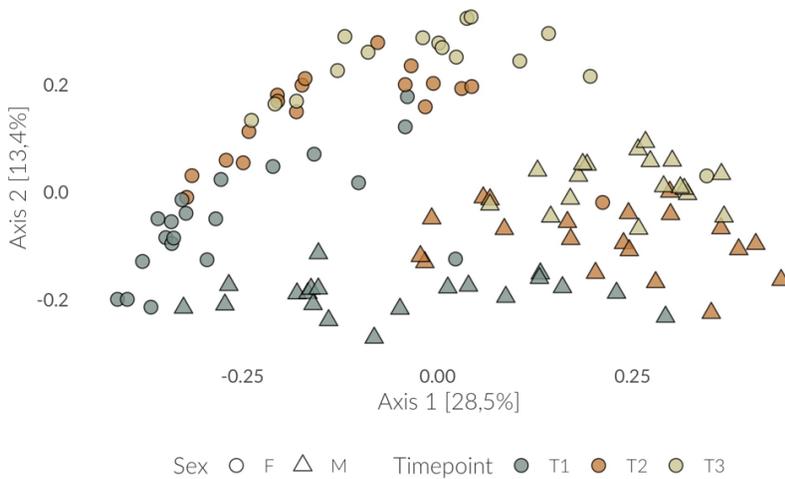


Figure 5.3: Gut microbiota similarity of northern elephant seal pups across sex and time. Multidimensional scaling (MDS) plot of Bray-Curtis distances between 113 Northern elephant seal samples across three time points (color) and the two sexes (shape). All samples were normalized using the variance stabilizing normalization implemented in DeSeq2 and axes were length-scaled to reflect the Eigenvalues of the underlying principle coordinates.

abundant across time points and sexes using the DESeq2 method (Love, Huber, & Anders, 2014). We provide a detailed description of all differential abundances including figures in the Supplementary Material 2. Overall, the majority of significant changes in bacterial abundance of both males and females happens between T1 and T2 (F: $n = 100$, M: $n = 106$) with less than half as many taxa changing in abundance from T2 to T3 (F: $n = 43$, M: $n = 26$). Most of these changes happen in taxa belonging to the *Clostridia* and *Bacteroidia* in both sexes and both age transitions (see Supplementary Figure 7).

The number of taxa with significant differential taxa between males and females was high at all time points (T1: $n = 96$, T2: $n = 102$, T3: $n = 80$, see Supplementary Figure 8), and more than a third of them belonged to the *Clostridia* Family *XI* and the family *Ruminococcaceae*. These numbers also reflect that microbiota differences between the sexes were fairly stable during the sampling period (cf Figure 5.2).

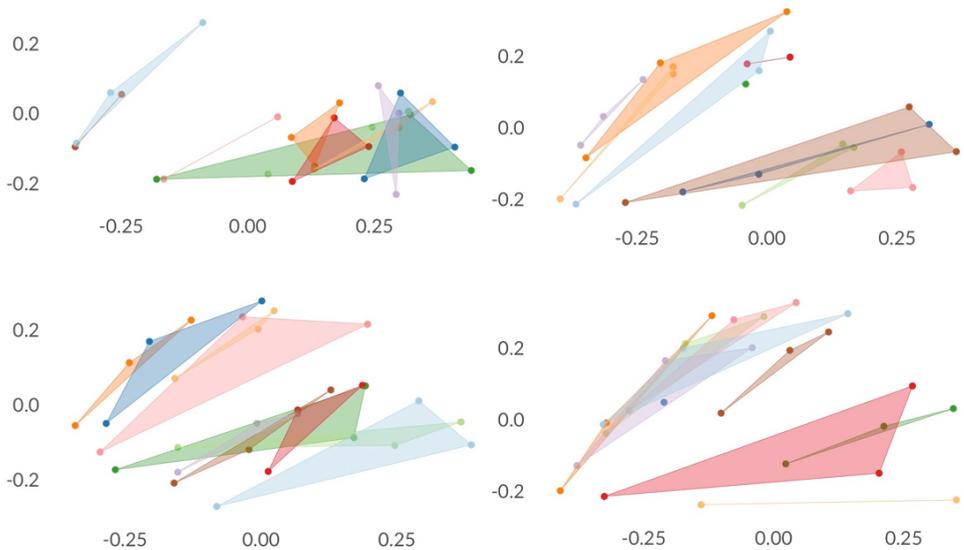


Figure 5.4: Gut microbiota similarity of samples from the same host. Shown is the same MDS as in Figure 5.1, which was split into four plots to avoid over-plotting. Consequently, each plot is showing only a quarter of the samples, while samples from the same individual are shown in the same color within each subplot.

Beta diversity of gut microbiota is repeatable for individual hosts

To investigate the stability of individual gut microbiota during development, we calculated the repeatability of microbiota beta diversity. We found that the gut microbiota beta diversity of individuals shows a low but significant repeatability ($r = 0.32$, 95% CI [0.1, 0.54]) which is also apparent in the similarity between samples from the same host depicted in Figure 5.3. Consequently, Northern elephant seal pups host individually distinct gut microbiotas, but those are still changing drastically throughout the weaning period, as detailed in Figure 5.2 and below.

Alpha diversity is constant over time, higher for males and not repeatable

Bacterial diversity is frequently quantified in microbiota studies and is usually found to change quite drastically during the development of mammals (Clark et al., 2015; O'Toole and Jeffery, 2015). As a measure of alpha diversity we quantified the Shannon diversity, which takes into account both species richness but also the relative abundances of different species. To investigate the factors impacting microbial diversity, we constructed a Gaussian mixed model of Shannon

diversity with sex and time point as fixed effects as well as host as random effect to investigate patterns of change across sexes. The model explained little variation overall ($R^2 = 0.06$, 95% CI [0.01, 0.18]), but revealed a higher diversity for males than for females ($\beta = 0.20$, 95% CI [0.03, 0.39]). Shannon diversity however did not change between any two time points (T2 vs. T1: ($\beta = 0.12$, 95% CI [-0.10, 0.34], T3 vs. T1: $\beta = 0.12$, 95% CI [-0.07, 0.34]). These patterns are also shown as boxplots alongside the raw data in Figure 5.5. Contrary to individual microbial composition (beta diversity), the alpha diversity of individuals was also not repeatable ($r = 0.1$, 95% CI [0.00, 0.3]).

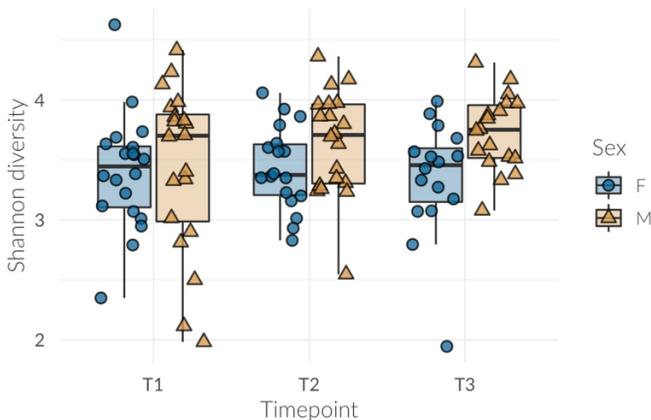


Figure 5.5: Gut microbiota diversity of Northern elephant seal pups across sex and time. Shown is the Shannon diversity of untransformed and unfiltered reads across three successive sampling-points during the weaning period of northern elephant seals, split by sex. The beige triangles are samples from males and the blue circles are samples from females. Also shown are Tukey boxplots.

Genetic relatedness is correlated with bacterial similarity, but only in males

A fundamental theme in microbial ecology is the importance of host genotype for the formation of the gut microbiota. We approached this question by quantifying the correlation between host genetic relatedness and microbial similarity (Figure 5.6). Surprisingly, Mantel tests showed a significant association in males ($r = 0.26$, CI [0.17, 0.34], $p = 0.0013$), and this relationship was visible across all three time points (Supplementary Figure 1). In contrast, we found no relationship in females ($r = 0.06$, CI [0.00, 0.12], $p = 0.41$) and this also appeared to hold true within each time point (Supplementary Figure 4). We followed up with a more in-depth

investigation and explored how many taxa might be influenced by host genetics. Therefore, we calculated the mantel correlation between genetic relatedness and microbial similarity based on an increasing number of taxa, starting with the two most abundant (relative abundance) and iteratively increasing the number by the next two most abundant taxa until we reached the full dataset (Figure 5.7). For females, the pattern across all subsets reflects the results from the full dataset and does not show a significant association between genetic relatedness and bacterial similarity. For males however, an interesting pattern was revealed. Relatively few taxa contribute strongly to the overall correlation and a peak was not reached until the 300 most abundant taxa were included in the analysis. This might reflect that a large proportion of taxa are at least slightly impacted by the host genotype as they contribute iteratively to an increasingly strong correlation.

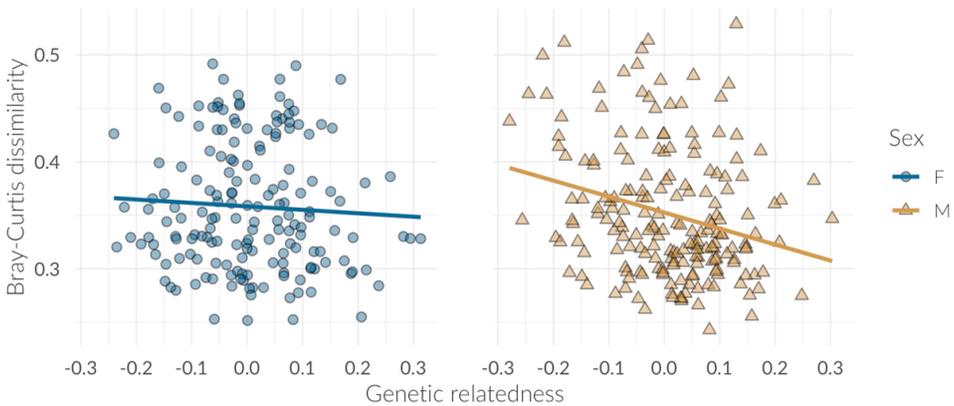


Figure 5.6: Relationship between pairwise gut microbiota dissimilarities and genetic relatedness of Northern elephant seal pups. For every individual, the microbial data across all time points have been merged by summing up taxa abundances. These data were then transformed using the variance-stabilising transformation in DEseq2 before calculating Bray-Curtis dissimilarities. Genetic relatedness was calculated based on microsatellite markers using the Loiselle estimator, with higher values representing higher genetic relatedness.

Discussion

Microbiota studies in wild populations are key to gain an understanding of the eco-evolutionary role of animal-microbe relationships (Hird, 2017). However, these studies can be difficult to interpret as environmental factors, such as diet (David et al., 2014b) can overshadow biological effects. Here, we conducted an in-depth investigation of the gut microbiota, its development,

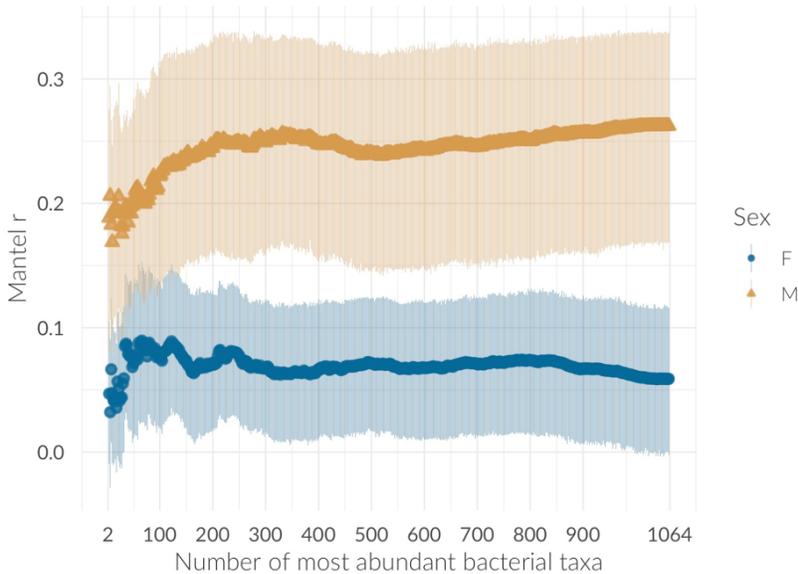


Figure 5.7: Correlation between gut microbial similarity and genetic relatedness of Northern elephant seal pups for an increasing number of bacterial taxa, split by pup sex. Each data point shows the correlation between the microbial Bray-Curtis dissimilarity and genetic relatedness with 95% confidence intervals calculated by non-parametric bootstrapping of samples. Bray-Curtis dissimilarities were calculated based on an increasing number of bacterial taxa, starting with the two taxa yielding the highest relative abundances across all samples and iteratively increasing the number always by the next two most abundant taxa up to 1064, the complete dataset.

host-genetic effects, and sex-specific composition in weaned northern elephant seals. In particular, we designed our study to control for differences in diet, habitat, and behaviour, making it possible to shed light on the more intricate and intrinsic factors shaping the gut microbiota. We found four main factors which impact early gut microbiota, including age, sex, host, and the genetic relatedness between hosts.

First of all, we showed that the gut microbiota at the time of weaning is already fairly complex with an average of nearly 300 taxa per individual from 14 different phyla. Four of these phyla are highly abundant, the *Bacteroidetes*, *Firmicutes*, *Fusobacteria*, and *Proteobacteria*, and have previously been shown to be the main phyla in most pinniped gut microbiota (Bik et al., 2016; Glad et al., 2010; Nelson et al., 2013; Numberger et al., 2016). However, their relative contribution varies drastically across studies, and this is likely to be a consequence of differences in sampling methods, age of the study individuals, and a lack of longitudinal studies.

The patterns of development of early gut microbiota are difficult to compare with other

mammals, as very few studies have conducted longitudinal sampling of gut microbiota during early life. One mammalian system where the gut microbiota has been studied before and during the weaning period is that of pigs, and the similarities with our results are striking. First of all, the most abundant phyla in weaned pigs, as in this study, were *Firmicutes* and *Bacteroidetes* (Alain B Pajarillo et al., 2014). Second, the dietary transition from nursing to weaning was found to be reflected in a strong decrease in the genus *Bacteroides* combined with a substantial increase in *Prevotella* (Frese et al., 2015; Alain B Pajarillo et al., 2014). *Bacteroides* have been shown to break down milk oligosaccharides and are therefore important during nursing (Marcobal and Sonnenburg, 2012; Marcobal et al., 2011), while *Prevotella* are associated with plant polysaccharide consumption and are potentially important for the digestion of solid food (Ivarsson et al., 2014). We found precisely the same pattern, with strongly decreasing *Bacteroides* and substantially increasing *Prevotella*, which became the second most abundant genus at T3. However, there are several interesting points here: (1) while *Bacteroides* decreased in both males and females, *Prevotella* increased much more markedly in males than in females. This might simply be due to a slower gut microbiota development in females but could also reflect a sex-specific adaptation. (2) In pigs, an increase in *Prevotella* was associated with the transition to a non-milk diet (Frese et al., 2015), but elephant seal weaners only start feeding at sea and have not yet transitioned to a solid diet yet. Consequently, diet cannot be the cause of an increase in *Prevotella*. This hints towards a case of co-evolution between host and microbe but will require further study. (3) Unlike pigs, northern elephant seals are not omnivorous but rather feed on squid and fish, which is why the increase in *Prevotella* is puzzling in the light of current knowledge about the genus as a harvester of plant based fibers and polysaccharides (Gorvitovskaia et al., 2016; Wu et al., 2011). However, we do not know how different the *Prevotella* species and their associated functions are between these two studies.

Despite these substantial changes in the structure of gut microbial communities, which include abundance changes, but also colonisation and extinction events, the alpha diversity of gut microbiota was highly stable during the weeks after weaning. This is surprising as alpha diversity across longitudinal microbe samples is usually found to change drastically within both short and longer time scales (Ballou et al., 2016; Frese et al., 2015; Kundu et al., 2017; Videvall et al., 2018). Consequently, the stability observed in this study might reflect the absence of diet during our sampling period, which can be a major source of new microbial diversity (Pantoja-Feliciano et al., 2013).

Patterns of beta and alpha diversity are frequently used to gain fundamental insights into the structure of microbial communities and the factors impacting them. Nevertheless, none of these measures provide direct insights into the stability of individual gut microbiota across

time. This is a question that is usually addressed by calculating intra-class coefficients or repeatabilities. However, while the repeatability of alpha diversity can be readily calculated using existing methods (Nakagawa and Schielzeth, 2010), the repeatability of beta diversity is a more complex measure quantified as a distance matrix rather than a one-dimensional vector. Here, we established a novel method by using the estimated multivariate sum of squares from a PERMANOVA (Anderson, 2001) model to calculate an ANOVA based repeatability (Nakagawa and Schielzeth, 2010). Interestingly, while the individual repeatability of gut microbial composition, measured as beta diversity, was relatively low but clearly present, the alpha diversity of individual gut microbiota was not at all repeatable. As a consequence, each individual elephant seal pup appears to harbour a unique core gut microbiota, although this uniqueness is not reflected in its diversity but rather in the continuous presence of core microbiota.

To understand the role of host-microbe interaction for ecology and evolution it is critical to examine not just the environmental effects on microbiota composition, but also individual specific effects that could provide mechanisms for co-evolution. A particularly strong factor among which individuals of a species differ is their sexual identity. However, sex differences on gut microbiota seem to be negligible or non-existent in wild populations (Bennett et al., 2016; Bobbie et al., 2017; Maurice et al., 2015; Ren et al., 2017; Tung et al., 2015; Videvall et al., 2018). This is surprising, given that the two sexes usually face very different selection pressures which manifest in strong sex-specific adaptations. In stark contrast to the literature, we found sex to be a strong and early determinant of gut microbiota variation in elephant seals and that this difference is likely to be caused by intrinsic rather than environmental factors. Before our first sampling, all elephant seal pups remained close to their mothers to nurse, so there would be little variation in behaviour, diet, or social interactions. However, even directly after weaning, males and females host very different gut microbiotas, despite the sexes not yet differing in body size or secondary sexual features. Consequently, sex-specific microbiotas could be due to very early intestinal or immunological adaptations related to extremely different life-histories of male and female elephant seals. Whether the sex-specific bacteria are directly related to different adult feeding strategies or immunological challenges will need to be examined in future studies. However, given that sexually mature males presumably face more non-native bacteria and pathogens while feeding along their longer migratory routes than females, it is possible that our results are indicative of higher immune tolerance in the enteric mucosa of male pups compared to females.

An unanswered question to date is how strongly host genetics impact the composition of the gut microbiota in wild populations. In humans and mice, genome-wide association studies revealed that at least a small proportion of the microbiota is genetically determined (Goodrich

et al., 2016; Kurilshikov et al., 2017). At a broader scale, it has been found that genetically more similar humans harbour more similar gut microbial communities (Ak, 2001), a pattern which seems to be difficult to replicate in wild populations (e.g. Degnan et al. 2012). In the wild, however, environmental factors such as diet, habitat or social behaviour are strong and are likely to mask the smaller and more intricate effects of genetics. We found that host genetic effects are strong and sex-specific. More closely related males hosted a more similar gut microbiota. However, genetic relatedness and microbial similarity were uncorrelated in females. To exclude the chance of a false positive finding, we showed that given our number of genetic markers, the Loiselle relatedness estimator was unbiased (Supplementary Figure 6). Moreover, the sex-specific correlation between genetic relatedness and microbiota similarity held true even when replicated within each sampling time point. One explanation for the observed pattern could be that males show a faster development of traits relevant to their future life-histories, and that these include the gut microbiota. In this case, a correlation should also occur in females, but later in life. Another explanation could be that males and females have different selection pressures on gut microbiota. In particular, as a consequence of a highly-skewed mating system, males face great energetic and immunological challenges during the breeding season. This might facilitate stronger balancing selection on genes and potentially microbiota among males than among females, which is reflected in the strong correlation found in this study.

Overall, our study paves the way for future research on the more intricate ecological, evolutionary and genetic bases of host-microbe interactions in the wild. By controlling for variation in environmental factors such as diet and habitat, we were able to attribute microbial variation to intrinsic factors such as sex and genetics. Previous studies have failed to find differences in gut microbiome diversity between sexes. However, detectable effects are much more likely to occur in organisms with extensive intra-specific differences in life-histories such as the northern elephant seal, which provide an exceptional opportunity to study potential adaptive functions of the microbiota. Our results add light to current understanding of the symbiotic relationship between animals and their microbiota.

Materials and Methods

Study design and sample collection

We marked forty northern elephant seal pups and their mothers during the breeding season in February/March 2017 at Benito del Oeste, the westernmost island of the San Benito Archipelago

off the west coast of Baja California, Mexico. We observed mother-offspring pairs to determine pup weaning date, which occurs close to 28 days after birth (Reiter et al., 1978). At this moment we sampled the newly weaned pup (time point T1). To analyse the distal gut microbiota composition, we took rectal swabs using FLOQSwabs™, which were immediately stored in 70% EtOH, frozen at -20°C within a few hours after collection and subsequently stored at -80°C two months after the end of the field season. To analyse the microbiota composition with respect to the genetic relatedness of the individuals, we collected a small skin sample (9 mm^2) from the flipper of each pup and stored it in sterile cryogenic vials with 70% EtOH. The vials were frozen at -20°C within a few hours after collection and subsequently stored at -80°C . During the T1 sampling period, we collected rectal swabs and skin samples from 40 pups, which were marked with plastic flipper tags with a unique ID number. Subsequently, we observed the pups on a daily basis and captured them after 15 days (T2) and 30 days (T3) to collect two additional second rectal swabs. The entire sampling scheme spanned the around two months long fasting period during which the weaned pups stay ashore (Reiter et al., 1978). Throughout the field season, we lost six of the marked pups, as one died between T1 and T2, one between T2 and T3, one was not found after T1 and three pups were lost after T2, despite intensive searching effort. Thus, sample sizes were 40 pups at T1, 38 at T2, and 34 at T3. All sampling was conducted by approval of the Bioethics Committee and IACUC of the Autonomous University of Queretaro, and all capture and sampling procedures were carried out under permit DGVS 00091/17 issued by the Mexican Secretariat of the Environment and Natural Resources.

Host DNA extraction and microsatellite genotyping

Total genomic DNA was extracted from a sample of skin tissue for each individual using a standard chloroform extraction protocol and genotyped at 21 previously developed microsatellite loci (details in the Supplementary Information). We tested all microsatellite loci for deviations from Hardy-Weinberg equilibrium (HWE) using exact tests based on Monte Carlo simulations implemented in pegas (Paradis, 2010) and applied a false discovery rate correction (Benjamini and Hochberg, 1995) of the resulting *p*-values. All 21 loci were retained in the final dataset as no locus was out of HWE.

Bacterial DNA extraction, library preparation and sequencing

We extracted DNA from 112 swabs using the QIAamp PowerFecal DNA Kit (Qiagen), and amplified a 300 bp of the V3 and V4 regions of the 16S rRNA gene. The amplicon libraries were prepared as follows: 1-10 ng of DNA extract (total volume 1 μ l), 15 pmol of each forward primer 341F 5'-NNNNNNNNNTCCTACGGGNGGCWGCAG and reverse primer 785R 5'-NNNNNNNNNTGACTACHVGGGTATCTAAKCC in 20 μ L volume of 1 x MyTaq buffer containing 1.5 units MyTaq DNA polymerase (Bioline) and 2 μ l of BioStabII PCR Enhancer (Sigma). For each sample, the forward and reverse primers had the same 10-nt barcode sequence. PCRs were carried out for 30 cycles of 1 min 96°C pre-denaturation; 96°C for 15 s, 50°C for 30 s, 70°C for 90 s. The DNA concentration of the amplicons of interest was determined by gel electrophoresis. About 20 ng amplicon DNA of each sample were pooled for up to 48 samples carrying different barcodes. The amplicon pools were purified with one volume of AMPure XP beads (Agencourt) to remove primer dimer and other small miss-priming products, followed by an additional purification on MiniElute columns (Qiagen). About 100 ng of each purified amplicon pool DNA was used to construct Illumina libraries using the Ovation Rapid DR Multiplex System 1-96 (NuGEN). Illumina libraries were pooled and size selected by preparative gel electrophoresis. Sequencing was performed on an Illumina MiSeq platform using V3 Chemistry - 2x300 bp read length (Illumina). DNA extraction, library preparation and sequencing were carried out by LGC Genomics in Berlin.

Bioinformatics pipeline

The 16s sequences in FASTQ format were demultiplexed using the Illumina bcl2fastq 2.16.1.14 software while allowing up to 2 mismatches or Ns in the barcode. Reads were sorted according to barcodes, allowing up to 1 mismatch per barcode and the barcode was then clipped from the sequence. Reads with missing, one-sided, or conflicting barcode pairs were discarded. Adapters were clipped using cutadapt 1.13 (Martin, 2011) and all reads smaller than 100 bp were filtered out. Amplicon primers were detected while allowing for up to three mismatches, and primer pairs (Forward-Reverse or Reverse-Forward) had to be present in the sequence fragments. If primer-dimers were detected, the outer primer copies were clipped from the sequence and the sequence fragments were turned into forward-reverse primer orientation after removing the primer sequence. We used DADA2 1.8 (Callahan et al., 2016a) for further filtering, processing and inferring Amplicon sequence variants (ASV), following the authors' published workflow (Callahan et al., 2016b). Unlike the traditional grouping into operational taxonomic units (OTUs), ASVs are exact sequence variants and have the compelling advantages of higher

taxonomic resolution as well as reproducibility and reusability across studies (Callahan et al., 2017). After visually inspecting the quality profiles of all reads, we used DADA2's `filterAndTrim` function to trim R1 and R2 sequences to 220 and 230 base pairs respectively, and to filter all reads with more than 2 expected errors (Edgar and Flyvbjerg, 2015). As DADA2 relies on a parametric error model, we used the `learnErrors` function to learn the error rate from the data and visually confirmed that the estimated error rates provided a good fit to the observed rates using `plotErrors` (Callahan et al., 2016b). After dereplication with `derepFastq`, we used the `dada` function for correcting substitution and indel errors, and for sample inference based on the pooled samples. Subsequently, we merged forward and reverse reads with a minimum overlap of 12 bp using `mergePairs` and constructed a sequence table with `makeSequenceTable`. After inspecting the distribution of sequence lengths across samples and considering a median full amplicon size of around 460 bp prior to primer clipping (Klindworth et al., 2013), primer-clipped sequences of lengths between 380 and 450 bp were retained. As a last filtering step, we removed chimeras with `removeBimeraDenovo` using the consensus method. We assigned taxa to the ASVs using the `assignTaxonomy` and `addSpecies` functions based on the SILVA database v128 (Quast et al., 2012). The resulting ASV table contained 2809 taxa in 112 samples.

Data processing and analysis

Microbial data

All subsequent analyses were conducted in R version 3.4.3 (R Core Team). As a first filtering step after taxonomic assignment, we discarded taxa classified as Mitochondria ($n = 3$) or Chloroplasts ($n = 8$) and taxa which could not be identified at the Class level ($n = 77$), as these are likely to contain sequencing errors. Depending on the analysis, we then used different filters and transformations for the ASV count data. To evaluate alpha diversity, we used the untransformed and unfiltered abundance of ASVs, and calculated Shannon diversities in `phyloseq` (McMurdie and Holmes, 2013). Based on a visual assessment of taxa abundance and prevalence (Supplementary Figure 5), we removed taxa that did not appear in at least three samples ($n = 982$) or had a total read count below 30 across all samples ($n = 683$).

Overall, 1063 taxa were retained in the final filtered ASV dataset. Before analysing microbiota similarities across groups, we applied the variance stabilising transformation (VST) in `DESeq2` (Love et al., 2014), which uses a negative binomial mixed modelling approach to account for differences in library size across samples and to disentangle the relationship between the variance and the mean inherent to count data. Compared to other normalisation and transformation methods traditionally applied to microbiota data, the VST has the advantage

of using all available data and is therefore preferable both to rarefying approaches (McMurdie and Holmes, 2014) but also to a transformation into relative abundances, which still yield the problem of heteroscedasticity (Love et al., 2014). Based on the VS transformed data, we calculated Bray-Curtis dissimilarities (Bray and Curtis, 1957) between samples to visualise group differences using principle coordinate analysis (PCoA). We then statistically evaluated the microbiota composition across sexes, time points and individuals using a permutational multivariate analysis of variance (PERMANOVA, Anderson, 2001) and 1000 permutations with the *adonis* function in *vegan* (Oksanen et al., 2017). This approach is analogous to a parametric analysis of variance in that it partitions distance matrices into sources of variation and produces a pseudo-F value, the significance of which can be tested using a permutation test. As group differences detected using a PERMANOVA can be caused by variation in dispersion across groups rather than mean values (Anderson, 2001), we tested for homogeneity of group dispersions using *betadisper* in *vegan* (Anderson, 2001; Oksanen et al., 2017) with post-hoc comparisons between specific contrasts tested with Tukey's 'Honest Significant Differences' method.

A main interest in microbial research is to determine the specific bacterial taxa that differ among groups. To calculate these differential abundances, we used the filtered but untransformed ASV data in combination with the DESeq2 method (Love et al., 2014). DESeq2 models abundance data such as microbial counts using a negative binomial distribution, estimates log fold changes between groups based on the specified model, and corrects the resulting p-values with a Benjamini and Hochberg false-discovery rate correction (Benjamini and Hochberg, 1995). As our ASV count matrix contained at least one zero in every row, we calculated the underlying size factors using the 'poscounts' estimator, which excludes zeros when calculating the geometric mean. To extract the appropriate group-specific contrasts, we fitted three different models and used a threshold of $p < 0.01$ to detect significant taxa. Specifically, for analysing differential abundances between time points but within sex, the first two models contained ASV data for just females and just males, respectively, while fitting both individual and time point in the model. To analyse and extract between-sex contrasts within each sampling time point, we constructed a third model by creating a new grouping factor as a combination of time point and sex, which was then fitted as predictor variable in the model.

To assess the factors that shape alpha diversity we calculated Shannon indices based on unfiltered and untransformed reads. We used a Gaussian mixed model in *lme4* (Bates et al., 2015) with Shannon diversity as response, sex and time point as fixed effects and individual as random effect. We calculated the R^2 based on Nakagawa and Schielzeth 2013 and assessed 95% confidence intervals around the R^2 and the model estimates using parametric bootstrapping

with 1000 replications. The individual adjusted repeatability including 95% CI was estimated with rptR (Stoffel et al., 2017), using the same model structure and 1000 bootstraps.

Genetic relatedness and microbial similarity

Pairwise genetic relatedness based on 21 microsatellite loci was calculated using Demerelate 0.93 (Kraemer & Gerlach, 2013). We used the Loiselle estimator (Loiselle et al., 1995) which is unbiased for small sample sizes and converged towards stable values for the number of loci used in this study (Supplementary Figure 6). To match the microbial data to the pairwise genetic relatedness matrix containing 40 individuals for further analyses, we merged the microbial data across the three time points for every individual by summing up the abundances of taxa. The 40 merged microbiota samples were then transformed using the variance-stabilising transformation in DESeq2 before calculating Bray-Curtis dissimilarities. Both the genetic relatedness matrix and the microbial dissimilarity matrix were then split by sex to calculate their correlation with a Mantel test using the ecodist 2.0.1 package (Goslee and Urban, 2007) using 10,000 bootstraps with the default resampling level of 0.9 to calculate confidence intervals and 10,000 permutations to test for statistical significance. As it is of major interest how many taxa are impacted by host genetics, we did a subsampling exercise. Specifically, we started calculating microbial similarities from the two most abundant taxa and calculated the correlation of the resulting microbial similarity matrix with genetic relatedness. Iteratively, we repeated this procedure while always adding the next two most abundant taxa up to the complete dataset containing 1064 taxa. Lastly, we wanted to know whether the correlation between genetic relatedness and microbial similarity changed across the three time points and if it was different between sexes. We therefore used the original unmerged dataset and subsetted both microbial data and genetic data six times to calculate and visualize the correlation for all three time points and both sexes.

Data code and availability

The documented analysis pipeline along with the raw data can be accessed via GitHub (https://github.com/mastoffel/nes_microbiome) and is fully reproducible.

Acknowledgements

This research was supported by a PhD studentship from Liverpool John Moores University. Field work was funded by a Genetics Society Heredity fieldwork grant to MAS and by a

grant from the Autonomous University of Queretaro (Fund for the Advancement of Scientific Research, I42908) awarded to KAW. Sequencing and laboratory work was funded by a DFG standard grant to JIH and by the Animal Behaviour Department at Bielefeld University.

Supplementary Information

Early sex-specific differentiation, development and genetic basis of the gut microbiota in a highly polygynous marine mammal, the northern elephant seal

Martin A Stoffel, Karina Acevedo-Whitehouse, Nami Morales-Duran, Stefanie Grosser, Nayden Chakarov, Oliver Krüger, Hazel J Nichols, Fernando R Elorriaga-Verplancken & Joseph I Hoffman

in preparation

This Supplementary Information contains Tables (5.1 - 5.4) and Figures (5.8 - 5.15) in Supplementary Material 1, the analyses of differential abundance in Supplementary Material 2 and the genotyping methods in Supplementary Material 3.

Supplementary Material 1 - Tables and Figures

Kingdom	Phylum	Class	Order	Family	Genus	Species	Mean rel. abundance %
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	NA	17.22
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	NA	7.84
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	mortiferum	2.04
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	mortiferum	1.99
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	fragilis	1.67
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	NA	1.22
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Anaerococcus	NA	1.04
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	NA	NA	0.97
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Odoribacter	NA	0.93
Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales	Succinivibrionaceae	Anaerobiospirillum	NA	0.89
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia/Shigella	NA	0.72
Bacteria	Proteobacteria	Gammaproteobacteria	Aeromonadales	Succinivibrionaceae	Anaerobiospirillum	NA	0.60
Bacteria	Firmicutes	Negativicutes	Selenomonadales	Acidaminococcaceae	Phascolarctobacterium	NA	0.51
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Peptoniphilus	NA	0.49
Bacteria	Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	Peptoclostridium	NA	0.49
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	NA	0.48
Bacteria	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Dialister	NA	0.44
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Anaerotruncus	NA	0.35
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	NA	0.14
Bacteria	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Collinsella	NA	0.11
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Parabacteroides	merdae	0.08

Table 5.1: Core microbiome shared between at least 95% of samples during sampling time point one (T1). Every row represents an ASV, and a full table including the exact sequences is provided as Supplementary data (core_microbiome_T1.txt). In some cases, a taxonomic level could not be assigned (NA). Shown is also the mean relative abundance of each core ASV across all samples at T1.

Kingdom	Phylum	Class	Order	Family	Genus	Species	Mean rel. abundance %
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Ezakiella	NA	4.37
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	NA	3.22
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	NA	2.75
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	fragilis	2.30
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Odoribacter	NA	1.40
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	mortiferum	1.02
Bacteria	Actinobacteria	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Lawsonella	NA	0.97
Bacteria	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Dialister	NA	0.94
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Peptoniphilus	NA	0.86
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	mortiferum	0.73
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Anaerococcus	NA	0.69
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcaceae_UCG-005	NA	0.50
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Anaerococcus	NA	0.48
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium	NA	0.35
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	NA	0.24

Table 5.2: Core microbiome shared between at least 95% of samples during sampling time point two (T2). Every row represents an ASV, and a full table including the exact sequences is provided as Supplementary data (core_microbiome_T2.txt). In some cases, a taxonomic level could not be assigned (NA). Shown is also the mean relative abundance of each core ASV across all samples at T2

Kingdom	Phylum	Class	Order	Family	Genus	Species	Mean rel. abundance %
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Ezakiella	NA	4.23
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	NA	4.22
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	NA	3.05
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Proteiniphilum	NA	3.02
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	NA	2.85
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	NA	2.68
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Porphyromonas	NA	2.00
Bacteria	Actinobacteria	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Lawsonella	NA	1.37
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	NA	1.11
Bacteria	Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Dialister	NA	0.90
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcaceae_UCG-005	NA	0.90
Bacteria	Actinobacteria	Actinobacteria	Corynebacteriales	Corynebacteriaceae	Lawsonella	NA	0.86
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Anaerococcus	NA	0.86
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Anaerococcus	NA	0.83
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	Odoribacter	NA	0.79
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	mortiferum	0.75
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcaceae_UCG-005	NA	0.71
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	mortiferum	0.66
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Peptoniphilus	NA	0.51
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	fragilis	0.48
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	NA	NA	0.46
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Anaerotruncus	NA	0.43
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcaceae_UCG-005	NA	0.40
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcaceae_UCG-005	NA	0.30
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	NA	0.28
Bacteria	Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	NA	NA	0.27
Bacteria	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	NA	0.14
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Anaerotruncus	NA	0.11
Bacteria	Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Ruminococcaceae_UCG-005	NA	0.11
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XIII	NA	NA	0.10
Bacteria	Firmicutes	Clostridia	Clostridiales	Family_XI	Peptoniphilus	NA	0.10
Bacteria	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	NA	0.10
Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Psychrobacter	NA	0.09
Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Psychrobacter	NA	0.09
Bacteria	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	NA	0.07

Table 5.3: Core microbiome shared between at least 95% of samples during sampling time point three (T3). Every row represents an ASV, and a full table including the exact sequences is provided as Supplementary data (core_microbiome_T3.txt). In some cases, a taxonomic level could not be assigned (NA). Shown is also the mean relative abundance of each core ASV across all samples at T3.

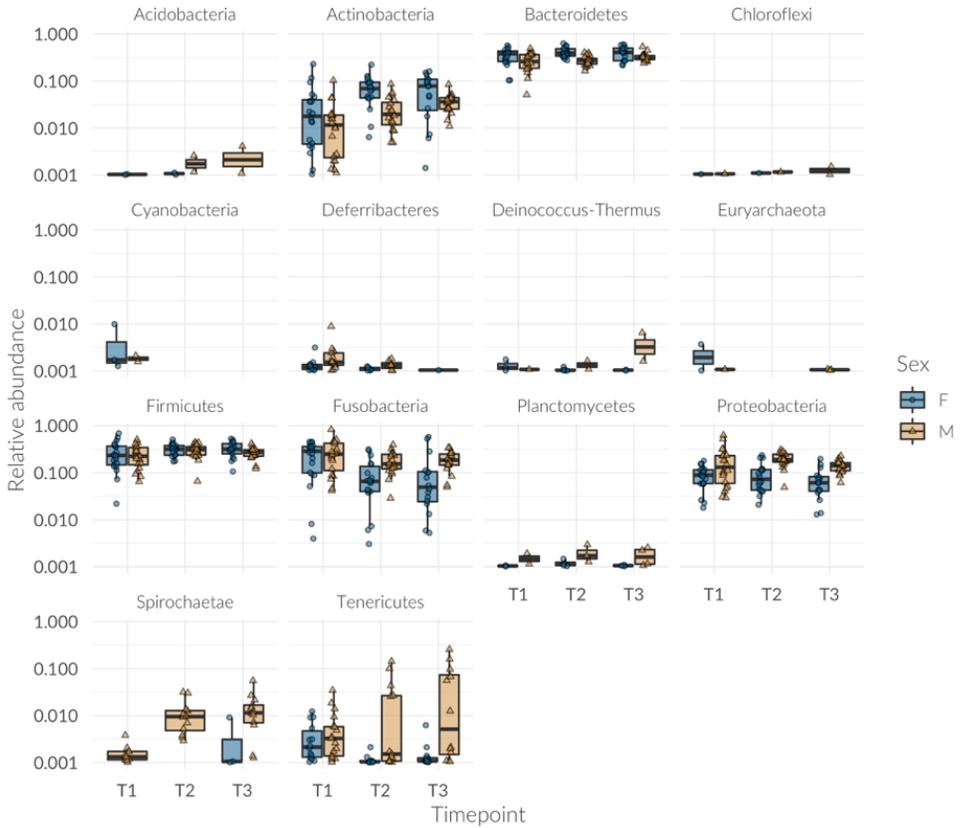


Figure 5.8: Relative abundance of all bacterial taxa analysed in this study on the Phylum rank across time and sex. Before visualization on the log scale, taxa with zero abundance were discarded and 0.001 added to all remaining relative abundances.

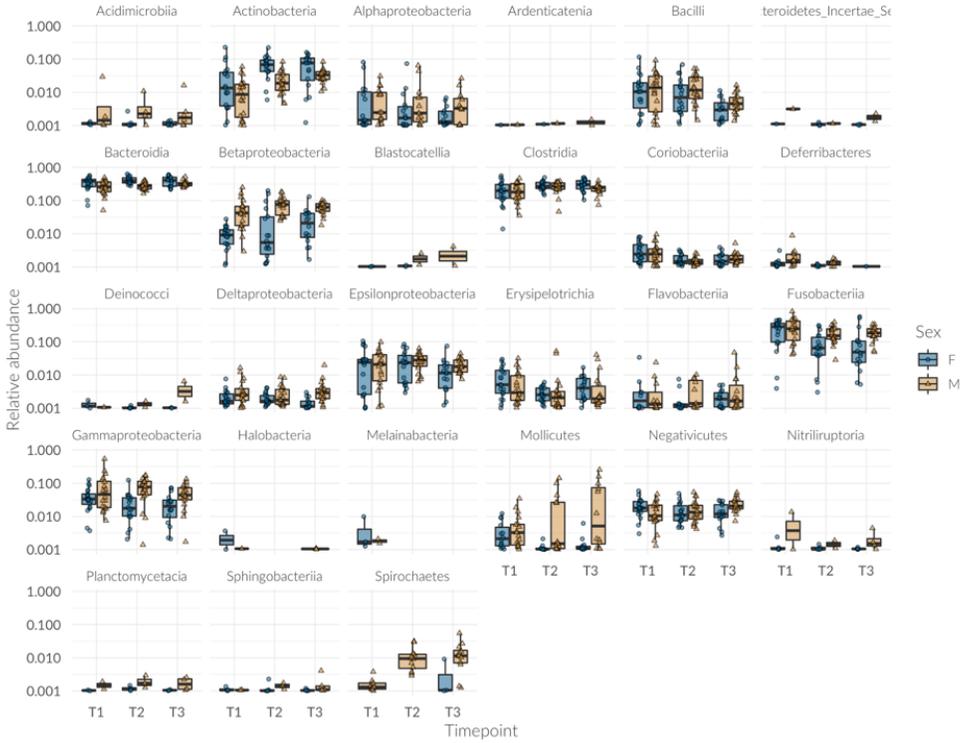


Figure 5.9: Relative abundance of all bacterial taxa analysed in this study on the Class rank across time and sex. Before visualization on the log scale, taxa with zero abundance were discarded and 0.001 added to all remaining relative abundances.

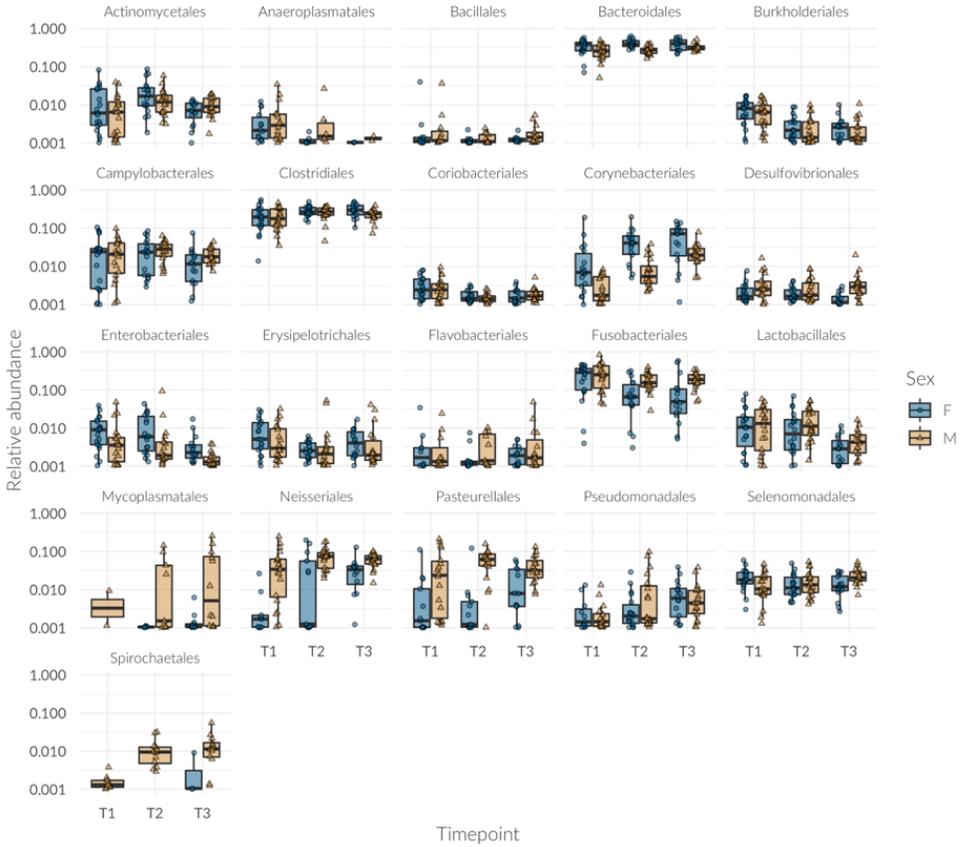


Figure 5.10: Relative abundance of bacterial taxa analysed in this study on the Order rank across time and sex. Before visualization on the log scale, taxa with zero abundance were discarded and 0.001 added to all remaining relative abundances. Shown is a subset of bacterial orders with interesting patterns and/or high prevalence across samples.

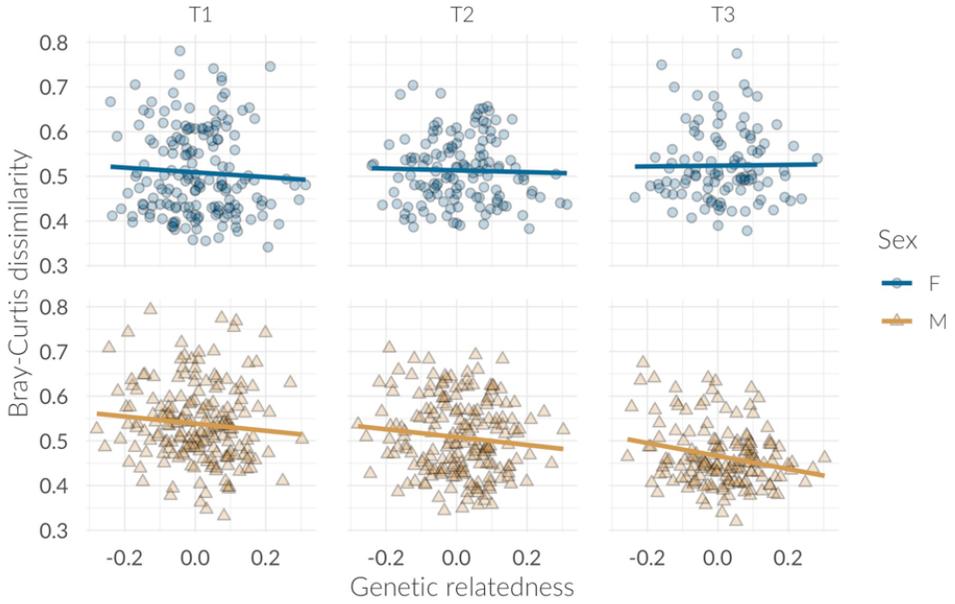


Figure 5.11: Correlation between microbial similarity and genetic relatedness at three time points, split by sex.

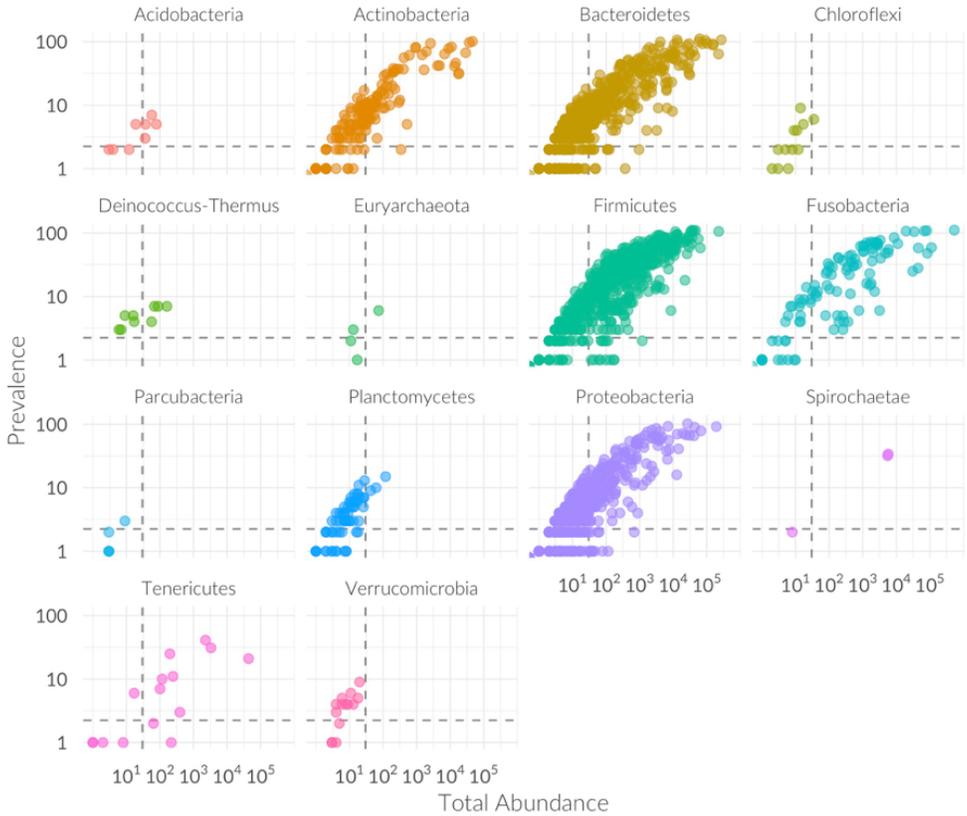


Figure 5.12: Prevalence and total abundance of taxa split by phylum. The horizontal and vertical dashed line represent the cut-offs for filtering, with taxa present in less than three individuals and/or an overall read count lower than 30 were discarded.

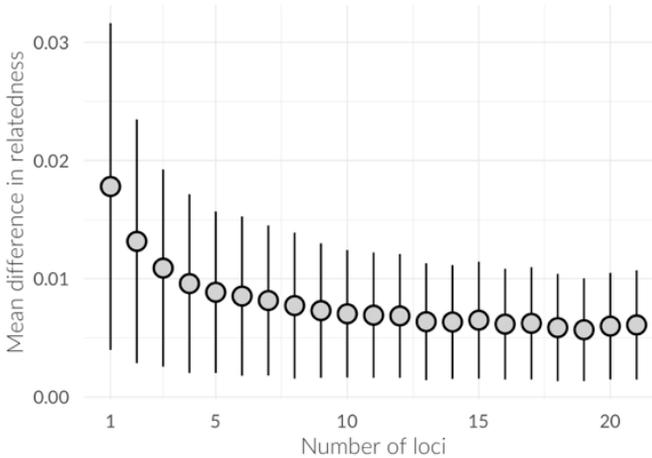


Figure 5.13: Sensitivity of the Loisel relatedness estimator to the number of loci used. Plotted are the mean and standard deviation (SD) of differences in pairwise genetic relatedness against the number of loci used. SDs were calculated from 1000 bootstrap replicates per locus number.

Supplementary Material 2 - Differential abundances of specific taxa

Differential abundance of specific taxa with age

Despite the apparent similarity of phyla across all three time points (Figure 1), on a finer scale a large number of bacterial taxa changed in abundance over time (Supplementary Figure 7 and 8). Most significant changes happen early on, with a large number of taxa for each sex varying from T1 to T2 (F: $n = 100$, M: $n = 106$) followed by a smaller number of significantly different abundances of taxa between T2 and T3 (F: $n = 43$, M: $n = 26$). On a taxonomic scale, most bacterial classes change drastically (Supplementary Figure 7). Between T1 and T2 most fluctuating bacteria belong to the Clostridia in both sexes (F: 47%, M: 44%), followed by Bacteroidia (F: 18%, M: 20%) and Fusobacteria (F: 13%, M: 12%), a pattern which is very similar for the second transition between T2 and T3 in males (Clostridia 35%, Bacteroidia 19%, Fusobacteria 15%) while in females the Bacteroidia (37%) change most drastically, more so than the Clostridia (30%) and Gammaproteobacteria (14%). Several interesting changes happen also in the less abundant bacterial classes. While Deferribacteres go extinct over time, the Spirochaetes increase largely in abundance in males (Supplementary Figure 3) and start to colonise also females at T3. The Bacilli and the Fusobacteria deplete quickly over time, while the Actinobacteria increase in their relative abundances by nearly 10-fold in females and more

than five-fold in males (Supplementary Figure 2).

Sex specific patterns of change

Bacterial communities in both sexes show similar dynamics throughout the weaning period, although the 'baseline' abundances of many species differ substantially (Supplementary Figures 1-3, Supplementary Figure 8). On the phylum level, the microbial shift from T1 to T2 in both females and males consists mostly of taxa belonging to the Firmicutes (F: 51%, M: 48%), followed by Bacteroidetes (F: 18%, M: 20%) and Fusobacteria in males (13%) but Proteobacteria in females (14%). Interestingly, a few bacterial families change drastically in abundance from T1 to T2 and make up a large part of significantly different taxa, especially the Ruminococcaceae (F: 22%, M: 19%), followed by Fusobacteriaceae (F: 12%, M: 10%) and Lachnospiraceae in females (12%) but Porphyromonadaceae in males (9%). Bacterial changes between T2 and T3 mainly occurred in the phyla Bacteroidetes (37%), Firmicutes (32%) and Proteobacteria (19%) in females and Firmicutes (46%), Bacteroidetes (19%), Fusobacteria (15%) and Proteobacteria (15%) in males. Most differentially abundant taxa belonged, similar to the first transition, to the Ruminococcaceae (F:12%, M:23%), Porphyromonadaceae (F:16%, M:12%), and the Lachnospiraceae (12%) in females as well as the Leptotrichiaceae (12%) in males.

Differential abundance of taxa across sexes

Despite similar dynamics over time, many taxa were significantly differentially abundant in males and females within all three time points (T1: n = 96, T2: n = 102, T3: n = 80, see Figure 5 and Figure 3). Although many phylogenetically different taxa contribute to these sex differences, three families contributed disproportionately much. The Clostridiales Family XI contributed 15% of differentially abundant taxa at T1, 16% at T2, and 18% at T3. The Ruminococcaceae contributed 15% of the taxa at T1, 19% at T2 and 13% at T3. The Porphyromonadaceae make up large differences at T1 (13%) and T2 (12%) but much less so at T3 (4%).

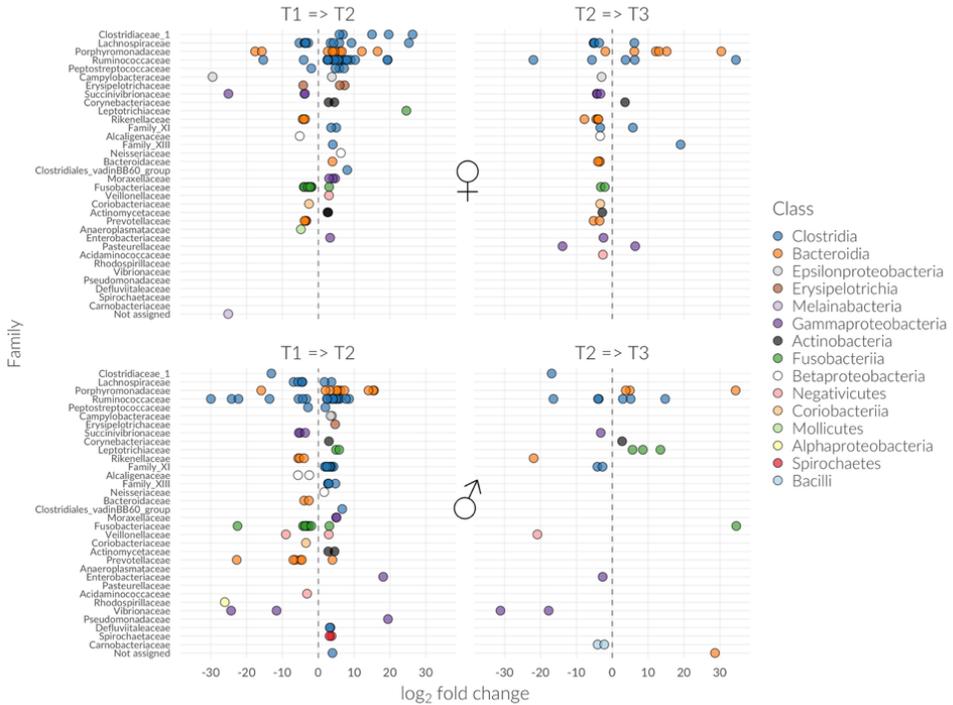


Figure 5.14: Differential abundance of taxa between sampling points, split by sex.

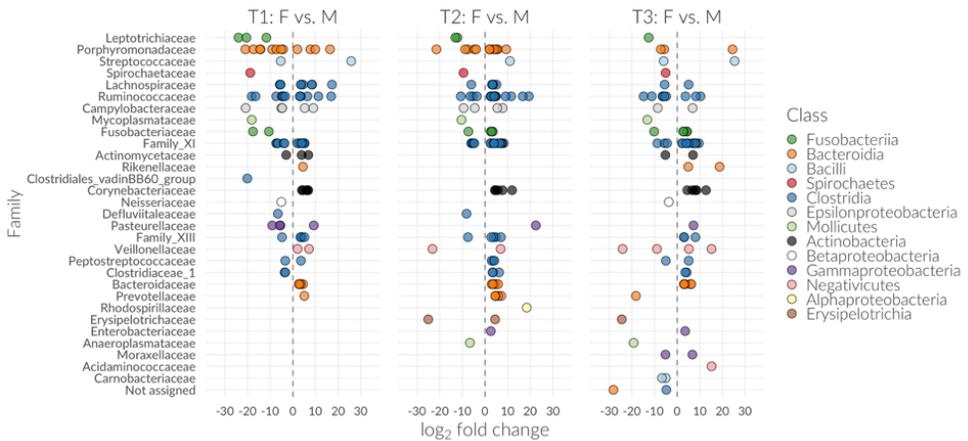


Figure 5.15: Differential abundance of microbes between sex, split by sampling points.

Supplementary Material 3 - Genotyping methods

Total genomic DNA of 40 *Mirounga angustirostris* samples was extracted from each sample using silica-gel membrane technology (DNeasy Blood and Tissue kit, Qiagen) and genotyped at 21 previously developed microsatellite loci (see Supplementary Table 4 for details). The microsatellite loci were amplified in singleplex or multiplex reactions. The following PCR profile was used: one cycle of 3 min at 94°C; 30 cycles of 30 s at 94°C, 30 s at T_a °C and 40 s at 72°C; 8 cycles of 30 s at 94°C, 30 s at 47°C and 40 s at 72°C; and one final cycle of 10 min at 72°C (see Supplementary Table 4 for T_a). Magnesium concentrations varied among the PCR mastermixes as shown in Supplementary Table 4. Fluorescently labelled PCR products were resolved by electrophoresis on an ABI 3730xl capillary sequencer and allele sizes were scored automatically using GeneMarker v1.85. To ensure high genotype quality, all traces were manually inspected and any incorrect calls were adjusted accordingly.

Table 5.4: Microsatellite loci genotyped in the Northern elephant seal. 'Multiplex' denotes the PCR mastermix into which each locus was multiplexed, 'Mg' denotes the concentration of magnesium used in the PCR mastermix and 'T_a' denotes the annealing temperature used. Loci not assigned to PCR multiplexes were amplified individually.

Locus	Literature source	Mg (mM)	T _a (°C)
71HDZ441	(Huebinger et al., 2007)	1.5	54
Hg4.2	(Allen et al., 1995)	1.5	56
Lw-8	(Davis et al., 2002)	1.5	47
ZcCgDh4.7	(Hernandez-Velazquez et al., 2005)	1.75	56
PV9	(Goodman, 1997)	2	53
ZcCgDh3.6	(Hernandez-Velazquez et al., 2005)	2	39
HI-8	(Davis et al., 2002)	2	53
PVC1	(Garza and Williamson, 2001)	1.5	52
71HDZ301	(Huebinger et al., 2007)	1.5	42
ZcCgDh1.8	(Hernandez-Velazquez et al., 2005)	1.5	42
ZcwA12	(Hoffman et al., 2007b)	1.75	49
ZcwF07	(Hoffman et al., 2007b)	1.75	49
Ag-9	(Hoffman et al., 2008)	2	57
ZcwC01	(Hoffman et al., 2007b)	2	57
ZcwE04	(Hoffman et al., 2007b)	2	52
ZcwG04	(Hoffman et al., 2007b)	2	52
Mango01	(Sanvito et al., 2013)	1.5	55
Mango44	(Sanvito et al., 2013)	1.5	55
Mango43	(Sanvito et al., 2013)	1.5	55
Mango35	(Sanvito et al., 2013)	1.5	53
Mango06	(Sanvito et al., 2013)	1.5	55
Mango09E19	(Sanvito et al., 2013)	1.5	52
PV9.1	This study	1.5	53



An old male elephant seal is lying far apart from the colony. He is not strong enough anymore to compete for a territory.

6

GAlignR: R PACKAGE

GAlignR: An R package for aligning gas-chromatography data for ecological and evolutionary studies

Meinolf Ottensmann*, Martin A Stoffel*, Hazel J Nichols & Joseph I Hoffman

PLOS ONE. 13: e0198311. 2018 *contributed equally

Abstract

Chemical cues are arguably the most fundamental means of animal communication and play an important role in mate choice and kin recognition. Consequently, there is growing interest in the use of gas chromatography (GC) to investigate the chemical basis of eco-evolutionary interactions. Both GC-MS (mass spectrometry) and FID (flame ionization detection) are commonly used to characterise the chemical composition of biological samples such as skin swabs. The resulting chromatograms comprise peaks that are separated according to their retention times and which represent different substances. Across chromatograms of different samples, homologous substances are expected to elute at similar retention times. However, random and often unavoidable experimental variation introduces noise, making the alignment of homologous peaks challenging, particularly with GC-FID data where mass spectral data are lacking. Here we present `GCalignR`, a user-friendly R package for aligning GC-FID data based on retention times. The package was developed specifically for ecological and evolutionary studies that seek to investigate similarity patterns across multiple and often highly variable biological samples, for example representing different sexes, age classes or reproductive stages. The package also implements dynamic visualisations to facilitate inspection and fine-tuning of the resulting alignments and can be integrated within a broader workflow in R to facilitate downstream multivariate analyses. We demonstrate an example workflow using empirical data from Antarctic fur seals and explore the impact of user-defined parameter values by calculating alignment error rates for multiple datasets. The resulting alignments had low error rates for most of the explored parameter space and we could also show that `GCalignR` performed equally well or better than other available software. We hope that `GCalignR` will help to simplify the processing of chemical datasets and improve the standardization and reproducibility of chemical analyses in studies of animal chemical communication and related fields.

Introduction

Chemical cues are arguably the most common mode of communication among animals (Wyatt, 2014*b*). In the fields of animal ecology and evolution, increasing numbers of studies have therefore been using approaches like gas chromatography (GC) to characterise the chemical composition of body odours and scent marks. These studies have shown that a variety of cues are chemically encoded, including phylogenetic relatedness (De Meulemeester et al., 2011), breeding status (Caspers et al., 2011), kinship (Bonadonna and Sanz-Aguilar, 2012; Krause et al., 2012; Stoffel et al., 2015) and genetic quality (Stoffel et al., 2015; Charpentier et al., 2010;

Leclaire et al., 2012).

GC vaporises a chemical sample and retards its components differentially based on their chemical properties while passing a gas through a column. The chemical composition of the sample can then be resolved using a number of approaches such as GC coupled to a flame ionization detector (GC-FID) or GC coupled to a mass spectrometer (GC-MS). GC-FID produces a chromatogram in which each substance is represented by a peak, the area of which is proportional to the concentration of that substance in the sample (McNair and Miller, 2011). Although GC-FID is a relatively inexpensive and high-throughput approach, the substances themselves can only be characterised according to their retention times, so their chemical composition remains effectively unknown. GC-MS similarly generates a chromatogram, but additionally provides spectral profiles corresponding to each peak, thereby allowing putative identification by comparison to databases of known substances. Both approaches have distinct advantages and disadvantages, but the low cost of GC-FID, coupled with the fact that most chemicals in non-model organisms do not reveal matches to databases containing known chemicals, has led to an increasing uptake of GC-FID in studies of wild populations (Boulay et al., 2007; Foitzik et al., 2007; Johnson et al., 2008; Reichle et al., 2013). GC-FID is particularly appropriate for studies seeking to characterise broad patterns of chemical similarity without reference to the exact nature of the chemicals involved.

As a prerequisite for any downstream analysis, homologous substances across samples need to be matched. Therefore, an important step in the processing of the chemical data is to construct a so called peak list, a matrix containing the relative abundances of each homologous substance across all of the samples. With GC-MS, homologous substances can be identified on the basis of both their retention times and the accompanying spectral information. However, with GC-FID, homologous substances can only be identified based on their retention times. This can be challenging because these retention times are often perturbed by subtle, random and often unavoidable experimental variation including changes in ambient temperature, flow rate of the carrier gas and column ageing (Scott, 2003; Pierce et al., 2005).

Numerous algorithms have been developed for aligning MS data (reviewed by (Lange et al., 2008) and (Smith et al., 2013)). To provide an overview of breadth of currently available software that provide implementations of these algorithms for users, we conducted a literature search. First, we screened the review papers described above and selected all peer-reviewed manuscripts reporting programs that are publicly available. We excluded publications reporting algorithms that are not implemented in software, that are described as 'available on request' from the authors, or which could only be accessed via expired web links. Furthermore, we conducted Web of Science searches in October 2017 using the search terms 'retention time

`align*`, `'peak align*'` and `'peak match*'` and used the same search terms to interrogate the list of packages deposited on CRAN and Bioconductor. We recovered a total of 25 programs, which we characterised according to a number of relevant criteria, ranging from the type of data for which they were designed through the programming environment to the dimensions that are used for aligning peaks (S1 File). We found that the majority (92%) of these programs were developed specifically for aligning MS data. Among these, a large proportion (87%) make use of spectral information either by binning the data according to mass-over-charge values or by directly taking mass information into consideration for the alignment method. Consequently, these programs will not support GC-FID data due to the lack of spectral information, which is a required part of the input.

Only three of the programs described in S1 File claim to support a peak list format lacking MS data, thereby making them potentially suitable for aligning GC-FID data. However, two of these programs (`amsrpm` (Kirchner et al., 2007) and `ptw` (Bloemberg et al., 2010)) may not be well suited to GC-FID data for two main reasons. First, they conduct alignments strictly pairwise with respect to a pre-defined reference sample, because in general the focus is on a relatively small pool of substances that are expected to be present in most if not all samples (Johnson et al., 2003). However, applied to wild animal populations, GC-FID often yields high diversity datasets in which only a small subset of chemicals may be common to all individuals (Stoffel et al., 2015; Jordan et al., 2011). Second, these algorithms are known to be sensitive to variation in peak intensity, which is expected in GC-FID datasets and may contain important biological information (Stoffel et al., 2015; Jordan et al., 2011; Breed et al., 1995; Wong et al., 2014).

To tackle the above issues, a third program called GALIGNER was recently written in Java for aligning GC-FID data (Dellicour and Lecocq, 2013). This program appears to perform well based on three test datasets, each corresponding to a different bumblebee species (*Bombus spp.*). However, the underlying algorithm compares each peak with the following peak in the same sample and therefore cannot align the last peak (Dellicour and Lecocq, 2013). Moreover, with the increasing popularity of open source environments such as R, there is a growing need for software that can be easily integrated into broader workflows, where the source code can be modified and potentially further extended by the user, and where related tools like `rmarkdown` (Allaire et al., 2016) can be applied to maximise transparency and reproducibility (Hoffman, 2016). Furthermore, especially for GC-FID data where spectral data are not available, a useful addition would be to integrate dynamic visualisation tools into software to facilitate the evaluation and subsequent fine-tuning of alignment parameters. However, the vast majority of currently available software (80%) lack such tools (S1 File).

In order to determine which alignment tools are commonly used in the fields of ecology and evolution, we conducted a bibliographic survey, focusing on the journals ‘Animal Behaviour’ and ‘Proceedings of the Royal Society B’, which recovered a total of 38 studies using GC-FID or GC-MS to investigate scent profiles (see S2 File for details). None of these studies used any form of alignment tool but rather aligned and called the peaks manually (e.g. (Greene and Drea, 2014)), a time-consuming process that can be prone to bias (van Wilgenburg and Elgar, 2013) and detrimental to reproducibility.

To address the above issues, we developed `GCalignR`, an R package for aligning GC-FID data, but which can also align data generated using other detectors that allow to characterise peaks by retention times. The package implements a fast and objective method to cluster putatively homologous substances prior to multivariate statistical analyses. Using sophisticated visualisations, the resulting alignments can then be fine tuned. Finally, the package provides a seamless transition from the processing of the peak data through to downstream analysis within other widely used R packages for multivariate analysis, such as `vegan` (Oksanen et al., 2017).

In this paper, we present `GCalignR` and describe the underlying algorithms and their implementation within a suite of R functions. We provide an example workflow using a previously published chemical dataset of Antarctic fur seals (*Arctocephalus gazella*) that shows a clear distinction between animals from two separate breeding colonies (Stoffel et al., 2015). We then compare the performance of `GCalignR` with `GCALIGNER` based on the same three bumblebee datasets given in (Dellicour and Lecocq, 2013) and explore the sensitivity of `GCalignR` to user-defined alignment parameter values. Finally, we compared our alignment procedure with a very different approach —parametric time warping— which is commonly used in the fields of proteomics and metabolomics (Bloemberg et al., 2010; Wehrens et al., 2015).

Materials & Methods

Overview of the package

Figure 6.1 shows an overview of `GCalignR` in the context of a workflow for analysing GC-FID data within R. A number of steps are successively implemented, from checking the raw data through aligning peak lists and inspecting the resulting alignments to normalising the peak intensity measures prior to export into `vegan` (Oksanen et al., 2017). In brief, the alignment procedure is implemented in three consecutive steps that start by accounting for systematic shifts in retention times among samples and subsequently align individual peaks based on

variation in retention times across the whole dataset. For simplicity, this procedure is embedded within a single function `align_chromatograms` that allows the customisation of peak alignments by adjusting a combination of three parameters. The package vignettes provide a detailed description of all of the functions and their arguments and can be accessed via `browseVignettes('GCalignR')` after the package has been installed.

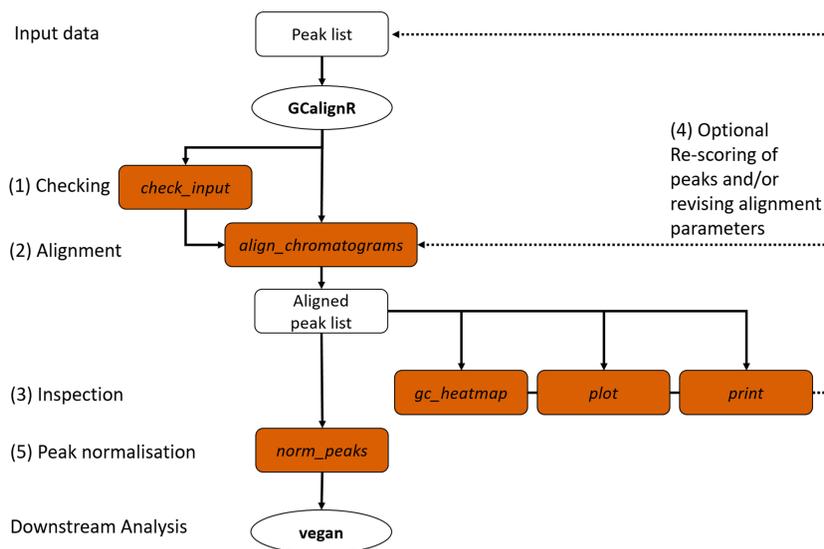


Figure 6.1: Overview of the GCalignR workflow. The steps listed in the main text are numbered from one to five and the filled boxes represent functions of the package (see main text for details).

Raw data format and conversion to working format

GC-FID produces raw data in the form of individual chromatograms that show the measured electric current over the time course of a separation run. Proprietary software provided by the manufacturers of GC-FID machines (e.g. ‘LabSolutions’, Shimadzu; ‘Xcalibur’, Thermo Fisher and ‘ChemStation’, Agilent Technologies) are then used to integrate and export peaks in the format of a table containing retention times and intensity values (e.g. peak area and height). Figure 6.2A shows chromatograms of three hypothetical samples where peaks have been integrated and annotated with retention times and peak heights. The corresponding input format comprising a table of retention times and peak heights is also shown. The working format of GCalignR is a retention time matrix in which each sample corresponds to a column and each peak corresponds to a row (see Figure 6.2B).

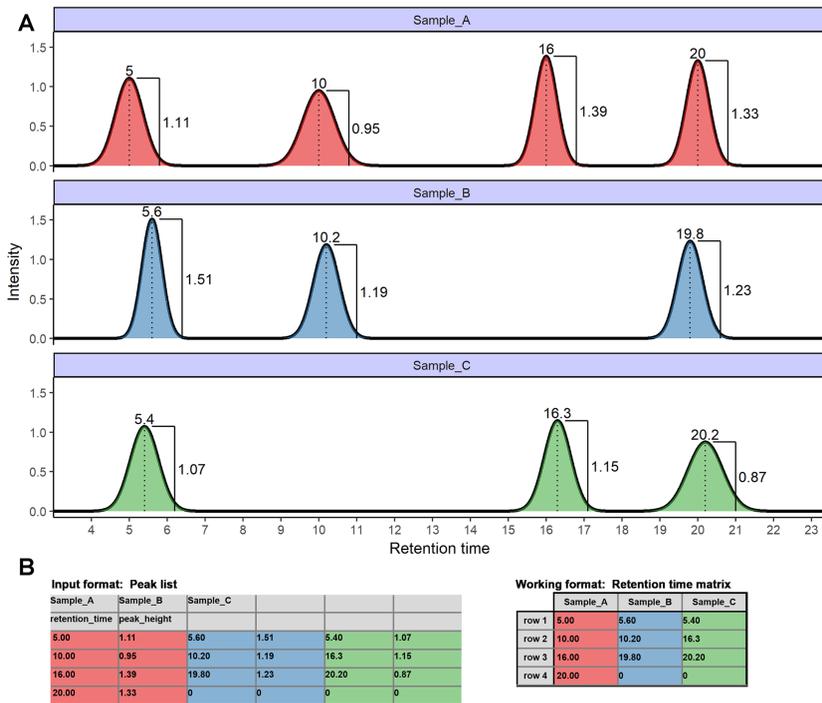


Figure 6.2: GC-FID data formats. A. Three hypothetical chromatograms are shown corresponding to samples A, B and C. Integrated peaks (filled areas) are annotated with retention times and peak heights. B. Using proprietary software (see main text), retention times and quantification measures like the peak height can be extracted and written to a peak list that contains sample identifiers ('Sample_A', 'Sample_B' and 'Sample_C'), variable names ('retention_time' and 'peak_height') and respective values. Computations described in this manuscript use a retention matrix as the working format.

Overview of the alignment algorithm

We developed an alignment procedure based on dynamic programming (Eddy, 2004) that involves three sequential steps to align and finally match peaks belonging to putatively homologous substances across samples (see Figure 6.4 for a flowchart and Figure 6.3 for a more detailed schematic representation). All of the raw code for implementing these steps is available via GitHub and CRAN and each step is described in detail below. The first step is to align each sample to a reference sample while maximising overall similarity through linear shifts of retention times. This procedure is often described in the literature as 'full alignment' (Bloembergen et al., 2010). In the second step, individual peaks are sorted into rows based on close similarity of their retention times, a procedure that is often referred to as 'partial alignment'

(Bloemberg et al., 2010). Finally, there is still a chance that homologous peaks can be sorted into different, but adjacent, rows in different samples, depending on the variability of their retention times (for empirical examples, see S3 File). Consequently, a third step merges rows representing putatively homologous substances.

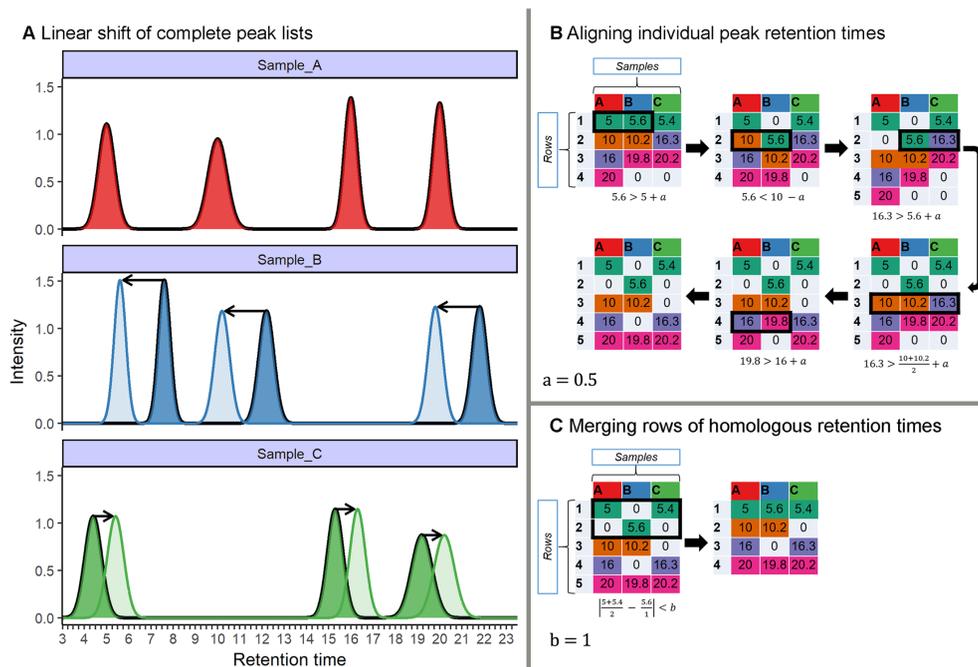


Figure 6.3: Overview of the three-step alignment algorithm implemented in GCalignR using a hypothetical dataset. A. Linear shifts are implemented to account for systematic drifts in retention times between each sample and the reference (Sample_A). In this hypothetical example, all of the peaks within Sample_B are shifted towards smaller retention times, while the peaks within Sample_C are shifted towards larger retention times. B and C work on retention time matrices, in which rows correspond to putative substances and columns correspond to samples. For illustrative purposes, each cell is colour coded to refer to the putative identity of each substance in the final alignment. B. Consecutive manipulations of the matrices are shown in clockwise order. Here, black rectangles indicate conflicts that are solved by manipulations of the matrices. Zeros indicate absence of peaks and are therefore not considered in computations. Peaks are aligned row by row according to a user-defined criterion, a (see main text for details). C. Rows of similar mean retention time are subsequently merged according to the user-defined criterion, b (see main text for details).

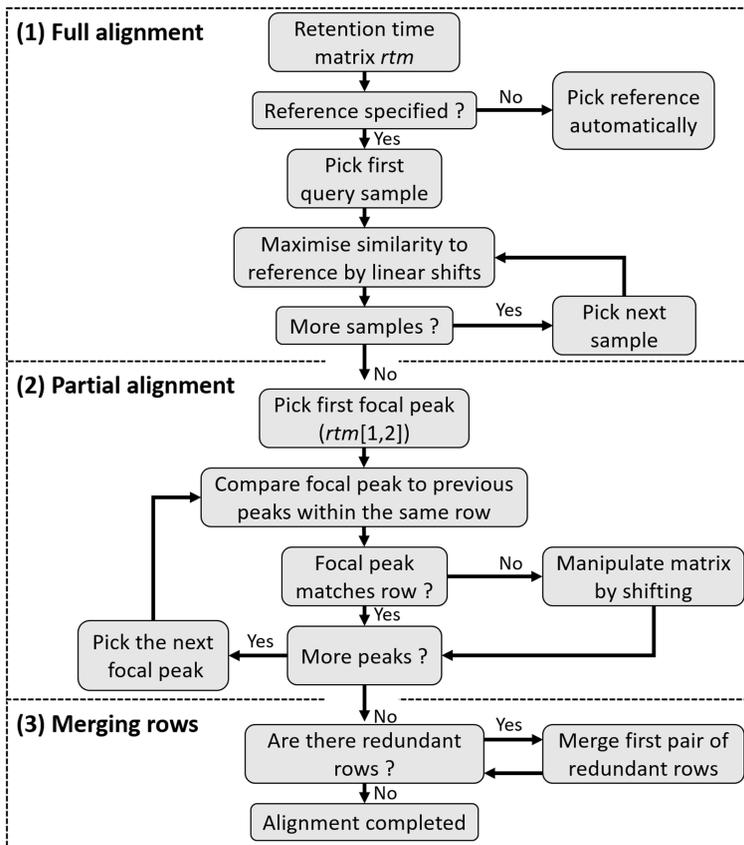


Figure 6.4: A flow chart showing the three sequential steps of the alignment algorithm of the peak alignment method.

Full alignment of peaks lists.

The first step in the alignment procedure consists of an algorithm that corrects systematic linear shifts between peaks of a query sample and a fixed reference to account for systematic shifts in retention times among samples (Figure 6.3A). Following the approach of Daszykowski et al. (Daszykowski et al., 2010), the sample that is most similar on average to the other samples can be automatically selected as a reference by choosing the sample with the lowest median deviation score weighted by the number of peaks to avoid a bias towards samples with few peaks:

$$\frac{1}{n} \sum_{i=1}^n [\min(\text{Ref}_i - \text{Query})] \quad (\text{Eq. 6.1})$$

where n is the number of retention times in the reference sample. Alternatively, the reference can be specified by the user. Using a simple warping method (Bloemberg et al., 2013), the complete peak list of the query is then linearly shifted within an user-defined retention time window with an interval of 0.01 minutes. For all of the shifts, the summed deviation in retention times between each reference peak and the nearest peak in the query is used to approximate similarity as follows:

$$\sum_{i=1}^n [\min(\text{Ref}_i - \text{Query})] \quad (\text{Eq. 6.2})$$

where n is the number of retention times in the reference sample. With increasing similarity, this score will converge towards zero the more homologous peaks are aligned, whereas peaks that are unique to either the query or the reference are expected to behave independently and will therefore have little effect on the overall score. The shift yielding to the smallest score is selected to transform retention times for the subsequent steps in the alignment (Figure 6.3B, C). As the effectiveness of this approach relies on a sufficient number of homologous peaks that can be used to detect linear drift, the performance of the algorithm may vary between datasets.

Partial alignment of peaks.

The second step in the alignment procedure aligns individual peaks across samples by comparing the peak retention times of each sample consecutively with the mean of all previous samples (Figure 6.3B) within the same row. If the focal cell within the matrix contains a retention time that is larger than the mean retention time of all previous cells within the same

row plus a user-defined threshold (Eq. 6.3), that cell is moved to the next row.

$$rt_m > \left(\frac{\sum_{i=1}^{m-1} rt_i}{m-1} \right) + a \quad (\text{Eq. 6.3})$$

where rt is the retention time; m is the focal cell and a is the user-defined threshold deviation from the mean retention time. If the focal cell contains a retention time that is smaller than the mean retention time of all previous cells within the same row minus a user-defined threshold (Eq. 6.4), all previous retention times are then moved to the next row.

$$rt_m < \left(\frac{\sum_{i=1}^{m-1} rt_i}{m-1} \right) - a \quad (\text{Eq. 6.4})$$

After the last retention time of a row has been evaluated, this procedure is repeated for the next row until the end of the retention time matrix is reached (Figure 6.3B).

Merging rows.

The third step in the alignment procedure accounts for the fact that a number of homologous peaks will be sorted into multiple rows that can be subsequently merged (Figure 6.3C). However, this results in a clear pattern whereby some of the samples will have a retention time in one of the rows while the other samples will have a retention time in an adjacent row (see S3 File). Consequently, pairs of rows can be merged when this does not cause any loss of information, an assumption that is true as long as no sample exists that contains peaks in both rows, (Figure 6.3C). The user can define a threshold value in minutes (i.e. parameter b in Figure 6.3C) that determines whether or not two such adjacent rows are merged. While the described pattern is unlikely to occur in large datasets purely by chance for non-homologous peaks, small datasets may require more strict threshold values to be selected.

Implementation of the alignment method

The alignment algorithms that are described above are all executed by the core function `align_chromatograms` based on the user-defined parameters shown in Table 6.1. Of these, parameters (`max_linear_shift`, `max_diff_peak2mean` and `min_diff_peak2peak`) can be adjusted by the user to fine-tune the alignment procedure. There are several additional parameters that allow for optional processing and filtering of the data independently of the alignment

Table 6.1: Mandatory arguments of the function `align_chromatograms`.

Parameter	Description
<code>data</code>	Path to a tab-delimited text file containing the chemical data. See the vignettes for examples including alternative input formats
<code>max_diff_peak2mean</code>	Numeric value defining the allowed deviation of the retention time of a focal peak from the mean of the corresponding row during partial peak alignment (See Eq. 6.3 and Eq. 6.4)
<code>max_linear_shift</code>	Numeric value defining the range that is considered for the adjustment of linear shifts in peak retention time across samples
<code>min_diff_peak2peak</code>	Numeric value defining the expected minimum difference in retention times across substances. Rows that are more similar than the threshold value will be merged as long as no conflict emerges due to the presence of peaks in more than one row within a single sample
<code>rt_col_name</code>	Name of the variable containing peak retention times. The name needs to correspond to variable included in the input file
<code>reference</code>	Name of the sample that will be used as reference to adjust linear shifts in peak retention times across samples. By default, a reference is automatically selected (see Materials and methods)
<code>sep</code>	Field separator character. By default, a tab-delimited text file is expected. Within R, type <code>?read.table</code> for a list of supported separators

procedure. For further details, the reader is referred to the accompanying vignettes (see S4 and S5 Files) and helpfiles of the R package.

Demonstration of the workflow

Here, we demonstrate a typical workflow in `GCalignR` using chemical data from skin swabs of 41 Antarctic fur seal (*Arctocephalus gazella*) mother-pup offspring pairs from two neighbouring breeding colonies at South Georgia in the South Atlantic. Sample collection and processing are described in detail in Stoffel et al. 2015. In brief, chemical samples were obtained by rubbing the cheek, underneath the eye, and behind the snout with a sterile cotton wool swab and preserved in ethanol stored prior to analysis. In order to account for possible contamination, two blank samples (cotton wool with ethanol) were processed and analysed using the same methodology. Peaks were integrated using ‘Xcalibur’ (Thermo Scientific). The chemical data associated with these samples are provided in the file `peak_data.txt`, which is distributed together with `GCalignR`. Additional data on colony membership and age-class are provided in the data frame `peak_factors.RData`.

Prior to peak alignment, the `check_input` function interrogates the input file for typical formatting errors and missing data. We encourage the use of unique names for samples consisting only of letters, numbers and underscores. If the data fail to pass this quality test, indicative warnings will be returned to assist the user in error correction. As this function is executed internally prior to alignment, the data need to pass this check before the alignment

can begin.

```
# load GCalignR
library(GCalignR)
# set the path to the input data
fpath <- system.file(dir = "extdata",
file = "peak_data.txt", package = "GCalignR")
# check for formatting problems
check_input(fpath)
```

In order to begin the alignment procedure, the following code needs to be executed:

```
aligned_peak_data <- align_chromatograms(data = peak_data,
rt_col_name = "time", max_diff_peak2mean = 0.02,
min_diff_peak2peak = 0.08, max_linear_shift = 0.05,
delete_single_peak = TRUE, blanks = c("C2", "C3"))
```

Here, we set `max_linear_shift` to 0.05, `max_diff_peak2mean` to 0.02 and `min_diff_peak2peak` to 0.08. By defining the argument `blanks`, we implemented the removal of all substances that are shared with the negative control samples from the aligned dataset. Furthermore, substances that are only present in a single sample were deleted from the dataset using the argument `delete_single_peak = TRUE` as these are not informative in analysing similarity pattern (Clark, 2006). Afterwards, a summary of the alignment process can be retrieved using the printing method, which summarises the function call including defaults that were not altered by the user. This provides all of the relevant information to retrace every step of the alignment procedure.

```
# verbal summary of the alignment
print(aligned_peak_data)
```

As alignment quality may vary with the parameter values selected by the user, the `plot` function can be used to output four diagnostic plots. These allow the user to explore how the parameter values affect the resulting alignment and can help to flag issues with the raw data.

```
# produces Fig 5
plot(aligned_peak_data)
```

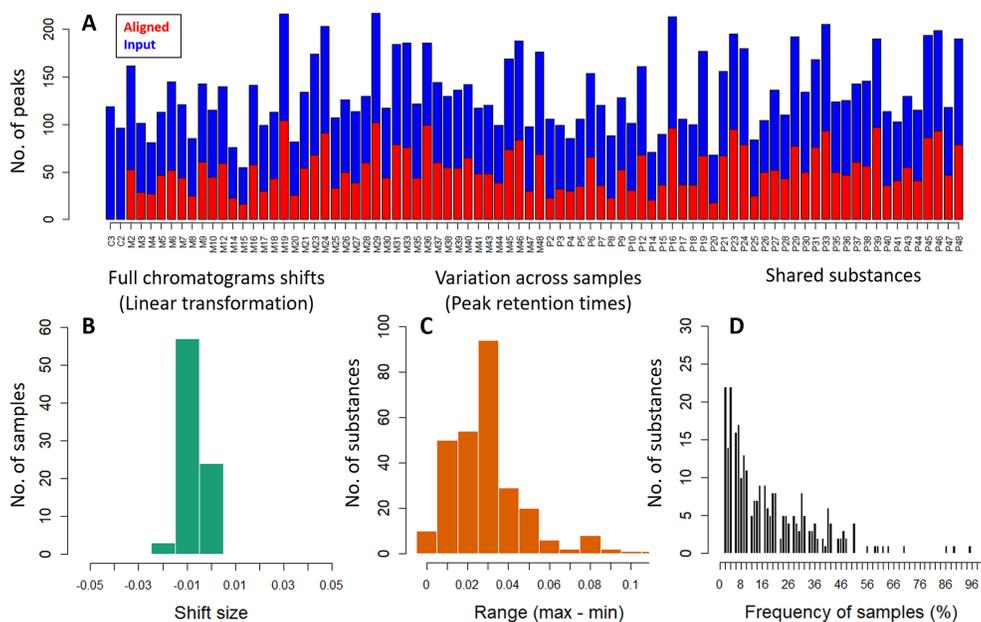


Figure 6.5: Diagnostic plots summarising the alignment of the Antarctic fur seal chemical dataset. A shows the number of peaks both prior to and after alignment; B shows a histogram of linear shifts across all samples; C shows the variation across samples in peak retention times; and D shows a frequency distribution of substances shared across samples.

The resulting output for the Antarctic fur seal chemical dataset, shown in Figure 6.5, reveals a number of pertinent patterns. Notably, the removal of substances shared with the negative controls or present in only one sample resulted in a substantial reduction in the total number of peaks present in each sample (Figure 6.5A). Furthermore, for the majority of the samples, either no linear shifts were required, or the implemented transformations were very small compared to the allowable range (Figure 6.5B). Additionally, the retention times of putatively homologous peaks in the aligned dataset were left-skewed, indicating that the majority of substances vary by less than 0.05 minutes (Figure 6.5C) but there was appreciable variation in the number of individuals in which a given substance was found (Figure 6.5D).

Additionally, the aligned data can be visualised using a heat map with the function `gc_heatmap`. Heat maps allow the user to inspect the distribution of aligned substances across samples and assist in fine-tuning of alignment parameters as described within the vignettes (see S4 and S5 Files).

```
gc_heatmap(aligned_peak_data)
```

Peak normalisation and downstream analyses

In order to account for differences in sample concentration, peak normalisation is commonly implemented as a pre-processing step in the analysis of olfactory profiles (Burgener et al., 2009; Setchell et al., 2011; Cristina Lorenzi et al., 2011). The GCalignR function `normalise_peaks` can therefore be used to normalise peak abundances by calculating the relative concentration of each substance in a sample. The abundance measure (e.g. peak area) needs to be specified as `conc_col_name` in the function call. By default, the output is returned in the format of a data frame that is ready to be used in downstream analyses.

```
# extract normalised peak area values
scent <- norm_peaks (data = aligned_peak_data, rt_col_name = "time",
conc_col_name = "area", out = "data.frame")
```

The output of GCalignR is compatible with other functionalities in R, thereby providing a seamless transition between packages. For example, downstream multivariate analyses can be conducted within the package `vegan` (Oksanen et al., 2017). To visualise patterns of chemical similarity within the Antarctic fur seal dataset in relation to breeding colony membership, we used non-metric multidimensional scaling (NMDS) based on a Bray-Curtis dissimilarity matrix in `vegan` after normalisation and log-transformation of the chemical data.

```
# log + 1 transformation
scent <- log (scent + 1)
# sorting by row names
scent <- scent[match(row.names(peak_factors),row.names(scent)),]
# Non-metric multidimensional scaling
scent_nmnds <- vegan::metaMDS(comm = scent, distance = "bray")
scent_nmnds <- as.data.frame(scent_nmnds[["points"]])
scent_nmnds <- cbind(scent_nmnds, colony = peak_factors[["colony"]])
```

The results results of the NMDS analysis are outputted to the data frame `scent_nmnds` and can be visualised using the package `ggplot2` (Wickham, 2009).

```
# load ggplot2
library(ggplot2)
# create the plot (see Fig 6)
ggplot(data = scent_nmnds, aes(MDS1,MDS2, color = colony)) +
geom_point() + theme_void() + scale_color_manual(values = c("blue", "red")) +
theme(panel.background = element_rect(colour = "black", size = 1.25, fill = NA),
aspect.ratio = 1,legend.position = "none")
```

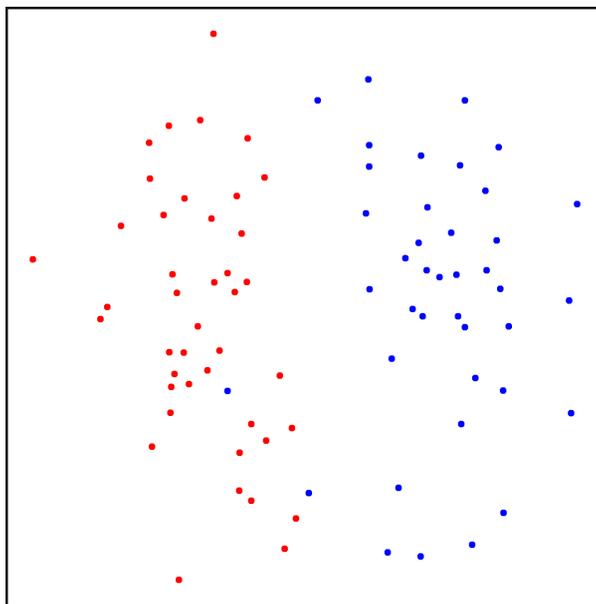


Figure 6.6: Two-dimensional nonmetric multidimensional scaling plot of chemical data from 41 Antarctic fur seal mother-offspring pairs. Bray-Curtis dissimilarity values were calculated from standardized and $\log(x+1)$ transformed abundance data (see main text for details). Individuals from the two different breeding colonies described in (Stoffel et al., 2015) are shown in blue and red respectively.

The resulting NMDS plot shown in Figure 6.6 reveals a clear pattern in which seals from the two colonies cluster apart based on their chemical profiles, as shown also by Stoffel et al. 2015. Although a sufficient number of standards were lacking in this example dataset to calculate the internal error rate (as shown below for the three bumblebee datasets), the strength of the overall pattern suggests that the alignment implemented by GCalignR is of high quality.

Evaluation of the performance of GCalignR

We evaluated the performance of GCalignR in comparison to GCALIGNER (Dellicour and Lecocq, 2013). For this analysis, we focused on three previously published bumblebee datasets that were published together with the GCALIGNER software (Dellicour and Lecocq, 2013). These data are well suited to the evaluation of alignment error rates because subsets of chemicals within each dataset have already been identified using GC-MS (Dellicour and Lecocq, 2013). Hence, by focusing on these known substances, we can test how the two alignment programs perform. Furthermore, these datasets allow us to further investigate the performance

of GCalignR by evaluating how the resulting alignments are influenced by parameter settings.

Comparison with GCALIGNER

To facilitate comparison of the two programs, we downloaded raw data on cephalic labial gland secretions from three bumblebee species (Dellicour and Lecocq, 2013) from <http://onlinelibrary.wiley.com/wo11/doi/10.1002/jssc.201300388/supinfo>. Each of these datasets included data on both known and unknown substances, the former being defined as those substances that were identified with respect to the NIST database (Linstrom and Mallard, 2009). The three datasets are described in detail by (Dellicour and Lecocq, 2013). Briefly, the first dataset comprises 24 *Bombus bimaculatus* individuals characterised for a total of 41 substances, of which 32 are known. The second dataset comprises 20 *B. ephippiatus* individuals characterised for 64 substances, of which 42 are known, and the third dataset comprises 11 *B. flavifrons* individuals characterised for 58 substances, of which 44 are known.

To evaluate the performance of GCALIGNER, we used an existing alignment provided by (Dellicour and Lecocq, 2013). For comparison, we then separately aligned each of the full datasets within GCalignR as described in detail in S3 File. We then evaluated each of the resulting alignments by calculating the error rate, based only on known substances, as the ratio of the number of incorrectly assigned retention times to the total number of retention times (Eq. 6.5).

$$\text{Error} = \left[\frac{\text{Number of misaligned retention times}}{\text{Total number of retention times}} \right] \quad (\text{Eq. 6.5})$$

where retention times that were not assigned to the row that defines the mode of a given substance were defined as being misaligned. Figure 6.7 shows that both programs have low alignment error rates (i.e. below 5%) for all three datasets. The programs performed equally well for one of the species (*B. flavifrons*), but overall GCalignR tended to perform slightly better, with lower alignment error rates being obtained for *B. bimaculatus* and *B. ephippiatus*.

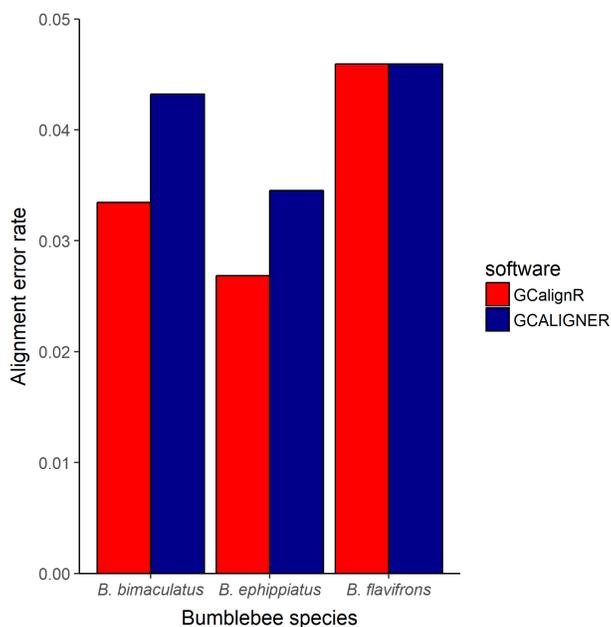


Figure 6.7: Alignment error rates for three bumblebee datasets using GCalignR and GCALIGNER. Error rates were calculated based only on known substances as described in the main text.

Effects of parameter values on alignment results

The first step in the alignment procedure accounts for systematic linear shifts in retention times. As most datasets will require relatively modest linear transformations (illustrated by the Antarctic fur seal dataset in Figure 6.5), the parameter `max_linear_shift` (Table 6.1), which defines the range that is considered for applying linear shifts (i.e. window size), is unlikely to appreciably affect the alignment results. By contrast, two user-defined parameters need to be chosen with care. Specifically, the parameter `max_diff_peak2mean` determines the variation in retention times that is allowed for sorting peaks into the same row, whereas the parameter `min_diff_peak2peak` enables rows containing homologous peaks that show larger variation in retention times to be merged (see Material and methods for details and Table 6.1 for definitions). To investigate the effects of different combinations of these two parameters on alignment error rates, we again used the three bumblebee datasets, calculating the error rate as described above for each conducted alignment. Figure 6.8 shows that for all three datasets, relatively low alignment error rates were obtained when `max_diff_peak2mean` was low (i.e. around 0.01 to 0.02 minutes). Error rates gradually increased with larger values of `max_diff_peak2mean`,

reflecting the incorrect alignment of non-homologous substances that are relatively similar in their retention times. In general, alignment error rates were relatively insensitive to parameter values of `min_diff_peak2peak` (see Figure 6.8). Higher error rates were only obtained when `max_diff_peak2mean` was larger than or the same as `min_diff_peak2peak`, in which case merging of homologous rows is not possible.

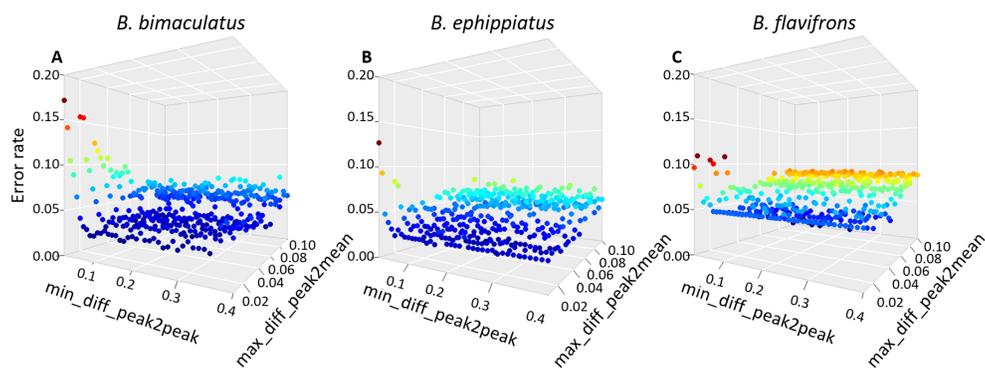


Figure 6.8: Effects of different parameter combinations on alignment error rates for three bumblebee datasets (see main text for details). Each point shows the alignment error rate for a given combination of `max_diff_peak2mean` and `min_diff_peak2peak`.

Comparison with parametric time warping

In the fields of proteomics and metabolomics, several methods (usually referred to as ‘time warping’ (Bloemberg et al., 2010)) for aligning peaks have been developed that aim to transform retention times in such a way that the overlap with the reference sample is maximised (Wehrens et al., 2015). The R package `ptw` (Bloemberg et al., 2010) implements parametric warping and supports a peak list containing retention times and intensity values for each peak of a sample, making it in principle suitable for aligning GC-FID data. However, parametric time warping of a peak list within `ptw` is based on strictly pairwise comparisons of each sample to a reference (Wehrens et al., 2015). Therefore, the sample and reference should ideally resemble one another and share all peaks (Johnson et al., 2003; Bloemberg et al., 2013). By comparison, `GCalignR` only requires a reference for the first step of the alignment procedure and should therefore be better able to cope with among-individual variability. Additionally, although `ptw` transforms individual peak lists relative to the reference, it does not provide a function to match homologous substances across samples.

In order to evaluate how these differences affect alignment performance, we analysed GC-MS data on cuticular hydrocarbon compounds of 330 European earwigs (*Forficula auricularia*) (Wong et al., 2014) using both GCalignR and ptw. This dataset was chosen for two main reasons. First, alignment success can be quantified based on twenty substances of known identity. Second, all of the substances are present in every individual, the only differences being their intensities. Hence, among-individual variability is negligible, which should minimise issues that may arise from samples differing from the reference. As a proxy for alignment success, we compared average deviations in the retention times of homologous peaks in the raw and aligned datasets, with the expectation that effective alignment should reduce retention time deviation.

For this analysis, we downloaded the earwig dataset from <https://datadryad.org/resource/doi:10.5061/dryad.73180> (Wong et al., 2014) and constructed input files for both GCalignR and ptw. We then aligned this dataset using both packages as detailed in supporting information S3 File. Following fine-tuning of alignment parameters within GCalignR, we obtained twenty substances in the aligned dataset and all of the homologous peaks were matched correctly (i.e. every substance had a retention time deviation of zero). Consequently, GCalignR consistently reduced retention time deviation across all substances relative to the raw data (Figure 6.9). By comparison, parametric time warping resulted in higher deviation in retention times for all but two of the substances (Figure 6.9). These differences in the performance of the two programs probably reflect differential sensitivity to variation in peak intensities.

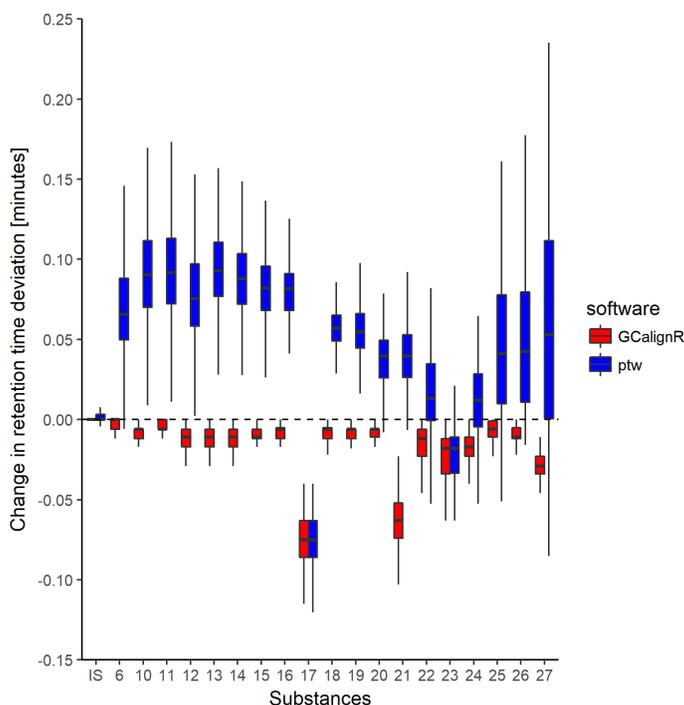


Figure 6.9: Boxplot showing changes in retention time deviation of twenty homologous substances relative to the raw data after having aligned a dataset of 330 European earwigs within GCalignR and ptw respectively (see main text for details).

Conclusions

GCalignR is primarily intended as a pre-processing tool in the analysis of complex chemical signatures of organisms where overall patterns of chemical similarity are of interest as opposed to specific (i.e. known) chemicals. We have therefore prioritised an objective and fast alignment procedure that is not claimed to be free of error. Nevertheless, our alignment error rate calculations suggest that GCalignR performs well with a variety of example datasets. GCalignR also implements a suite of diagnostic plots that allow the user to visualise the influence of parameter settings on the resulting alignments, allowing fine-tuning of both the pre-processing and alignment steps (Figure 6.1). For tutorials and worked examples illustrating the functionalities of GCalignR, we refer to the vignettes that are distributed with the package and are available as supporting information S4 and S5 Files.

Acknowledgements

We are grateful to Barbara Caspers and Sarah Goltke for helpful discussions and providing data for testing purposes during the development of the package. This research was supported by a Deutsche Forschungsgemeinschaft (DFG) standard Grant (HO 5122/3-1) to J.I.H. together with a dual PhD studentship from Liverpool John Moores University to M.A.S.

Supplementary Information

GCalignR: An R package for aligning gas-chromatography data for ecological and evolutionary studies

Meinolf Ottensmann*, Martin A Stoffel*, Hazel J Nichols & Joseph I Hoffman

PLOS ONE. 13: e0198311. 2018 *contributed equally

This Supplementary Information can be found online at
<https://doi.org/10.1371/journal.pone.0198311>



Only a week after weaning, these young elephant seals differ substantially in their body weight. The thin pup looking towards the camera has probably very little chance of surviving the following weeks of fasting.

7

inbreedR: R PACKAGE

inbreedR: an R package for the analysis of inbreeding based on genetic markers

Martin A Stoffel, Mareike Esser, Marty Kardos, Emily Humble, Hazel Nichols, Patrice David & Joseph I Hoffman

Methods in Ecology & Evolution. 7:1331–1339. 2016

Summary

1. Heterozygosity-fitness correlations (HFCs) have been widely used to explore the impact of inbreeding on individual fitness. Initially, most studies used small panels of microsatellites, but more recently with the advent of next-generation sequencing, large SNP datasets are becoming increasingly available and these provide greater power and precision to quantify the impact of inbreeding on fitness.
2. Despite the popularity of HFC studies, effect sizes tend to be rather small. One reason for this may be low variation in inbreeding levels among individuals. Using genetic markers, it is possible to measure variance in inbreeding through the strength of correlation in heterozygosity across marker loci, termed identity disequilibrium (ID).
3. ID can be quantified using the measure g_2 formula, which is also a central parameter in HFC theory that can be used within a wider framework to estimate the direct impact of inbreeding on both marker heterozygosity and fitness. However, no software exists to calculate g_2 for large SNP datasets nor to implement this framework.
4. `inbreedR` is an R package that provides functions to calculate g_2 based on microsatellite and SNP markers with associated P -values and confidence intervals. Within the framework of HFC theory, `inbreedR` also estimates the impact of inbreeding on marker heterozygosity and fitness. Finally, `inbreedR` implements user-friendly simulations to explore the precision and magnitude of estimates based on different numbers of genetic markers. We hope this package will facilitate good practice in the analysis of HFCs and help to deepen our understanding of inbreeding effects in natural populations.

Introduction

Offspring of close relatives often show reduced fitness, a phenomenon referred to as inbreeding depression (Charlesworth and Charlesworth, 1987; Charlesworth and Willis, 2009). This decline in fitness among inbred individuals is a result of the increased proportion of loci in the genome that are identical by descent (IBD). A homozygous locus is IBD or autozygous when it carries two alleles that both originate from a single copy in a common ancestor. An increased proportion of loci in the genome that are identical by descent IBD_G may lead to the unmasking of deleterious recessive alleles and a reduction in heterozygote advantage by decreasing genomewide heterozygosity (Charlesworth and Charlesworth, 1987; Charlesworth and Willis, 2009). In populations with unknown pedigrees, many studies have used genetic

marker heterozygosity as a measure of IBD_G . The result is a large and expanding literature describing heterozygosity-fitness correlations (HFCs) across a range of species and traits (Coltman and Slate, 2003; Chapman et al., 2009; Szulkin et al., 2010).

Despite the large and growing number of HFC studies, effect sizes are usually small (Chapman et al., 2009) and there has been debate over their mechanistic basis (Balloux et al., 2004; Slate et al., 2004; Hansson and Westerberg, 2008; Szulkin et al., 2010). This reflects the fact that under many circumstances multilocus heterozygosity based on the 10-20 microsatellite markers employed by most studies provides little power to estimate IBD_G (Hansson and Westerberg, 2002; Balloux et al., 2004; Szulkin et al., 2010; Hoffman et al., 2014). This is why the pedigree-derived inbreeding coefficient (F_P) has long been the gold standard for estimating IBD_G (Pemberton, 2004, 2008). F_P is defined as the probability of a given locus in an individual's genome being autozygous based on its pedigree. However, an individual's F_P will differ from its IBD_G as F_P can be imprecise due to linkage among loci and downwardly biased due to incomplete pedigree information (Hill and Weir, 2011; Keller et al., 2011; Kardos et al., 2015). Consequently, IBD_G can vary substantially among individuals with the same F_P (Franklin, 1977; Hill and Weir, 2011; Forstmeier et al., 2012). In other words, even F_P derived from a perfect pedigree cannot fully capture the variance in genomic autozygosity ($\sigma^2(IBD_G)$) among individuals, as it does not incorporate variation due to linkage.

Recent advances in next-generation sequencing technology (e.g. Baird et al. 2008; Peterson et al. 2012) now allow many tens or even hundreds of thousands of single-nucleotide polymorphisms (SNPs) to be genotyped in virtually any organism. Applied to HFCs, these dense marker panels provide much greater power than a small panel of microsatellites to quantify the impact of inbreeding on fitness (Hoffman et al. 2014). Recent simulation and empirical studies also suggest that inbreeding coefficients based on genomewide SNP data provide more precise measures of IBD_G and inbreeding depression than F_P (Keller et al., 2011; Pryce et al., 2014; Kardos et al., 2015; Huisman et al., 2016).

HFC theory

For marker loci to indicate inbreeding depression, their heterozygosity must be correlated with the heterozygosity of functional loci in the genome (Szulkin et al., 2010). Such correlations between marker loci and functional loci have been proposed to occur through two possible mechanisms: The 'general effect hypothesis' on the one hand assumes that multilocus heterozygosity (MLH) reflects genomewide heterozygosity. This association emerges because variation in inbreeding causes heterozygosity to be correlated across loci, a phenomenon termed identity

disequilibrium (ID) (Weir and Cockerham, 1973). Alternatively, the ‘local effect hypothesis’ states that one or a few of the markers are in linkage disequilibrium (LD) with a trait locus under balancing selection, which creates a pattern whereby heterozygosity at the gene and marker are correlated. However, ID and LD do not necessarily have to be considered as competing hypotheses to explain HFCs as ID is a consequence and LD is a cause of variation in IBD_G (Bierne et al., 2000; Szulkin et al., 2010). Both mechanisms can therefore be united under an inbreeding or general effect model (Bierne et al., 2000).

Variance in individual inbreeding levels can be caused by a variety of scenarios other than systematic consanguineous matings (Szulkin et al., 2010). For example, in small or bottlenecked populations, $\sigma^2(IBD_G)$ and therefore ID can occur as a consequence of variation in the relatedness of mating partners. Similarly, immigration and admixture can result in the offspring of parents from different populations being relatively outbred, leading to an increased $\sigma^2(IBD_G)$ within a population (Tsitroni et al., 2001; Szulkin et al., 2010). In addition, in small randomly mating populations, both genetic drift and immigration generate LD (Hill and Robertson, 1968; Sved, 1968; Bierne et al., 2000), which in turn leads to ID (Szulkin et al., 2010). All of these scenarios ultimately increase $\sigma^2(IBD_G)$ and lead to ID, which is the fundamental cause of HFCs according to the general effect model.

The general effect model assumes that HFCs arise due to the simultaneous effects of inbreeding on variation among individuals in marker heterozygosity and fitness (David et al., 1995; David, 1998; Bierne et al., 2000; Hansson and Westerberg, 2002). Specifically, inbreeding affects the genome including the panel of genetic markers by increasing the proportion of loci that are IBD and by causing ID. When the aim of a study is to infer the effects of inbreeding on fitness from a panel of genetic markers, two related questions arise: (i) How well does MLH at genetic markers reflect IBD_G ? and (ii) How large is the inbreeding load, that is the correlation between inbreeding and fitness? These questions led to the development of a model to estimate these relationships based on the inbreeding coefficient f defined as individual IBD_G (Bierne et al., 2000). This model was developed further to estimate how well marker heterozygosity reflects F_P , which itself is an imprecise measure of IBD_G , but the best that existed in pre-genomic times (Slate et al., 2004). Within this framework, Szulkin et al. (2010) used g_2 (David et al., 2007), a point estimate of ID, to measure $\sigma^2(IBD_G)$. This allows the derivation of formulae to estimate the correlations between inbreeding, MLH and fitness purely from genetic marker data.

Quantifying effects of inbreeding on heterozygosity and fitness

The general effect model assumes that heterozygosity at genetic markers (h , here defined as standardized MLH, Coltman et al. 1999), is correlated with genomic heterozygosity through variation in individual inbreeding levels (f) and that individual fitness (W) declines as a linear function of f , which is expected if deleterious mutations have non-epistatic effects (Bierne et al., 2000). In other words, the correlation between W and h arises through the simultaneous effects of inbreeding level on fitness ($r(W, f)$) and marker heterozygosity ($r(h, f)$) (Bierne et al., 2000; Slate et al., 2004; Szulkin et al., 2010):

$$r(W, h) = r(h, f)r(W, f) \quad (\text{Eq. 5.1})$$

Although F_P has been used as a measure of f in the above formula (Slate et al., 2004; Szulkin et al., 2010), here we define the inbreeding coefficient f as a variable that explains all of the variance in genomic heterozygosity ($\sigma^2(IBM_G)$) and therefore includes both variance depending on an individual's pedigree and the degree of linkage among loci (Bierne et al., 2000). When it is not possible to directly measure an individual's inbreeding level f , we can use ID to characterize the distribution of f in a population. A measure of ID that can be related to HFC theory is $g2$ (David et al., 2007), which quantifies the extent to which heterozygosities are correlated across pairs of loci (see Appendix S1 for details). Based on $g2$ as an estimate of ID, it is then possible to calculate the expected correlation between h and inbreeding level f as follows (Szulkin et al., 2010):

$$r^2(h, f) = \frac{g2}{\sigma^2(h)} \quad (\text{Eq. 5.2})$$

Finally, the expected squared correlation between a fitness trait W and inbreeding level f can be derived by rearranging Eq. 5.1 (Szulkin et al., 2010):

$$r^2(W, f) = \frac{r^2(W, h)}{r^2(h, f)} \quad (\text{Eq. 5.3})$$

Software is already available for calculating $g2$ from microsatellite datasets (David et al., 2007). However, for larger (e.g. SNP) datasets, the original formula is not computationally practical, as it requires a double summation over all pairs of loci. For example, with 15,000 loci, the double summations take of the order of 0.2×10^9 computation steps. For this reason, it is necessary to implement a computationally more feasible formula to calculate $g2$, which assumes that the distribution of true heterozygosity is the same in missing data as in non-missing data, i.e. that the frequency of missing values does not vary much between pairs of loci (Hoffman et al., 2014). In turn, the $g2$ parameter builds the foundation for the implementation of the above framework to analyse HFCs, which is recommended to be routinely computed in future HFC studies (Szulkin et al., 2010; Kardos et al., 2014).

The package

`inbreedR` is an R package (R Core Team, 2015) that provides functions for analysing inbreeding and HFCs based on microsatellite and SNP data. The main aims of the package are to (i) calculate $g2$ and its confidence interval and P -value for both microsatellites and large SNP datasets, (ii) estimate the influence of inbreeding on marker heterozygosity and fitness through the derivation of $r^2(h, f)$ and $r^2(W, f)$ and (iii) explore the sensitivity of $g2$ and $r^2(h, f)$ to marker number through user-friendly simulations. The overall workflow is shown in Figure 7.1 and described below. For a more detailed description of the package and the functions, we have supplied a vignette for the package that can be accessed via `browseVignettes("inbreedR")` once the package is installed.

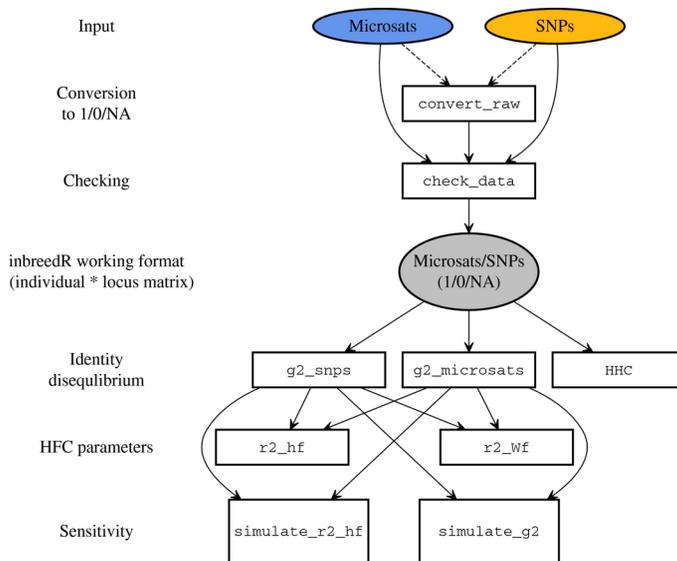


Figure 7.1: *inbreedR* workflow. For both microsatellite and SNP datasets, the program provides utilities for data conversion and checking, estimation of identity disequilibrium, derivation of key parameters relating to HFC theory and exploration of sensitivity to the number of loci deployed. Further details are provided in the main text.

Example datasets

The functionality of *inbreedR* is illustrated using genetic and phenotypic data from an inbred captive population of oldfield mice (*Peromyscus polionotus*, Hoffman et al. 2014). These mice were paired over six laboratory generations to produce offspring with F_P ranging from 0 to 0453. Example files are provided containing the genotypes of 36 *P. polionotus* individuals at 12 microsatellites and 13,198 SNPs respectively. Data on body mass at weaning, a fitness proxy, are also available for the same individuals.

```

library(inbreedR)
data("mouse_msats") # microsatellite data
data("mouse_snps") # snp data
data("bodyweight") # fitness data

```

Data conversion and checking

The working format of *inbreedR* is an *individual* \times *locus* matrix or `data.frame` in which rows represent individuals and each column represents a locus. If an individual is heterozygous at a given locus, it is coded as 1, whereas a homozygote is coded as 0, and missing data are

coded as NA. We provide a converter function from a common two-column-per-locus (allelic) format to the working format, as well as a function to check for common formatting errors within the input matrix. Guidelines for extracting genotype data from VCF files are given in the vignette.

```
# transforms microsatellite data into (0/1)
mouse_microsats <- convert_raw(mouse_msats)
# check the data
check_data(mouse_microsats, num_ind = 36, num_loci = 12)
#> [1] TRUE
check_data(mouse_snps, num_ind = 36, num_loci = 13198)
#> [1] TRUE
```

Identity disequilibrium

The package provides functions to calculate g_2 for both microsatellites and SNPs. The `g2_microsats()` function implements the formula given in David et al. (2007). For large datasets (e.g. SNPs) the `g2_snps()` function implements a computationally feasible formula described in Appendix S1. For both microsatellites and SNPs, `inbreedR` also calculates confidence intervals by bootstrapping over individuals (Table 71). It also permutes the genetic data to generate a P-value for the null hypothesis of no variance in inbreeding in the sample (i.e. $g_2 = 0$). The `g2_snps()` function provides an additional argument for parallelization which distributes bootstrapping and permutation across cores.

```
g2_mouse_microsats <- g2_microsats(mouse_microsats,
  nperm = 1000, nboot = 1000, CI = 0.95)
g2_mouse_snps <- g2_snps(mouse_snps, nperm = 100,
  nboot = 100, CI = 0.95, parallel = FALSE, ncores = NULL)
```

Table 7.1: Output of the g_2 functions showing g_2 values and their 95% confidence intervals, standard errors and P-values for 36 mice genotyped at 12 microsatellites and 13,198 SNPs.

	g_2	Lower CI	Upper CI	SE	P-value
Microsats	0.022	-0.008	0.065	0.019	0.076
SNPs	0.035	0.022	0.050	0.008	0.010

Results of both functions can be plotted as histograms with CIs (Figure 72).

```
par(mfrow=c(1,2))
plot(g2_microsats, main = "Microsatellites",
     col = "cornflowerblue", cex.axis = 0.85)
plot(g2_snps, main = "SNPs",
     col = "darkgoldenrod1", cex.axis = 0.85)
```

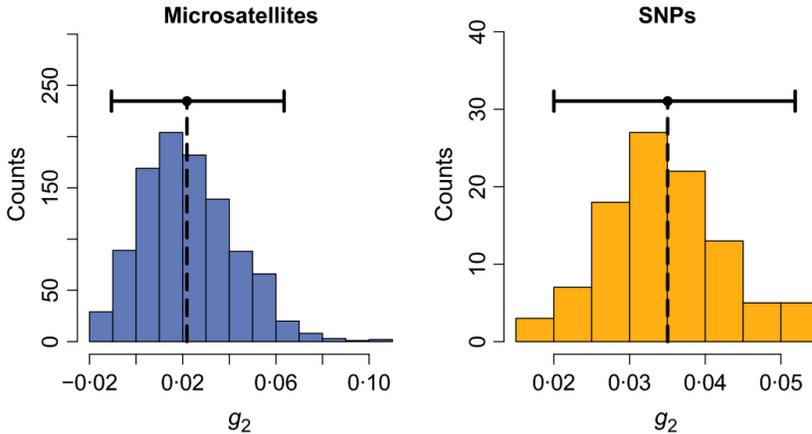


Figure 7.2: Output of the g_2 functions for the microsatellite and SNP datasets showing the distribution of g_2 estimates from bootstrap samples over individuals together with their 95% CIs. The empirical g_2 estimate is marked as a black dot along the CI.

Another approach for estimating ID is to divide the marker panel into two random subsets, compute the correlation in heterozygosity between the two and repeat this hundreds or thousands of times in order to obtain a distribution of heterozygosity-heterozygosity correlation coefficients (Balloux et al., 2004). This approach is intuitive and has been shown to be equivalent to g_2 in its power to detect non-zero variance in inbreeding (Kardos et al., 2014) although it can be criticized on the grounds that samples within the HHC distribution are non-independent. Moreover, g_2 is preferable because it directly relates to HFC theory (Eq. 5.2). The `HHC()` function in `inbreedR` calculates HHCs together with confidence intervals, specifying how often the dataset is randomly split into two halves with the `reps` argument.

```
HHC_mouse_microsats <- HHC(mouse_microsats, reps = 1000)
HHC_mouse_snps <- HHC(mouse_snps, reps = 100)
```

The results can be outputted as text (Table 7.2) or plotted as histograms with CIs (Figure 7.3).

Table 7.2: Output of the HHC function, showing mean HHCs with 95% confidence intervals and standard deviations for 36 mice genotyped at 12 microsatellites and 13 198 SNPs.

	Mean	Lower CI	Upper CI	SD
Microsats	0.194	-0.062	0.453	0.128
SNPs	0.976	0.961	0.987	0.007

```
par(mfrow=c(1,2))
plot(HHC_microsats, main = "Microsatellites",
     col = "cornflowerblue", cex.axis = 0.85)
plot(HHC_snps, main = "SNPs",
     col = "darkgoldenrod1", cex.axis = 0.85)
```

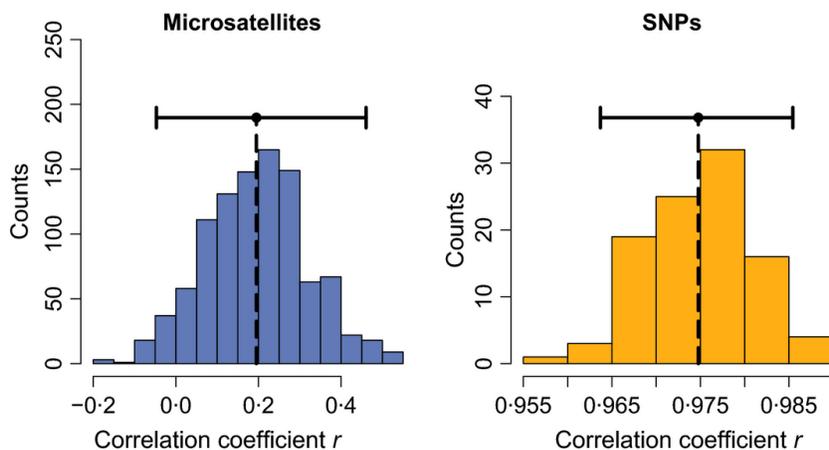


Figure 7.3: Output of the HHC function showing the distribution of heterozygosity-heterozygosity correlation coefficients for the microsatellite and SNP datasets. Also shown are the mean HHCs as black dots and their 95% CIs. The two distributions are very different, microsatellites being positive but with the 95% CI overlapping zero, and SNPs being well in excess of 0.9 with a much greater precision. This reflects the enhanced power of the larger SNP dataset to capture variance in f among individuals.

HFC parameters

Assuming that HFCs are due to inbreeding depression, it is possible to calculate both the expected correlation between heterozygosity and inbreeding level ($r^2(h, f)$) and the expected correlation between a fitness trait and inbreeding ($r^2(W, f)$) as described in Eq. 5.1. These calculations are implemented in `inbreedR` using the functions `r2_hf()` and `r2_Wf()`. Both functions include an `nboot` argument to run bootstrapping over individuals and estimate confidence intervals. Similar to the `glm()` function, the distribution of the fitness trait can be specified using the `family` argument, as shown below:

```
# r^2 between inbreeding and heterozygosity
hf <- r2_hf(genotypes = mouse_microsats, nboot = 100, type = "msats")
# r^2 between inbreeding and fitness
Wf <- r2_Wf(genotypes = mouse_microsats, trait = bodyweight,
family = gaussian, nboot = 100, type = "msats")
```

Workflow for estimating the impact of inbreeding on fitness using HFC

Szulkin et al. (2010) in their Appendix S1 provide a worked example of how to estimate the impact of inbreeding on fitness within an HFC framework. Below, we show how the required calculations can be implemented in `inbreedR`. We start with the estimation of identity disequilibrium (g_2) and calculation of the variance of standardized multilocus heterozygosity ($\sigma^2(h)$), followed by the estimation of the three σ^2 correlations from Eq. 5.1. Example code for the microsatellite dataset is shown below and the results for both microsatellites and SNPs are given in Table 7.3.

Table 7.3: Parameters central to interpreting HFCs for the microsatellite and SNP datasets. \hat{g}_2 is the empirical point estimate of g_2 , $\hat{\sigma}^2(h)$ is the variance in sMLH, $\hat{\beta}_{Wh}$ is the regression slope of sMLH in a linear model of the fitness trait, \hat{r}_{Wh}^2 is the squared correlation of the fitness trait and sMLH, \hat{r}_{hf}^2 is the expected squared correlation of sMLH and inbreeding, and \hat{r}_{Wf}^2 is the expected squared correlation between sMLH and fitness. 95% confidence intervals are shown in squared brackets for the estimates from the package. Note that \hat{r}_{hf}^2 is an expected correlation derived from the ratio of $\hat{g}_2/\hat{\sigma}^2(h)$ and may slightly exceed one due to missing values; we therefore constrain the estimate between 0 and 1.

	\hat{g}_2	$\hat{\sigma}^2(h)$	$\hat{\beta}_{Wh}$	\hat{r}_{Wh}^2	\hat{r}_{hf}^2	\hat{r}_{Wf}^2
Microsats	0.022 [-0.01, 0.06]	0.078	1.601	0.121	0.280 [0, 0.52]	0.434 [0, 0.88]
SNPs	0.035 [0.02, 0.05]	0.033	2.634	0.139	1 [0.89, 1]	0.132 [0, 0.14]

```
# g2 and bootstrap to estimate CI
g2 <- g2_microsats(mouse_microsats, nboot = 1000)
# calculate sMLH
het <- sMLH(mouse_microsats)
# variance in sMLH
het_var <- var(het)
#linear model
mod <- lm(bodyweight ~ het)
# regression slope
beta <- coef(mod)[2]
# r2 between fitness and heterozygosity
Wh <- cor(bodyweight, predict(mod))^2
# r2 between inbreeding and sMLH including bootstraps to estimate CI
hf <- r2_hf(genotypes = mouse_microsats, type = "msats"),
nboot = 1000)
# r2 between inbreeding and fitness including bootstraps to estimate CI
Wf <- r2_Wf(genotypes = mouse_microsats, trait = bodyweight,
family = gaussian, type = "msats", nboot = 1000))
```

Sensitivity to the number of markers

Sampling subsets of loci from an empirical genetic dataset and estimation of a statistic of interest based on these subsets can give insights into the power provided by a given marker panel (Hoffman et al., 2014; Stoffel et al., 2015). However, although subsampling markers (with replacement) from an empirical dataset allows exploration of trends in the magnitude of a statistic, the precision (variation) of the same statistic will be biased. This is due to the increasing non-independence of resampled marker sets as they approach the total number of markers. For example, given a dataset of 20 genetic markers, repeatedly subsampling 18 markers and calculating $g2$ will always lead to lower variation in the estimates than subsampling sets of 5 markers. To circumvent this problem, the `simulate_g2()` function simulates genotypes from which subsets of loci can be sampled independently. The simulations can be used to evaluate the effects of the number of individuals and loci on the precision and magnitude of $g2$. The user specifies the number of simulated individuals (`n_ind`), the subsets of loci (`subsets`) to be drawn, the heterozygosity of non-inbred individuals (`H_nonInb`, i.e. expected heterozygosity in the base population) and the distribution of f among the simulated individuals. The f values of the simulated individuals are sampled randomly from a beta distribution with mean (`meanF`) and variance (`varF`) specified by the user (e.g. as in Wang 2011). This enables the simulation to mimic populations with known inbreeding characteristics or to simulate hypothetical scenarios of interest. For computational simplicity, allele frequencies are assumed to be constant across

loci and the simulated loci are unlinked. Genotypes (i.e. heterozygosity/homozygosity at each locus) are assigned stochastically based on the f values of the simulated individuals. Specifically, the probability of an individual being heterozygous at any given locus (H) is expressed as $H = H_0(1 - f)$, where H_0 is the user-specified heterozygosity of a non-inbred individual and f is an individual's inbreeding coefficient drawn from the beta distribution.

```
sim_g2_mouse_microsats <- simulate_g2(n_ind = 50, H_nonInb = 0.5,
meanF = 0.2, varF = 0.03, subsets = c(5, 10, 15, 20, 25, 30, 35, 40, 45, 50),
reps = 100, type = "msats")
sim_g2_mouse_snps <- simulate_g2(n_ind = 50, H_nonInb = 0.5, meanF = 0.2,
varF = 0.03, subsets = seq(from = 1000, to = 10000, by = 1000), reps = 100,
type = "snps")
```

The results can be visualized by showing the mean and CI of g^2 plotted against the number of loci used (Figure 7.4).

```
par(mfrow = c(1,2), mar = c(5,5.15,3,1.2))
plot(sim_g2_microsats, main = "Microsatellites", cex.axis = 1.5,
cex.main = 1.5, cex.lab = 1.5)
plot(sim_g2_snps, main = "SNPs", cex.axis = 1.5,
cex.main = 1.5, cex.lab = 1.5)
```

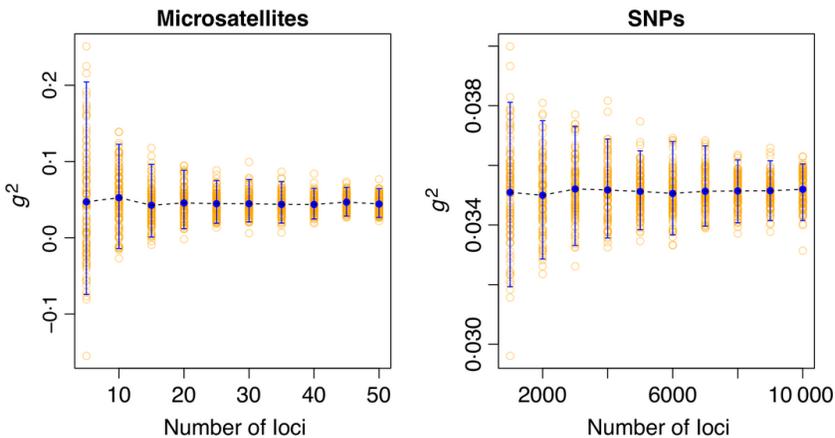


Figure 7.4: Output of the `simulate_g2()` function. Different sets of microsatellites and SNPs were simulated and stochastically drawn from distributions based on a mean (SD) inbreeding level f of 0.2 (0.03) assuming that a non-inbred individual has a heterozygosity of 0.5. The two plots show the g^2 statistics from all samples including their means and 95% CIs.

Bear in mind that g^2 values calculated from the simulated data may overestimate precision due to the assumption of unlinked loci. However, in practice, the number of linked SNPs in most

real datasets will be small compared to the number of unlinked SNPs (Szulkin et al., 2010) and hence g^2 should not be substantially affected.

Finally, it is of interest to infer how well genetic marker heterozygosity reflects the inbreeding level f and whether this correlation could be increased by genotyping individuals at a larger set of markers. The `simulate_r2_hf()` function can be used to compare the precision and magnitude of the expected squared correlation between heterozygosity and inbreeding ($r^2(h, f)$) for a given number of genetic markers.

```

sim_r2_mouse_microsats <- simulate_r2(n_ind = 50, H_nonInb = 0.5,
meanF = 0.2, varF = 0.03, subsets = c(5, 10, 15, 20, 25, 30,
35, 40, 45, 50), reps = 100, type = "msats")

sim_r2_mouse_snps <- simulate_r2(n_ind = 50, H_nonInb = 0.5, meanF = 0.2,
varF = 0.03, subsets = seq(from = 1000, to = 10000, by = 1000), reps = 100,
type = "snps")

```

The results can again be plotted as a series of $r^2(h, f)$ estimates together with their means and CIs (Figure 75).

```

par(mfrow = c(1,2), mar = c(5,5.15,3,1.2))
plot(sim_r2_microsats, main = "Microsatellites", cex.axis = 1.5,
cex.main = 1.5, cex.lab = 1.5)
plot(sim_r2_snps, main = "SNPs", cex.axis = 1.5, cex.main = 1.5,
cex.lab = 1.5)

```

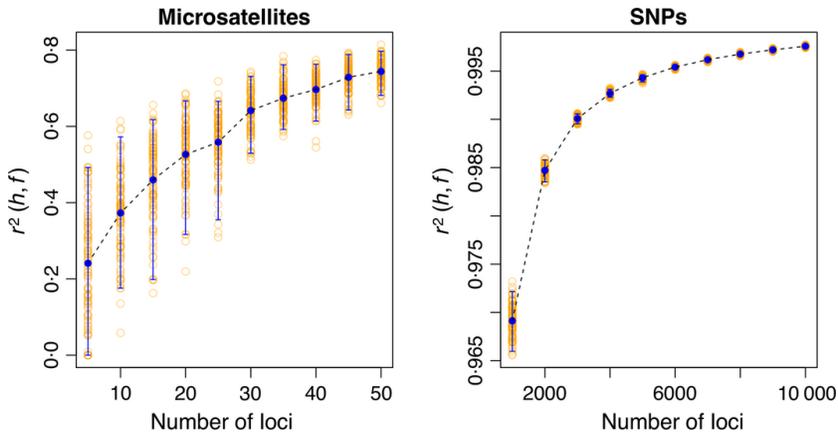


Figure 7.5: Output of the `simulate_r2_hf()` function. Different sets of microsatellites and SNPs were simulated and stochastically drawn from distributions based on a mean (SD) inbreeding level f of 0.2 (0.03) assuming that a non-inbred individual has a heterozygosity of 0.5. The two plots show the $r^2(W, f)$ values for an increasing number of markers including their means and 95% CIs. The expected correlation between inbreeding and marker heterozygosity increases and is estimated with higher precision when the number of markers is increased.

Effects of LD under the general effect model

LD may affect the strength of an HFC because it increases $\sigma^2(IBD_G)$ (Bierne et al., 2000). This is because the variance in individual IBD_G is explained by (i) a component that reflects the different pedigrees of individuals and (ii) a component that reflects variation among individuals with the same pedigree (Bierne et al., 2000). In the absence of linkage (i.e. if there were infinitely many unlinked loci), an individual's IBD_G would solely depend on the pedigree. However, loci do not segregate independently and LD and especially physical linkage will therefore cause variation in IBD_G among individuals with the same pedigree. Calculating g^2 and derived HFC statistics based on large SNP datasets, which are likely to include linked markers, is therefore not a problem *per se*. As g^2 does not incorporate any pedigree information but purely quantifies correlated heterozygosity among genetic marker pairs, it is a direct measure of $\sigma^2(IBD_G)$. The only assumption needed is that IBD is equally frequent among marker loci and fitness loci that are responsible for inbreeding depression. Put another way, the fitness loci should have an equivalent genomic distribution to the genetic markers.

Increasing the total number of genetic markers should not affect the proportion of linked markers and should thus not affect g^2 . To test this, we evaluated the sensitivity of g^2 to marker number by repeatedly sampling random subsets of between 100 and 13,000 SNPs from the full mouse dataset and calculating the respective g^2 values. For each subset, markers were sampled without replacement to avoid non-independence, which is why the number of repetitions decreases with increasing marker number. The mean g^2 was found to be stable across all subset sizes, suggesting that, for our dataset, the expected g^2 does not vary appreciably with marker density (Figure 76).

In general, the number of locus pairs in strong linkage is expected to be very low compared to the number of non-linked pairs (Szulkin et al., 2010). As g^2 averages over all pairs of loci, this point estimate should therefore be relatively insensitive to the inclusion of linked markers as long as all markers are broadly distributed across the genome. To test this, we conducted LD pruning of our SNP dataset at various stringency thresholds to determine how linkage among SNPs affects g^2 estimates and their confidence intervals. We used the indep-pairphase function in PLINK version 1.09 (Purcell et al., 2007) to remove one SNP from each pair with an r^2 above thresholds ranging from 0.5 to 0.99 with increments of 0.05 and a last increment of 0.04. In order to account for our SNPs being on unplaced contigs, we assumed that all SNPs were on the same 'chromosome' and used a sliding window spanning the full dataset. The magnitude and precision of g^2 estimates was found to be stable across all LD pruned datasets (Figure 77), suggesting that, for our dataset, g^2 is relatively insensitive to the inclusion

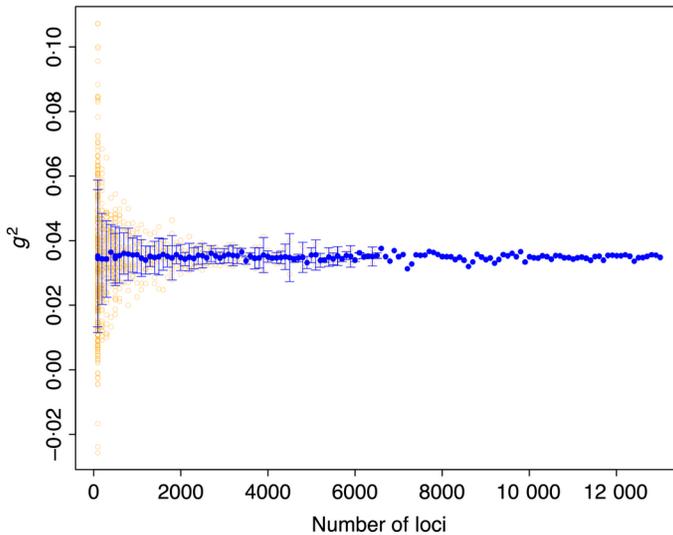


Figure 7.6: Mean and standard deviation of g^2 derived from an increasing number of SNPs drawn at random from the empirical mouse dataset (13,198 SNPs). The distribution of data points for each subset size is based on sampling without replacement to obtain non-overlapping marker sets. For this reason, the number of data points decreases from 131 for 100 markers to 1 for subsets larger than 6599 SNPs. The mean g^2 is stable across all subset sizes, which suggests that estimating g^2 from larger numbers of markers does not introduce bias for our dataset.

of strongly linked SNPs.

Final remarks

The `inbreedR` package implements a framework to estimate the impact of variation in inbreeding on marker heterozygosity and fitness, which has been suggested to be routinely reported in HFC studies (Szulkin et al., 2010; Kardos et al., 2014). A good example is a recent study of red deer, in which Huisman et al. (2016) quantified identity disequilibria through g^2 in several datasets to estimate the power of a genomic inbreeding measure to detect inbreeding depression. In addition to the quantification of ID and HFCs for empirical data, straightforward simulations within `inbreedR` provide a way to explore the effect of the number of genetic markers on g^2 and the expected correlation between marker heterozygosity and inbreeding. This is important for evaluating the power of a given dataset to measure inbreeding depression and could also facilitate the planning of future projects by allow exploration of the effects of sample size and marker number on the power to detect ID and HFCs.

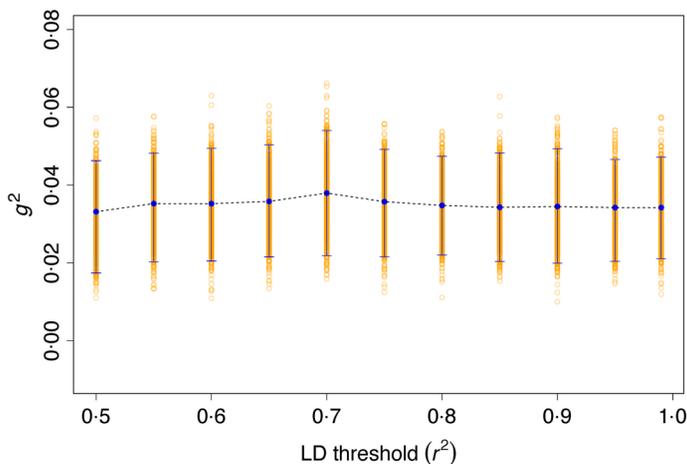


Figure 7.7: Estimates of g^2 with confidence intervals for subsets of SNPs pruned based on different LD thresholds. We used PLINK to remove one SNP from each marker pair with an r^2 above the respective threshold. As we used a sliding window spanning the full dataset instead of local regions on a chromosome, the retained datasets contained a maximum of 4363 ($r^2 > 0.99$) and a minimum of 1095 ($r^2 > 0.5$) SNPs. The magnitude and precision of g^2 does not vary noticeably for our dataset when pruning strongly linked SNPs.

Although g^2 and related parameters can provide insights into whether an HFC is due to inbreeding or not, the user should be aware that spurious HFCs can occur due to population structure (Slate et al., 2004), which should therefore be appropriately dealt with beforehand. For instance, genetically distinct populations could be analysed separately. Also, it is worthwhile considering whether SNPs should be filtered based on their minor allele frequencies (MAF) prior to analysis. On the one hand, genotyping by sequencing approaches rely on sufficient depth of coverage to call SNPs with reasonable confidence. Thus, low MAF SNPs may be disproportionately error prone when the depth of sequence coverage is not high enough to capture multiple copies of the minor allele. On the other hand, filtering out low MAF SNPs may distort the allele frequency spectrum and lead to the loss of valuable information (Hoffman et al., 2014).

Finally, LD and ID have been seen as alternative hypotheses to explain HFCs (Hansson and Westerberg, 2008). However, LD often goes hand in hand with ID and is therefore a relevant variance component when the aim is to estimate $\sigma^2(IBD_G)$ (Bierne et al., 2000; Szulkin et al., 2010). As most HFC studies should be interested in estimating $\sigma^2(IBD_G)$ through g^2 , linked markers need not be pruned as long as the genomic distributions of the marker and trait loci are comparable. However, if the goal of a study is to infer characteristics

of a pedigree from g_2 (such as self-fertilization rates), it might be useful to reduce physical linkage among markers using PLINK (Purcell et al., 2007) or other methods to ensure their independence (David et al., 2007). Further investigation would be needed to evaluate the impact of pruning linked markers on selfing or inbreeding rates estimated through g_2 .

Computation times

Computation times will be negligible for most microsatellite datasets but somewhat longer for very large SNP datasets. On a standard laptop (Intel Core I5 2.60 GHz, 8 GB RAM), running the `g2_snps()` function for our example SNP dataset (36 individuals genotyped at 13,198 loci) with 1000 bootstraps takes 1 min 12 s without parallelization and 38 s with parallelization on three cores. For comparison, we also simulated a large SNP dataset with 3500 individuals at 37,000 loci (similar to Huisman et al. (2016) and ran this on a 40 core server with 1000 bootstraps, which took 73 h.

Availability

The current stable version of the package requires R 3.2.1 and can be downloaded from CRAN as follows:

```
install.packages("inbreedR")
```

In the future, we will aim to extend the functionality of `inbreedR` and the latest development version can be downloaded from GitHub.

```
install.packages("devtools")  
devtools::install_github("mastoffel/inbreedR")
```

Data accessibility

Both example datasets are included in the R package.

Supplementary Information

inbreedR: an R package for the analysis of inbreeding based on genetic markers

Martin A Stoffel, Mareike Esser, Marty Kardos, Emily Humble, Hazel Nichols, Patrice David & Joseph I Hoffman

Methods in Ecology & Evolution. 7:1331–1339. 2016

This Supplementary Information can be found online at
<https://doi.org/10.1111/2041-210X.12588>



A harbour seal pup yawns. In the background is a female adult Northern elephant seal. Four different pinniped species live on San Benitos, also including California sea lions and Guadalupe fur seals.

8

rptR: R PACKAGE

rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models

Martin A Stoffel, Shinichi Nakagawa & Holger Schielzeth

Methods in Ecology & Evolution. 8:1639–1644. 2017

Summary

1. Intra-class correlations (*ICC*) and repeatabilities (*R*) are fundamental statistics for quantifying the reproducibility of measurements and for understanding the structure of biological variation. Linear mixed effects models offer a versatile framework for estimating *ICC* and *R*. However, while point estimation and significance testing by likelihood ratio tests is straightforward, the quantification of uncertainty is not as easily achieved.
2. A further complication arises when the analysis is conducted on data with non-Gaussian distributions because the separation of the mean and the variance is less clear-cut for non-Gaussian than for Gaussian models. Nonetheless, there are solutions to approximate repeatability for the most widely used families of generalized linear mixed models (GLMMs).
3. Here, we introduce the R package rptR for the estimation of *ICC* and *R* for Gaussian, binomial and Poisson-distributed data. Uncertainty in estimators is quantified by parametric bootstrapping and significance testing is implemented by likelihood ratio tests and through permutation of residuals. The package allows control for fixed effects and thus the estimation of adjusted repeatabilities (that remove fixed effect variance from the estimate) and enhanced agreement repeatabilities (that add fixed effect variance to the denominator). Furthermore, repeatability can be estimated from random-slope models. The package features convenient summary and plotting functions.
4. Besides repeatabilities, the package also allows the quantification of coefficients of determination R^2 as well as of raw variance components. We present an example analysis to demonstrate the core features and discuss some of the limitations of rptR.

Introduction

Whenever quantitative measurements are hierarchically organized at multiple levels, intra-class correlations *ICC* can be used to express the average correlation among measurements taken from the same hierarchical level (McGraw and Wong, 1996). A classical application is the quantification of the reproducibility of measurements from the same study object, and the *ICC* is therefore also known as the repeatability *R*. In the context of ecology and evolution, the repeatability typically represents the fraction of the total phenotypic variance *VP* in the population of interest that can be attributed to variation among groups *VG*. The ratio of these

variance components and can be written as $R = \frac{VG}{VG+VR}$, where VR represents the within-group (residual) variance and $VG + VR = VP$. The term ‘group’ is used here in the statistical sense and could represent various biological grouping, e.g. within individuals, families, social groups, plots or years. In the study of animal behaviour, for example, repeatabilities are frequently used as a measure of individual consistency with individual identities being the grouping factor (Boake, 1989; Bell et al., 2009).

Notably, there can be repeatabilities at multiple levels in the same dataset: for example, the repeatability of phenotypes in different years of measurements and the repeatability of phenotypes within individuals. The coefficient of determination R^2 is a similar statistic that quantifies the proportion of variance explained by fixed effects (marginal R^2 sensu Nakagawa and Schielzeth 2013). R^2 is therefore a complementary statistic for decomposing the phenotypic variance. Both R and R^2 represent standardized statistics in the sense that they are variance components divided by the total phenotypic variance and are thus expressed as proportions of the phenotypic variance. Sometimes, however, it is relevant to estimate the variance components as they are, without standardization by the phenotypic variance. A convenient software tool for decomposing phenotypic variances will thus allow for the estimating of repeatabilities at multiple levels, offer flexibility in controlling for fixed effects, estimate the marginal R^2 for fixed effects and allow for the estimation of raw (unstandardized) variance components.

Repeatabilities can be estimated from a variety of sampling designs by a variety of statistical tools. The most widely used statistical framework for analyzing various designs are mixed effects models that allow the estimation of the relevant variance components with the possibility to control for fixed effect covariates (Nakagawa and Schielzeth, 2010). Non-normal error distributions have represented a challenge for estimating repeatabilities, because of the nonlinearity induced by the link function (de Villemereuil et al., 2016). However, we have previously reviewed the equations for estimating repeatabilities and R^2 from generalized linear mixed effects models (GLMMs) (Nakagawa and Schielzeth, 2013). The challenge here is that not all variance components necessary for calculating repeatabilities are part of the model output. Another challenge is the quantification of uncertainty of variance components and ratios of variance components (such as repeatabilities and R^2). We have suggested the use of parametric bootstrapping for quantifying uncertainty of repeatabilities (Nakagawa and Schielzeth, 2010) and this can be easily applied to R^2 , too. Parametric bootstrapping quantifies the design-specific sampling variance by simulating response values from the fitted model followed by a re-estimation of the repeatability or R^2 . Under the assumption that the model is correctly specified, the variance among replicated simulations thus represents the sampling uncertainty of the estimate.

There are a number of other R packages that can estimate certain aspects of *ICC* and repeatabilities for more specific problems. The packages `irr` (Gamer et al., 2012), `psy` (Falissard, 2012) and `psych` (Revelle, 2016) with their functions `icc` and `ICC` allow the estimation of rater agreement and consistency for simple (one- or two-way) designs. From the field of ecology and evolution, the `icc` package (Wolak et al., 2012) allows the estimation of ANOVA-based repeatability with a single response vector and a single grouping vector. Notably, the `icc` package allows the exploration of optimized sampling designs. However, none of these packages can estimate repeatabilities and their uncertainties for non-Gaussian models or control for confounding effects. The `MuMIn` package (Barton, 2016) allows the quantification of R^2 from fitted generalized linear models, but does not allow the quantification of intra-class correlations.

Here, we introduce the `rptR` package for the free software environment R (R Core Team, 2015). `rptR` provides general utilities for estimating adjusted and agreement repeatabilities for Gaussian, binomial (binary and proportion) and Poisson models. The package relies on mixed-effects models fitted by the `lmer` and `glmer` functions from the `lme4` package (Bates et al. 2015). Confidence intervals for repeatabilities are estimated by parametric bootstrapping and statistical significance against $H_0: R = 0$ is tested by likelihood ratio and permutation tests. Moreover, it is also possible to estimate marginal R^2 and raw variance components along with their uncertainties and statistical significance.

Features

The package `rptR` as it is introduced here represents a complete rewrite of the `rptR` developmental package that we had written as part of our original repeatability review (Nakagawa and Schielzeth, 2010). Compared to the developmental version of the package, the current version now provides several new features: (i) The package allows for the estimation of adjusted repeatabilities for Gaussian and well as non-Gaussian data. Adjusted repeatabilities are repeatabilities that control (adjust) for fixed effects (table 6 in Nakagawa and Schielzeth 2010 provides an overview of different types of repeatabilities). The variance explained by fixed effect is excluded from the denominator. (ii) The package allows estimating what we here call enhanced agreement repeatabilities. Enhanced agreement repeatabilities fit fixed effects, but include their variance in the denominator. (iii) The package allows estimating marginal R^2 along with repeatabilities. (iv) The package allows repeatability and marginal R^2 estimation

from random-slope models, using the approach introduced by Johnson 2014. (v) The package allows the estimation of raw (unstandardized) variances. (vi) The package provides new plotting and summary functions as well as a detailed documentation in the form of a vignette (which is accessible via `vignette("rptR")`). (vii) We added the option to conduct bootstrapping and permutations in parallel for reduced waiting time, a progress bar for monitoring and an update option for stepwise increases of bootstrapping and/or permutations.

The package now features the four core functions `rptGaussian` for Gaussian error distributions with identity link, `rptPoisson` for Poisson error distributions with log or square root link, `rptBinary` for binary data with logit or probit link functions and `rptProportion` for proportion data (in the form of counts of success and failure events) following binomial distributions with logit or probit link. All functions can be called via the general `rpt` function by specifying the datatype argument. Results are returned as S3 class `rpt` objects and can be conveniently displayed via the generic `print`, `summary`, and `plot` functions. The package now has a simplified and unified formula interface and is based on the popular `lme4` package (Bates et al. 2015) as the central model-fitting engine. Grouping factors of interest are fitted as random effects and potentially confounding variables can be fitted as fixed effects. For non-Gaussian models, the package internally adds an observational level random effect (with the reserved term ‘Overdispersion’) for estimating overdispersion.

Likelihood ratio tests are returned by default. Randomization procedures are also implemented, but since randomization is time-consuming, this option is deactivated in the default state (argument `npermut = 0`). Randomization in the `rptR` package is implemented as a permutation of residuals of the fitted null model (excluding the grouping factor of interest). This ensures that the remaining data structure and effects are represented in the simulated data while the dependence with the grouping factor of interest is broken.

Example Analysis

We will illustrate the features of `rptR` by estimating adjusted repeatabilities for Poisson data with log link for a dataset that was generated for estimating R^2 in GLMMs (Nakagawa and Schielzeth, 2013). The data represent counts of eggs for an imaginary species of beetle and we estimate repeatabilities at two hierarchical level of organization, among populations and among housing containers, while controlling for an experimental food manipulation. The data are distributed with `rptR` and can be loaded and analyzed by the following:

```
library(rptR)
data(BeetlesFemale)
```

The repeatability can be estimated through the general `rpt` function or directly with the more specialized `rptPoisson` function. The syntax for the formula argument is the same as for the `glmer` function from the `lme4` package. We here specify two random effects (population and container) and aim to estimate the agreement repeatabilities at both levels simultaneously (controlled by the `grname` argument).

```
rptPoisson(Egg~ 1+(1|Container)+(1|Population),
grname=c("Container", "Population"),
data=BeetlesFemale, link="log")
```

The `nboot` argument controls the number of parametric bootstrap iterations for confidence interval estimation and defaults to 500 bootstraps. It may be advisable to reduce this number initially, but to increase the number to 1000 or more for the final analysis. The `npermut` argument controls the number of randomizations for permutation-based null hypothesis testing and defaults to 0. We recommend starting with the default and to use a substantial number of permutations (e.g. 1000) for the final analysis. More iterations for bootstrapping and permutation can be added, using the `update` and `rptObj` arguments (see package documentation for details).

With the example dataset, it is advisable to control for important design effects, in particular the nutritional treatment that may have inflated the phenotypic variance experimentally. Therefore, we estimate adjusted repeatabilities by including a dummy coded binary predictor as a covariate in the fixed part on the right-hand side of the formula argument (while using the option `parallel = TRUE` for reduced waiting time).

```
rep1 <- rptPoisson(Egg ~Treatment+(1|Container)+
(1|Population), grname=c("Container", "Population"),
data=BeetlesFemale, link="log",nboot=1000,
parallel=TRUE)
```

The output can be viewed using the generic `print`, `summary` and `plot` functions (Figure 8.1). Unlike the agreement repeatability estimated above, the adjusted repeatability has removed the variance explained by `Treatment` from the repeatability estimation. Adjusted repeatabilities are also known as consistency repeatabilities, particularly in the psychological literature (Shrout and Fleiss, 1979; McGraw and Wong, 1996), or as narrow sense repeatabilities (Biro and Stamps, 2015). Depending on the sampling designs, adjusted repeatabilities are often, but not always, larger than unadjusted (agreement) repeatabilities. The unadjusted repeatability without

any control for systematic effects is also known as the broad sense or agreement repeatability (Shrout and Fleiss, 1979; Nakagawa and Schielzeth, 2010; Biro and Stamps, 2015).

Plots can be customized by additional arguments handed over to `plot.default` (Figure 8.1):

```
print(rep1)
summary(rep1)
plot(rep1, grname="Population", scale="link",
     cex.main=0.8, main="Population variance", las=1)
plot(rep1, grname="Container", scale="link",
     cex.main=0.8, main="Container variance", las=1)
```

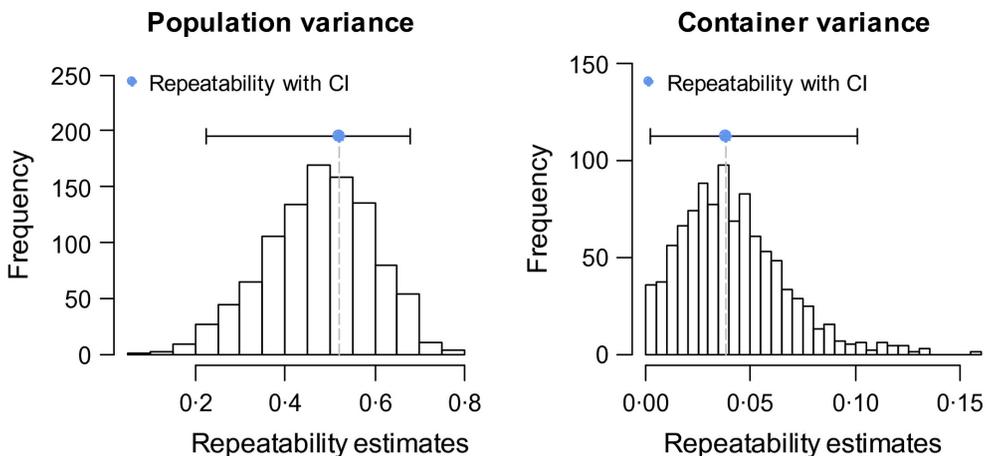


Figure 8.1: Examples for plots of rpt objects. The data show the analysis of a toy dataset of Poisson distributed count data analyzed with log link. Link scale repeatabilities are shown. CI = confidence interval.

Sometimes it is desired to fit models that include fixed effects, but to add the variance explained by fixed effects in the denominator of the repeatability estimation. In rptR, it is possible to estimate the variance explained by fixed effects as the variance in the linear predictor (i.e. the variance in fitted values) and to include this variance in the repeatability estimation. This functionality is controlled by the `adjusted` argument. By setting `adjusted = FALSE`, the model estimates repeatabilities with the variance explained by fixed effect in the denominator. We call this enhanced agreement repeatabilities.

```
rep2 <- rptPoisson (Egg~Treatment+(1 | Container)+
  (1 | Population), grname=c("Container",
  "Population"), data=BeetlesFemale, link="log",
  nboot=1000, parallel=TRUE, adjusted=FALSE)
```

Print	Summary																																																																																																																		
Repeatability estimation using the glmm method and log link Repeatability for Container ----- Link-scale approximation: R = 0-038 SE = 0-026 CI = [0-003, 0-107] P = 0-00874 [LRT] 0-086 [Permutation] Original-scale approximation: R = 0-032 SE = 0-024 CI = [0-003, 0-097] P = 0-00874 [LRT] 0-041 [Permutation] ----- Repeatability for Population ----- Link-scale approximation: R = 0-52 SE = 0-124 CI = [0-234, 0-684] P = 1-19e-16 [LRT] 0-001 [Permutation] Original-scale approximation: R = 0-504 SE = 0-124 CI = [0-221, 0-672] P = 1-19e-16 [LRT] 0-001 [Permutation]	Repeatability estimation using glmer method Call = rptPoisson(formula = Egg ~ Treatment + (1 Container) + (1 Population), grname = c("Container", "Population"), data = BeetlesFemale, link = "log", nboot = 1000, npermut = 1000, parallel = TRUE) Data: 480 observations ----- Container (60 groups) Repeatability estimation overview: <table> <thead> <tr> <th></th> <th>R</th> <th>SE</th> <th>2-5%</th> <th>97-5%</th> <th>P_permut</th> </tr> </thead> <tbody> <tr> <td>Org</td> <td>0-0319</td> <td>0-0237</td> <td>0-00276</td> <td>0-0966</td> <td>0-041</td> </tr> <tr> <td>Link</td> <td>0-0380</td> <td>0-0261</td> <td>0-00333</td> <td>0-1071</td> <td>0-086</td> </tr> </tbody> </table> Bootstrapping: <table> <thead> <tr> <th></th> <th>N</th> <th>Mean</th> <th>Median</th> <th>2-5%</th> <th>97-5%</th> </tr> </thead> <tbody> <tr> <td>Org</td> <td>1000</td> <td>0-0369</td> <td>0-0334</td> <td>0-00276</td> <td>0-0966</td> </tr> <tr> <td>Link</td> <td>1000</td> <td>0-0428</td> <td>0-0392</td> <td>0-00333</td> <td>0-1071</td> </tr> </tbody> </table> Permutation test: <table> <thead> <tr> <th></th> <th>N</th> <th>Mean</th> <th>Median</th> <th>2-5%</th> <th>97-5%</th> <th>P_permut</th> </tr> </thead> <tbody> <tr> <td>Org</td> <td>1000</td> <td>0-00742</td> <td>0-00132</td> <td>0</td> <td>0-0347</td> <td>0-041</td> </tr> <tr> <td>Link</td> <td>1000</td> <td>0-01126</td> <td>0-00203</td> <td>0</td> <td>0-0518</td> <td>0-086</td> </tr> </tbody> </table> Likelihood ratio test: logLik full model = -1178-331 logLik red. model = -1181-155 D = 5-65, d.f. = 1, P = 0-00874 ----- Population (12 groups) Repeatability estimation overview: <table> <thead> <tr> <th></th> <th>R</th> <th>SE</th> <th>2-5%</th> <th>97-5%</th> <th>P_permut</th> </tr> </thead> <tbody> <tr> <td>Org</td> <td>0-504</td> <td>0-124</td> <td>0-221</td> <td>0-672</td> <td>0-001</td> </tr> <tr> <td>Link</td> <td>0-520</td> <td>0-124</td> <td>0-234</td> <td>0-684</td> <td>0-001</td> </tr> </tbody> </table> Bootstrapping: <table> <thead> <tr> <th></th> <th>N</th> <th>Mean</th> <th>Median</th> <th>2-5%</th> <th>97-5%</th> </tr> </thead> <tbody> <tr> <td>Org</td> <td>1000</td> <td>0-459</td> <td>0-463</td> <td>0-221</td> <td>0-672</td> </tr> <tr> <td>Link</td> <td>1000</td> <td>0-473</td> <td>0-478</td> <td>0-234</td> <td>0-684</td> </tr> </tbody> </table> Permutation test: <table> <thead> <tr> <th></th> <th>N</th> <th>Mean</th> <th>Median</th> <th>2-5%</th> <th>97-5%</th> <th>P_permut</th> </tr> </thead> <tbody> <tr> <td>Org</td> <td>1000</td> <td>0-184</td> <td>0-184</td> <td>0-129</td> <td>0-240</td> <td>0-001</td> </tr> <tr> <td>Link</td> <td>1000</td> <td>0-244</td> <td>0-246</td> <td>0-176</td> <td>0-313</td> <td>0-001</td> </tr> </tbody> </table> Likelihood ratio test: logLik full model = -1178-331 logLik red. model = -1211-963 D = 67-3, d.f. = 1, P = 1-19e-16 -----		R	SE	2-5%	97-5%	P_permut	Org	0-0319	0-0237	0-00276	0-0966	0-041	Link	0-0380	0-0261	0-00333	0-1071	0-086		N	Mean	Median	2-5%	97-5%	Org	1000	0-0369	0-0334	0-00276	0-0966	Link	1000	0-0428	0-0392	0-00333	0-1071		N	Mean	Median	2-5%	97-5%	P_permut	Org	1000	0-00742	0-00132	0	0-0347	0-041	Link	1000	0-01126	0-00203	0	0-0518	0-086		R	SE	2-5%	97-5%	P_permut	Org	0-504	0-124	0-221	0-672	0-001	Link	0-520	0-124	0-234	0-684	0-001		N	Mean	Median	2-5%	97-5%	Org	1000	0-459	0-463	0-221	0-672	Link	1000	0-473	0-478	0-234	0-684		N	Mean	Median	2-5%	97-5%	P_permut	Org	1000	0-184	0-184	0-129	0-240	0-001	Link	1000	0-244	0-246	0-176	0-313	0-001
	R	SE	2-5%	97-5%	P_permut																																																																																																														
Org	0-0319	0-0237	0-00276	0-0966	0-041																																																																																																														
Link	0-0380	0-0261	0-00333	0-1071	0-086																																																																																																														
	N	Mean	Median	2-5%	97-5%																																																																																																														
Org	1000	0-0369	0-0334	0-00276	0-0966																																																																																																														
Link	1000	0-0428	0-0392	0-00333	0-1071																																																																																																														
	N	Mean	Median	2-5%	97-5%	P_permut																																																																																																													
Org	1000	0-00742	0-00132	0	0-0347	0-041																																																																																																													
Link	1000	0-01126	0-00203	0	0-0518	0-086																																																																																																													
	R	SE	2-5%	97-5%	P_permut																																																																																																														
Org	0-504	0-124	0-221	0-672	0-001																																																																																																														
Link	0-520	0-124	0-234	0-684	0-001																																																																																																														
	N	Mean	Median	2-5%	97-5%																																																																																																														
Org	1000	0-459	0-463	0-221	0-672																																																																																																														
Link	1000	0-473	0-478	0-234	0-684																																																																																																														
	N	Mean	Median	2-5%	97-5%	P_permut																																																																																																													
Org	1000	0-184	0-184	0-129	0-240	0-001																																																																																																													
Link	1000	0-244	0-246	0-176	0-313	0-001																																																																																																													

The data show the analysis of a toy dataset of Poisson distributed count data analyzed with log link, 1000 parametric bootstrap iterations for estimating confidence intervals and 1000 permutations for significance testing. R = repeatability, SE = standard error, CI = confidence interval, P = error probability (P-value), LRT = likelihood ratio test.

Table 8.1: Examples of print and summary displays of an rpt object

Furthermore, it is sometimes of interest to estimate the variances directly rather than as ratios of variances as represented by the repeatability. By setting the argument `ratio = FALSE`, rptR will estimate raw variances rather than repeatabilities. There are three reserved terms

to the `gname` argument, ‘Overdispersion’, ‘Residual’, and ‘Fixed’ that allow estimating the overdispersion variance (on the latent scale modelled as an observation specific random effect), residual variance (the sum of overdispersion and distribution-specific variance) and the variance explained by fixed effects.

```
rep3 <- rptPoisson (Egg ~ Treatment +(1 | Container) +
(1 |Population), gname=c("Container", "Population",
"Fixed", "Overdispersion", "Residual"),
data=BeetlesFemale, link="log", nboot=1000,
parallel=TRUE, ratio=FALSE, adjusted=FALSE)
```

Limitations

While point estimation is convenient and fast using rptR, the two Monte Carlo simulation steps can be slow. For a large dataset with multiple random effects such as the beetle toy dataset, a decent number of bootstraps or permutations can take several minutes. This time can be substantially shortened by conducting the simulations on multiple cores in parallel. For many applications, we suggest to initially fit the model with bootstrapping and permutations switched off (`nboot = 0`, `npermut = 0`). The final analysis should be done with larger numbers of bootstraps and permutations (possibly in separated steps, since computation times add up). It is also possible to increase the number of bootstraps and permutations by calling the function again with the arguments `update = TRUE` and the rpt object from the previous function call being handed over to the `rptObj` argument.

With the focus on `glmer` from the `lme4` package (Bates et al. 2015) as the central mixed-model fitting engine, we are limited to modelling additive overdispersion models. We have previously described repeatability estimation also for multiplicative overdispersion models (Nakagawa and Schielzeth, 2010), but since we find the model fits of `glmer` accurate, fast and efficient, we currently do not see the need for implementing estimation based on multiplicative overdispersion models.

The package does not work properly for data exhibiting underdispersion relative to the chosen GLMM family. Since the additive overdispersion term cannot drop below zero, the total residual variance cannot become smaller than the distribution specific variance. What seems like a limitation is in fact an unavoidable feature. While overdispersion can easily arise from unmodelled confounding effects, underdispersion signifies that the GLMM family is simply not appropriate. There is no formal warning build in in rptR for potential underdispersion, hence users are advised to check their data themselves. The lack of any link-scale residual variance

might, however, suggest potential underdispersion.

Availability

A stable version of the package (rptR 0.9.2 at the time of writing) can be downloaded from Comprehensive R Archive Network (CRAN, <https://cran.r-project.org/package=rptR>). This version will run with the current version of R (R 3.3.3 at the time of writing) and will provide the features introduced in this publication. At the same time, we aim to further develop the package, potentially improving plot and display features as well as supporting additional GLMM families. A developmental version will be available on Github (<https://github.com/mastoffel/rptR>), where it will be tested before being released on CRAN. The current stable Github version can be downloaded manually or via `install_github` from the devtools suite of functions (Wickham & Chang 2016): `install_github("mastoffel/rptR", build_vignettes = TRUE)`.

Acknowledgements

The package was motivated by the continued interest in and requests concerning an initial draft package on r-forge. We greatly appreciate the input and suggestions that we received from many colleagues. The manuscript and vignette have benefited from valuable input by Paul Johnson and two anonymous reviewers. The programming of the package was financially supported by the Research Centre for Mathematical Modelling (RCM2). H.S. was supported by an Emmy Noether fellowship from the German Research Foundation (SCHI 1188/1 2). S.N. is supported by a Future Fellowship, Australia (FT130100268).

Data accessibility

The simulated data used for illustration is distributed with the package. Once the package has been installed as described above, the data can be viewed by a call to `data(package="rptR")`. Data deposition: <https://cran.r-project.org/package=rptR>

Supplementary Information

rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models

Martin A Stoffel, Shinichi Nakagawa & Holger Schielzeth

Methods in Ecology & Evolution. 8:1639–1644. 2017

This Supplementary Information can be found **online** at
<https://doi.org/10.1111/2041-210X.12797>



Dawn on the archipelago.

9

GENERAL DISCUSSION

‘The great thing about population genetics is that you get to play in everyone’s back yard’¹. From the demographic past, ecology and life-history of a species to the complex chemical and microbial patterns associated with all living organisms, everything can be viewed in the light of genetic variation. The highly mathematical discipline of population genetics in the time of Wright, Fisher and Haldane has now transformed into a universal approach for understanding biological phenomena which requires a whole range of scientific efforts, from field and lab work to bioinformatics, mathematical modeling and statistics. To hold up with the demands of modern biological science in general, and high-throughput technology in particular, both reproducible research and methodological developments are the key to guarantee quality and progress of the field. On that note, these are the main findings of my thesis:

Main findings

One third of pinnipeds show signatures of genetic bottlenecks, which are mediated by variation in ecology and life-history.

Large scale commercial exploitation of pinnipeds around the globe is known to have decimated several species, yet it was largely unclear to what degree (IUCN, 2018; Wilson and Mittermeier, 2014). Although efforts had been made to infer population bottlenecks from genetic data for several species (e.g de Oliveira et al. 2009; Hedrick 1995; Hoffman et al. 2011; Osborne et al. 2016), most of these studies only estimate the presence or absence of genetic bottlenecks without any inference of their strength or associated population sizes. In Chapter 2, we showed that genetic demographic inference based on the very same demographic model is possible for a whole group of species and that one-third of pinnipeds underwent severe genetic bottlenecks as a consequence of overexploitation, sometimes with effective population sizes of only a few dozens of individuals. Furthermore, we provide evidence that a species’ demography in the Anthropocene can be highly dependent on its ecology and life-history, with land breeding species and highly polygynous species exhibiting stronger genetic bottleneck signals. While land breeders were likely to simply be more accessible than ice-breeders, highly polygynous species usually have very predictable breeding seasons, which probably made them easier targets for the sealing industry. Lastly, N_e is known to be lower in highly polygynous species (Nunney, 1993), which will also might slow down a genetic recovery after a bottleneck as compared to monogamous species.

¹Words of the eminent statistician John Tukey, while ‘statistics’ was replaced with ‘population genetics.’

Recent demography shapes genetic diversity across species.

Changes in N_e over time are predicted to strongly affect genetic diversity across species (Alcala and Vuilleumier, 2014; Ellegren and Galtier, 2016). Despite this prediction, the old riddle about the role of demography as a determinant of genetic diversity has been neglected for a long time (Leffler et al., 2012), which is likely due to the complexity and multitude of drivers of genetic diversity which can only be disentangled with large-scale comparative genetic data. In the largest effort so far, Romiguier et al. (2014) compared the genome-wide diversity of 76 animal species and revealed that life-history in the form of reproductive strategy is the main driver of diversity across species, while recent fluctuations in N_e did not seem to play a strong role. However, the relationship between demography and diversity has never been investigated for a group of closely related species which show substantial variation in their recent N_e , such as the pinnipeds (Leffler et al., 2012). In Chapter 2, we started filling this gap and showed that only the most severe genetic bottlenecks substantially reduced genetic diversity and that both demography and current population sizes explain much of the variation in genetic diversity across the pinnipeds.

The Northern elephant seal was very close to extinction, but probably expanded beforehand.

The question of how severely population bottlenecks affect genetic diversity is among the oldest in the field of evolutionary genetics (Nei et al., 1975). The empirical evidence for a severe loss of diversity as a consequence of bottlenecks is mixed, with some studies reporting lower diversity (Hoelzel et al., 2002; Houlden et al., 1996; Pinsky and Palumbi, 2014) while others still find unexpectedly high diversity after known population declines (Busch et al., 2007; Dinerstein and McCracken, 1990; Hailer et al., 2006). The Northern elephant seal has probably experienced one of the strongest declines among any vertebrate due to commercial overexploitation, and previous genetic studies of the Northern elephant seal indeed found very low genetic diversity (Hoelzel, 1999; Hoelzel et al., 2002; Weber et al., 2000). In Chapter 3, we used a newly sequenced genome and RAD sequencing data for 80 individuals to study the elephant seal bottleneck from a genomics perspective. In line with previous studies, we found that only an extreme bottleneck can cause a distortion of the site frequency spectrum as extreme as observed in the Northern elephant seal. Moreover, for the first time, genomic data made it possible to precisely estimate that a bottleneck effective population size of only two individuals over a time period of around ten generations is necessary to reduce diversity strongly enough to fit to the observed pattern.

Lastly, genomic analyses based on the SFS allowed us to investigate the pre-sealing demography of the species, which is unknown except for some archaeological evidence of the species' scarcity during the Holocene (Rick et al., 2011). We showed that the excess of rare alleles in the SFS of this species can be explained by a 10-fold growth of the population starting after the last glacial-period and lasting until the time of commercial sealing during the 19th century. Although in principle excessive rare alleles can also be caused by different demographic histories, there is strong evidence for post-glacial expansions in a wide range of species in the Nearctic (Burbrink et al., 2016), and such a scenario also explains the archaeological evidence of a small population size during the Holocene, with an increase in specimens during the last 3,500 years (Rick et al., 2011).

Skin chemical profiles encode a variety of information.

The fact that odour-based mate choice is real appears to be common knowledge, at least since the first studies on human preferences for genetically complementary mates using sweaty t-shirt experiments (Wedekind and Fürti, 1997; Wedekind et al., 1995). Later on, evidence for MHC-dependent mate choice came from a variety of organisms, in particular mice and rats (Lanyon et al., 2007; Singer et al., 1997; Yamazaki et al., 1999) but also sticklebacks (Aeschlimann et al., 2003) and house sparrows Bonneaud et al. (2006). In the wild, similar olfactory mechanisms to identify genotypes are expected not only for mate choice, but also for kin recognition, a critical mechanism in many species (Hurst and Beynon, 2010). However, a functional understanding of how genotypes can be chemically encoded in wild animals is largely lacking (Hurst and Beynon, 2010). In Chapter 4, we investigated the potential chemical underpinnings of olfactory communication in two colonies of Antarctic fur seals which likely have a strong olfactory recognition mechanism for their offspring (Dobson and Jouventin, 2003) and potentially even for finding mates (Hoffman et al., 2007*a*). We show that the skin chemicals of mothers and pups are similar, and therefore provide a potential self-referent recognition mechanism. Moreover, these chemical profiles, or fingerprints, appear to also encode individual genotypes through associations with both heterozygosity and genetic relatedness. While experimental studies are needed to confirm specific functions of skin chemicals, our analysis shows that complex chemical phenotypes in the wild have the potential to encode information about genotypes which could probably not be signalled by any other means.

Gut microbes as early life-history adaptations.

The idea that all animals live in a microbial world has taken the scientific world by storm (McFall-Ngai et al., 2013). Evidence for the immense importance of symbiotic microbes for development, health and function of their hosts is unequivocal (Diaz Heijtz et al., 2011; Lathrop et al., 2011; Pedersen et al., 2016; Zhu et al., 2011), but most research is based on humans or laboratory mice whilst wild microbiome studies are lacking (Hird, 2017). In Chapter 5, we examined the development of gut microbiota in young Northern elephant seals across a critical developmental period of several weeks after weaning. Although elephant seals are fasting during this period, we show that their gut microbiota are still complex and changing drastically, with several species going extinct and others colonising the gut. Unlike virtually all other studies in wild populations, we show that sex-differences in gut microbial communities are substantial, which could be an early adaptation to vastly different life-histories in Northern elephant seal males and females. Moreover, across all three sampling times, genotype is associated with gut bacterial communities in males but not in females. Although this pattern could be a reflection of a slower gut microbial development in females than in males, there is also a more functional explanation. If certain parts of the males' but not the females' genomes are under balancing selection, and these genes are linked to certain gut microbiota, this would only cause an association between male genotypes and gut microbiota. Such genes could be involved in any part of the male specific life-history, such as immunity or feeding patterns. Overall, our study highlights the advantages of a diet-controlled microbiota study in the wild, as the results are in stark contrast to existing studies in natural populations and reveal that intrinsic factors such as sex and genotype can strongly shape gut microbiota in the wild. Our study gives some of the first insights into the development of gut microbiota in a wild population with implications for understanding host-microbe co-evolution.

Major methodological advancements of this thesis.

The studies in this thesis often dealt with unusual and complex data, which required the development of novel methods. In the light of the current reproducibility crisis and to make these methods readily available for the community, the complete and documented analytical pipelines for all chapters are available at <https://github.com/mastoffel>. Moreover, some of the analyses which seemed particularly useful for the community are summarized in the R packages GCalignR, inbreedR and rptR and have together already been downloaded over 20,000 times (September 2018) from the official platform CRAN.

A reproducible and fast method for aligning Gas-Chromatography data.

Chemical communication is emerging as a field of increasing interest in animal ecology and evolution, and several studies have revealed chemical cues underlying fundamental signals and behaviours (Wyatt, 2014a). Gas-Chromatography (GC) coupled with either a flame ionization detector (GC-FID) or a mass spectrometer (GC-MS) is often the method of choice to quantify the molecules in complex chemical samples obtained from wild animals. Before statistically analysing GC samples however, a major obstacle must be overcome: Aligning homologous substances across many samples. This is particularly demanding in samples obtained from the field, which can differ substantially in their concentration. Due to a lack of appropriate alignment software, manual alignment is common (see Chapter 6), which is time consuming, prone to bias and makes a study impossible to reproduce. To address these problems, in Chapter 6 we presented GCalignR, an R package based on an alignment algorithm developed for Chapter 4, which allows the user to automatically align GC samples obtained from field studies. GCalignR not only provides a tool to make chemical communication studies reproducible but will decrease the time usually spent on manual alignment from weeks to only a few hours. Moreover, GCalignR includes sophisticated visualisation tools to optimise and correct alignments and evidently performs better than existing tools.

A framework for estimating inbreeding and inbreeding depression based on genetic markers.

Inbreeding depression is a phenomenon recognized for centuries (Charlesworth and Charlesworth, 1987), but individual inbreeding coefficients could traditionally only be measured when a detailed pedigree was available (Wright, 1922). When genetic markers became widely available, marker heterozygosity was often used as a proxy for genome-wide homozygosity and resulted a slew of so-called heterozygosity fitness correlations across many species (Chapman et al., 2009). However, the degree to which a few markers represent genome-wide homozygosity or identity by descent (IBDG) were unclear (Pemberton, 2004), and so were the explanations for heterozygosity fitness correlations (Szulkin et al., 2010). Although a large number of genomic markers certainly provide a more accurate representation of IBDG than a small number of genetic markers (Kardos et al., 2015), it is still not straightforward to elucidate how strongly individual inbreeding coefficients in a sample vary and how well this variation is reflected by a homozygosity at a given set of markers. The inbreedR package described in Chapter 7 implements a theoretical framework around the concept of identity disequilibrium (Slate et al., 2004) to address these questions. An identity disequilibrium is essentially a correlation of

heterozygosity across loci in a given sample of individuals and occurs when there is variation in inbreeding (David et al., 2007). Based on this framework, *inbreedR* provides functions to analyse variance in inbreeding and to explore the power of a set of genetic and genomic markers to estimate inbreeding depression. Finally, the package implements simulations to plan the number of genetic markers and sample sizes of future studies on inbreeding depression.

Calculating repeatabilities for Gaussian and non-Gaussian data.

The repeatability or intra-class coefficient is a very general statistic to describe how strongly units of a group resemble each other, where a group can for example consist of different individuals at one time point or measurements of one individual across many time points. The repeatability is commonly used to quantify the reproducibility of measurements or in biology to quantify animal personality. While formulas to calculate repeatabilities particularly for non-Gaussian traits have previously been described (Nakagawa and Schielzeth, 2010), they have not been implemented in user friendly software until the *rptR* package (Chapter 8) was published. Moreover, rather than focussing on point estimates, *rptR* provides sophisticated functions for parametric bootstrapping to quantify confidence intervals and permutation of residuals to estimate statistical significance. Lastly, we also implemented the calculation of repeatabilities in random slope models (Johnson, 2014), which will hopefully further increase the usability of the package.

Limitations and future directions

The neutral theory is dead. Long live the neutral theory ².

The genetic and genomic methods for demographic inference used in Chapter 2 and Chapter 3 are among the most cutting edge methods in the field (Beaumont, 2010; Excoffier et al., 2013). However, all genetic methods for demographic inference rely on the assumption that variation in the marker panel is shaped by changes in N_e rather than natural selection. While Kimura's neutral theory stated that most variation is selectively neutral (Kimura, 1983) and hence shaped by N_e , it has recently become clear that many putatively neutral loci are linked to loci under selection (Lohmueller et al., 2011). The degree to which the neutral theory is supported by recent genomic evidence has been severely questioned in a recent paper (Kern and Hahn, 2018) which caused a hot debate between leading scientists in the field ³. Overall, it is not

²Quote from the title of Martin Kreitman's paper from 1996 in *BioEssays*.

³<https://www.molecularecologist.com/2018/05/is-the-neutral-theory-dead/>

entirely clear to what degree genetic markers are linked to loci under selection, and how this might impact genetic demographic inference. To remedy this concern, we used a comparative approach and compared the same two demographic models across 30 pinniped species in Chapter 2. We had clear expectations of the extent to which land and ice-breeding species had undergone population bottlenecks and these were largely matched by the genetic demographic inference. Therefore, it seems unlikely that linked loci had much of an impact on the marker panel in this study. Moreover, when using genomic data to infer the genetic bottleneck of the Northern elephant seal in Chapter 3, the empirical site frequency spectrum (SFS) matched the expected SFS from the simulations very closely, again indicating that selection had little impact on the genetic markers.

I conclude that there are ways to escape the neutral marker dilemma, including comparative approaches and precise theoretical expectations. However, future studies should aim to validate genomic demographic inferences in an experimental setting using rapidly reproducing organisms where population histories can be manipulated and are therefore precisely known. Overall, the increasing ease of obtaining large scale genomic data will facilitate our knowledge of the unknown parameters governing genetic variation, such as mutation and recombination rates. Incorporating these parameters together with knowledge about loci under selection into demographic simulations of neutral variation will certainly increase the reliability of genetic demographic inference. To sum up, even if the major assumptions of the neutral theory might not hold true in the genomics age (Kern and Hahn, 2018), it will still have its place in empirical population genomics studies aiming to explore the demography of species.

The genetic underpinnings of chemical communication

The search for molecules involved in chemical communication, so called semiochemicals, and their genetic basis, is not easy (Hurst and Beynon, 2010; Wyatt, 2014a). The discovery of semiochemicals involves many steps, starting with field observations of a potential chemically mediated behavior to bioassays to prove that a substance can actually induce a behavior (see Figure 91). In Chapter 4, we only went half the way, from recognizing interesting behaviors in the Antarctic fur seal to identifying a potential set of substances (black colored part in Figure 91). Consequently, future studies should aim to synthesize these substances and use behavioural experiments to confirm their role as semiochemicals. Moreover, the subset of chemicals found in the Antarctic fur seal is likely to be incomplete, as the cotton swab sampling method does mostly capture non-volatile substances rather than volatiles (Kuecklich et al., 2017). I would hence recommend future studies to sample body odours using a set of complementary methods,

including cotton swabs for non-volatile substance and methods such as thermal desorption tubes for volatiles (Weiss et al., 2018). Lastly, after specific semiochemicals have been identified, modern genomic methods such as genome-wide association studies (GWAS) can be used to explore their underlying genetic basis (Figure 91).

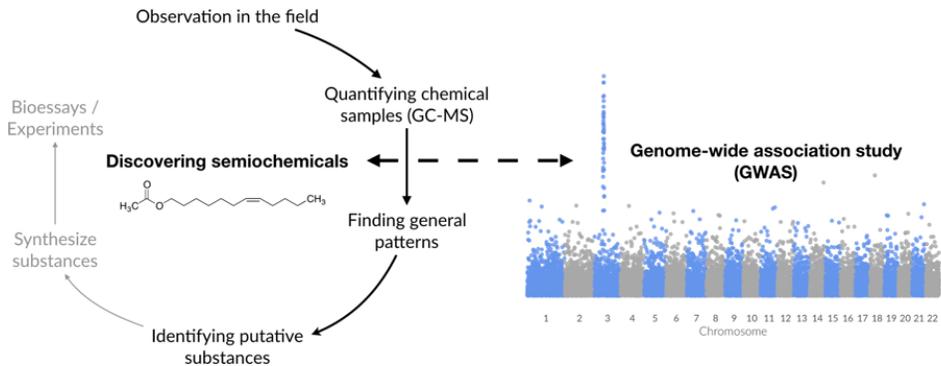


Figure 9.1: On the left side: The scientific efforts required to unambiguously identify semiochemicals. The depicted chemical structure is 7-dodecen-1-yl acetate, a sex pheromone in female Asian elephants (*Elephas maximus*). On the right side: A Manhattan plot representing a GWAS as a potential genomic approach to map the genetic basis of semiochemicals.

The host-genetic underpinnings of gut microbiota in the wild

While human studies now use large genomic datasets together with metagenomics analyses of microbiota to unravel the genetic basis of microbial variation (Kurilshikov et al., 2017), some questions can simply not be answered by studying humans or laboratory animals. These questions include the role of host-microbe associations for life-history or ecological adaptations, which need to be studied in wild animals. Nevertheless, studies in the wild will rarely reach the sample sizes or financial power of human studies and hence need more sophisticated sampling designs and approaches than the brute force large-throughput studies often conducted in human populations. In Chapter 5, we found that gut microbial communities of young Northern elephant seals differ strongly between the sexes and are linked to genetic relatedness in males. As we designed the study to control for potential confounding effects such as differences in diet, we concluded that differences in microbial communities might reflect different adaptations to sex-specific life-histories. While this is a novel finding in a wild population, it will be important for future studies to go one step further and elucidate the functional aspects of microbial diversity. Using approaches such as metagenomics or meta-transcriptomics make it

possible to quantify the genes or transcripts of gut microbiota, and their potential function can then be identified using online databases such as Gene ontology (Ashburner et al., 2000). The identified functional gene variation of microbiota can subsequently be linked to variation in their host's genome through GWAS and similar approaches. However, our study hopefully paved the way for future research in showing that appropriate sampling designs (e.g. controlling for environmental variation) and study species (e.g. sex-specific microbiota are expected in the highly sexual dimorphic elephant seal) will drastically decrease the necessary sample sizes to identify associations between gut microbiota and host genomes and will therefore make it possible to gain a functional understanding of gut microbiota even in wild populations.

A last reflection

Science has changed fundamentally in the last centuries, from a science where a single polymath such as Aristotle or Leonardo da Vinci could be a leading expert in physics, biology, geology and the fine arts, to a science of highly specialized sub-disciplines, where a researcher can work a whole career deciphering a single gene's function. This dissertation probably falls somewhere in between these two extremes. I certainly used a specific set of methods and perspectives, but the studies in this thesis are nevertheless connected by very broad questions: Where does genetic diversity come from (Chapters 2 and 3)? Is a biological unit only genes and proteins or a much more complex structure, including chemicals and bacteria (Chapters 4 and 5)? What are the consequences of human disturbances for the genetic variability in natural populations (Chapters 2 and 3)? Science is also becoming more complex, data intensive and collaborative. Scientific progress relies more than ever on the transparency and reproducibility of scientific findings and their methodological advancements. However, the academic reward system does not yet sufficiently encourage open research practices but puts scientists which spend time on documenting and publishing code and data in a disadvantage (Nosek et al., 2015). Nevertheless, the awareness for a new way of doing science, an open and truly collaborative way, where quality weighs more than quantity, is on the rise. The open science movement is growing and the studies in this dissertation are hopefully making a little contribution.

References

- Aarts, A. A., J. E. Anderson, C. J. Anderson, et al. 2015. Estimating the reproducibility of psychological science. *Science* **349**:aac4716.
- Aeschlimann, P. B., M. A. Haberli, T. Reusch, T. Boehm, and M. Milinski. 2003. Female sticklebacks *Gasterosteus aculeatus* use self-reference to optimize MHC allele number during mate selection. *Behavioral Ecology and Sociobiology* **54**:119–126.
- Ak, E. G. Z. A. D. L. 2001. The host genotype affects the bacterial community in the human gastrointestinal tract. *Microbial Ecology in Health and Disease* **13**:129–134.
- Alain B Pajarillo, E., J.-P. Chae, M. P. Balolong, H. Bum Kim, and D.-K. Kang. 2014. Assessment of fecal bacterial diversity among healthy piglets during the weaning transition. *The Journal of General and Applied Microbiology* **60**:140–146.
- Alcala, N., and S. Vuilleumier. 2014. Turnover and accumulation of genetic diversity across large time-scale cycles of isolation and connection of populations. *Proceedings of the Royal Society B: Biological Sciences* **281**:20141369.
- Allaire, J. J., J. Cheng, Y. Xie, et al., 2016. rmarkdown: Dynamic Documents for R. URL <https://CRAN.R-project.org/package=rmarkdown>.
- Allen, P. J., W. Amos, P. P. Pomeroy, and S. D. Twiss. 1995. Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two British breeding colonies. *Molecular Ecology* **4**:653–662.
- Allendorf, F. W. 2017. Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* **26**:420–430.
- Altshuler, D. M., R. M. Durbin, G. R. Abecasis, et al. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**:56–65.
- Anderson, M. J. 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**:32–46.
- Archer, F. I., P. E. Adams, and B. B. Schneiders. 2016. stratag: An rpackage for manipulating, summarizing and analysing population genetic data. *Molecular Ecology Resources* **17**:5–11.
- Arnold, C., L. J. Matthews, and C. L. Nunn. 2010. The 10kTrees website: A new online resource for primate phylogeny. *Evolutionary Anthropology: Issues, News, and Reviews* **19**:114–118.
- Ashburner, M., C. A. Ball, J. A. Blake, et al. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**:25–29.
- Atkinson, S., D. P. Demaster, and D. G. Calkins. 2008. Anthropogenic causes of the western Steller sea lion *Eumetopias jubatus* population decline and their threat to recovery. *Mammal Review* **38**:1–18.
- Avery, O. T., C. M. MacLeod, and M. McCarty. 1944. Studies on the chemical nature of the substance inducing transformation of Pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *Journal of Experimental Medicine* **79**:137–158.
- Baker, J. D., and T. C. Johanos. 2004. Abundance of the Hawaiian monk seal in the main Hawaiian Islands. *Biological Conservation* **116**:103–110.

References

- Baker, M. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* **533**:452–454.
- Ballou, A. L., R. A. Ali, M. A. Mendoza, et al. 2016. Development of the chick microbiome: How early exposure influences future microbial diversity. *Frontiers in Veterinary Science* **3**:395.
- Balloux, F., W. Amos, and T. Coulson. 2004. Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* **13**:3021–3031.
- Barton, K., 2016. MuMIn: multi-model inference. R package version 1.15.6. URL <http://CRAN.R-project.org/package=MuMIn>.
- Bates, D., M. Maechler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**:1–48.
- Beaumont, M. A. 2010. Approximate Bayesian Computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics* **41**:379–406.
- Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian Computation in population genetics. *Genetics* **162**:2025–2035.
- Begley, C. G., and L. M. Ellis. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* **483**:531.
- Bell, A. M., S. J. Hankison, and K. L. Laskowski. 2009. The repeatability of behaviour: a meta-analysis. *Animal Behaviour* **77**:771–783.
- Benjamin, D. J., J. O. Berger, M. Johannesson, et al. 2018. Redefine statistical significance. *Nature Human Behaviour* pages 1–5.
- Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate - A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**:289–300.
- Bennett, G., M. Malone, M. L. Sauter, et al. 2016. Host age, social group, and habitat type influence the gut microbiota of wild ring-tailed lemurs (*Lemur catta*). *American Journal of Primatology* **78**:883–892.
- Benson, A. K., S. A. Kelly, R. Legge, et al. 2010. Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**:18933–18938.
- Besnier, F., and K. A. Glover. 2013. ParallelStructure: A R Package to Distribute Parallel Runs of the Population Genetics Program STRUCTURE on Multi-Core Computers. *PLoS ONE* **8**:e70651.
- Bierne, N., A. Tsitronis, and P. David. 2000. An inbreeding model of associative overdominance during a population bottleneck. *Genetics* **155**:1981–1990.
- Bik, E. M., E. K. Costello, A. D. Switzer, et al. 2016. Marine mammals harbor unique microbiotas shaped by and yet distinct from the sea. *Nature Communications* **7**:10516.
- Biro, P. A., and J. A. Stamps. 2015. Using repeatability to study physiological and behavioural traits: ignore time-related change at your peril. *Animal Behaviour* **105**:223–230.
- Blaustein, A. R. 1983. Kin recognition mechanisms: Phenotypic matching or recognition alleles? *The American Naturalist* **121**:749–754.
- Bloemberg, T. G., J. Gerretzen, A. Lunshof, R. Wehrens, and L. M. C. Buydens. 2013. Warping methods for spectroscopic and chromatographic signal alignment: a tutorial. *Analytica Chimica Acta* **781**:14–32.
- Bloemberg, T. G., J. Gerretzen, H. J. P. Wouters, et al. 2010. Improved parametric time warping for proteomics. *Chemometrics and Intelligent Laboratory Systems* **104**:65–74.

- Boake, C. R. B. 1989. Repeatability: Its role in evolutionary studies of mating behavior. *Evolutionary Ecology* 3:173–182.
- Bobbie, C. B., N. C. S. Mykytczuk, and A. I. Schulte-Hostedde. 2017. Temporal variation of the microbiome is dependent on body region in a wild mammal (*Tamiasciurus hudsonicus*). *FEMS Microbiology Ecology* 93:434.
- Boeuf, B. J. L., D. E. Crocker, D. P. Costa, et al. 2000. Foraging ecology of northern elephant seals. *Ecological Monographs* 70:353.
- Boinski, S. 1988. Sex differences in the foraging behavior of squirrel monkeys in a seasonal habitat. *Behavioral Ecology and Sociobiology* 23:177–186.
- Bonadonna, F., and A. Sanz-Aguilar. 2012. Kin recognition and inbreeding avoidance in wild birds: the first evidence for individual kin-related odour recognition. *Animal Behaviour* 84:509–513.
- Boness, D. J., and W. D. Bowen. 1996. The evolution of maternal care in Pinnipeds. *Bioinformatics* 46:645–654.
- Bonin, C. A., M. E. Goebel, J. Forcada, R. S. Burton, and J. I. Hoffman. 2013. Unexpected genetic differentiation between recently recolonized populations of a long-lived and highly vagile marine mammal. *Ecology and Evolution* 3:3701–3712.
- Bonneaud, C., O. Chastel, P. Federici, H. Westerdahl, and G. Sorci. 2006. Complex MHC-based mate choice in a wild passerine. *Proceedings of the Royal Society B: Biological Sciences* 273:1111–1116.
- Bonnel, M. L., and R. K. Selander. 1974. Elephant seals: genetic variation and near extinction. *Science* 184:908–909.
- Bordenstein, S. R., and K. R. Theis. 2015. Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biology* 13.
- Boulay, R., X. Cerdá, T. Simon, M. Roldan, and A. Hefetz. 2007. Intraspecific competition in the ant *Camponotus cruentatus*: should we expect the ‘dear enemy’ effect? *Animal Behaviour* 74:985–993.
- Boulet, M., M. J. Charpentier, and C. M. Drea. 2009. Decoding an olfactory mechanism of kin recognition and inbreeding avoidance in a primate. *BMC Evolutionary Biology* 9:281.
- Bray, J. R., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs* 27:325–349.
- Breed, M. D., M. F. Garry, A. N. Pearce, et al. 1995. The role of wax comb in honey bee nestmate recognition. *Animal Behaviour* 50:489–496.
- Burbrink, F. T., Y. L. Chan, E. A. Myers, et al. 2016. Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates. *Ecology Letters* 19:1457–1467.
- Burbrink, F. T., F. Fontanella, R. Alexander Pyron, T. J. Guiher, and C. Jimenez. 2008. Phylogeography across a continent: The evolutionary and demographic history of the North American racer (Serpentes: Colubridae: *Coluber constrictor*). *Molecular Phylogenetics and Evolution* 47:274–288.
- Burgener, N., M. Dehnhard, H. Hofer, and M. L. East. 2009. Does anal gland scent signal identity in the spotted hyaena? *Animal Behaviour* 77:707–715.
- Busch, J. D., P. M. Waser, and J. A. DeWoody. 2007. Recent demographic bottlenecks are not accompanied by a genetic signature in banner-tailed kangaroo rats (*Dipodomys spectabilis*). *Molecular Ecology* 16:2450–2462.
- Bush, W. S., and J. H. Moore. 2012. Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology* 8:e1002822.
- Callahan, B. J., P. J. McMurdie, and S. P. Holmes. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* 11:2639–2643.

References

- Callahan, B. J., P. J. McMurdie, M. J. Rosen, et al. 2016a. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**:581–583.
- Callahan, B. J., K. Sankaran, J. A. Fukuyama, P. J. McMurdie, and S. P. Holmes. 2016b. Bioconductor workflow for microbiome data analysis: from raw reads to community analyses. *F1000Research* **5**:1492.
- Candon, S., A. Perez-Arroyo, C. Marquet, et al. 2015. Antibiotics in early life alter the gut microbiome and increase disease incidence in a spontaneous mouse model of autoimmune insulin-dependent diabetes. *PLoS ONE* **10**:e0125448.
- Caspers, B. A., F. C. Schroeder, S. Franke, and C. C. Voigt. 2011. Scents of adolescence: The maturation of the olfactory phenotype in a free-ranging mammal. *PLoS ONE* **6**:e21162.
- Cassini, M. H. 1999. The evolution of reproductive systems in pinnipeds. *Behavioral Ecology* **10**:612–616.
- Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**:3124–3140.
- Cattell, R. B. 1966. The scree test for the number of factors. *Multivariate Behavioral Research* **1**:245–276.
- Ceballos, G., P. R. Ehrlich, and R. Dirzo. 2017. Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America* **47**:201704949–E6096.
- Chan, Y. L., C. N. K. Anderson, and E. A. Hadly. 2006. Bayesian estimation of the timing and severity of a population bottleneck from ancient DNA. *PLoS Genetics* **2**:e59.
- Chan, Y. L., D. Schanzenbach, and M. J. Hickerson. 2014. Detecting concerted demographic response across community assemblages using hierarchical Approximate Bayesian Computation. *Molecular Biology and Evolution* **31**:2501–2515.
- Chapman, J. R., S. Nakagawa, D. W. Coltman, J. Slate, and B. C. Sheldon. 2009. A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology* **18**:2746–2765.
- Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**:195–205.
- Charlesworth, B., and D. Charlesworth. 1999. The genetic basis of inbreeding depression. *Genetics Research* **74**:329–340.
- Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology & Systematics* **18**:237–268.
- Charlesworth, D., and J. H. Willis. 2009. The genetics of inbreeding depression. *Nature Reviews Genetics* **10**:783–796.
- Charpentier, M. J. E., M. Boulet, and C. M. Drea. 2008. Smelling right: the scent of male lemurs advertises genetic quality and relatedness. *Molecular Ecology* **17**:3225–3233.
- Charpentier, M. J. E., J. C. Crawford, M. Boulet, and C. M. Drea. 2010. Message ‘scent’: lemurs detect the genetic relatedness and quality of conspecifics via olfactory cues. *Animal Behaviour* **80**:101–108.
- Cheesman, S. E., J. T. Neal, E. Mitge, B. M. Seredick, and K. Guillemin. 2011. Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88. *Proceedings of the National Academy of Sciences of the United States of America* **108**:4570–4577.
- Cho, I., S. Yamanishi, L. Cox, et al. 2012. Antibiotics in early life alter the murine colonic microbiome and adiposity. *Nature* **488**:621–626.
- Clark, R. I., A. Salazar, R. Yamada, et al. 2015. Distinct shifts in microbiota composition during drosophila aging impair intestinal function and drive mortality. *Cell Reports* **12**:1656–1667.

- Clark, T. B. 2006. Population Structure of *Manta birostris* (Chondrichthyes: Mobulidae) From the Pacific and Atlantic Oceans. Masters Thesis pages 1–78.
- Clarke, K. R. 1999. Nonmetric multivariate analysis in community-level ecotoxicology. *Environmental Toxicology and Chemistry* **18**:118.
- Clarke, K. R., and R. M. Warwick. 2001. *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*. 2nd edition. Plymouth Marine Laboratory, Plymouth, UK.
- Coltman, D. W., J. G. Pilkington, J. A. Smith, and J. M. Pemberton. 1999. Parasite-mediated selection against inbred soay sheep in a free-living island population. *Evolution* **53**:1259–1267.
- Coltman, D. W., and J. Slate. 2003. Microsatellite measures of inbreeding: a meta-analysis. *Evolution* **57**:971–983.
- Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* .
- Coventry, A., L. M. Bull-Ottersson, X. Liu, et al. 2010. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. *Nature Communications* **1**:131.
- Cox, L. M., S. Yamanishi, J. Sohn, et al. 2014. Altering the intestinal microbiota during a critical developmental window has lasting metabolic consequences. *Cell* **158**:705–721.
- Coyer, J. A., A. F. Peters, W. T. Stam, and J. L. Olsen. 2003. Post-ice age recolonization and differentiation of *Fucus serratus* L. (Phaeophyceae; Fucaceae) populations in Northern Europe. *Molecular Ecology* **12**:1817–1829.
- Crawford, J. C., M. Boulet, and C. M. Drea. 2010. Smelling wrong: hormonal contraception in lemurs alters critical female odour cues. *Proceedings of the Royal Society B: Biological Sciences* **278**:122–130.
- Crawley, M. J. 2002. *Statistical Computing, an Introduction to Data Analysis Using S-plus*. John Wiley and Sons, Chichester, UK.
- Cristina Lorenzi, M., R. Cervo, and A.-G. Bagnères. 2011. Facultative social parasites mark host nests with branched hydrocarbons. *Animal Behaviour* **82**:1143–1149.
- Crnokrak, P., and D. A. Roff. 1999. Inbreeding depression in the wild. *Heredity* **83**:260–270.
- Csilléry, K., M. G. B. Blum, O. E. Gaggiotti, and O. François. 2010. Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* **25**:410–418.
- Csilléry, K., O. François, and M. G. B. Blum. 2012. abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3**:475–479.
- Dahm, R. 2005. Friedrich Miescher and the discovery of DNA. *Developmental Biology* **278**:274–288.
- Darwin, C. 1859. *On the Origin of Species*. Routledge.
- Daszykowski, M., Y. Vander Heyden, C. Boucon, and B. Walczak. 2010. Automated alignment of one-dimensional chromatographic fingerprints. *Journal of Chromatography. A* **1217**:6127–6133.
- David, L. A., A. C. Materna, J. Friedman, et al. 2014a. Host lifestyle affects human microbiota on daily timescales. *Genome Biology* **15**.
- David, L. A., C. F. Maurice, R. N. Carmody, et al. 2014b. Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **505**:559–563.
- David, P. 1998. Heterozygosity-fitness correlations: new perspectives on old problems. *Heredity* **80**:531–537.
- David, P., B. Delay, P. Berthou, and P. Jarne. 1995. Alternative models for allozyme-associated heterosis in the marine bivalve *Spisula ovalis*. *Genetics* **139**:1719–1726.

References

- David, P., B. Pujol, F. Viard, V. Castella, and J. Goudet. 2007. Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology* **16**:2474–2487.
- Davis, C. S., T. S. Gelatt, D. Siniff, and C. Strobeck. 2002. Dinucleotide microsatellite markers from the Antarctic seals and their use in other Pinnipeds. *Molecular Ecology Notes* **2**:203–208.
- De Meulemeester, T., P. Gerbaux, M. Boulvin, A. Coppée, and P. Rasmont. 2011. A simplified protocol for bumble bee species identification by cephalic secretion analysis. *Insectes Sociaux* **58**:227–236.
- de Oliveira, L. R., D. Meyer, J. Hoffman, P. Majluf, and J. S. Morgante. 2009. Evidence of a genetic bottleneck in an El Niño affected population of South American fur seals, *Arctocephalus australis*. *Journal of the Marine Biological Association of the United Kingdom* **89**:1717–1725.
- de Villemereuil, P., H. Schielzeth, S. Nakagawa, and M. Morrissey. 2016. General methods for evolutionary quantitative genetic inference from generalized mixed models. *Genetics* **204**:1281–1294.
- Degnan, P. H., A. E. Pusey, E. V. Lonsdorf, et al. 2012. Factors associated with the diversification of the gut microbial communities within chimpanzees from Gombe National Park. *Proceedings of the National Academy of Sciences of the United States of America* **109**:13034–13039.
- Dellicour, S., and T. Lecocq. 2013. GCALIGNER 1.0: An alignment program to compute a multiple sample comparison data matrix from large eco-chemical datasets obtained by GC. *Journal of Separation Science* **105**:n/a–n/a.
- Di Rienzo, A., A. C. Peterson, J. C. Garza, et al. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the United States of America* **91**:3166–3170.
- Díaz Heijtz, R., S. Wang, F. Anuar, et al. 2011. Normal gut microbiota modulates brain development and behavior. *Proceedings of the National Academy of Sciences of the United States of America* **108**:3047–3052.
- Dinerstein, E., and G. F. McCracken. 1990. Endangered greater one-horned rhinoceros carry high levels of genetic variation. *Conservation Biology* **4**:417–422.
- Dobson, F. S., and P. Jouventin. 2003. How mothers find their pups in a colony of Antarctic fur seals. *Behavioural Processes* **61**:77–85.
- Dobzhansky, T. 1957. *Evolution, Genetics & Man*. 2nd edition. John Wiley and Sons.
- Doidge, D. W., J. P. Croxall, and C. Ricketts. 1984. Growth rates of Antarctic fur seal *Arctocephalus gazella* pups at South Georgia. *Journal of Zoology* **203**:87–93.
- Dornburg, A., M. C. Brandley, M. R. McGowen, and T. J. Near. 2012. Relaxed clocks and inferences of heterogeneous patterns of nucleotide substitution and divergence time estimates across whales and dolphins (Mammalia: Cetacea). *Molecular Biology and Evolution* **29**:721–736.
- Duchen, P., D. Zivkovic, S. Hutter, W. Stephan, and S. Laurent. 2013. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. *Genetics* **193**:291–301.
- Eddy, S. R. 2004. What is dynamic programming? *Nature Biotechnology* **22**:909–910.
- Edgar, R. C., and H. Flyvbjerg. 2015. Error filtering, pair assembly and error correction for next-generation sequencing reads. *Bioinformatics* **31**:3476–3482.
- Eid, M., M. Gollwitzer, and M. Schmitt. 2010. *Statistik und Forschungsmethoden*. Beltz, Weinheim, Germany.
- Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* **5**:435–445.
- Ellegren, H., and N. Galtier. 2016. Determinants of genetic diversity. *Nature Reviews Genetics* **17**:422–433.
- Erlandson, J. M., M. A. Tveskov, and R. Scott Bryam. 1998. The development of maritime adaptations on the southern Northwest Coast of North America. *Journal of California and Great Basin Anthropology* **35**:6–22.

- Etter, P. D., S. Bassham, P. A. Hohenlohe, E. A. Johnson, and W. A. Cresko. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods in Molecular Biology* 772:157–178.
- Evans, S. N., Y. Shvets, and M. Slatkin. 2007a. Non-equilibrium theory of the allele frequency spectrum. *Theoretical Population Biology* 71:109–119.
- Evans, S. N., Y. Shvets, and M. Slatkin. 2007b. Non-equilibrium theory of the allele frequency spectrum. *Theoretical Population Biology* 71:109–119.
- Excoffier, L., I. Dupanloup, E. Huerta-Sánchez, V. C. Sousa, and M. Foll. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* 9:e1003905.
- Excoffier, L., and M. Foll. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27:1332–1334.
- Fabrigar, L. R., D. T. Wegener, R. C. MacCallum, and E. J. Strahan. 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 4:272–299.
- Fagundes, N. J. R., N. Ray, M. Beaumont, et al. 2007. Statistical evaluation of alternative models of human evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104:17614–17619.
- Falissard, B., 2012. psy: Various procedures used in psychometry. R package version 1.1. URL <http://cran.r-project.org/package=psy>.
- Ferguson, S. H., and J. W. Higdson. 2006. How seals divide up the world: environment, life history, and conservation. *Oecologia* 150:318–329.
- Fisher, R. A. 1931. XVII.—The distribution of gene ratios for rare mutations. *Proceedings of the Royal Society of Edinburgh* 50:204–219.
- Foitzik, S., H. Sturm, K. Pusch, P. D’Etorre, and J. Heinze. 2007. Nestmate recognition and intraspecific chemical and genetic variation in *Temnothorax* ants. *Animal Behaviour* 73:999–1007.
- Foote, A. D., N. Vijay, M. C. Ávila-Arcos, et al. 2016. Genome-culture coevolution promotes rapid divergence of killer whale ecotypes. *Nature Communications* 7:11693.
- Forcada, J., and J. I. Hoffman. 2014. Climate change selects for heterozygosity in a declining fur seal population. *Nature* 511:462–465.
- Forstmeier, W., H. Schielzeth, J. C. Mueller, H. Ellegren, and K. Bart. 2012. Heterozygosity-fitness correlations in zebra finches: microsatellite markers can be better than their reputation. *Molecular Ecology* 21:3237–3249.
- Fountain, E. D., J. N. Pauli, B. N. Reid, P. J. Palsbøll, and M. Z. Peery. 2016. Finding the right coverage: the impact of coverage and sequence quality on single nucleotide polymorphism genotyping error rates. *Molecular Ecology Resources* 16:966–978.
- Frankham, R. 2005. Genetics and extinction. *Biological Conservation* 126:131–140.
- Frankham, R., K. Lees, M. E. Montgomery, et al. 1999. Do population size bottlenecks reduce evolutionary potential? *Animal Conservation* 2:255–260.
- Franklin, I. R. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theoretical Population Biology* 11:60–80.
- Franklin, R. E., and R. G. Gosling. 1953. Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate. *Nature* 172:156–157.
- Frese, S. A., K. Parker, C. C. Calvert, and D. A. Mills. 2015. Diet shapes the gut microbiome of pigs during nursing and weaning. *Microbiome* 3:1237439.

References

- Fu, Y. X. 1995. Statistical properties of segregating sites. *Theoretical Population Biology* **48**:172–197.
- Gamer, M., J. Lemon, I. Fellows, and P. Singh. 2012. irr: Various coefficients of interrater reliability and agreement. R package version 0.84. URL <http://cran.r-project.org/package=irr>.
- Garza, J. C., and E. G. Williamson. 2001. Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10**:305–318.
- Gehara, M., A. A. Garda, F. P. Werneck, et al. 2017. Estimating synchronous demographic changes across populations using hABC and its application for a herpetological community from northeastern Brazil. *Molecular Ecology* **26**:4756–4771.
- Gelman, A. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* **27**:2865–2873.
- Gelman, A., J. Carlin, H. Stern, et al. 1995. *Bayesian Data Analysis*. Chapman and Hall/CRC.
- Gentleman, R. C., V. J. Carey, D. M. Bates, et al. 2004. Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology* **5**.
- Gentry, R. L., U. S. N. M. F. Service, and J. R. Holt. 1982. Equipment and techniques for handling Northern fur seals. Technical report, Seattle, WA, USA.
- Gilbert, S. F., J. Sapp, and A. I. Tauber. 2012. A symbiotic view of life: We have never been individuals. *Quarterly Review of Biology* **87**:325–341.
- Giongo, A., K. A. Gano, D. B. Crabb, et al. 2011. Toward defining the autoimmune microbiome for type 1 diabetes. *The ISME Journal* **5**:82–91.
- Glad, T., V. F. Kristiansen, K. M. Nielsen, et al. 2010. Ecological characterisation of the colonic microbiota in arctic and sub-arctic seals. *Microbial Ecology* **60**:320–330.
- Goodman, S. J. 1997. Dinucleotide repeat polymorphisms at seven anonymous microsatellite loci cloned from the European harbour seal (*Phoca vitulina vitulina*). *Animal Genetics* **28**:310–311.
- Goodrich, J. K., E. R. Davenport, M. Beaumont, et al. 2016. Genetic determinants of the gut microbiome in UK twins. *Cell Host & Microbe* **19**:731–743.
- Gorvitovskaia, A., S. P. Holmes, and S. M. Huse. 2016. Interpreting Prevotella and Bacteroides as biomarkers of diet and lifestyle. *Microbiome* **4**:174.
- Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**:1–19.
- Gravel, S., B. M. Henn, R. N. Gutenkunst, et al. 2011. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America* **108**:11983–11988.
- Greene, L. K., and C. M. Drea. 2014. Love is in the air: sociality and pair bondedness influence siaka reproductive signalling. *Animal Behaviour* **88**:147–156.
- Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics* **43**:1031–U151.
- Gutenkunst, R. N., R. D. Hernandez, and S. H. Williamson. 2010. Diffusion approximations for demographic inference: DaDi. *Nature Precedings*.
- Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*.
- Hailer, F., B. Helander, A. O. Folkestad, et al. 2006. Bottlenecked but long-lived: high genetic diversity retained in white-tailed eagles upon recovery from population decline. *Biology Letters* **2**:316–319.

- Halsey, L. G., D. Curran-Everett, S. L. Vowler, and G. B. Drummond. 2015. The fickle P value generates irreproducible results. *Nature Methods* **12**:179–185.
- Hamilton, J. E. 1956. Scent of Otariids. *Nature* **177**:900–900.
- Han, E., J. S. Sinsheimer, and J. Novembre. 2014. Characterizing bias in population genetic inferences from low-coverage sequencing data. *Molecular Biology and Evolution* **31**:723–735.
- Hansson, B., and L. Westerberg. 2002. On the correlation between heterozygosity and fitness in natural populations. *Molecular Ecology* **11**:2467–2474.
- Hansson, B., and L. Westerberg. 2008. Heterozygosity-fitness correlations within inbreeding classes: local or genome-wide effects? *Conservation Genetics* **9**:73–83.
- Hardy, M. H., E. Roff, T. G. Smith, and M. Ryg. 1991. Facial skin glands of ringed and grey seals, and their possible function as odoriferous organs. *Canadian Journal of Zoology* **69**:189–200.
- Harris, H. 1966. Enzyme polymorphisms in man. *Proceedings of the Royal Society Series B-Biological Sciences* **164**:298–8.
- Hartl, D. L., and A. G. Clark. 1997. *Principles of Population Genetics*. Sinauer Associates, Sunderland.
- Hauber, M. E., and P. W. Sherman. 2001. Self-referent phenotype matching: theoretical considerations and empirical evidence. *Trends in Neurosciences* **24**:609–616.
- Hedrick, P. W. 1995. Elephant seals and the estimation of a population bottleneck. *Journal of Heredity* **86**:232–235.
- Hernandez-Velazquez, F. D., C. E. Galindo-Sanchez, M. I. Taylor, et al. 2005. New polymorphic microsatellite markers for California sea lions (*Zalophus californianus*). *Molecular Ecology Notes* **5**:140–142.
- Hershey, A. D., and M. Chase. 1952. Independent functions of viral protein and nucleic acid in growth of Bacteriophage. *Journal of General Physiology* **36**:39–56.
- Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. *Nature* **405**:907–913.
- Hewitt, G. M. 1999. Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society* **68**:87–112.
- Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**:226–231.
- Hill, W. G., and B. S. Weir. 2011. Variation in actual relationship as a consequence of Mendelian sampling and linkage. *Genetics Research* **93**:47–64.
- Hird, S. M. 2017. Evolutionary biology needs wild microbiomes. *Frontiers in Microbiology* **8**:689.
- Hoban, S. M., O. E. Gaggiotti, and G. Bertorelle. 2013. The number of markers and samples needed for detecting bottlenecks under realistic scenarios, with and without recovery: a simulation-based study. *Molecular Ecology* **22**:3444–3450.
- Hoelzel, A. 1999. Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biological Journal of the Linnean Society* **68**:23–39.
- Hoelzel, A. R., R. C. Fleischer, C. Campagna, B. J. Le Boeuf, and G. Alvord. 2002. Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology* **15**:567–575.
- Hoelzel, A. R., J. Halley, S. J. O'Brien, et al. 1993. Elephant Seal Genetic Variation and the Use of Simulation Models to Investigate Historical Population Bottlenecks. *Journal of Heredity* **84**:443–449.
- Hoffman, J. I. 2016. Reproducibility: Archive computer code with raw data. *Nature* **534**:326–326.

References

- Hoffman, J. I., I. L. Boyd, and W. Amos. 2003. Male reproductive strategy and the importance of maternal status in the Antarctic fur seal *Arctocephalus gazella*. *Evolution* 57:1917.
- Hoffman, J. I., I. L. Boyd, and W. Amos. 2004. Exploring the relationship between parental relatedness and male reproductive success in the Antarctic fur seal *Arctocephalus gazella*. *Evolution* 58:2087.
- Hoffman, J. I., I. L. Boyd, and W. Amos. 2005. Exploring the relationship between parental relatedness and male reproductive success in the Antarctic fur seal *Arctocephalus gazella*. *Evolution* pages 1–13.
- Hoffman, J. I., K. K. Dasmahapatra, and H. J. Nichols. 2008. Permanent genetic resources: Ten novel polymorphic dinucleotide microsatellite loci cloned from the Antarctic fur seal *Arctocephalus gazella*. *Molecular Ecology Resources* 8:459–461.
- Hoffman, J. I., J. Forcada, and W. Amos. 2010. Exploring the mechanisms underlying a heterozygosity–fitness correlation for canine size in the Antarctic fur seal *Arctocephalus gazella*. *Journal of Heredity* 101:539–552.
- Hoffman, J. I., J. Forcada, P. N. Trathan, and W. Amos. 2007a. Female fur seals show active choice for males that are heterozygous and unrelated. *Nature* 445:912–914.
- Hoffman, J. I., S. M. Grant, J. Forcada, and C. D. Phillips. 2011. Bayesian inference of a historical bottleneck in a heavily exploited marine mammal. *Molecular Ecology* 20:3989–4008.
- Hoffman, J. I., F. Simpson, P. David, et al. 2014. High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America* 111:3775–3780.
- Hoffman, J. I., S. Steinfartz, and J. B. W. Wolf. 2007b. Ten novel dinucleotide microsatellite loci cloned from the Galápagos sea lion (*Zalophus californianus wollebaeki*) are polymorphic in other pinniped species. *Molecular Ecology Notes* 7:103–105.
- Houlden, B. A., P. R. England, A. C. Taylor, W. D. Greville, and W. B. Sherwin. 1996. Low genetic variability of the koala *Phascolarctos cinereus* in south-eastern Australia following a severe population bottleneck. *Molecular Ecology* 5:269–281.
- Hudson, R. R. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
- Huebinger, R. M., E. E. Louis, T. Gelatt, L. D. Rea, and J. W. Bickham. 2007. Characterization of eight microsatellite loci in Steller sea lions (*Eumetopias jubatus*). *Molecular Ecology Notes* 7:1097–1099.
- Huisman, J., L. E. B. Kruuk, P. A. Ellis, T. Clutton-Brock, and J. M. Pemberton. 2016. Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America* 113:3585–3590.
- Humble, E., K. K. Dasmahapatra, A. Martinez-Barrio, et al. 2018. RAD sequencing and a hybrid Antarctic fur seal genome assembly reveal rapidly decaying linkage disequilibrium, global population structure and evidence for inbreeding. *G3* 8:2709–2722.
- Hurst, J. L., and R. J. Beynon. 2010. Making progress in genetic kin recognition among vertebrates. *Journal of Biology* 9:13.
- Innsley, S. J., A. V. Phillips, and I. Charrier. 2003. A review of social recognition in pinnipeds. *Aquatic Mammals* 29:181–201.
- Ioannidis, J. 2005. Why most published research findings are false. *Plos Medicine* 2:696–701.
- IUCN. 2018. The IUCN Red List of Threatened Species .
- Ivarsson, E., S. Roos, H. Y. Liu, and J. E. Lindberg. 2014. Fermentable non-starch polysaccharides increases the abundance of *Bacteroides*–*Prevotella*–*Porphyromonas* in ileal microbial community of growing pigs. *Animal* 8:1777–1787.

- Johnson, C. A., P. L. Phelan, and J. M. Herbers. 2008. Stealth and reproductive dominance in a rare parasitic ant. *Animal Behaviour* 76:1965–1976.
- Johnson, K. J., B. W. Wright, K. H. Jarman, and R. E. Synovec. 2003. High-speed peak matching algorithm for retention time alignment of gas chromatographic data for chemometric analysis. *Journal of Chromatography. A* 996:141–155.
- Johnson, P. C. D. 2014. Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution* 5:944–946.
- Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* 10:551–555.
- Jones, T. L., W. R. Hildebrandt, D. J. Kennett, and J. F. Porcasi. 2002. Prehistoric marine mammal overkill in the northeastern Pacific: A review of new evidence. *Journal of California and Great Basin Anthropology* 24:69–80.
- Jordan, N. R., M. B. Manser, F. Mwanguhya, et al. 2011. Scent marking in wild banded mongooses: 1. Sex-specific scents and overmarking. *Animal Behaviour* 81:31–42.
- Joshi, P. K., T. Esko, H. Mattsson, et al. 2015. Directional dominance on stature and cognition in diverse human populations. *Nature* 523:459–U176.
- Karamanlidis, A. A., P. Dendrinou, P. F. de Larrinoa, et al. 2016. The Mediterranean monk seal *Monachus monachus*: status, biology, threats, and conservation priorities. *Mammal Review* 46:92–105.
- Kardos, M., F. W. Allendorf, and G. Luikart. 2014. Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium? *Molecular Ecology Resources* 14:519–530.
- Kardos, M., G. Luikart, and F. W. Allendorf. 2015. Measuring individual inbreeding in the age of genomics: marker-based measures are better than pedigrees. *Heredity* 115:63–72.
- Kardos, M., A. Qvarnström, and H. Ellegren. 2017. Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. *Genetics* 205:1319–1334.
- Keinan, A., and A. G. Clark. 2012. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* 336:740–743.
- Keller, L. 2002. Inbreeding effects in wild populations. *Trends in Ecology & Evolution* 17:230–241.
- Keller, M. C., P. M. Visscher, and M. E. Goddard. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189:237–249.
- Kern, A. D., and M. W. Hahn. 2018. The neutral theory in light of natural selection. *Molecular Biology and Evolution* 35:1366–1371.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University Press.
- Kingman, J. F. C. 1982. The coalescent. *Stochastic Processes and their Applications* 13:235–248.
- Kirchner, M., B. Saussen, H. Steen, J. A. J. Steen, and F. A. Hamprecht. 2007. amsrpm: Robust Point Matching for Retention Time Alignment of LC/MS Data with R. *Journal of Statistical Software* 18:1–12.
- Kishida, T., S. Kubota, Y. Shirayama, and H. Fukami. 2007. The olfactory receptor gene repertoires in secondary-adapted marine vertebrates: evidence for reduction of the functional proportions in cetaceans. *Biology Letters* 3:428–430.
- Klein, S. L., and K. L. Flanagan. 2016. Sex differences in immune responses. *Nature Reviews Immunology* 16:626–638.
- Klindworth, A., E. Pruesse, T. Schweer, et al. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research* 41:e1–e1.

References

- Korneliusson, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**:356.
- Kováts, E. 1958. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil I: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helvetica Chimica Acta* **41**:1915–1932.
- Kowalewsky, S., M. Dambach, B. Mauck, and G. Dehnhardt. 2006. High olfactory sensitivity for dimethyl sulphide in harbour seals. *Biology Letters* **2**:106–109.
- Krause, E. T., O. Krüger, P. Kohlmeier, and B. A. Caspers. 2012. Olfactory kin recognition in a songbird. *Biology Letters* **8**:327–329.
- Krüger, O., J. B. W. Wolf, R. M. Jonker, J. I. Hoffman, and F. Trillmich. 2014. Disentangling the contribution of sexual selection and ecology to the evolution of size dimorphism in pinnipeds. *Evolution* **68**:1485–1496.
- Kryukov, G. V., L. A. Pennacchio, and S. R. Sunyaev. 2007. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *The American Journal of Human Genetics* **80**:727–739.
- Kuecklich, M., M. Moeller, A. Marcillo, et al. 2017. Different methods for volatile sampling in mammals. *PLoS ONE* **12**.
- Kundu, P., E. Blacher, E. Elinav, and S. Pettersson. 2017. Our gut microbiome: The evolving inner self. *Cell* **171**:1481–1493.
- Kurilshikov, A., C. Wijmenga, J. Fu, and A. Zhernakova. 2017. Host genetics and gut microbiome: Challenges and perspectives. *Trends in Immunology* **38**:633–647.
- Lande, R. 1988. Genetics and demography in biological conservation. *Science* **241**:1455–1460.
- Lande, R. 1994. Risk of population extinction from fixation of new deleterious mutations. *Evolution* **48**:1460–1469.
- Lange, E., R. Tautenhahn, S. Neumann, and C. Gröpl. 2008. Critical assessment of alignment procedures for LC-MS proteomics and metabolomics measurements. *BMC Bioinformatics* **9**:375.
- Langille, M. G., C. J. Meehan, J. E. Koenig, et al. 2014. Microbial shifts in the aging mouse gut. *Microbiome* **2**:260.
- Lanyon, C. V., S. P. Rushton, A. G. O'Donnell, et al. 2007. Murine scent mark microbial communities are genetically determined. *FEMS Microbiology Ecology* **59**:576–583.
- Lathrop, S. K., S. M. Bloom, S. M. Rao, et al. 2011. Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478**:250–254.
- Le Boeuf, B. J., and R. M. Laws. 1994. *Elephant Seals: Population Ecology, Behavior, and Physiology*. University of California Press.
- LeBoeuf, B. J., and C. L. Ortiz. 1977. Composition of elephant seal milk. *Journal of Mammalogy* **58**:683–685.
- Leclaire, S., T. Merklings, C. Raynaud, et al. 2012. Semiochemical compounds of preen secretion reflect genetic make-up in a seabird species. *Proceedings of the Royal Society B: Biological Sciences* **279**:1185–1193.
- Leffler, E. M., K. Bullaughey, D. R. Matute, et al. 2012. Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biology* **10**:e1001388.
- Lemaire, L., F. Jay, I.-H. Lee, K. Csilléry, and M. G. B. Blum. 2016. Goodness-of-fit statistics for approximate Bayesian computation .
- Lewis, S., S. Benvenuti, L. Dall-Antonia, et al. 2002. Sex-specific foraging behaviour in a monomorphic seabird. *Proceedings of the Royal Society B: Biological Sciences* **269**:1687–1693.
- Lewontin, R. C. 1974. *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

- Lewontin, R. C., and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**:595–609.
- Ley, R. E., M. Hamady, C. Lozupone, et al. 2008. Evolution of mammals and their gut microbes. *Science* **320**:1647–1651.
- Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**:2987–2993.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv .
- Li, H., and R. Durbin. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**:493–496.
- Li, H., J. Xiang-Yu, G. Dai, et al. 2016. Large numbers of vertebrates began rapid population decline in the late 19th century. *Proceedings of the National Academy of Sciences of the United States of America* **113**:14079–14084.
- Lindensfors, P., B. S. Tullberg, and M. Biuw. 2002. Phylogenetic analyses of sexual selection and sexual size dimorphism in pinnipeds. *Behavioral Ecology and Sociobiology* **52**:188–193.
- Ling, J. K., 1972. The integument of marine mammals. *in* *Functional Anatomy of Marine Mammals*. London.
- Linstrom, P., and W. Mallard, 2009. NIST Chemistry WebBook, NIST Standard Reference Database Number 69. National Institute of Standards and Technology, Gaithersburg MD. URL <https://webbook.nist.gov/>.
- Liu, J.-X., T.-X. Gao, Z.-M. Zhuang, et al. 2006. Late Pleistocene divergence and subsequent population expansion of two closely related fish species, Japanese anchovy (*Engraulis japonicus*) and Australian anchovy (*Engraulis australis*). *Molecular Phylogenetics and Evolution* **40**:712–723.
- Lohmueller, K. E., A. Albrechtsen, Y. Li, et al. 2011. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genetics* **7**.
- Loiselle, B. A., V. L. Sork, J. Nason, and C. Graham. 1995. Spatial genetic structure of a tropical understory shrub, *psychotria officinalis* (Rubiaceae). *American Journal of Botany* **82**:1420–1425.
- Lotze, H. K., and B. Worm. 2009. Historical baselines for large marine animals. *Trends in Ecology & Evolution* **24**:254–262.
- Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:550.
- Lowry, M. 2014. Abundance, distribution, and population growth of the northern elephant seal (*Mirounga angustirostris*) in the United States from 1991 to 2010. *Aquatic Mammals* **40**:20–31.
- Lukic, S., and J. Hey. 2012. Demographic inference using spectral methods on snp data, with an analysis of the human out-of-africa expansion. *Genetics* **192**:619–+.
- Lyman, R. L., 2011. A History of Paleoecological Research on Sea Otters and Pinnipeds of the Eastern Pacific Rim. Pages 18–40 *in* *Human Impacts on Seals, Sea Lions, and Sea Otters Integrating Archaeology and Ecology in the Northeast Pacific*. University of California Press.
- MacLeod, I. M., D. M. Larkin, H. A. Lewin, B. J. Hayes, and M. E. Goddard. 2013. Inferring demography from runs of homozygosity in whole-genome sequence, with correction for sequence errors. *Molecular Biology and Evolution* **30**:2209–2223.
- Macpherson, A. J., and N. L. Harris. 2004. Interactions between commensal intestinal bacteria and the immune system. *Nature Reviews Immunology* **4**:478–485.

References

- Malaspinas, A.-S., M. C. Westaway, C. Muller, et al. 2016. A genomic history of Aboriginal Australia. *Nature* **538**:207–214.
- Marcobal, A., M. Barboza, E. D. Sonnenburg, et al. 2011. Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host & Microbe* **10**:507–514.
- Marcobal, A., and J. L. Sonnenburg. 2012. Human milk oligosaccharide consumption by intestinal microbiota. *Clinical Microbiology and Infection* **18**:12–15.
- Markle, J. G. M., D. N. Frank, S. Mortin-Toth, et al. 2013. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. *Science* **339**:1084–1088.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**:10–12.
- Maurice, C. F., S. CL Knowles, J. Ladau, et al. 2015. Marked seasonal variation in the wild mouse gut microbiota. *The ISME Journal* **9**:2423–2434.
- McFall-Ngai, M., M. G. Hadfield, T. C. G. Bosch, et al. 2013. Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences of the United States of America* **110**:3229–3236.
- McFerrin, L., 2015. HDMD: Statistical analysis tools for high dimension molecular data (HDMD). R package version 1.2. URL <http://cran.r-project.org/web/packages/HDMD/index.html>.
- McGraw, K. O., and S. P. Wong. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* **1**:30–46.
- McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **8**:e61217.
- McMurdie, P. J., and S. Holmes. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology* **10**:e1003531.
- McNair, H. M., and J. M. Miller. 2011. *Basic gas chromatography*. 2nd edition. John Wiley & Sons, Hoboken, New Jersey.
- McShane, B. B., D. Gal, A. Gelman, C. Robert, and J. L. Tackett. 2017. Abandon Statistical Significance. *arXiv* **170907588**.
- Mendel, G. 1886. Versuche über Pflanzen-Hybriden. *Journal of Heredity* **42**:3–47.
- Miescher, F. 1871. Ueber die chemische Zusammensetzung der Eiterzellen. *Medizinisch-chemische Untersuchungen*, **4**.
- Mills, L. S., and P. E. Smouse. 1994. Demographic consequences of inbreeding in remnant populations. *The American Naturalist* **144**:412–431.
- Mitchell-Olds, T., J. H. Willis, and D. B. Goldstein. 2007. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* **8**:845–856.
- Moeller, A. H., Y. Li, E. M. Ngole, et al. 2014. Rapid changes in the gut microbiome during human evolution. *Proceedings of the National Academy of Sciences of the United States of America* **111**:16431–16435.
- Muegge, B. D., J. Kuczynski, D. Knights, et al. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**:970–974.
- Myers, S., C. Fefferman, and N. Patterson. 2008. Can one learn history from the allelic spectrum? *Theoretical Population Biology* **73**:342–348.
- Nakagawa, S., and I. C. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews* **82**:591–605.

- Nakagawa, S., and H. Schielzeth. 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. *Biological Reviews* 136:no–no.
- Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4:133–142.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America* 70:3321–3323.
- Nei, M., T. Maruyama, and R. Chakraborty. 1975. The bottleneck effect and genetic variability in populations. *Evolution* 29:1–10.
- Nelson, T. M., T. L. Rogers, and M. V. Brown. 2013. The gut bacterial community of mammals from marine and terrestrial habitats. *PLoS ONE* 8:e83655.
- Nielsen, R., I. Hellmann, M. Hubisz, C. Bustamante, and A. G. Clark. 2007. Recent and ongoing selection in the human genome. *Nature Reviews Genetics* 8:857–868.
- Nielsen, R., M. J. Hubisz, I. Hellmann, et al. 2009. Darwinian and demographic forces affecting human protein coding genes. *Genome Research* 19:838–849.
- Nosek, B. A., G. Alter, G. C. Banks, et al. 2015. Promoting an open research culture: Author guidelines for journals could help to promote transparency, openness, and reproducibility. *Science* 348:1422–1425.
- Numberger, D., D. P. R. Herlemann, K. Jürgens, G. Dehnhardt, and H. Schulz-Vogt. 2016. Comparative analysis of the fecal bacterial community of five harbor seals (*Phoca vitulina*). *MicrobiologyOpen* 5:782–792.
- Nunney, L. 1993. The influence of mating system and overlapping generations on effective population-size. *Evolution* 47:1329–1341.
- Nunziata, S. O., and D. W. Weisrock. 2018. Estimation of contemporary effective population size and population declines using RAD sequence data. *Heredity* 120:196–207.
- O'Brien, S. J. 1994. A role for molecular genetics in biological conservation. *Proceedings of the National Academy of Sciences of the United States of America* 91:5748–5755.
- Ohta, T. 1973. Slightly deleterious mutant substitutions in evolution. *Nature* 246:96–98.
- Oksanen, J., F. G. Blanchet, M. Friendly, et al., 2017. vegan: Community Ecology R package. R package version 2.4-5. URL <https://cran.r-project.org/web/packages/vegan/index.html>.
- Olsson, M., T. Madsen, J. Nordby, et al. 2003. Major histocompatibility complex and mate choice in sand lizards. *Proceedings of the Royal Society B: Biological Sciences* 270:S254–S256.
- Osborne, A. J., S. S. Negro, B. L. Chilvers, et al. 2016. Genetic evidence of a population bottleneck and inbreeding in the endangered New Zealand sea lion, *Phocartos hookeri*. *Journal of Heredity* 107:392–402.
- O'Toole, P. W., and I. B. Jeffery. 2015. Gut microbiota and aging. *Science* 350:1214–1215.
- Pantoja-Feliciano, I. G., J. C. Clemente, E. K. Costello, et al. 2013. Biphasic assembly of the murine intestinal microbiota during early development. *The ISME Journal* 7:1112–1115.
- Paradis, E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26:419–420.
- Pedersen, C.-E. T., A. Albrechtsen, P. D. Etter, et al. 2018. A southern African origin and cryptic structure in the highly mobile plains zebra. *Nature Ecology & Evolution* 2:491–498.
- Pedersen, H. K., V. Gudmundsdottir, H. B. Nielsen, et al. 2016. Human gut microbes impact host serum metabolome and insulin sensitivity. *Nature* 535:376–+.

References

- Peery, M. Z., R. Kirby, B. N. Reid, et al. 2012. Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology* **21**:3403–3418.
- Pemberton, J. 2004. Measuring inbreeding depression in the wild: the old ways are the best. *Trends in Ecology & Evolution* **19**:613–615.
- Pemberton, J. M. 2008. Wild pedigrees: the way forward. *Proceedings of the Royal Society B: Biological Sciences* **275**:613–621.
- Peng, R. 2015. The reproducibility crisis in science: A statistical counterattack. *Significance* **12**:30–32.
- Peng, R. D. 2011. Reproducible research in computational science. *Science* **334**:1226–1227.
- Perez-Muñoz, M. E., M.-C. Arrieta, A. E. Ramer-Tait, and J. Walter. 2017. A critical assessment of the "sterile womb" and "in utero colonization" hypotheses: implications for research on the pioneer infant microbiome. *Microbiome* **5**:243.
- Perofsky, A. C., R. J. Lewis, L. A. Abondano, A. Di Fiore, and L. A. Meyers. 2017. Hierarchical social networks shape gut microbial composition in wild Verreaux's sifaka. *Proceedings of the Royal Society B: Biological Sciences* **284**:20172274.
- Phillips, A. V. 2003. Behavioral cues used in reunions between mother and pup South American fur seals (*Arctocephalus australis*). *Journal of Mammalogy* **84**:524–535.
- Pierce, K. M., J. L. Hope, K. J. Johnson, B. W. Wright, and R. E. Synovec. 2005. Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis. *Journal of Chromatography. A* **1096**:101–110.
- Pinsky, M. L., and S. R. Palumbi. 2014. Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology* pages n/a–n/a.
- Piry, S., G. Luikart, and J. M. Cornuet. 1999. Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* **90**:502–503.
- Pitcher, B. J., H. Ahonen, I. Charrier, and R. G. Harcourt. 2011a. Allosuckling behavior in the Australian sea lion (*Neophoca cinerea*): An updated understanding. *Marine Mammal Science* **27**:881–888.
- Pitcher, B. J., R. G. Harcourt, B. Schaal, and I. Charrier. 2011b. Social olfaction in marine mammals: wild female Australian sea lions can identify their pup's scent. *Biology Letters* **7**:60–62.
- Preacher, K. J., G. Zhang, C. Kim, and G. Mels. 2013. Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research* **48**:28–56.
- Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* **16**:1791–1798.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.
- Pryce, J. E., M. Haile-Mariam, M. E. Goddard, and B. J. Hayes. 2014. Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution* **46**:71.
- Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**:559–575.
- Quast, C., E. Pruesse, P. Yilmaz, et al. 2012. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41**:D590–D596.
- Queller, D. C., and K. F. Goodnight. 1989. Estimating relatedness using genetic markers. *Evolution* **43**:258.

- R Core Team. 2015. R: A language and environment for statistical computing .
- Radwan, J., A. Tkacz, and A. Kloch. 2008. MHC and preferences for male odour in the bank vole. *Ethology* **114**:827–833.
- Ray-Mukherjee, J., K. Nimon, S. Mukherjee, et al. 2014. Using commonality analysis in multiple regressions: a tool to decompose regression effects in the face of multicollinearity. *Methods in Ecology and Evolution* **5**:320–328.
- Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**:248–249.
- Reichle, C., I. Aguilar, M. Ayasse, et al. 2013. Learnt information in species-specific ‘trail pheromone’ communication in stingless bees. *Animal Behaviour* **85**:225–232.
- Reiter, J., N. L. Stinson, and B. J. Le Boeuf. 1978. Northern elephant seal development: The transition from weaning to nutritional independence. *Behavioral Ecology and Sociobiology* **3**:337–367.
- Ren, T., S. Boutin, M. M. Humphries, et al. 2017. Seasonal, spatial, and maternal effects on gut microbiome in wild red squirrels. *Microbiome* **5**:55.
- Revelle, W. 2016. psych: Procedures for psychological, psychometric, and personality research. R package version 1.84. URL <http://cran.r-project.org/package=psych>.
- Rick, T. C., R. L. DeLong, J. M. Erlandson, et al. 2011. Where were the northern elephant seals? Holocene archaeology and biogeography of *Mirounga angustirostris*. *The Holocene* **21**:1159–1166.
- Robertson, B. C., and B. L. Chilvers. 2011. The population decline of the New Zealand sea lion *Phocarctos hookeri*: a review of possible causes. *Mammal Review* **41**:253–275.
- Roman, J., and S. R. Palumbi. 2003. Whales before whaling in the North Atlantic. *Science* **301**:508–510.
- Romiguier, J., P. Gayral, M. Ballenghien, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**:261–263.
- Round, J. L., and S. K. Mazmanian. 2009. The gut microbiota shapes intestinal immune responses during health and disease. *Nature Reviews Immunology* **9**:313–323.
- Russell, S. L., M. J. Gold, M. Hartmann, et al. 2012. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO reports* **13**:440–447.
- Saccheri, I., M. Kuussaari, M. Kankare, et al. 1998. Inbreeding and extinction in a butterfly metapopulation. *Nature* **392**:491–494.
- Salmona, J., R. Heller, M. Lascoux, and A. Shafer, 2017. Inferring demographic history using genomic data. Pages 1699–27 . Springer International Publishing, Cham.
- Sanvito, S., A. Dueñas Meza, Y. Schramm, et al. 2013. Isolation and cross-species amplification of novel microsatellite loci in a charismatic marine mammal species, the Northern elephant seal (*Mirounga angustirostris*). *Conservation Genetics Resources* **5**:93–96.
- Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* **1**:103–113.
- Scott, R. P. 2003. Principles and Practice of Chromatography. Chrom-Ed Book Series.
- Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**:615–629.
- Setchell, J. M., S. Vaglio, K. M. Abbott, et al. 2011. Odour signals major histocompatibility complex genotype in an Old World monkey. *Proceedings of the Royal Society B: Biological Sciences* **278**:274–280.

References

- Shafer, A. B. A., L. M. Gattepaille, R. E. A. Stewart, and J. B. W. Wolf. 2015a. 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. *Molecular Ecology* **24**:328–345.
- Shafer, A. B. A., J. B. W. Wolf, P. C. Alves, et al. 2015b. Genomics and the challenging translation into conservation practice. *Trends in Ecology & Evolution* **30**:78–87.
- Shrout, P. E., and J. L. Fleiss. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* **86**:420–428.
- Singer, A. G., G. K. Beauchamp, and K. Yamazaki. 1997. Volatile signals of the major histocompatibility complex in male mouse urine. *Proceedings of the National Academy of Sciences of the United States of America* **94**:2210–2214.
- Sipilä, T., and H. Hyvärinen. 2014. Status and biology of Saimaa (*Phoca hispida saimensis*) and Ladoga (*Phoca hispida ladogensis*) ringed seals. *NAMMCO Scientific Publications* **1**:83–99.
- Slate, J., P. David, K. G. Dodds, et al. 2004. Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **93**:255–265.
- Smith, P., D. Willemsen, M. Popkes, et al. 2017. Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. *eLife* **6**:e1002352.
- Smith, S., L. Bernatchez, and L. B. Beheregaray. 2013. RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics* **14**:375.
- Spielman, D., B. W. Brook, and R. Frankham. 2004. Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Sciences of the United States of America* **101**:15261–15264.
- Stephens, W. Z., A. R. Burns, K. Stagaman, et al. 2016. The composition of the zebrafish intestinal microbial community varies across development. *The ISME Journal* **10**:644–654.
- Stoffel, M. A., B. A. Caspers, J. Forcada, et al. 2015. Chemical fingerprints encode mother-offspring similarity, colony membership, relatedness, and genetic quality in fur seals. *Proceedings of the National Academy of Sciences of the United States of America* **112**:E5005–12.
- Stoffel, M. A., E. Humble, K. Acevedo-Whitehouse, et al. 2018. Recent demographic histories and genetic diversity across pinnipeds are shaped by anthropogenic interactions and mediated by ecology and life-history. *bioRxiv* **293894**.
- Stoffel, M. A., S. Nakagawa, and H. Schielzeth. 2017. rptR: repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution* **8**:1639–1644.
- Storey, J. D. 2002. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**:479–498.
- Sun, L., and D. Müller-Schwarze. 1998. Anal gland secretion codes for family membership in the beaver. *Behavioral Ecology and Sociobiology* **44**:199–208.
- Sunyaev, S., V. Ramensky, I. Koch, et al. 2001. Prediction of deleterious human alleles. *Human Molecular Genetics* **10**:591–597.
- Sved, J. A. 1968. The stability of linked systems of loci with a small population size. *Genetics* **59**:543–563.
- Zsulkin, M., N. Bierne, and P. David. 2010. Heterozygosity-fitness correlations: a time for reappraisal. *Evolution* **64**:1202–1217.
- Tajima, F. 1989. The effect of change in population size on DNA polymorphism. *Genetics* **123**:597–601.
- Tarka, M., A. Guenther, P. T. Niemelä, S. Nakagawa, and D. W. A. Noble. 2018. Sex differences in life history, behavior, and physiology along a slow-fast continuum: a meta-analysis. *Behavioral Ecology and Sociobiology* **72**:a017566.

- Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. Inferring coalescence times from DNA sequence data. *Genetics* **145**:505–518.
- Taylor, M., 2015. *sinkr*: A collection of functions featured on the blog 'Me nugget'. R package version 1.0. URL <https://github.com/marchtaylor/sinkr>.
- Theis, K. R., A. Venkataraman, J. A. Dycus, et al. 2013. Symbiotic bacteria appear to mediate hyena social odors. *Proceedings of the National Academy of Sciences of the United States of America* **110**:19832–19837.
- Trucchi, E., P. Gratton, J. D. Whittington, et al. 2014. King penguin demography since the last glaciation inferred from genome-wide data. *Proceedings. Biological sciences / The Royal Society* **281**:20140528–20140528.
- Tsitronis, A., F. Rousset, and P. David. 2001. Heterosis, Marker Mutational Processes and Population Inbreeding History. *Genetics* **159**:1845–1859.
- Tung, J., L. B. Barreiro, M. B. Burns, et al. 2015. Social networks predict gut microbiome composition in wild baboons. *eLife* **4**.
- Turnbaugh, P. J., F. Bäckhed, L. Fulton, and J. I. Gordon. 2008. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host & Microbe* **3**:213–223.
- Vaglio, S. 2009. Chemical communication and mother-infant recognition. *Communicative & integrative biology* **2**:279–281.
- van Wilgenburg, E., and M. A. Elgar. 2013. Confirmation bias in studies of nestmate recognition: a cautionary note for research into the behaviour of animals. *PLoS ONE* **8**:e53548.
- Venter, J. C., M. D. Adams, E. W. Myers, et al. 2001. The sequence of the human genome. *Science* **291**:1304–+.
- Videvall, E., S. J. Song, H. M. Bensch, et al. 2018. The development of gut microbiota in ostriches and its association with juvenile growth. *bioRxiv* page 270017.
- Wang, J. 2011. A new likelihood estimator and its comparison with moment estimators of individual genome-wide diversity. *Heredity* **107**:433–443.
- Watson, J. D., and F. H. Crick. 1953. Molecular structure of nucleic acids - A structure for deoxyribose nucleic acid. *Nature* **171**:737–738.
- Weber, D. S., B. S. Stewart, J. C. Garza, and N. Lehman. 2000. An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Current Biology* **10**:1287–1290.
- Weber, D. S., B. S. Stewart, J. Schienman, and N. Lehman. 2004. Major histocompatibility complex variation at three class II loci in the northern elephant seal. *Molecular Ecology* **13**:711–718.
- Wedekind, C., and S. Fürti. 1997. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proceedings of the Royal Society B: Biological Sciences* **264**:1471–1479.
- Wedekind, C., T. Seebeck, F. Bettens, and A. J. Paepke. 1995. MHC-dependent mate preferences in humans. *Proceedings of the Royal Society B: Biological Sciences* **260**:245–249.
- Wegmann, D., and L. Excoffier. 2010. Bayesian inference of the demographic history of chimpanzees. *Molecular Biology and Evolution* **27**:1425–1435.
- Wehrens, R., T. G. Bloemberg, and P. H. C. Eilers. 2015. Fast parametric time warping of peak lists. *Bioinformatics* **31**:3063–3065.
- Weir, B. S., and C. C. Cockerham. 1973. Mixed self and random mating at two loci. *Genetics Research* **21**:247–262.
- Weiss, B. M., A. Marcillo, M. Manser, et al. 2018. A non-invasive method for sampling the body odour of mammals. *Methods in Ecology and Evolution* **9**:420–429.

References

- Wicherts, J. M., C. L. S. Veldkamp, H. E. M. Augusteijn, et al. 2016. Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology* 7:108.
- Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, USA.
- Wickham, H., R. François, L. Henry, K. Müller, and Rstudio, 2014. *dplyr: A grammar of data manipulation*.
- Wickham, H., J. Hester, W. Chang, Rstudio, and R Core Team, 2013. *devtools: Tools to make developing R packages easier: R package version 1.13.6*.
- Wilson, D. E., and R. A. Mittermeier, editors. 2014. *Handbook of the Mammals of the World*. Lynx Edicions, Barcelona.
- Wolak, M. E., D. J. Fairbairn, and Y. R. Paulsen. 2012. Guidelines for estimating repeatability. *Methods in Ecology and Evolution* 3:129–137.
- Wong, J. W. Y., J. Meunier, C. Lucas, and M. Kölliker. 2014. Paternal signature in kin recognition cues of a social insect: concealed in juveniles, revealed in adults. *Proceedings of the Royal Society B: Biological Sciences* 281:20141236–20141236.
- Wright, S. 1922. Coefficients of inbreeding and relationship. *The American Naturalist* 56:330–338.
- Wu, G. D., J. Chen, C. Hoffmann, et al. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334:105–108.
- Wyatt, T. D. 2014a. *Pheromones and Animal Behaviour*. 2nd edition. Cambridge University Press, Cambridge, UK.
- Wyatt, T. D. 2014b. Proteins and peptides as pheromone signals and chemical signatures. *Animal Behaviour* 97:273–280.
- Wyatt, T. D. 2015. The search for human pheromones: the lost decades and the necessity of returning to first principles. *Proceedings of the Royal Society B: Biological Sciences* 282:20142994–20142994.
- Xie, Y., 2014. *knitr: A Comprehensive Tool for Reproducible Research in R*. Chapman and Hall/CRC.
- Xue, A. T., and M. J. Hickerson. 2015. The aggregate site frequency spectrum for comparative population genomic inference. *Molecular Ecology* 24:6223–6240.
- Yamazaki, K., G. K. Beauchamp, A. Singer, J. Bard, and E. A. Boyse. 1999. Odortypes: Their origin and composition. *Proceedings of the National Academy of Sciences of the United States of America* 96:1522–1525.
- Yamazaki, K., M. Yamaguchi, L. Baranoski, et al. 1979. Recognition among mice. Evidence from the use of a Y-maze differentially scented by congenic mice of different major histocompatibility types. *Journal of Experimental Medicine* 150:755–760.
- Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T.-Y. Lam. 2017. *ggtree: an rpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data*. *Methods in Ecology and Evolution* 8:28–36.
- Zhu, L., Q. Wu, J. Dai, S. Zhang, and F. Wei. 2011. Evidence of cellulose metabolism by the giant panda gut microbiome. *Proceedings of the National Academy of Sciences of the United States of America* 108:17714–17719.
- Zilber-Rosenberg, I., and E. Rosenberg. 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews* 32:723–735.

Acknowledgements

First and foremost, I would like to thank my supervisor Joe Hoffman. He gave me the chance to work on pinniped genetics and a lot of other things (nearly all of which I had no idea of) and supported me with his endless enthusiasm and ideas and always showed a great deal of trust when I pursued new scientific paths. I think this scientific freedom and Joe's openness for new fields, ideas and methods gave me the chance to become the scientist that I am now, and I feel very fortunate to have started my scientific career in his lab. I would also like to thank my second supervisor, Hazel Nichols. Although, after all, the distance between Liverpool and Bielefeld was a bit bigger than we thought, Hazel was always incredibly dedicated and supportive, both scientifically and with the countless administrative issues.

Working at the VHF was (and is) great, and this is mainly due to the lovely people working there and the relaxed atmosphere. I thank Elke Hippauf, Katrin Lehmann and Ann-Christin Polikeit for the support when I had to get my hands wet in the lab (which was not so often) or when contracts had to be dealt with and obscure things had to be ordered for the field. I am grateful to Anja Wildberg, Susanne Teipel and Claudia Petersen for always providing knowledge and skills on issues where I did not have the slightest ideas how to solve them. Especially, Claudia, if scientists everywhere would be backed up by such hands-on and good-humored people like you, academia would be a better place. Ich bin außerdem froh, die Menschen hinter der Bühne kennengelernt zu haben. Vielen Dank an Ulla, für die vielen netten Gespräche im Kaffeeraum. Monika, danke, dass du es irgendwie geschafft hast, unser Büro sauber zu halten, und dafür, dass du mich jeden Donnerstag Morgen ein wenig aus der Versenkung geholt hast und wir über Gott und die Welt quatschen konnten. Viktor, danke auch dir für deine gute Laune, die ich immer sehr inspirierend fand. Natürlich auch ein herzliches Dankeschön an Hassan und Sinan, das mit Abstand beste Kaffee-Team, das ich kenne. Thanks to Oliver Krüger, Fritz Trillmich, Klaus Reinhold and all the other senior scientists for keeping the hierarchies low and creating an atmosphere in the VHF and during the seminars where students and Professors can talk and discuss eye-to-eye, which is surely not

the standard among German universities. A young researchers' views on science and academia are strongly shaped by their colleagues. I want to say a huge thank you to Peter Korsten, Steve Ramm, Klaus Reinhold and Holger Schielzeth for many discussions and feedback around science and statistics, and the challenges of academic life in general, happening between the Bielefeld Mensa, Jena and the forests of central Greece. I also want to thank Tim Schmoll and Steve for organizing the legendary Stats club.

Next, I would like to thank all my friends and all the inspiring people from the VHF and beyond, who joined me along the way. From the sandy grounds of the Beachvolleyball court, the dusty football fields close to the Uni, the muggy atmosphere in the Irish pub, to the hiking paths of the Teutoburger Wald and the Bavarian Alps all the way to the confusing cycling roads on the way to Dijon and the icy-winds of Iceland, the adventures and activities we had were widespread and numerous. In particular, I would like to thank: Abhijeet Shah, Alfredo Sanchez-Tojar, Astrid Potiek, Anna Müller, Anneke Pajmans and Sela Danisman, Athina Giannakara and Stefan Becker, Bahar Patlar, David Vendrami and Elena Fissenewert, David Wells, Geno DeRango, Jan Sauer, Jonas Schwartz, Kati Michalek, Lucienne Eweleit, Lucy Peters, Luke L-dawg Eberhart-Phillips, Matthias Galipaud, Matthias Spangenberg, Meinolf Ottensmann, Michele DeNoia and Miquel Joan Servera Lauder, Michael Weber, Nayden Chakarov, Rienk Fokkema, Ralf Jochmann, Rebecca Nagel, Steffie Bölting, Steffie Großer, Stephe Kalberer and Stephen Salazar.

Wie jeder mittlerweile weiß, weint man immer zweimal in Bielefeld, einmal wenn man ankommt, und einmal, wenn man wieder gehen muss. In der ersten Zeit wurde ich unglaublich warmherzig von Christof Taube und den Wasserfreunden Bielefeld aufgenommen, wodurch ich nicht nur wieder Spaß am Schwimmsport hatte, sondern was mich auch durch eine schwierige Anfangsphase in Bielefeld gebracht hat. Vielen Dank dafür! Außerdem haben sich aus dieser Zeit Freundschaften entwickelt, die immer noch bestehen und wohl noch lange bestehen werden, insbesondere mit Tobias Lange, Matthias Zinram und Mareike Esser, Martin Wolf, und über Umwege dann auch mit Immanuel 'Emma' Block und Marvin 'Marvatron' Kather. An dieser Stelle nochmal vielen Dank für spannende Schwimmwettkämpfe, Festivals, epische Laufevents und Billiardabende.

I am really grateful for the fantastic people who worked with me in the field. First of all, I want to thank Karina Acevedo-Whitehouse for making my fieldtrip to the San Benitos Archipelago possible and for teaching me how to work with elephant seals with an unparalleled enthusiasm and patience. My other colleagues in the field, namely Nami Morales Duran, Fernando Elorriaga, Luis Soto and Cecilia Barragan showed me that Mexico does not just host some great scientists but also some of the most warm-hearted people I know. The Pescadores,

especially Jopo, Grande, Nery and Chito welcomed me like one of their own and introduced me to the art of *hacer un nudo* (knot-making), how to eat an extremely spicy chilly as a side dish for basically every meal of the day, and how to dive for Abalones.

During my PhD, I was also lucky enough to participate in the first part of a great adventure, the ACE expedition, sailing all the way from Bremerhaven to Cape Town on the rustic Russian Icebreaker Akademik Tryoshnikov. I met a lot of young scientists from disciplines ranging from oceanography and climate modelling to atmospheric sciences and biology on board, which was, in its diversity, probably one of the most inspiring and horizon-broadening experiences possible. For ensuring me, that wherever in the world you go, be it France, Russia, the UK, Africa or the USA, you will find young people that truly care about the world and about science, I want to thank: (can't remember all the last names, but its not that important anyway) Kyle, Pike, Jordan, Jen, Manon, Ivo, Pari, Daniel, Marcel, Marie, Thor, Mark, Ismael, Eduard, James and all the others!

Most importantly, I want to thank Emily for being an incredible partner and friend, throughout great times such as our amazing adventures in Canada and California but also during the tougher times of the PhD. I am very grateful to have you, and for your constant support when it was really necessary, like during the last few weeks.

Schlussendlich wäre diese Arbeit nicht möglich gewesen ohne meinen Bruder Roman, der mir bei den großen Lebensentscheidungen immer mit Rat und Tat zur Seite stand, und vor allem nicht ohne meine Eltern, die mich und meinen Lebensweg immer mit großem Vertrauen und Verständnis unterstützt haben. Dankeschön!