Sea Parrot Genomics:

Linking past and present population structure and demography

of the Atlantic puffin (*Fratercula arctica*)

Oliver Kersten

Dissertation presented for the degree of

Philosophiae Doctor (PhD)

2023

Centre for Ecological and Evolutionary Synthesis

Department of Biosciences

Faculty of Mathematics and Natural Sciences

University of Oslo

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 2614

ISSN 1501-7710

Cover: UiO.
Print production: Graphic Center, University of Oslo.

CONTENTS

ACKNOWLEDGEMENTS

The last 4.5 years have been an unforgettable adventure and an incredible journey. Being a Ph.D. student, university-level teaching assistant, and on-site ancient DNA (aDNA) lab manager, while “properly” settling in Norway (moving from Oslo to Kongsberg), navigating work and life during a global pandemic, and becoming/being a dad, was a multi-faceted and challenging, but most-of-the-time extremely pleasant, experience. Yet, none of it would have been possible without the people and groups mentioned below.

THANK YOU !

I am grateful for the support and guidance of my supervisors during my Ph.D. I would like to especially thank **Sanne** for being a tremendous main supervisor and a prime example of what a supervisor should be. Your structured approach and open ear were instrumental in guiding me through my Ph.D. with a clear vision. I appreciated our colleague-level relationship, and I am thankful that you stayed away from micro-management, which created an environment in which I thrived. **Kjetill**, as my co-supervisor, I thank you for always finding time to chat and to provide valuable and constructive comments on all of my manuscripts despite being the head of the Centre for Ecological and Evolutionary Synthesis (CEES). Your support was truly invaluable, especially for generating the Atlantic puffin reference genome(s). **Bastiaan**, you could have easily had the role as a third co-supervisor. I am grateful for your close, pleasant, and efficient collaboration and interaction throughout my Ph.D. Conversations about methods, analyses, and life in general, especially our shared love of beer, made this time all the more enjoyable. I look forward to working even more closely with you in the near future.

My Ph.D. experience would definitely not have been the same without the Archaeogenomics research group. **Anneke**, you were there from day one and were the perfect office mate. Our time together was truly enjoyable, and I haven’t met many people with your artistic talent. **Lane**, thank you for all the gossip when I needed it and

our discussions on bioinformatics, as well as the one running tour we managed to go on together :). Also, it was inspiring how you always remained humble despite your well-traveled background. **Lulú**, you are one of the nicest, kindest and most positive people I have ever met, and I will always remember the Mexican food you let us taste. Let's be real: You are the reason my office plants have survived the last 4.5 years. **Emma**, thank you for making the course in Bodø and, essentially, the start of the Covid-19 pandemic memorable, as well as for the delightful conversations over pizza. I am grateful for your help with social media outreach and for your Norwegian translations. Finally, I want to thank **Giada** for being part of this journey as the only German-speaking (fluent) person in the group. We despised and refused to speak German to each other, and I will always remember your humor and our shared interest in whiskey. Also, thank you to the rest of the wider Archaeogenomics circle – **Adriana, Albína, Angélica, Aurélie, Guðný, Håvard, Jason, Lauren, Lydia, Marianne** and **Sam** – you made this research group an extremely pleasant environment to work in!

I would like to extend my gratitude to several other groups and individuals who joined me on this journey and made this thesis possible. The **Norwegian Sequencing Centre** (especially Ave, Spyros and Morten) were instrumental in providing the necessary resources and support to complete this work. I also appreciate the nice environment created by **all the people at CEES**, especially during coffee breaks and events, such as the bowling night and several "Holmenkollenstaffeten". **Nanna** and the **Innkjøpskontoret-Team**, thank you for your help and guidance, which made life as the on-site aDNA lab manager relatively easy. The **SEAPOP/SEATRACK** network, especially **Tycho**, played a huge role in placing the genomic results into an ecological background by providing samples and ecological knowledge. I would also like to thank **all my co-authors** for their hard work and valuable contributions, as well as their timely responses to my emails. Finally, I want to thank **Annemarie Loof, Kevin Morgans,** and **Anneke (again)** for providing all the lovely images, pictures, and drawings of Atlantic puffins that I have used in my presentations, posters, and this thesis.

Last but not least, I am grateful for family and friends who were part of my life during the Ph.D. **Emil** entered this journey after the first year and has provided a delightful

change of scenery outside of work. He has reminded me every day of what is actually important in life. No offense to all the other people mentioned so far … **Inger and Håvard** have been crucial supporting pillars who have always helped when needed, which certainly allowed me to finish the Ph.D. on time. Emil is lucky to have such wonderful grandparents. The **Kongsberg crew**, including "**Awesomestreet**" & Co., provided numerous occasions to celebrate life with a more than sufficient number of beverages and, on the other side of the spectrum, forced me to not completely neglect my health/shape. **My parents, Kerstin** and **Sven**, have been ever-supporting and encouraging, and their numerous visits have made them the best long-distance Omi and Opi, Emil could have hoped for. Happy child, happy dad!

To cap it all off and most important of all! I would like to thank **Guri** for her ever-present support, for joining me on this journey, and for filling the days with happiness and joy. Despite being hesitant, you even *tried* to show interest in what I do at work :D As my soulmate, you have been the perfect partner-in-crime during this journey and the last 10 years!
Elsker deg, liebe dich, love you <3 <3 <3

SUMMARY

Seabird populations worldwide have been declining dramatically over the last decades as a result of a range of environmental and anthropogenic stressors. Nevertheless, management of threatened seabirds is arguably hampered by the severe underutilization of whole genome sequencing (WGS) combined with a limited understanding of the interplay of complex ecological factors affecting population connectivity and contributing to the genetic population structure. By providing detailed genomic data, WGS allows to assess levels of connectivity and gene flow between distinct breeding populations and, thus, helps to identify relevant conservation units for seabirds.

Atlantic puffins (*Fratercula arctica*) have been designated as *vulnerable to extinction* globally and listed as *endangered* in Europe. A lack of genetic data for puffins at all spatial scales obstructs efforts towards an assessment of dispersal barriers, limits our understanding of cause-and-effect dynamics between population trends, ecology and the marine ecosystem, and hinders the development of adapted large-scale conservation actions.

Here, I present the first whole genome analysis of population structure, gene flow, demographic history and structural DNA variation of a pelagic, North Atlantic seabird. The analysis of 13 Atlantic puffin colonies throughout the majority of the species' breeding range revealed four large, genetically distinct clusters, which broadly overlap with the currently recognized taxonomy that includes three subspecies (*F. a. naumanni, F. a. arctica* and *F. a. grabae)* (**Paper I**). Additionally, I found a hybrid population in the High Arctic resulting from interbreeding between the High Arctic, large-bodied subspecies *F. a. naumanni* and the temperate and smaller subspecies *F. a. arctica* (**Paper I & Paper III**). Using whole genome data from contemporary and museum specimens, I provide evidence that this hybridization started as recent as six to seven generations ago resulting from a southward range shift of *F. a. naumanni* and coinciding with a period of rapid ecological change in the Arctic (**Paper III**). The presence of a hybrid population may also be a forecast of future scenarios throughout other parts of the Arctic illustrated by the sympatry of genetically distinct, but non-admixing, puffin subspecies within a single High Arctic colony on the west coast of

Greenland (**Paper II**). While genomic-based demographic reconstructions suggest that *F. a. naumanni* and *F. a. arctica* diverged due to climatic oscillations in the Pleistocene (**Paper III**), our understanding of the genomic basis of puffin subspecies differentiation and potential adaptive divergence is limited. Hence, I used single nucleotide polymorphisms, structural variants and short tandem repeats to identify genomic outlier loci that potentially contribute to intraspecific gene flow barriers and phenotypic differences between the subspecies (**Paper IV**). The results of this thesis highlight the importance of historical and modern whole genome data in understanding population structure and gene flow in seabirds, as well as the genomic basis of intraspecific, phenotypic differences and local adaptation. In light of a global biodiversity loss occurring at unprecedented rates, these findings should have implications for future seabird research and conservation management.

ZUSAMMENFASSUNG

Die weltweiten Seevögelpopulationen sind in den letzten Jahrzehnten aufgrund von Umweltbelastungen und anthropogenen Stressfaktoren dramatisch zurückgegangen. Trotzdem wird der Schutz bedrohter Seevögel durch die unzureichende Nutzung der Gesamtgenomsequenzierung (WGS) in Verbindung mit einem begrenzten Verständnis des Zusammenspiels komplexer ökologischer Faktoren, die die Konnektivität von Populationen beeinflusst und zur genetischen Populationsstruktur beitragen, behindert. Durch das Erstellen von detaillierten genetischen Daten ermöglicht die WGS die Bewertung der Konnektivität und des Austauschs von Erbgut zwischen verschiedenen Brutpopulationen und hilft somit bei der Identifizierung von wichtigen Schutzgebieten für Seevögel.

Der Papageientaucher (*Fratercula arctica*) wurde weltweit als *vom Aussterben bedroht* und in Europa als *gefährdet* eingestuft. Ein Mangel an genetischen Daten für Papageientaucher auf allen räumlichen Skalen behindert das Finden von Barrieren für den Austauch von Erbgut, beschränkt unser Verständnis der Ursache-Wirkungs-Dynamik zwischen Populationstrends, Ökologie und dem Meeresökosystem und schränkt die Entwicklung eines angepassten großflächigen Schutzes ein.

Ich präsentiere hier die erste vollständige Genomanalyse von Populationsstruktur, Genfluss, Demografie und struktureller DNA-Variation eines pelagischen nordatlantischen Seevogels. Die Analyse von 13 Papageientaucher-Kolonien im Großteil des Brutgebiets der Art ergab vier große, genetisch unterschiedliche Gruppen, die sich weitgehend mit der derzeit anerkannten Taxonomie überschneiden, welche aus drei Unterarten (*F. a. naumanni, F. a. arctica* und *F. a. grabae*) besteht (**Paper I**). Außerdem fand ich in der Hocharktis eine Hybridpopulation, die aus der Kreuzung zwischen der hocharktischen, großwüchsigen Unterart *F. a. naumanni* und der borealen und kleineren Unterart *F. a. arctica* (**Paper I & III**) enstand. Mit der Verwendung von Gesamtgenomdaten von zeitgenössischen Individuen und Museumsexemplaren liefere ich Beweise dafür, dass diese Hybridisierung erst vor sechs bis sieben Generationen begann und durch eine südliche Verschiebung des Verbreitungsgebiets von *F. a. naumanni*, die mit einer Periode rascher ökologischer Veränderungen in der Arktis zusammenhängt, veursacht wurde

(**Paper III**). Die Präsenz einer Hybridpopulation kann auch eine Vorhersage zukünftiger Szenarien in anderen Teilen der Arktis sein, die durch die Sympatrie genetisch unterschiedlicher, aber sich noch nicht vermischender Papageientaucher-Unterarten innerhalb einer einzigen hocharktischen Kolonie an der Westküste Grönlands veranschaulicht wird (**Paper II**). Während genombasierte demografische Rekonstruktionen darauf hindeuten, dass *F. a. naumanni* und *F. a. arctica* aufgrund klimatischer Schwankungen im Pleistozän divergierten (**Paper III**), ist unser Verständnis der genomischen Grundlage der Differenzierung von Papageientaucher-Unterarten und der möglichen adaptiven Divergenz begrenzt. Daher habe ich Einzelnukleotid-Polymorphismen, strukturelle Varianten und Mikrosatelliten verwendet, um genomische Ausreißer-Loci (eng. «outlier loci») zu identifizieren, da diese möglicherweise zu intraspezifischen Genflussbarrieren und phänotypischen Unterschieden zwischen den Unterarten beitragen (**Paper IV**). Die Ergebnisse dieser Dissertation unterstreichen die Bedeutung historischer und moderner Gesamtgenomdaten für die Erfassung und Analyse der Populationsstruktur und des Genflusses bei Seevögeln sowie der genomischen Grundlage von intraspezifischen, phänotypischen Unterschieden und lokaler, evolutionärer Anpassung. Angesichts des globalen Verlusts der Biodiversität sollten diese Ergebnisse Auswirkungen auf die zukünftige Seevögelforschung und den Naturschutz haben.

SAMMENDRAG

Verdens sjøfuglbestander har opplevd en kraftig nedgang de siste tiårene som følge av mangfoldige miljø- og menneskeskapte stressfaktorer. Beskyttelse av truede sjøfugl avhenger av kunnskap om den genetiske populasjonsstrukturen, men forvaltningen hemmes av dårlig utnyttelse av helgenomsekvensering (WGS) kombinert med begrenset kunnskap om samspillet mellom komplekse, økologiske faktorer. WGS gjør det mulig å vurdere tilknytning innad - og genflyt mellom - ulike populasjoner basert på detaljert genomisk data, og kan på denne måten bidra til å identifisere relevante bevaringsenheter for sjøfugl.

Lundefuglen (*Fratercula arctica*) anses som sårbar for utryddelse globalt og står oppført som truet i Europa. Manglende genetisk data på alle romlige skalaer hindrer innsats for å vurdere spredningsbarrierer, begrenser vår forståelse av årsak-virkning-dynamikk mellom populasjonsvekst, økologi og marine økosystemer, og hindrer utviklingen av tilpassede, storskala bevaringsplaner.

Her presenterer jeg den første helgenomanalysen av populasjonsstruktur, genflyt, demografisk historie og strukturell DNA-variasjon som har blitt gjort av en pelagisk, nordatlantisk sjøfugl. Analyser av 13 atlantiske lundefuglkolonier gjennom det meste av artens hekkeområde viste fire store, genetisk distinkte grupper som i stor grad overlapper den nåværende, anerkjente taksonomien som inkluderer tre underarter (*F. a. naumanni*, *F. a. arctica* og *F. a. grabae*) (**Paper I**). I tillegg fant jeg en hybridpopulasjon i øvre Arktis, som viste seg å være en krysning mellom den høyarktiske, storvokste underarten *F. a. naumanni* og den tempererte, mindre underarten *F. a. arctica* (**Paper I & Paper III**). Ved hjelp av helgenomdata fra både nåværende individer og museumseksemplarer, gir jeg bevis for at denne hybridiseringen startet så nylig som seks til syv generasjoner siden som resultat av en sørlig utbredelsesendring av *F. a. naumanni* og skjedde samtidig med en periode med rask økologisk endring i Arktis (**Paper III**). Tilstedeværelsen av en hybridpopulasjon kan også være en prognose for fremtidige scenarier i andre deler av Arktis. For eksempel identifiserte jeg genetisk distinkte underarter av lundefugl innenfor en enkelt høyarktisk koloni på Vest-kysten av Grønland, som ikke hadde genutveksling enda til tross sympatrisk (overlappende) utbredelse (**Paper II**). Genom-baserte demografiske

rekonstruksjoner tyder på at *F. a. naumanni* og *F. a. arctica* divergerte på grunn av klimatiske svingninger i Pleistocen (**Paper III**). Det genetiske grunnlaget for differensiering av lundefugl-underarter og potensiell tilstedeværelse av adaptiv divergens er imidlertid lite undersøkt. Derfor brukte jeg enkeltbasepolymorfier, strukturelle varianter og korte, tandemrepeterte sekvenser for å identifisere avvikende genetiske loci som potensielt bidrar til intraspesifikke barrierer for genflyt og fenotypiske forskjeller mellom *F. a. naumanni* og *F. a. arctica* (**Paper IV**). Resultatene i denne avhandlingen understreker betydningen av historiske og moderne helgenomdata for å forstå populasjonsstruktur og genflyt hos sjøfugl, samt det genetiske grunnlaget for intraspesifikke, fenotypiske forskjeller og lokal tilpasning. Med tanke på den globale nedgangen i biodiversitet som skjer med ekstrem hastighet, bør disse funnene ha implikasjoner for fremtidig forskning og forvaltning av sjøfugl.

LIST OF PAPERS

This thesis is composed of two published papers, one submitted and one unsubmitted manuscript:

1. **Kersten, O.,** Star, B., Leigh, D. M., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps, S., Erikstad, K. E., Fitzsimmons, M. G., Fort, J., Hansen, E. S., Harris, M. P., Irestedt, M., Kleven, O., Mallory, M. L., Jakobsen, K. S., & Boessenkool, S. (2021). Complex population structure of the Atlantic puffin revealed by whole genome analyses. *Communications Biology*, *4*(1), 922. doi: 10.1038/s42003-021-02415-4

2. Leigh, D. M.*, **Kersten, O.***, Star, B., Anker-Nilssen, T., Burnham, K., Johnson, J., Provencher, J., & Boessenkool, S. (2022). Sympatry of genetically distinct Atlantic Puffins (*Fratercula arctica*) in the High Arctic. *IBIS*. doi: 10.1111/ibi.13153
*Contributed equally

   Nominated for the best Early Career Researcher (ECR) publication in IBIS in 2022 by the British Ornithologists' Union

3. **Kersten, O.,** Star, B., Krabberød, A. K., Atmore, L. M., Tørresen, O. K., Anker-Nilssen, T., Descamps, S., Strøm, H., Johansson, U. S., Sweet, P. R., Jakobsen, K. S., & Boessenkool, S. (submitted). Hybridization of Atlantic puffins in the Arctic coincides with 20th-century climate change. *Submitted to Science Advances*

4. **Kersten, O.,** Star, B., Jakobsen, K. S., Anker-Nilssen, T., Strøm, H., & Boessenkool, S. (manuscript). The Genomic Basis of Differentiation in the Atlantic Puffin. *To be submitted to Molecular Ecology.*

INTRODUCTION

Background

Seabirds constitute 3-4% of all avian species and are dependent on marine ecosystems for survival. They exhibit specialized morphological adaptations, are generally characterized by high longevity, and display a strong site fidelity during the breeding season and high mobility during the non-breeding season (e.g. Croxall et al. 2012, BirdLife International 2018). Recent studies have revealed that seabird population sizes worldwide have decreased by 47-70% over the past few decades (Croxall et al. 2012, Paleczny et al. 2015). These declines have been attributed to a variety of direct and indirect human-induced threats, such as invasive alien species, fisheries bycatch, and human disturbance (e.g. Chardine and Mendenhall 1998, Croxall et al. 2012, Fauchald et al. 2015, Paleczny et al. 2015, Keogan et al. 2018). Climate change likely amplifies most if not all of these threats and introduces novel dangers, indirectly through cumulative effects on changes in food availability and invasive predators, and directly through sea level rise and increases in extreme weather events that can reduce foraging efficiency and the quality of nest sites (Burger 2018). Indeed, various climate parameters have already been correlated to the availability and variation in food resources (Descamps et al. 2017, Keogan et al. 2018), which has important repercussions for the reproduction and survival of seabird populations (Cury et al. 2011). As a result, climate change is predicted to become the most pressing threat to many seabird species in the near future (e.g. Durant et al. 2004, Croxall et al. 2012, Poloczanska et al. 2013).

Given their deteriorating status, it is imperative to acknowledge the ecological, cultural and economic value of seabirds. Ecologically, seabirds act as indicator species occupying the highest trophic level in the food web and function as biological pumps that transfer large amounts of nutrients between marine and terrestrial ecosystems (e.g. Piatt et al. 2007, Parsons et al. 2008, Otero et al. 2018). From a cultural and economic standpoint, seabirds also play a distinctive role in our society. While contemporary seabird tourism is generating substantial revenues for local economies, seabirds have traditionally been harvested for meat, oil, and feathers for thousands of

years and continue to be harvested commercially today (Chardine and Mendenhall 1998, Denlinger and Wohl 2001, Merkel and Barry 2008).

The observed ongoing declines in seabird populations will have far-reaching impacts on both marine ecosystems and human society. Therefore, conservation efforts must be intensified, particularly in regions such as the Arctic, where warming is expected to be exacerbated (Serreze and Barry 2011, Hoberg et al. 2013). However, the success of these efforts is contingent on a better understanding of environmental and ecological processes affecting seabirds, as well as of the spatiotemporal genetic structure of seabird species and populations (e.g. Croxall et al. 2012, Sydeman et al. 2012, Hoberg et al. 2013).

Thus, the aim of this thesis is to resolve an extensive gap in the spatiotemporal resolution of the genomic population structure of the Atlantic puffin, a pelagic North Atlantic seabird currently experiencing substantial declines globally. Ideally, this thesis eventually facilitates conservation programs by building a molecular framework that allows for the evaluation of short- and long-term impacts of population threats. In the first part of this thesis, general information on the Atlantic puffin and a short overview of the approaches I chose to study the genetic structure of Atlantic puffin populations on various spatiotemporal scales are provided in the introduction. This is followed by the research questions and outline of each presented manuscript. In the second part of this thesis, the four manuscripts, which present the results of the work, are provided followed by a general discussion addressing the new insights obtained on the genomic population structure of the Atlantic puffin and how to implement these results into a broader conservation framework.

The Atlantic Puffin

The Atlantic puffin (*Fratercula arctica*) is a small- to medium sized seabird within the family of alcids (Alcidae) in the order Charadriiformes (Harris and Wanless 2011, BirdLife International 2017). Within the genus *Fratercula*, the Atlantic puffin is joined by its two North Pacific “cousins” (BirdLife International 2017), the Horned puffin (*Fratercula corniculata*) and the Tufted puffin (*Fratercula cirrhata*). The Atlantic puffin is readily recognized by its eye-catching outer appearance, including a large triangular bill with a very distinct color pattern (Figure 1). This is also the foundation for its

Figure 1: Atlantic Puffin at Runde, Norway. By Annemarie Loof.

alternative names, such as sea parrot or clowns of the sea. During the breeding season, puffins nest on grassy or rocky slopes and sea cliffs throughout their North Atlantic distribution, which ranges from Spitsbergen and northern Greenland in the north, to France and Maine in the south (Figure 2; e.g. Harris and Wanless 2011, BirdLife International 2017). During the non-breeding season, puffins are found in the offshore pelagic realm resulting in a very extensive North Atlantic range (Figure 2; e.g. Harris and Wanless 2011, BirdLife International 2017, Fayet et al. 2017). Atlantic puffins are pursuit divers that forage within 10 - 100 km of their colony during the breeding season (e.g. Harris and Wanless 2011, Shoji et al. 2015). The diet of puffins

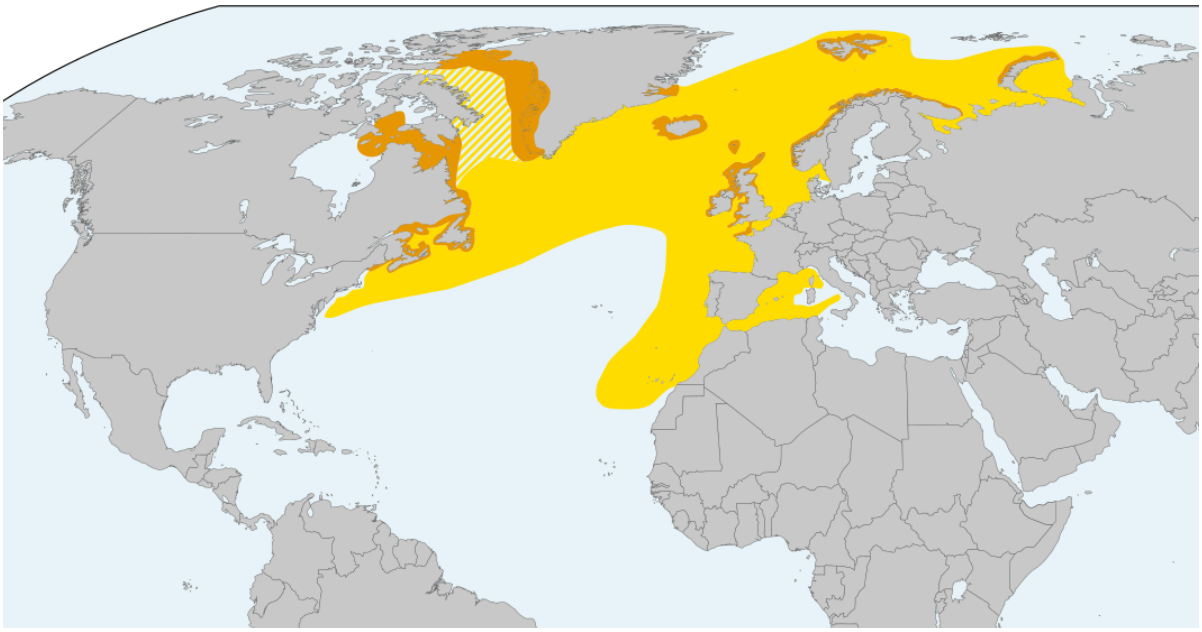


Figure 2: Atlantic Puffin distribution during breeding (orange) and non-breeding (yellow). Yellow stripes = Baffin Bay. Creative Commons 3.0.

essentially consists of sandeel (*Ammodytes spp.*), juvenile herring (*Clupea harengus*), capelin (*Mallotus villosus*) and small gadoids, but can vary substantially between seasons and across life-stages (e.g. Anker-Nilssen and Aarvak 2002, Fauchald et al. 2015, Harris et al. 2015, BirdLife International 2017).

The species' extensive range, recognizable outer appearance and its popularity as photo object has made the Atlantic puffin a culturally and economically important seabird species. The bird has been featured on a variety of stamps (Gibbins 1998) and currency (NorgesBank 2017), and has given the "Lundehund" - one of Norway's most ancient breed of dogs - its name (Melis et al. 2013). Furthermore, puffins have historically been exploited for their meat and down (Hodgetts 1999, Dove and Wickler 2016), which remains an important cultural tradition in Iceland and the Faroe Islands (Merkel and Barry 2008, Huijbens and Einarsson 2018). Puffins also provide a source of economic revenue to local areas via tourism (Harris and Wanless 2011, Huijbens and Einarsson 2018, Lund et al. 2018).

Nevertheless, the species has been designated as ***vulnerable to extinction*** globally and as ***endangered*** in Europe due to estimations that the European population (4.8-5.8 million breeding pairs, accounting for more than 90% of the global population) will decrease by 50-79% between 2000-2065 (e.g. Harris and Wanless

2011, Fauchald et al. 2015, BirdLife International 2017). Additionally, some colonies in Norway and Iceland that are home to 75-80% of the European population have experienced substantial breeding failure over the last decade (Lilliendahl et al. 2013, Anker-Nilssen et al. 2022). Most of these declines are presumed to be associated with food limitation as a result of climate change, as well as an interplay of threats such as hunting, bycatch, predation/parasitism, pollution, severe weather events and human disturbance (e.g. Anker-Nilssen and Aarvak 2002, Durant et al. 2006, Harris and Wanless 2011, Anker-Nilssen et al. 2017, Descamps et al. 2017, Hansen et al. 2021). In contrast, a few colonies in Iceland, the UK, and Norway have reported a stable or increasing population size. These colony-specific responses to an interplay of potential threats affecting breeding success, as well as chick and adult survival, indicate a high degree of ecological independence of the different colonies (e.g. Harris and Wanless 2011, BirdLife International 2017, Hansen et al. 2021, Anker-Nilssen et al. 2022). Such independence is also reflected in the colony-specific ranges with limited overlap during the non-breeding season compared to many other seabirds (Fayet et al. 2017). Yet, the cause-and-effect dynamics between population trends and species ecology, the marine ecosystem and the aforementioned stressors remain largely unclear.

Within recent years, various conservation actions surrounding the Atlantic puffin have been set in place, ranging from monitoring key breeding colonies to reintroductions and invasive alien species eradication (e.g. Anker-Nilssen and Aarvak 2002, Harris and Wanless 2011, Anker-Nilssen et al. 2022). Yet, knowledge of the genetic population structure to evaluate the geographic scale of conservation strategies is essentially absent and, surprisingly, even basic taxonomy within the species remains unresolved and controversial (Salomonsen 1944, Harris 1979, Moen 1991, Harris and Wanless 2011). It also remains unknown whether the apparent ecological independence of colonies has long-term evolutionary significance. Traditionally, the Atlantic puffin is separated into three subspecies based on size variation, i.e. *F. a. naumanni* (largest - High Arctic), *F. a. arctica* (intermediate - N. Norway, Iceland) and *F. a. grabae* (smallest - Britain and S. Norway; Harris and Wanless 2011). However, the size differences are clinal (varying with latitude and sea-surface temperatures) and size distributions overlap between subspecies (Salomonsen 1944, Harris 1979, Harris and Wanless 2011). Remarkably, the only prior

genetic study on this iconic seabird was conducted in the 1990s and is based on allozyme patterns combined with a limited spatial sampling scheme (Moen 1991). This study found low allelic differentiation and essentially no genetic structuring. Given that taxonomic classification and general genetic population structure are basic requirements for designing effective conservation measures (Funk et al. 2012), the complete absence of a thorough investigation of the population structure of Atlantic puffins on all spatial-temporal scales using appropriate genetic methods hinders effective large-scale conservation actions. Ultimately, it also limits our understanding of cause-and-effect dynamics between puffin population trends, ecology and the marine ecosystem.

Population Genomics in Seabirds

In light of an ongoing global biodiversity crisis including drastic population declines of many terrestrial and marine species (e.g. Sala et al. 2000, Dirzo et al. 2014, Jaureguiberry et al. 2022), population genetics has become an integral part of the status assessment of species of conservation concern by shedding light on inter- and intraspecific demographic histories, genetic variation and diversity, as well as taxonomic delineations (e.g. Kohn et al. 2006, Funk et al. 2012, Shafer et al. 2015, Fuentes-Pardo and Ruzzante 2017). Traditionally, population genetics has used few molecular markers - such as allozymes, mtDNA, or microsatellites - and mostly targeted neutral sites, which limited genome-wide parameter estimations (e.g. Shafer et al. 2015, Allendorf 2017, Fuentes-Pardo and Ruzzante 2017). In the last decade(s), advances in DNA sequencing technology and throughput have made it feasible to sequence entire genomes (whole genome (re)sequencing – WGS) of 10s to 100s of individuals. The resulting high genetic marker density, including both adaptive and neutral loci, has led to an enormous increase in the resolution, accuracy and power of genetic analyses (e.g. Allendorf et al. 2010, Shafer et al. 2015, Allendorf 2017, Fuentes-Pardo and Ruzzante 2017).

Population genomics assesses genomic variation within and among populations of species. It is a powerful tool to, for instance, investigate intraspecific hybridization, genomic erosion, local adaptation, demographic history, as well as population structure, admixture and gene flow or barriers to gene flow (e.g. Fuentes-Pardo and

Ruzzante 2017, Diez-del-Molino et al. 2018). Population genomics can therefore provide valuable insights for conservation management programs by identifying conservation units (CUs) and other genomic parameters used as input for systematic conservation planning (SCP; Funk et al. 2012, Fuentes-Pardo and Ruzzante 2017, Hohenlohe et al. 2021, Andrello et al. 2022, Hoban et al. 2022, Nielsen et al. 2022). Within the field of population genomics, the two dominating methods are genome-wide reduced representation techniques (such as RAD-Seq) and whole genome (re)sequencing (from hereinafter referred to as WGS). The former is arguably less suited for several population genomics analyses due to an incomplete representation of genetic variation, limiting the detection of local adaptation and the reconstruction of demographic histories (Lowry et al. 2017, Marandel et al. 2020). Many population genomics analyses also require an assembled reference genome to make robust inferences. For non-model taxa without an already existing reference genome this poses limitations and ultimately decelerates their implementation in conservation (e.g. Shafer et al. 2015, Allendorf 2017, Fuentes-Pardo and Ruzzante 2017, Grealy et al. 2017). Yet, sequencing and assembly of a reference genome is becoming progressively easier and cheaper, also for non-model species (Feng et al. 2020, Paez et al. 2022), highlighted by large genome assembly efforts, such as the Vertebrate Genomes Project (https://vertebrategenomesproject.org/), Earth Biogenome Project (https://www.earthbiogenome.org/), or B10K (https://b10k.genomics.cn/). B10K, specifically, is an initiative to generate reference genome sequences from all extant bird species. Bird genomes are relatively small in size (1.0-1.3 Gb) and less complex (fewer repetitive elements, introns etc.) compared to other vertebrate genomes (Organ et al. 2007, Zhang et al. 2014, Oyler-McCance et al. 2016), which has led to a rapid increase of published bird and seabird reference genomes within the last decade and will ultimately result in a growing number of avian population genomics studies (Oyler-McCance et al. 2016, Feng et al. 2020).

While the status of the world's seabirds has deteriorated at alarming rates (e.g. Croxall et al. 2012, Paleczny et al. 2015), population genomics in seabirds using WGS or other genome-wide techniques is still in an early phase (e.g. Friesen 2015). Yet, given the complex ecology of seabirds combined with large effective population sizes, detailed genomic data including thousands of loci, as opposed to the mitogenome or

a few microsatellites, provide great potential to assess levels of population connectivity and disentangle barriers to gene flow (e.g. Friesen 2015). The few studies that have investigated the genome-wide population structure in colonial philopatric seabirds (e.g. Dierickx et al. 2015, Tigano et al. 2017, Clucas et al. 2018, Colston-Nepali et al. 2019, 2020, Antaky et al. 2020) have found only low levels of intraspecific genetic variation using neutral loci, but the results of Tigano et al. (2017) suggest that outlier loci within the genome could be informative for the structure, adaptation, and/or demographic connectivity of breeding populations. Yet, to my knowledge, no seabird population genomics study has, to date, employed WGS and utilized its ability to potentially detect fine-scale structure and barriers to gene flow with a number of informative loci that is 1-2 orders of magnitude higher than in other genome-wide approaches. This highlights the potential of WGS in population genomics analyses in seabirds, as well as the gap in our knowledge regarding differences in detailed genomic structure between separated breeding populations of pelagic seabirds. Moreover, only a handful of studies have explored genomic information of historical seabird populations to investigate status and dynamics prior to human-induced impacts (e.g. Thomas et al. 2019). Generating such long-term perspectives is critical to determine baseline targets for seabird conservation programs and enables us to understand how seabirds have historically responded to past environmental change and anthropogenic pressures (Shafer et al. 2015, Díez-del-Molino et al. 2018, Jensen and Leigh 2022).

In summary, population genomics using WGS coupled with a high-quality reference genome has several crucial benefits for seabird conservation. It allows the assessment of genetic population structure, gene flow and demographic history, and is not limited to *a priori* selected candidate regions, therefore enabling an impartial estimate of genetic diversity. Moreover, WGS of historical or museum specimens could provide a unique opportunity to link historical and present seabird population dynamics with ecology, environmental parameters, and anthropogenic impacts.

Temporal Genomics with Avian Museum Specimens

Temporal genomics refers to genomic studies that analyze genetic variation within populations over various time scales, with the aim to detect and quantify changes in genetic diversity, allele frequencies, and population structure (Jensen and Leigh 2022).

However, one of the challenges of conducting temporal genomics studies is acquiring a good baseline (Díez-del-Molino et al. 2018, Jensen and Leigh 2022). This entails obtaining a sufficient number of historical samples of the same population of interest collected before an event of interest (Wandeler et al. 2007, Díez-del-Molino et al. 2018, Jensen and Leigh 2022). Additionally, such studies might be limited by the lack of historical or archeological specimens in museum collections or the lack of critical metadata (Wandeler et al. 2007, Holmes et al. 2016, Jensen and Leigh 2022). Despite these challenges, temporal genomics can provide unique insights and increased analytical power for four key genetic indicators of population responses to anthropogenic and environmental stressors. These include: 1) genomic erosion, 2) changes in population structure, 3) adaptation, and 4) hybridization (Wandeler et al. 2007, Habel et al. 2014, Holmes et al. 2016, Díez-del-Molino et al. 2018, Jensen and Leigh 2022). Specifically, temporal genomics can detect changes in genetic diversity over time, changes in population structure due to altered gene flow patterns, shifts in allele frequencies due to selection and onsets of hybridization driven by range expansions, and allows to assess whether and how these changes are related to anthropogenic stressors (Wandeler et al. 2007, Habel et al. 2014, Holmes et al. 2016, Díez-del-Molino et al. 2018, Jensen and Leigh 2022). In conclusion, temporal genomics is a valuable tool for understanding and addressing the impacts of anthropogenic and environmental pressure on natural populations, and, as such, is becoming increasingly important for the development of appropriate conservation measures (Díez-del-Molino et al. 2018, Jensen and Leigh 2022).

Analyzing DNA sequences obtained from museum specimens that are usually not older than 250 years (hDNA, Raxworthy and Smith 2021, Irestedt et al. 2022) has been applied to a wide diversity of species ranging from birds and mammals, to insect and plants (Raxworthy and Smith 2021). Among avian studies, hDNA is usually extracted from toe pads or museum skins, although sources like bones, eggshells or feathers have also proven to contain sufficient amounts of hDNA (Grealy et al. 2017, Raxworthy and Smith 2021, Irestedt et al. 2022). Benefitting from the advances of DNA sequencing technology and extraction methods, avian hDNA has been used to study a variety of topics, including phylogenetics, biogeography, taxonomy/classification, domestication, paleoenvironmental reconstruction, zooarchaeology and conservation

(Grealy et al. 2017, Raxworthy and Smith 2021, Irestedt et al. 2022). Yet, only a few avian, let alone seabird, hDNA studies have applied WGS (Irestedt et al. 2022) despite the fact that genome-wide reduced representation approaches, a more cost-effective alternative, are likely poorly suited for studying historic museum samples and genome-wide variation (see *Population Genomics in Seabirds*; Bi et al. 2013, Lowry et al. 2017, Marandel et al. 2020, Irestedt et al. 2022; but see Burrell et al. 2015).

Given the potential of temporal genomics and WGS of hDNA from museum specimens, a combination of both approaches allows for the comparison of genome-wide parameters of populations before and after human disturbances that have occurred within the last 200 years, and enables to assess the genetic changes that occurred in response to these disturbances (Wandeler et al. 2007, Habel et al. 2014, Holmes et al. 2016, Díez-del-Molino et al. 2018, Jensen and Leigh 2022). As a result, this approach is seemingly extremely valuable for seabird conservation, as climate change and human activities and their associated direct and indirect impacts on seabird populations have been most pronounced within this time span (e.g. Durant et al. 2004, Croxall et al. 2012, Poloczanska et al. 2013, Paleczny et al. 2015). Nevertheless, studies using WGS of seabird hDNA of the same populations across multiple points in time (temporal genomics) are rare, if not absent. For example, one study has investigated the impact of human hunting on populations of the great auk by using WGS of samples spanning the period 170-15,000 years before present (Thomas et al 2019). However, this study only recovered and analyzed whole mitogenomes thereby effectively only investigating one single genetic locus. Nevertheless, given the recent technological advances and the dramatic declines of seabirds, establishing baseline levels of genome-wide parameters (i.e. whole nuclear and mitochondrial genomes) prior to recent demographic declines, and resolving the causes of these declines will become increasingly relevant for the development of effective conservation strategies (e.g., Dietl and Flessa 2011, Díez-del-Molino et al. 2018, Jensen and Leigh 2022).

Structural Genomic Variants: Enlarging the Population Genomics Analyses Toolbox beyond SNPs

In addition to studying genetic structure, diversity and gene-flow over various spatiotemporal scales, population genomics also allows to assess local adaptation within and among populations. Traditionally, the focus has been on analyzing single nucleotide polymorphisms (SNPs), which allows for the detection of outlier loci by identifying regions along the genome that are characterized by elevated genetic differentiation between populations and species (Nachman and Payseur 2012, Cruickshank and Hahn 2014, Ravinet et al. 2017, Leigh et al. 2021). As a result, many of these loci presumably contribute to inter- or intraspecific gene flow barriers and play a role in the process of local adaptation and speciation via selection (Cruickshank and Hahn 2014, Ravinet et al. 2017, Leigh et al. 2021). The detected peaks of elevated differentiation might, however, represent false positives due to evolutionary forces other than selection, such as genetic drift, and it has therefore been advocated that outlier analyses are ideally based on whole genome sequencing, as opposed to other genome-wide methods, and should not just rely on relative divergence measures (Cruickshank and Hahn 2014, Lowry et al. 2017, Leigh et al. 2021). Consequently, methods combining SNP-based measures of relative genetic divergence, absolute genetic divergence, nucleotide diversity, and other genome-wide parameters have recently been developed to increase the detection power of outlier loci (Ma et al. 2015).

While there are countless studies investigating the genetic differentiation and local adaptation of species or populations using SNPs, it is important to note that the genetic basis of inter- and intraspecific divergence may be more complex than SNPs alone (Merot et al. 2020, Wold et al. 2021, Campagna and Toews 2022). Indeed, owing to technological advances in genomics, population genomic studies have now started to incorporate analyses of structural variants (SVs; Merot et al. 2020, Wold et al. 2021). Mounting evidence suggests that SVs – including insertions, deletions, duplications, and inversions of a length of > 50 bp – are taxonomically ubiquitous and key contributors to a multitude of evolutionary processes (Merot et al. 2020, Wold et al. 2021). They are specifically associated with adaptive phenotypes and the maintenance of differentiation between species and populations, as they can interfere with recombination and promote reproductive isolation, have shown to underlie fine-scale

population structure and influence the ability of species or populations to hybridize (Weissensteiner et al. 2020, Dorant et al. 2020, Cayuela et al. 2021, Tigano et al. 2021, Merot et al. 2022). As a result, SVs are thought to facilitate local adaptation and are presumably important drivers of speciation. Given that the detection of SVs enriches our understanding of the genetic diversity and adaptation of populations and species, SVs could become an integral part of conservation genomics. Yet, despite their potential importance, the application of SVs in conservation and population genomics remains challenging due to outstanding questions on how to cost-effectively detect and genotype them at the population scale (Wold et al. 2021, Merot et al. 2022). Also, the detection and usage of SVs within population genomics studies remains at its infancy (Weissensteiner et al. 2020, Dorant et al. 2020, Cayuela et al. 2021, Merot et al. 2022). Nevertheless, recent studies have demonstrated that an average read depth of 10x is sufficient for population-scale comparisons, given a representative sample size and a high-quality reference genome (reviewed in Wold et al. 2021). Overall, structural variants provide an exciting opportunity to complement SNP-based approaches and expand our understanding of genome-wide variation in population and conservation genomics.

References

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. Molecular Ecology, 26(2), 420–430. https://doi.org/10.1111/mec.13948

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. Nature Reviews. Genetics, 11, 697. https://doi.org/10.1038/nrg2844

Andrello, M., D'Aloia, C., Dalongeville, A., Escalante, M. A., Guerrero, J., Perrier, C., Torres-Florez, J. P., Xuereb, A., & Manel, S. (2022). Evolving spatial conservation prioritization with intraspecific genetic data. Trends in Ecology & Evolution, 37(6), 553–564. https://doi.org/10.1016/j.tree.2022.03.003

Anker-Nilssen, T., & Aarvak, T. (2002). The population ecology of puffins at Røst. Status after the breeding season 2001. Norsk Institutt for Naturforskning. Oppdragsmelding, 736, 1–40.

Anker-Nilssen, T., Hanssen, S. A., Moe, B., Systad, G. H. R., Barrett, R., Bustnes, J. O., Christensen-Dalsgaard, S., Dehnhard, N., Descamps, S., Erikstad, K. E., Follestad, A., Langset, M., Layton-Matthews, K., Lorentsen, S. H., Lorentzen, E., Reiertsen, T. K., & Strøm, H. (2022). Key-site monitoring in Norway 2021, including Svalbard and Jan Mayen. SEAPOP Short Report 1-2022.

Anker-Nilssen, T., Harris, M. P., Kleven, O., & Langset, M. (2017). Status, origin, and population level impacts of Atlantic Puffins killed in a mass mortality event in southwest Norway early 2016. Seabird, 30, 1–14.

Antaky, C. C., Conklin, E. E., Toonen, R. J., Knapp, I. S. S., & Price, M. R. (2020). Unexpectedly high genetic diversity in a rare and endangered seabird in the Hawaiian Archipelago. PeerJ, 8, e8463. https://doi.org/10.7717/peerj.8463

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: next-generation museum population genomics. Molecular Ecology, 22(24), 6018–6032. https://doi.org/10.1111/mec.12516

BirdLife International. (2017). Fratercula arctica. The IUCN Red List of Threatened Species 2017. https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T22694927A117606911.en.

BirdLife International. (2018). Seabirds and Marine - What are seabirds? BirdLife. https://www.birdlife.org/europe-and-central-asia/seabirds-and-marine-what-are-seabirds

Burger, J. (2018). Understanding population changes in seabirds requires examining multiple causal factors and developing science-based adaptive species conservation plans. Animal Conservation, 21(1), 17–18. https://onlinelibrary.wiley.com/doi/abs/10.1111/acv.12395

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. Journal of Human Evolution, 79, 35–44. https://doi.org/10.1016/j.jhevol.2014.10.015

Campagna, L., & Toews, D. P. L. (2022). The genomics of adaptation in birds. Current Biology: CB, 32(20), R1173–R1186. https://doi.org/10.1016/j.cub.2022.07.076

Cayuela, H., Dorant, Y., Mérot, C., Laporte, M., Normandeau, E., Gagnon-Harvey, S., Clément, M., Sirois, P., & Bernatchez, L. (2021). Thermal adaptation rather than demographic history drives genetic structure inferred by copy number variants in a marine fish. Molecular Ecology, 30(7), 1624–1641. https://doi.org/10.1111/mec.15835

Chardine, J., & Mendenhall, V. (1998). Human disturbance at Arctic seabird colonies (CAFF Technical Report 2). Conservation of Arctic Flora and Fauna (CAFF).

Clucas, G. V., Younger, J. L., Kao, D., Emmerson, L., Southwell, C., Wienecke, B., Rogers, A. D., Bost, C.-A., Miller, G. D., Polito, M. J., Lelliott, P., Handley, J., Crofts, S., Phillips, R. A., Dunn, M. J., Miller, K. J., & Hart, T. (2018). Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. Molecular Ecology, 27, 4680–4697. https://doi.org/10.1111/mec.14896

Colston-Nepali, L., Provencher, J. F., Mallory, M. L., Franckowiak, R. P., Sun, Z., Robertson, G. J., & Friesen, V. L. (2020). Using genomic tools to inform management of the Atlantic northern fulmar. Conservation Genetics. https://doi.org/10.1007/s10592-020-01309-y

Colston-Nepali, L., Tigano, A., Boyle, B., & Friesen, V. (2019). Hybridization does not currently pose conservation concerns to murres in the Atlantic. Conservation Genetics. https://doi.org/10.1007/s10592-019-01223-y

Croxall, J. P., Butchart, S. H. M., Lascelles, B., Stattersfield, A. J., Sullivan, B., Symes, A., & Taylor, P. (2012). Seabird conservation status, threats and priority actions: a global assessment. Bird Conservation International, 22(1), 1–34. https://doi.org/10.1017/S0959270912000020

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular

Ecology, 23(13), 3133–3157. https://doi.org/10.1111/mec.12796

Cury, P. M., Boyd, I. L., Bonhommeau, S., Anker-Nilssen, T., Crawford, R. J. M., Furness, R. W., Mills, J. A., Murphy, E. J., Österblom, H., Paleczny, M., Piatt, J. F., Roux, J.-P., Shannon, L., & Sydeman, W. J. (2011). Global Seabird Response to Forage Fish Depletion—One-Third for the Birds. Science, 334(6063), 1703–1706. https://doi.org/10.1126/science.1212928

Denlinger, L., & Wohl, K. (2001). Seabird harvest regimes in the circumpolar nations. Conservation of Arctic Flora and Fauna (CAFF). https://oaarchive.arctic-council.org/handle/11374/171

Descamps, S., Anker-Nilssen, T., Barrett, R. T., Irons, D. B., Merkel, F., Robertson, G. J., Yoccoz, N. G., Mallory, M. L., Montevecchi, W. A., Boertmann, D., & Others. (2017). Circumpolar dynamics of a marine top-predator track ocean warming rates. Global Change Biology, 23(9), 3770–3780. https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.13715

Dierickx, E. G., Shultz, A. J., Sato, F., Hiraoka, T., & Edwards, S. V. (2015). Morphological and genomic comparisons of Hawaiian and Japanese Black-footed Albatrosses (Phoebastria nigripes) using double digest RADseq: implications for conservation. Evolutionary Applications, 8(7), 662–678. https://doi.org/10.1111/eva.12274

Dietl, G. P., & Flessa, K. W. (2011). Conservation paleobiology: putting the dead to work. Trends in Ecology & Evolution, 26(1), 30–37. https://doi.org/10.1016/j.tree.2010.09.010

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying Temporal Genomic Erosion in Endangered Species. Trends in Ecology & Evolution, 33(3), 176–185. https://doi.org/10.1016/j.tree.2017.12.002

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J. B., & Collen, B. (2014). Defaunation in the Anthropocene. Science, 345(6195), 401–406. https://doi.org/10.1126/science.1251817

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., Normandeau, E., Rochette, R., & Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype-temperature association in a marine species. Molecular Ecology, 29(24), 4765–4782. https://doi.org/10.1111/mec.15565

Dove, C. J., & Wickler, S. (2016). Identification of bird species used to make a Viking age feather pillow. Arctic, 69(1), 29–36. http://www.jstor.org/stable/43871396

Durant, J. M., Anker-Nilssen, T., & Stenseth, N. C. (2006). Ocean climate prior to breeding affects the duration of the nestling period in the Atlantic puffin. Biology Letters, 2(4), 628–631. https://doi.org/10.1098/rsbl.2006.0520

Durant, J. M., Stenseth, N. C., Anker-Nilssen, T., Harris, M. P., Thompson, P. M., & Wanless, S. (2004). Marine birds and climate fluctuation in the North Atlantic. Marine Ecosystems and Climate Variation: The North Atlantic, 95–105.

Fauchald, P., Anker-Nilssen, T., Barrett, R. T., Bustnes, J. O., Bårdsen, B.-J., Christensen-Dalsgaard, S., Descamps, S., Engen, S., Erikstad, K. E., Hanssen, S. A., & Others. (2015). The status and trends of seabirds breeding in Norway and Svalbard.

Fayet, A. L., Freeman, R., Anker-Nilssen, T., Diamond, A., Erikstad, K. E., Fifield, D., Fitzsimmons, M. G., Hansen, E. S., Harris, M. P., Jessopp, M., Kouwenberg, A.-L., Kress, S., Mowat, S., Perrins, C. M., Petersen, A., Petersen, I. K., Reiertsen, T. K., Robertson, G. J., Shannon, P., … Guilford, T. (2017). Ocean-wide drivers of migration strategies and their influence on population breeding performance in a declining seabird. Current Biology: CB, 27(24), 3871–3878. https://doi.org/10.1016/j.cub.2017.11.009

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., … Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. Nature, 587(7833), 252–257. https://doi.org/10.1038/s41586-020-2873-9

Friesen, V. L. (2015). Speciation in seabirds: why are there so many species… and why aren't there more? Journal of Ornithology / DO-G, 156(1), 27–39. https://link.springer.com/article/10.1007/s10336-015-1235-0

Fuentes-Pardo, A. P., & Ruzzante, D. E. (2017). Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. Molecular Ecology, 26(20), 5369–5406. https://doi.org/10.1111/mec.14264

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. Trends in Ecology & Evolution, 27(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Gibbins, C. (1998). Atlantic Puffin stamps. Birds of the World on Postage Stamps. http://www.bird-stamps.org/cspecies/7002100.htm

Grealy, A., Rawlence, N. J., & Bunce, M. (2017). Time to Spread Your Wings: A Review of the Avian Ancient

DNA Field. Genes, 8(7). https://doi.org/10.3390/genes8070184

Habel, J. C., Husemann, M., Finger, A., Danley, P. D., & Zachos, F. E. (2014). The relevance of time series in molecular ecology and conservation biology. Biological Reviews of the Cambridge Philosophical Society, 89(2), 484–492. https://doi.org/10.1111/brv.12068

Hansen, E. S., Sandvik, H., Erikstad, K. E., Yoccoz, N. G., Anker-Nilssen, T., Bader, J., Descamps, S., Hodges, K., Mesquita, M. D. S., Reiertsen, T. K., & Varpe, Ø. (2021). Centennial relationships between ocean temperature and Atlantic puffin production reveal shifting decennial trends. Global Change Biology, 27(16), 3753–3764. https://doi.org/10.1111/gcb.15665

Harris, M. P. (1979). Measurements and weights of British Puffins. Bird Study: The Journal of the British Trust for Ornithology, 26(3), 179–186. https://doi.org/10.1080/00063657909476636

Harris, M. P., Leopold, M. F., Jensen, J.-K., Meesters, E. H., & Wanless, S. (2015). The winter diet of the Atlantic Puffin Fratercula arctica around the Faroe Islands. The Ibis, 157(3), 468–479. https://doi.org/10.1111/ibi.12272

Harris, M. P., & Wanless, S. (2011). The Puffin. T & AD Poyser, Bloomsbury Publishing.

Hoban, S., Archer, F. I., Bertola, L. D., Bragg, J. G., Breed, M. F., Bruford, M. W., Coleman, M. A., Ekblom, R., Funk, W. C., Grueber, C. E., Hand, B. K., Jaffé, R., Jensen, E., Johnson, J. S., Kershaw, F., Liggins, L., MacDonald, A. J., Mergeay, J., Miller, J. M., … Hunter, M. E. (2022). Global genetic diversity status and trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. Biological Reviews of the Cambridge Philosophical Society, 97(4), 1511–1538. https://doi.org/10.1111/brv.12852

Hoberg, E. P., Kutz, S. J., Cook, J. A., & Galaktionov, K. (2013). Arctic Biodiversity Assessment: Status and Trends in Arctic Biodiversity - Synthesis. Conservation of Arctic Flora and Fauna (CAFF).

Hodgetts, L. M. (1999). Animal bones and human society in the late younger stone age of arctic Norway [Doctoral, Durham University]. http://etheses.dur.ac.uk/4491/

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. Molecular Ecology, 30(1), 62–82. https://doi.org/10.1111/mec.15720

Holmes, M. W., Hammond, T. T., Wogan, G. O. U., Walsh, R. E., LaBarbera, K., Wommack, E. A., Martins, F. M., Crawford, J. C., Mack, K. L., Bloch, L. M., & Nachman, M. W. (2016). Natural history collections as windows on evolutionary processes. Molecular Ecology, 25(4), 864–881. https://doi.org/10.1111/mec.13529

Huijbens, E. H., & Einarsson, N. (2018). Feasting on friends: Whales, puffins, and tourism in Iceland. In Tourism Experiences and Animal Consumption (pp. 10–27). Routledge.

Irestedt, M., Thörn, F., Müller, I. A., Jønsson, K. A., Ericson, P. G. P., & Blom, M. P. K. (2022). A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. Molecular Ecology Resources, 22(7), 2672–2684. https://doi.org/10.1111/1755-0998.13660

Jaureguiberry, P., Titeux, N., Wiemers, M., Bowler, D. E., Coscieme, L., Golden, A. S., Guerra, C. A., Jacob, U., Takahashi, Y., Settele, J., Díaz, S., Molnár, Z., & Purvis, A. (2022). The direct drivers of recent global anthropogenic biodiversity loss. Science Advances, 8(45), eabm9982. https://doi.org/10.1126/sciadv.abm9982

Jensen, E. L., & Leigh, D. M. (2022). Using temporal genomics to understand contemporary climate change responses in wildlife. Ecology and Evolution, 12(9), e9340. https://doi.org/10.1002/ece3.9340

Keogan, K., Daunt, F., Wanless, S., Phillips, R. A., Walling, C. A., Agnew, P., Ainley, D. G., Anker-Nilssen, T., Ballard, G., Barrett, R. T., Barton, K. J., Bech, C., Becker, P., Berglund, P.-A., Bollache, L., Bond, A. L., Bouwhuis, S., Bradley, R. W., Burr, Z. M., … Lewis, S. (2018). Global phenological insensitivity to shifting ocean temperatures among seabirds. Nature Climate Change, 8(4), 313–318. https://doi.org/10.1038/s41558-018-0115-z

Kohn, M. H., Murphy, W. J., Ostrander, E. A., & Wayne, R. K. (2006). Genomics and conservation genetics. Trends in Ecology & Evolution, 21(11), 629–637. https://doi.org/10.1016/j.tree.2006.08.001

Leigh, D. M., Lischer, H. E. L., Guillaume, F., Grossen, C., & Günther, T. (2021). Disentangling adaptation from drift in bottlenecked and reintroduced populations of Alpine ibex. Molecular Ecology Resources, 21(7), 2350–2363. https://doi.org/10.1111/1755-0998.13442

Lilliendahl, K., Hansen, E. S., Bogason, V., Sigursteinsson, M., Magnúsdóttir, M. L., Jónsson, P. M., Helgason, H. H., Óskarsson, G. J., Óskarsson, P. F., & Sigurðsson, Ó. J. (2013). Recruitment failure of Atlantic puffins Fratercula arctica and sandeels Ammodytes marinus in Vestmannaeyjar Islands. Náttúrufræðingurinn, 83(1-2), 65–79.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of

adaptation. Molecular Ecology Resources, 17(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Lund, K. A., Kjartansdóttir, K., & Loftsdóttir, K. (2018). “Puffin love”: Performing and creating Arctic landscapes in Iceland through souvenirs. Tourist Studies, 18(2), 142–158. https://doi.org/10.1177/1468797617722353

Marandel, F., Charrier, G., Lamy, J.-B., Le Cam, S., Lorance, P., & Trenkel, V. M. (2020). Estimating effective population size using RADseq: Effects of SNP selection and sample size. Ecology and Evolution, 10(4), 1929–1937. https://doi.org/10.1002/ece3.6016

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., & Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. Heredity, 115(5), 426–436. https://doi.org/10.1038/hdy.2015.42

Melis, C., Borg, Å. A., Espelien, I. S., & Jensen, H. (2013). Low neutral genetic variability in a specialist puffin hunter: the Norwegian Lundehund. Animal Genetics, 44(3), 348–351. https://doi.org/10.1111/age.12000

Merkel, F., & Barry, T. (2008). Seabird harvest in the Arctic. Conservation of Arctic Flora and Fauna (CAFF).

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. Trends in Ecology & Evolution, 35(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Mérot, C., Stenløkk, K. S. R., Venney, C., Laporte, M., Moser, M., Normandeau, E., Árnyasi, M., Kent, M., Rougeux, C., Flynn, J. M., Lien, S., & Bernatchez, L. (2022). Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (Coregonus sp.) with long and short reads. Molecular Ecology. https://doi.org/10.1111/mec.16468

Moen, S. M. (1991). Morphologic and genetic variation among breeding colonies of the Atlantic puffin (Fratercula arctica). The Auk, 108(4), 755–763. http://www.jstor.org/stable/4088303

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 367(1587), 409–421. https://doi.org/10.1098/rstb.2011.0249

Nielsen, E. S., Hanson, J. O., Carvalho, S. B., Beger, M., Henriques, R., Kershaw, F., & von der Heyden, S. (2022). Molecular ecology meets systematic conservation planning. Trends in Ecology & Evolution. https://doi.org/10.1016/j.tree.2022.09.006

NorgesBank. (2017). New Banknote Series. https://www.norges-bank.no/en/notes-and-coins/New-banknote-series/

Organ, C. L., Shedlock, A. M., Meade, A., Pagel, M., & Edwards, S. V. (2007). Origin of avian genome size and structure in non-avian dinosaurs. Nature, 446(7132), 180–184. https://doi.org/10.1038/nature05621

Otero, X. L., De La Peña-Lastra, S., Pérez-Alberti, A., Ferreira, T. O., & Huerta-Diaz, M. A. (2018). Seabird colonies as important global drivers in the nitrogen and phosphorus cycles. Nature Communications, 9(1), 246. https://doi.org/10.1038/s41467-017-02446-8

Oyler-McCance, S. J., Oh, K. P., Langin, K. M., & Aldridge, C. L. (2016). A field ornithologist’s guide to genomics: Practical considerations for ecology and conservation. The Auk, 133(4), 626–648. https://doi.org/10.1642/AUK-16-49.1

Paez, S., Kraus, R. H. S., Shapiro, B., Gilbert, M. T. P., Jarvis, E. D., Vertebrate Genomes Project Conservation Group, Al-Ajli, F. O., Ceballos, G., Crawford, A. J., Fedrigo, O., Johnson, R. N., Johnson, W. E., Marques-Bonet, T., Morin, P. A., Mueller, R. C., Ryder, O. A., Teeling, E. C., & Venkatesh, B. (2022). Reference genomes for conservation. Science, 377(6604), 364–366. https://doi.org/10.1126/science.abm8127

Paleczny, M., Hammill, E., Karpouzi, V., & Pauly, D. (2015). Population trend of the world’s monitored seabirds, 1950-2010. PloS One, 10(6), e0129342. https://doi.org/10.1371/journal.pone.0129342

Parsons, M., Mitchell, I., Butler, A., & Ratcliffe, N. (2008). Seabirds as indicators of the marine environment. ICES Journal of Marine Science. https://academic.oup.com/icesjms/article-abstract/65/8/1520/714606

Piatt, J. F., Sydeman, W. J., & Wiese, F. (2007). Introduction: a modern role for seabirds as indicators. Marine Ecology Progress Series, 352, 199–204. https://www.jstor.org/stable/24872144

Poloczanska, E. S., Brown, C. J., Sydeman, W. J., Kiessling, W., Schoeman, D. S., Moore, P. J., Brander, K., Bruno, J. F., Buckley, L. B., Burrows, M. T., Duarte, C. M., Halpern, B. S., Holding, J., Kappel, C. V., O’Connor, M. I., Pandolfi, J. M., Parmesan, C., Schwing, F., Thompson, S. A., & Richardson, A. J. (2013). Global imprint of climate change on marine life. Nature Climate Change, 3, 919. https://doi.org/10.1038/nclimate1958

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene

flow. Journal of Evolutionary Biology, 30(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. Trends in Ecology & Evolution, 36(11), 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Sala, O. E., Chapin, F. S., 3rd, Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., Leemans, R., Lodge, D. M., Mooney, H. A., Oesterheld, M., Poff, N. L., Sykes, M. T., Walker, B. H., Walker, M., & Wall, D. H. (2000). Global biodiversity scenarios for the year 2100. Science, 287(5459), 1770–1774. https://doi.org/10.1126/science.287.5459.1770

Salomonsen, F. (1944). The Atlantic Alcidae - The seasonal and geographical variation of the auks inhabiting the Atlantic Ocean and the adjacent waters (Vol. 6). Elanders boktryckeri aktiebolag.

Serreze, M. C., & Barry, R. G. (2011). Processes and impacts of Arctic amplification: A research synthesis. Global and Planetary Change, 77(1), 85–96. https://doi.org/10.1016/j.gloplacha.2011.03.004

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., Dalén, L., De Meester, L., Ekblom, R., Fawcett, K. D., Fior, S., Hajibabaei, M., Hill, J. A., Hoezel, A. R., Höglund, J., Jensen, E. L., Krause, J., Kristensen, T. N., … Zieliński, P. (2015). Genomics and the challenging translation into conservation practice. Trends in Ecology & Evolution, 30(2), 78–87. https://doi.org/10.1016/j.tree.2014.11.009

Sydeman, W. J., Thompson, S. A., & Kitaysky, A. (2012). Seabirds and climate change: roadmap for the future. Marine Ecology Progress Series, 454, 107–117. https://doi.org/10.3354/meps09806

Thomas, J. E., Carvalho, G. R., Haile, J., Rawlence, N. J., Martin, M. D., Ho, S. Y., Sigfússon, A. Þ., Jósefsson, V. A., Frederiksen, M., Linnebjerg, J. F., Samaniego Castruita, J. A., Niemann, J., Sinding, M.-H. S., Sandoval-Velasco, M., Soares, A. E., Lacy, R., Barilaro, C., Best, J., Brandis, D., … Knapp, M. (2019). Demographic reconstruction from ancient DNA supports rapid extinction of the great auk. eLife, 8. https://doi.org/10.7554/eLife.47509

Tigano, A., Jacobs, A., Wilder, A. P., Nand, A., Zhan, Y., Dekker, J., & Therkildsen, N. O. (2021). Chromosome-Level Assembly of the Atlantic Silverside Genome Reveals Extreme Levels of Sequence Diversity and Structural Genetic Variation. Genome Biology and Evolution, 13(6). https://doi.org/10.1093/gbe/evab098

Tigano, A., Shultz, A. J., Edwards, S. V., Robertson, G. J., & Friesen, V. L. (2017). Outlier analyses to test for local adaptation to breeding grounds in a migratory arctic seabird. Ecology and Evolution, 7(7), 2370–2381. https://doi.org/10.1002/ece3.2819

Wandeler, P., Hoeck, P. E. A., & Keller, L. F. (2007). Back to the future: museum specimens in population genetics. Trends in Ecology & Evolution, 22(12), 634–642. https://doi.org/10.1016/j.tree.2007.08.017

Weissensteiner, M. H., Bunikis, I., Catalán, A., Francoijs, K.-J., Knief, U., Heim, W., Peona, V., Pophaly, S. D., Sedlazeck, F. J., Suh, A., Warmuth, V. M., & Wolf, J. B. W. (2020). Discovery and population genomics of structural variation in a songbird genus. Nature Communications, 11(1), 3403. https://doi.org/10.1038/s41467-020-17195-4

Wold, J., Koepfli, K.-P., Galla, S. J., Eccles, D., Hogg, C. J., Le Lec, M. F., Guhlin, J., Santure, A. W., & Steeves, T. E. (2021). Expanding the conservation genomics toolbox: Incorporating structural variants to enhance genomic studies for species of conservation concern. Molecular Ecology, 30(23), 5949–5965. https://doi.org/10.1111/mec.16141

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., … Wang, J. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. Science, 346(6215), 1311–1320. https://doi.org/10.1126/science.1251385

THESIS AIMS & OUTLINE

The main aim of this thesis is to resolve the extensive gap in the spatiotemporal resolution of the genomic structure of Atlantic Puffin populations and to build a molecular framework that allows for the evaluation of short- and long-term impacts of environmental and anthropogenic threats to this seabird. As a result, this thesis is attempting to answer the following two main research questions:

1. *What is the contemporary genomic population structure of the Atlantic puffin across the species' breeding range and what ecological factors potentially drive barriers to gene flow?*

2. *What is the genomic basis for the differentiation between the large-bodied, High Arctic subspecies, F. a. naumanni, and the smaller, temperate subspecies, F. a. arctica?*

During the process of finding answers to questions 1) and 2), a fundamental third question arose:

3. *What is the timing and the direction of gene flow that resulted in the hybrid population of puffins on the Arctic island of Bjørnøya?*

In order to answer all three questions, a broad range of genomic analyses were conducted, whose results were placed in an ecological context in four papers, as follows:

Paper I provides the first insights into the range-wide genomic population structure of the Atlantic puffin. I generated the first available reference genome of the Atlantic puffin using 10x Genomics data and sequenced whole genomes of 77 individuals across 13 breeding colonies (Figure 3). Given the medium-coverage data (average depth of coverage of 5-10X per ind.), I did not rely on called genotypes but calculated genotype likelihoods for conducting genomic analyses. I uncovered four large, genetically distinct clusters, demonstrating isolation-by-distance within these clusters and evidence of a hybrid population. These findings challenge the current taxonomy and suggest that biotic factors are limiting gene flow over varying distances. This paper highlights the importance of whole genome data in revealing genetic population structure in seabirds and its significance for taxonomy, evolution, and conservation efforts.

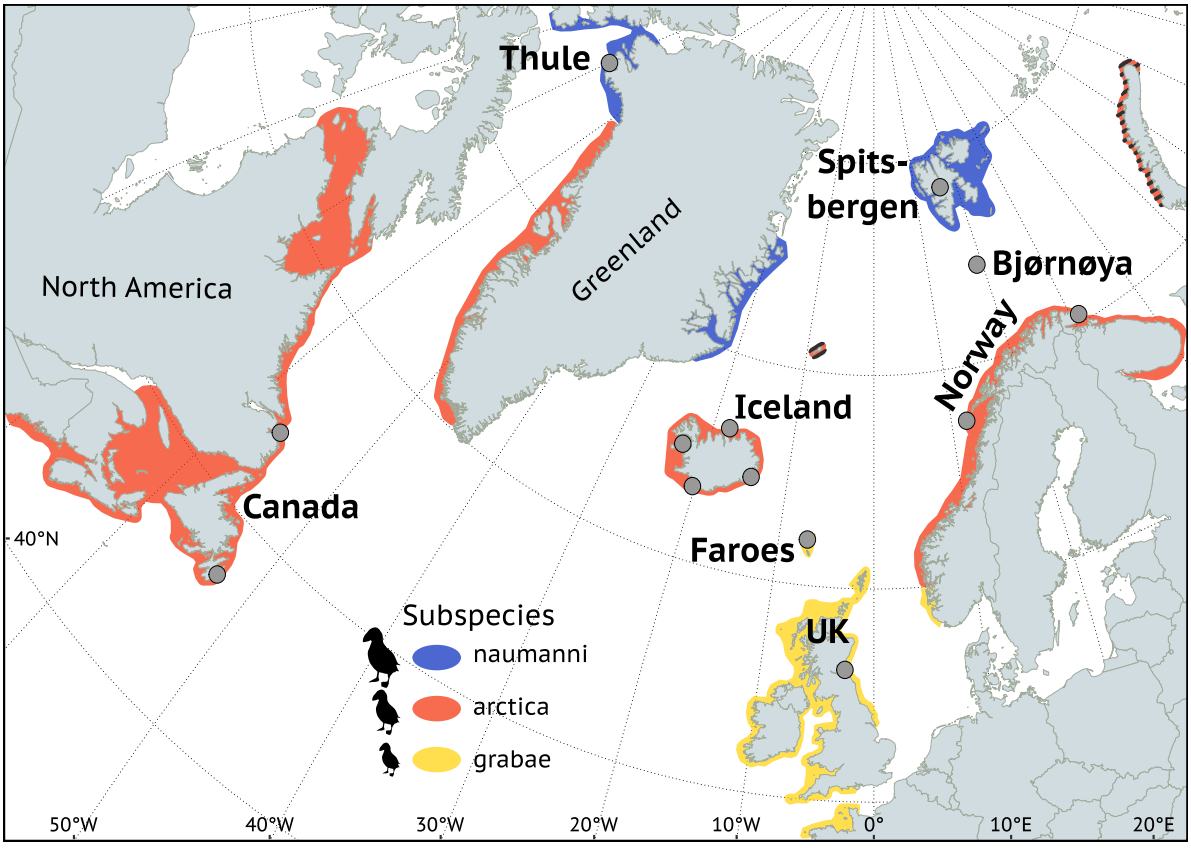
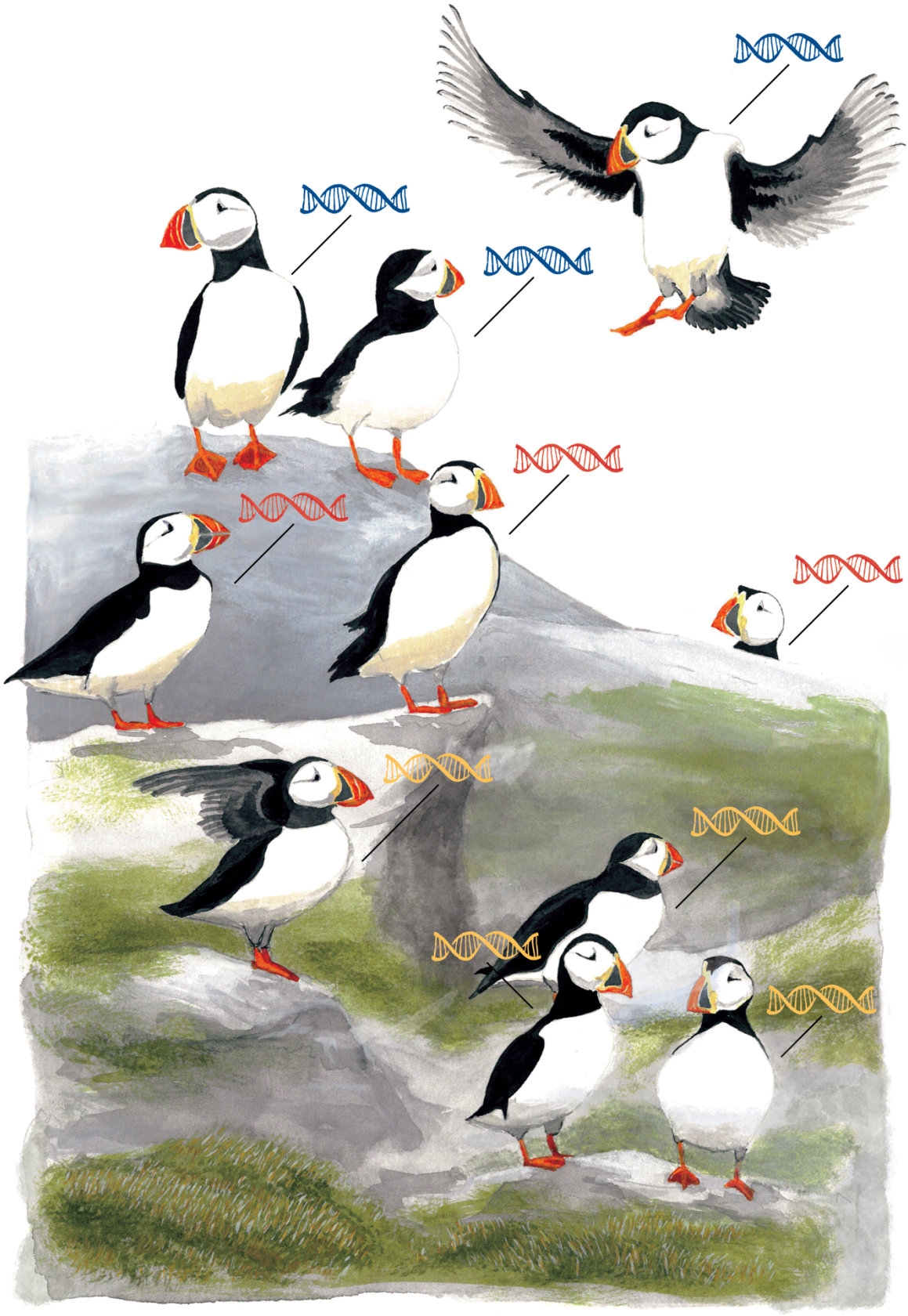


Figure 3: Breeding range of the Atlantic puffin including the 13 breeding colonies sampled for this thesis. Colonies are indicated as grey dots.

Paper II reveals an unexpected population structure at a single Atlantic puffin colony, Thule, in northwestern Greenland. I analyzed whole genome data (medium coverage) of six individuals from Thule. Although this colony comprises two discrete size phenotypes of Atlantic puffins, I found that Thule harbors individuals from three distinct clusters; a resident High Arctic cluster, as well as individuals from West and East Atlantic temperate clusters. Interestingly, no signs of recent interbreeding were visible in the sampled Thule puffins. Considering the population structure identified in Paper I, these findings suggest the beginnings of a potential northward shift of temperate puffins in the West Atlantic, consistent with responses to a warming climate.

Paper III sheds light on the origin of the hybrid population of Atlantic puffins on Bjørnøya. I assembled and annotated a chromosome-level reference genome using PacBio, 10x Genomics and Hi-C data and sequenced whole genomes of 18 contemporary individuals of the two parental populations and one hybrid population to an average depth of coverage of 20X. Additionally, I sequenced 22 historical specimens (from 1860 to 1910) from these three breeding colonies to an average depth of coverage of 5-10X. I estimated the timing of the onset of admixture and the direction of gene flow that led to the formation of the hybrid population on Bjørnøya by using the length of genomic tracts in modern hybrid individuals originating from one of the two parental populations, by generating demographic histories of the parental populations and by placing the historical individuals into the contemporary genomic population structure. The results of this paper show that the origin of the hybrid population falls within the last 100 years, coinciding with the rapid 20th century climate change in the Arctic, and is accompanied by substantial losses of genetic variation within the parental populations. These results highlight the power and importance of temporal genomics to assess the potential impact of rapid ecological changes on fragile ecosystems worldwide.

Paper IV characterizes intraspecific genomic variation using single nucleotide polymorphisms (SNPs), structural variants (SVs) and short tandem repeats (STRs) to reveal patterns of genomic subspecies differentiation and adaptation across two Atlantic puffin subspecies. I used the genomes of the 18 contemporary individuals of the two parental populations and one hybrid population sequenced to an average depth of coverage of 20X (Paper III). The parental populations are representative of the two different subspecies, *F. a. arctica* and *F. a. naumanni*. I applied state-of-the-art bioinformatics pipelines to detect and genotype SVs and STRs and subsequently identified outlier SNPs, SVs and STRs between the two subspecies. Genes falling within or in close proximity to outlier SNPs, SVs or STRs were run through a gene ontology analysis, manually inspected and placed into a biological context. This study revealed several genomic outliers near genes linked to phenotypic differences between subspecies, such as body size, skeletal development, and fat storage. Outliers also included loci related to the olfactory and visual systems, exposing previously unknown, potentially adaptive physiological differences between the subspecies. These findings are critical for assessing local adaptation and will aid conservation efforts aimed at preserving genetic diversity in the Atlantic puffin.



Paper I

Complex population structure of the Atlantic puffin revealed by whole genome analyses

communications biology

ARTICLE

https://doi.org/10.1038/s42003-021-02415-4 OPEN

Complex population structure of the Atlantic puffin revealed by whole genome analyses

Oliver Kersten[1✉], Bastiaan Star[1], Deborah M. Leigh[2], Tycho Anker-Nilssen[3], Hallvard Strøm[4], Jóhannis Danielsen[5], Sébastien Descamps[4], Kjell E. Erikstad[6,7], Michelle G. Fitzsimmons[8], Jérôme Fort[9], Erpur S. Hansen[10], Mike P. Harris[11], Martin Irestedt[12], Oddmund Kleven[3], Mark L. Mallory[13], Kjetill S. Jakobsen[1] & Sanne Boessenkool[1✉]

The factors underlying gene flow and genomic population structure in vagile seabirds are notoriously difficult to understand due to their complex ecology with diverse dispersal barriers and extensive periods at sea. Yet, such understanding is vital for conservation management of seabirds that are globally declining at alarming rates. Here, we elucidate the population structure of the Atlantic puffin (*Fratercula arctica*) by assembling its reference genome and analyzing genome-wide resequencing data of 72 individuals from 12 colonies. We identify four large, genetically distinct clusters, observe isolation-by-distance between colonies within these clusters, and obtain evidence for a secondary contact zone. These observations disagree with the current taxonomy, and show that a complex set of contemporary biotic factors impede gene flow over different spatial scales. Our results highlight the power of whole genome data to reveal unexpected population structure in vagile marine seabirds and its value for seabird taxonomy, evolution and conservation.

[1] Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway. [2] WSL Swiss Federal Research Institute, Birmensdorf, Switzerland. [3] Norwegian Institute for Nature Research (NINA), Trondheim, Norway. [4] Norwegian Polar Institute, Fram Centre, Langnes, Tromsø, Norway. [5] Faroe Marine Research Institute (FAMRI), Tórshavn, Faroe Islands. [6] Norwegian Institute for Nature Research (NINA), Fram Centre, Langnes, Tromsø, Norway. [7] Centre for Biodiversity Dynamics (CBD), Norwegian University of Science and Technology (NTNU), Trondheim, Norway. [8] Environment and Climate Change Canada, Newfoundland and Labrador, Canada. [9] Littoral, Environment et Sociétés (LIENSs), UMR 7266 CNRS—La Rochelle Université, La Rochelle, France. [10] South Iceland Nature Research Centre, Ægisgata 2, Vestmannaeyjar, Iceland. [11] UK Centre for Ecology & Hydrology, Penicuik, Midlothian, UK. [12] Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden. [13] Department of Biology, Acadia University, Wolfville, Nova Scotia, Canada. ✉email: oliver.kersten@ibv.uio.no; sanne.boessenkool@ibv.uio.no

Seabirds are important ecosystem indicators and drivers[1–3], and have long had an integral place in human culture and economy[4–6]. Nevertheless, global seabird numbers have deteriorated by an alarming 70% since the mid-20th century[7,8]. These declines pose a serious threat to marine ecosystems, human society, and culture[7,9,10], highlighting the importance of seabird conservation management. Within such management, the identification of distinct population units, i.e., demographically independent populations with restricted gene flow among them[11,12], is a fundamental first step towards optimized conservation[11,13,14]. Defining such units is, however, difficult for many seabirds because of their complex ecology[15]. Detailed genomic data including thousands of loci provide new possibilities to assess levels of connectivity and gene flow between distinct breeding populations and, thus, help identify relevant conservation units for seabirds[15,16]. Indeed, a few recent publications using reduced genomic representation approaches (e.g., RAD-seq) have reported fine-scale structure over various spatial scales[17–21]. These studies highlight the great potential of genomic data to disentangle barriers to gene flow that would otherwise remain undetected, but remain nonetheless limited due to incomplete sampling of the genome[22].

The Atlantic puffin (*Fratercula arctica*, Linnaeus, 1789, hereafter "puffin") is an iconic seabird species, prevalent in popular culture[23], important for tourism[24,25], and inherently valuable for the marine ecosystem[1]. Puffins were historically widely harvested for their meat and down[6,26,27] and exploitation remains an important cultural tradition in Iceland and the Faroe Islands[6,24]. Its breeding range stretches from the Arctic coast and islands of European Russia, Norway, Greenland, and Canada, southward to France and the USA[28] (Fig. 1a). Puffins have been designated as "vulnerable" to extinction globally and listed as "endangered" in Europe[29]. Notably, the once world's largest puffin colony (Røst, Norway) has experienced complete fledging failure during nine of the last 13 seasons and has lost nearly 80% of its breeding pairs during the last 40 years[29–31]. Similarly, Icelandic and Faroese puffins have experienced low productivity and negative population growth since 2003[32].

Puffins have been broadly classified into three taxonomic groups along a latitudinal gradient based on size, with the *smallest* puffins found around France, Britain, Ireland and southern Norway (*F. a. grabae*), *intermediate* sized puffins around Norway, Iceland, and Canada (*F. a. arctica*) and the *largest* puffins found in the High Arctic, e.g. Spitsbergen[33], Greenland[34], and northeastern Canada[35] (*F. a. naumanni*)[36] (Fig. 1a). Nevertheless, this broad classification into three subspecies has been controversial[28,37,38] and the population structure of puffins remains unresolved at all spatial scales[37]. This knowledge gap obstructs efforts towards an assessment of dispersal barriers, limits our understanding of cause-and-effect dynamics between population trends, ecology and the marine ecosystem, and hinders the development of adapted large-scale conservation actions.

Here, we present the, to the best of our knowledge, first whole-genome analysis of structure, gene flow, and taxonomy of a pelagic, North Atlantic seabird. We generated a de novo draft assembly for the puffin and resequenced 72 individuals across 12 colonies representing the majority of the species' breeding range (Fig. 1a). Our work suggests that a complex interplay of ecological factors contributes to the range-wide genomic population structure of this vagile seabird.

Results

Genome assembly and population sequencing. Based on synteny with the razorbill (*Alca torda*), a total of 13,328 puffin scaffolds were placed into 26 pseudo-chromosomes, leaving 17.06 Mbp (1.4%) unplaced and yielding an assembly of 1.294 Gbp (Supplementary Data 1, Table S1). This assembly contains 4,522 of the 4,915 genes (92.0%) of complete protein-coding sequences from the avian set of the OrthoDB v9 database (Supplementary Data 1). We also assembled the puffin mitogenome (length of 17,084 bp) with a similar arrangement of genomic elements as other members within the Alcidae[39,40] (Fig. S1, Table S2). For the 72 resequenced specimens, we analyzed a total of 5.77 billion paired reads, obtaining an average fold-coverage of 7X (range 3.0–10) for the nuclear genome and 591X (5.3–1800) for the mitochondrial genome per specimen (Fig. 1a, Supplementary Data 2). One individual (IOM001) was removed from both datasets (nuclear and mitochondrial) due to a substantially lower number of mapped reads (endogeny) relative to all other samples (Supplementary Data 2) resulting in a large proportion of missing sites (Fig. S2). Additional filtering produced a final genotype likelihood dataset of 1,093,765 polymorphic nuclear sites and 192 mitochondrial single-nucleotide polymorphisms (SNPs, Supplementary Data 3) in 71 birds (36 males and 35 females).

Genomic population structure. Genomic variation across 71 puffin mitogenomes defines 66 polymorphic haplotypes that indicate a recent global population expansion and show no significant population structure (Fig. 1b, Figs. S3, S4, Tables S3, S4). In contrast, we inferred four main population clusters using principal component analysis (PCA) of the nuclear whole-genome dataset (Fig. 1c). Puffins from Spitsbergen are most distinct, while puffins from Bjørnøya are located between Spitsbergen and a larger, central cluster consisting of populations from Norway, Iceland, and the Faroe Islands (Fig. 1c, Fig. S5a). Puffins from Canada form their own distinct cluster, as do those from the Isle of May, southeast Scotland (Fig. 1c, Fig. S5b). Hierarchical PCA analyses of the cluster comprising the mainland Norwegian, Icelandic and Faroese colonies reveal further fine-scale structure separating Norwegian (Hornøya and Røst) and Faroese/Icelandic colonies (Fig. S5c). Model-based clustering (ngsAdmix) agrees with the results from the PCA (Fig. 1d). The optimal model fit for the entire dataset is either $K = 2$ or $K = 4$ (Fig. S6a), as determined by the method of Evanno et al.[41]. At $K = 2$, ngsAdmix separates Spitsbergen from the other colonies, with Bjørnøya being admixed (following separation along PCA 1), whereas at $K = 4$, ngsAdmix reflects the structure of three additional distinct clusters representing Spitsbergen, Canada, the Isle of May, and a central group with more shared ancestry (Fig. 1d). The shared ancestry of the central group remains present in hierarchical admixture analyses excluding Spitsbergen and Bjørnøya individuals (Figs. S6b, S7). We find no fixed alleles and pairwise F_{ST} values between colonies and genomic clusters are low (<0.01) (Table S4), apart from any comparisons involving the Spitsbergen population, which show substantially higher F_{ST} values (0.03–0.08).

Phylogenetic reconstructions using individual-based Neighbor-Joining (NJ) and maximum likelihood (ML) methods (Fig. 2a, Fig. S8), as well as population-based analyses in Treemix (Fig. 2b), support the distinctiveness of the Spitsbergen, Canada, and the Isle of May puffins with each group forming monophyletic clades with 100% bootstrap support. In contrast, Bjørnøya forms a paraphyletic clade between Spitsbergen and northern Norway (Fig. 2a). The population clusters identified by the PCA and ngsAdmix at smaller spatial scales are also identified in the topologies of the NJ and ML trees, sorting individuals predominantly according to geographical location, although with low bootstrap support (>80) due to large inter-individual variability (Fig. 2a, Fig. S7). Allowing a single migration edge in the Treemix phylogeny identifies recent gene flow from

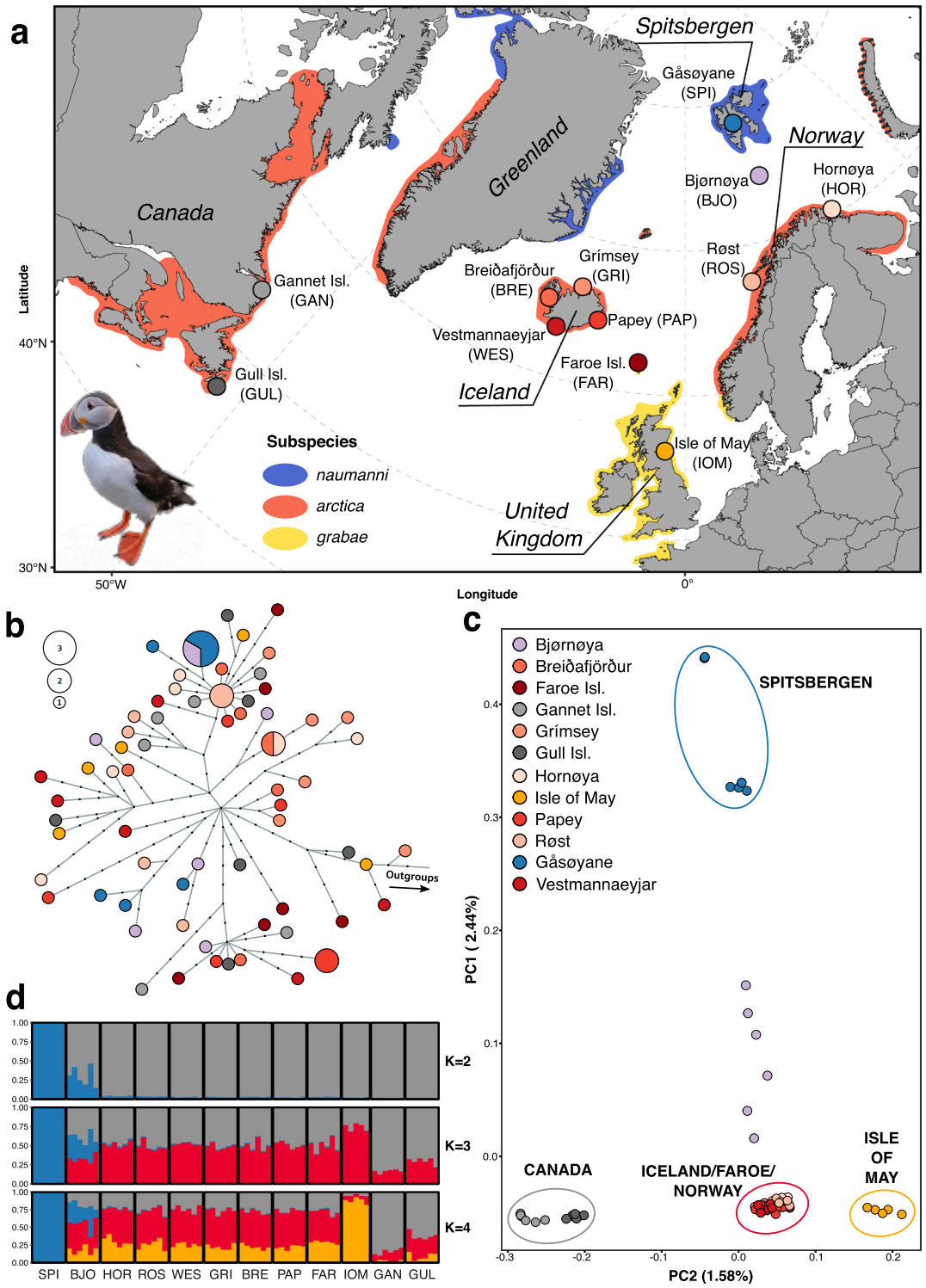


Fig. 1 Sampling distribution and genomic structure of 71 Atlantic puffin individuals across 12 colonies throughout the breeding range. a Map presenting the location of the 12 sampling sites. Color shading indicates the breeding range of the species as a whole, as well as the recognized subspecies. **b** Mitochondrial haplotype network based on a maximum likelihood tree generated with IQTree and visualized using Fitchi. It contains 66 unique haplotypes identified by 192 mitogenome-wide SNPs. Sizes of circles are proportional to haplotype abundance. Color legend is provided in (**c**). Black dots represent inferred haplotypes that were not found in the present sampling. **c** Principal component analysis (PCA) using genotype likelihoods at 1,093,765 polymorphic nuclear sites calculated in ANGSD to project the 71 individuals onto PC axes 1 and 2. Each circle represents a sample and colors indicate the different colonies. The percentage indicates the proportion of genomic variation explained by each axis. The color coding of the colonies is consistently used throughout the manuscript. **d** CLUMPAK-averaged admixture plots of the best K's using the same genotype likelihood panel as in (**c**). Each column represents a sample and colonies are separated by solid white lines. Optimal K's were determined by the method of Evanno et al.[41] (see Fig. S6a) and colors indicate the ancestry fraction to the different clusters. The dataset(s) needed to create this figure can be found at https://doi.org/10.6084/m9.figshare.14743242.v1.

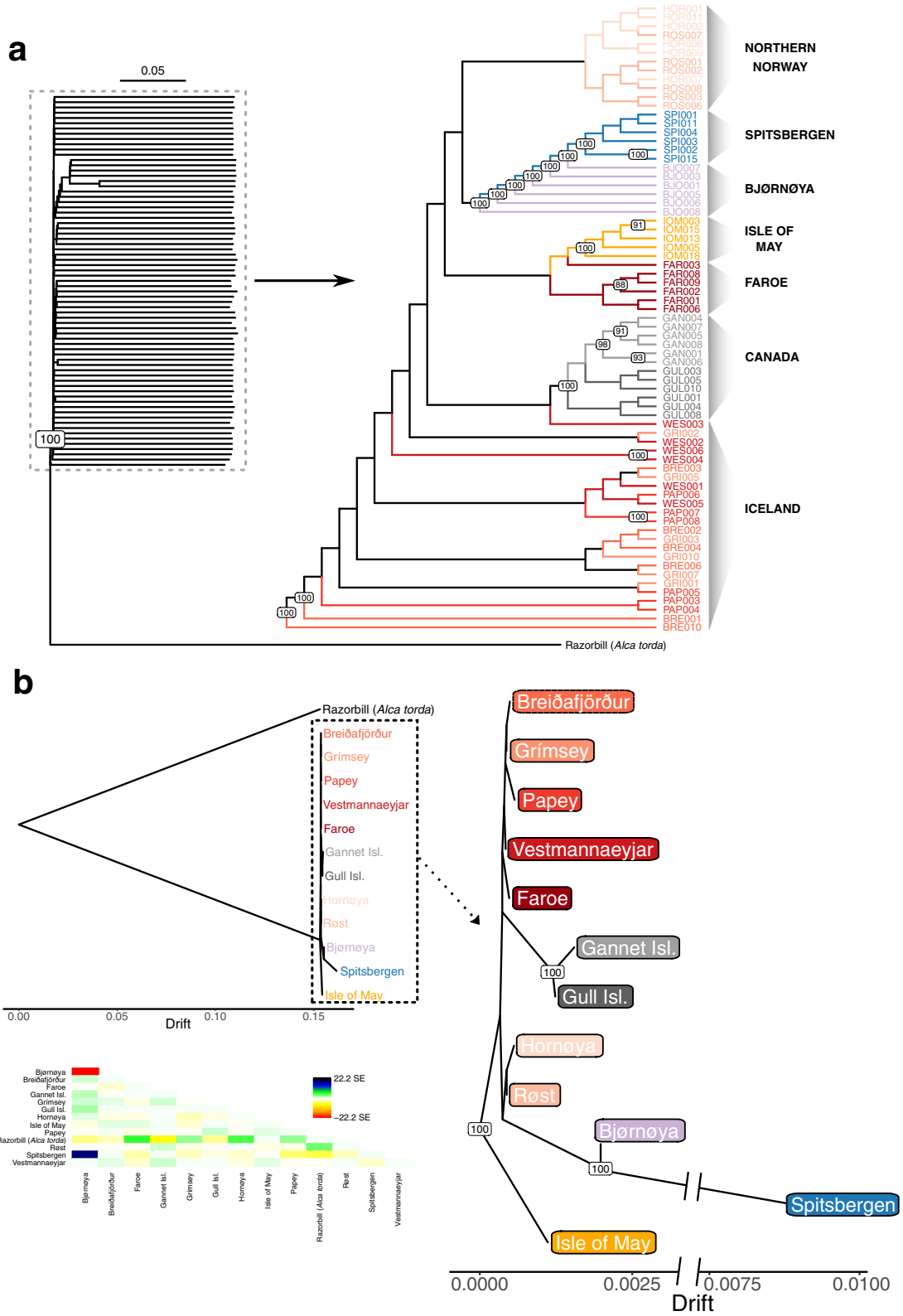


Fig. 2 Phylogenetic reconstruction of individual and colony relationships from 71 Atlantic puffin individuals sampled across 12 colonies throughout the species' breeding range. a An individual-based neighbor-joining tree constructed using pairwise p-distances calculated from genotype likelihoods at 1,093,765 polymorphic nuclear sites. Branch lengths and the outgroup were removed for the zoomed-in section to improve visualization. **b** A population-based maximum likelihood Treemix analysis using allele frequencies at the same 1,093,765 polymorphic nuclear sites as in (**a**). Both trees are rooted using the razorbill as an outgroup. The tree in (**b**) is visualized with and without the outgroup. Branch lengths are equivalent to a genetic drift parameter. The heatmap indicates the residual fit of the tree displaying the standard error of the covariance between populations. In (**a**) and (**b**), the color coding of the colonies is consistent with those in Fig. 1 and node labels show bootstrap support >80. The dataset(s) needed to create this Figure can be found at https://doi.org/10.6084/m9.figshare.14743299.v1.

Spitsbergen to Bjørnøya (likelihood = 792.106; Figs. S9, S10a). Adding additional migration edges to the population-based ML tree does not improve the model fit and such edges are therefore not further interpreted (Figs. S9-S11).

Genetic diversity, heterozygosity, and inbreeding. Tajima's D does not significantly deviate from neutral expectation per colony (Table S3). Nucleotide diversity (π) of puffins is significantly different between colonies, with the Spitsbergen population having significantly lower nucleotide diversity than the global median (Wilcoxon Rank Sum test, $U = 4824$, $n_{SPI} = 25$, $n_{Global} = 300$, $P = 0.017$, Table S3). Colonies also differ significantly in levels of heterozygosity (Kruskal–Wallis test, $n = 12$, $P = 1 \times 10^{-6}$; Fig. 3a) and inbreeding (Kruskal–Wallis test, $n = 12$, $P = 1 \times 10^{-7}$, Fig. 3b), whereby individual inbreeding (F_{RoH}) was approximated based on runs of homozygosity (RoH)[42]. Again, the Spitsbergen colony has significantly lower levels of heterozygosity (0.00220–0.00223) and significantly higher levels of F_{RoH} values (0.161–0.172), compared to the Faroese and Icelandic colonies (Dunn test with Holm correction, $P < 0.05$, $n_1 = 6$, $n_2 = 6$). The Faroese and Icelandic colonies contain the highest levels of heterozygosity and lowest F_{RoH} values (Figs. 3a, b, Fig. S12) overall. The remaining colonies display intermediate levels (Fig. 3a, b), although heterozygosity is significantly lower (Fig. 3a, Fig. S12) and inbreeding is significantly higher (Fig. 3b, Fig. S12) on Gull Island and Bjørnøya compared to the Icelandic and Faroese colonies (Dunn test with Holm correction, $P < 0.05$, $n_1 = 6$, $n_2 = 6$). Moreover, Spitsbergen harbors the most (an average of 718 per individual) and longest RoHs with eight being ≥2.3 Mbp long (4.21 ± 3.02% of respective chromosome), whereas none of the RoHs in the remaining colonies are >2.15 Mbp long (Fig. 3c). The only exception is a 9.65 Mbp long RoH on pseudo-chromosome 7 (18% of chromosome length) in an Isle of May individual (Fig. 3c).

Patterns of gene flow and isolation-by-distance (IBD). We investigated patterns of gene flow and IBD between the colonies using two-dimensional estimated effective migration surface (EEMS) analyses[43]. Levels of gene flow between the Icelandic and Faroese colonies and within the Canadian group is high (3–10× higher than the global average), while intermediate between the Norwegian mainland colonies (around the global average). In contrast, the Spitsbergen colony is split from the remaining colonies by migration rates up to 100× lower than the global average (Fig. 4a, Fig. S13), while additional regions of low gene flow (2–3× lower than the global average) separate the Isle of May, Canadian, and Bjørnøya colonies from the rest (Fig. 4a, Fig. S13). Geographic distance between all puffin colonies is a poor predictor of pairwise genetic distance, driven by high Slatkin's linearized F_{ST} values between Spitsbergen and the other colonies (Tables S5, S6, Fig. S14). Nevertheless, the geographic distance among a subset of puffin colonies is significantly associated with genetic distance as shown by Mantel tests, linear regression model analyses, and distance-based Redundancy Analysis (dbRDA) models (Fig. 4b, Fig. S14, Tables S5, S6). Specifically, by progressively removing the more distant colonies (Spitsbergen, Isle of May, Bjørnøya, Canada), which are characterized by high Slatkin's linearized F_{ST} values at relatively small geographic distances (Fig. S14), the fit of a linear IBD model is significantly improved and the proportion of variance of genetic dissimilarity explained by geographic distance is more than doubled (Spitsbergen removed: 37.58%; Spitsbergen/Isle of May/Bjørnøya/Gannet Isl. removed: 84.98%) (Fig. 4b, Fig. S14, Table S5). Similarly, the proportion of explained genetic variance by spatial features estimated in global dbRDA models is more than tripled (All colonies = 18.76%, Spitsbergen/Isle of May/Bjørnøya removed = 59.87%) (Table S5). In all optimized dbRDA models, geographic variables (IBD) contribute significantly to the genetic divergence, while the contribution of the mean sea surface temperature (isolation-by-environment, IBE) is minimal. IBE is only once significantly contributing to the observed genetic variance (when Spitsbergen was removed), yet accounts for less than half of the observed genetic variance (11.37%) compared to the geographic distance (28.66%) (Table S6).

Admixture on Bjørnøya. We specifically tested for patterns of admixture in Bjørnøya. Significantly negative *f*3 statistics (Z score < −3) are found for all unique combinations of the phylogeny (Spitsbergen, *X*; Bjørnøya) (Table S7), indicating an admixed colony on Bjørnøya caused by gene flow between Spitsbergen and the remaining colonies. Similarly, significantly positive D-statistics (Z score > 3) caused by an excess of ABBA sites reveal excessive allele sharing between Spitsbergen and Bjørnøya (Fig. S15a). The close association and gene flow from Spitsbergen to Bjørnøya is further confirmed by D-statistics not being significantly different from 0 for the (((Bjørnøya, Spitsbergen), H3), Razorbill) topology (Fig. S15b).

Genetic differentiation. We assessed genome-wide patterns of genetic differentiation by calculating pairwise F_{ST} between the four genomic clusters in 50 kb sliding windows. These analyses show that the differentiation between the clusters is driven by increased F_{ST} in windows across the entire genome, including the presence of several smaller regions with elevated F_{ST} (Fig. S16). Several of these elevated F_{ST} regions are present in all pairwise comparisons (Fig. S16), whereas others are specific for certain comparisons, and may be indicative of local adaptation (Fig. S16).

Discussion

Barriers to gene flow leading to population structure are notoriously difficult to identify and remain largely unknown for most seabirds[15,44]. Using whole-genome analyses, we here provide insights into the genetic structure of the Atlantic puffin. First, we identify four main puffin population clusters consisting of (1) Spitsbergen (High Arctic), (2) Canada, (3) Isle of May, and (4) multiple colonies in Iceland, the Faroe Islands, and Norway. Second, we find that within such clusters, genetic differentiation is driven by IBD. Finally, we find evidence for secondary contact between two clusters. These observations show that a complex set of drivers impacts gene flow over different spatial scales (100–1000s of km) between these clusters and the colonies within. In particular, the interplay between overwintering grounds, philopatry, natal dispersal, geographic distance, and potentially ocean regimes appears to explain the genomic differentiation between puffin colonies[45].

Mature puffins rarely, if ever, change their colonies, resulting in very high colony fidelity once they start breeding[28]. Immatures, however, have been observed to visit other nearby colonies during the summer and may breed in non-natal colonies[28,46]. Nevertheless, data on natal philopatry remain scarce, but existing evidence shows rates vary greatly (38–92%) between colonies[28,46]. If either breeding or natal philopatry alone drive the puffin population structure, each colony should constitute its own distinct genomic entity and substantial genomic differentiation across the puffin's entire breeding range would be observed. Yet, philopatry alone cannot explain the presence of the four large-scale population clusters we observe here. Additional factors must therefore promote the distinctiveness of the four clusters. For instance, the Isle of May birds have a largely separate overwintering distribution mainly in the North Sea (Fig. S17)[28,38,47]. Such potential

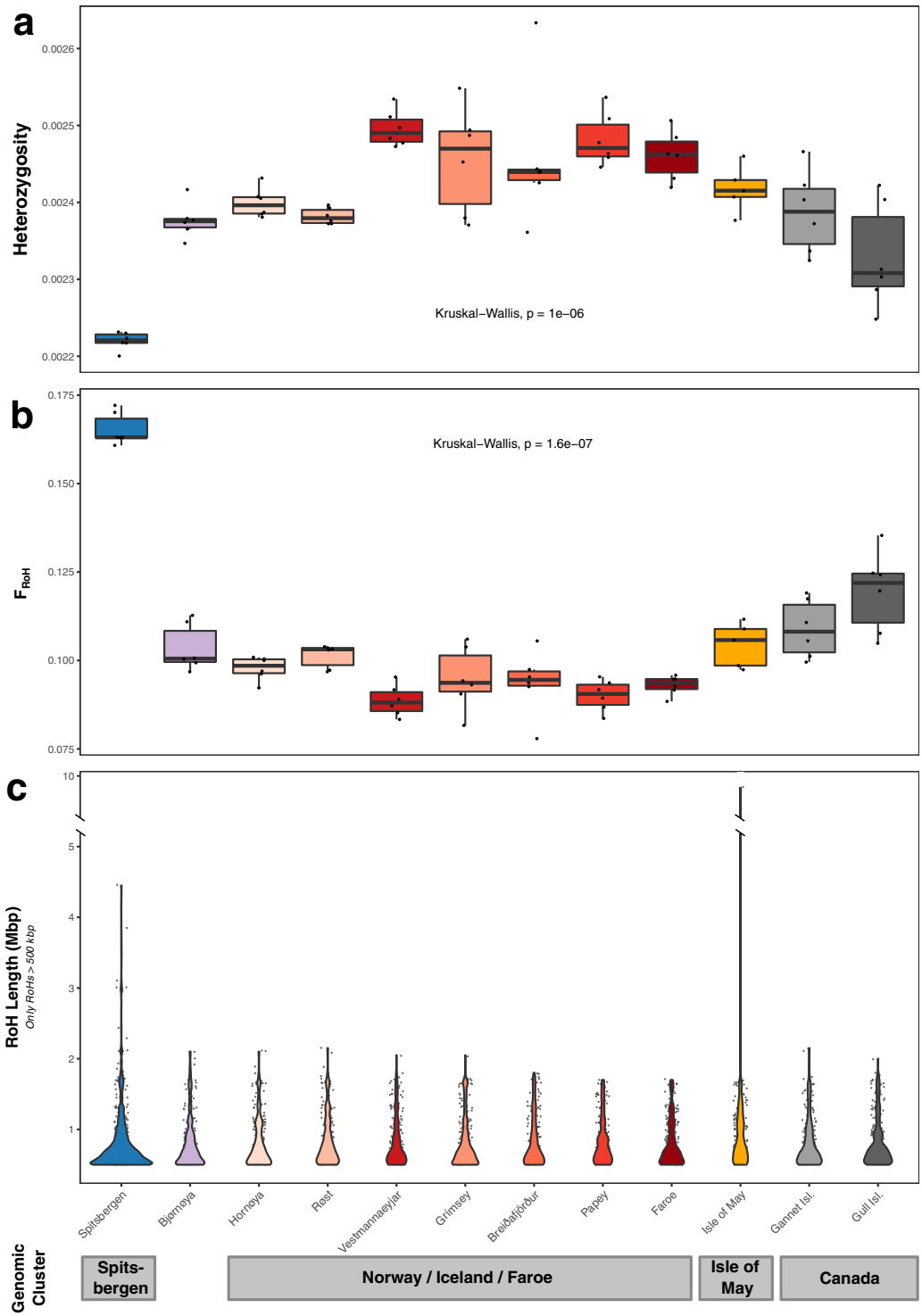


Fig. 3 Genome-wide heterozygosity, inbreeding, and Runs-of-Homozygosity (RoH) compared between 12 Atlantic puffin colonies across the species' breeding range. a Estimates of individual genome-wide heterozygosity based on the per-sample one-dimensional Site Frequency Spectrum calculated in ANGSD. **b** Individual inbreeding coefficients, F_{RoH}, defined as the fraction of the individual genomes falling into RoHs of a minimum length of 150 kb. RoHs were declared as all regions with at least two subsequent 100 kb windows harboring a heterozygosity below 1.435663×10^{-3}. **c** RoH length distribution across the 12 colonies only including RoHs longer than 500 kb. A single 9.65 Mbp long RoH on pseudo-chromosome 7 in an Isle of May individual required to introduce a break in the y-axis. In (**a**) and (**b**), black dots indicate individual sample estimates and black lines the median per colony, while in (**c**), black dots represent single RoHs. Statistical significance of differences in heterozygosity and F_{RoH} between populations was assessed with a global Kruskal-Wallis test ($n = 12$). The results of post hoc Dunn tests with Holm corrections are presented in Fig. S12. Error bars show range of values within 1.5 times the interquartile range. Different colonies in all three plots are indicated using the same color code as in Fig. 1. The dataset(s) needed to create this figure can be found at https://doi.org/10.6084/m9.figshare.14743317.v1.

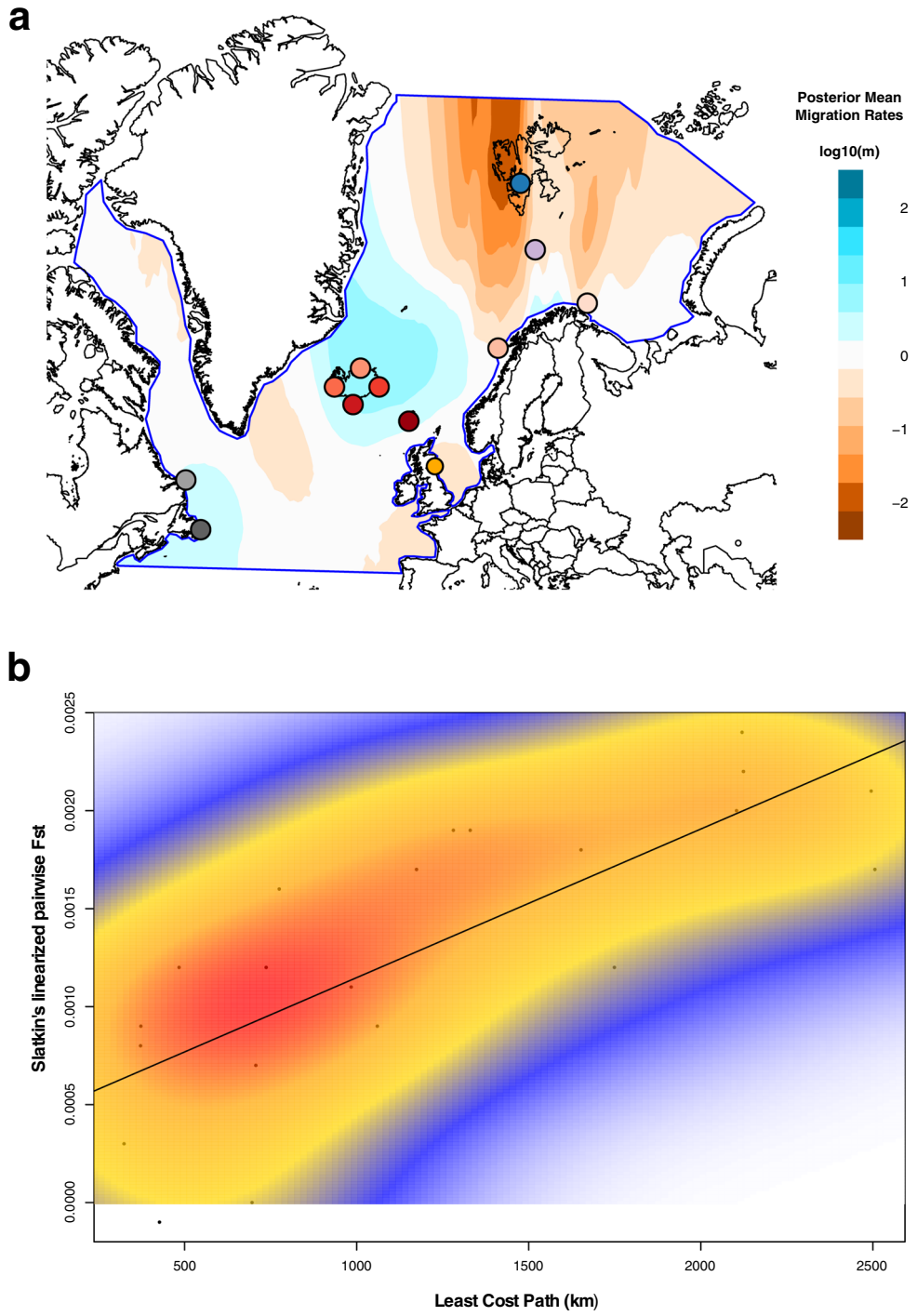


Fig. 4 Estimates of continuous long-distance gene flow and isolation by distance (IBD) across the breeding range of the Atlantic puffin estimated from 71 individuals across 12 colonies. a Effective migration surfaces inferred by the program EEMS using the average distance between pairs of individuals calculated in ANGSD by sampling the consensus base for each individual at 1,093,765 polymorphic nuclear sites. Darker reds indicate reduced migration across those areas, while darker blues highlight higher migration rates than the global mean. Different colonies are indicated using colors consistent with those in Fig. 1. **b** Correlation between genetic (Slatkin's linearized F_{ST}) and geographic (Least Cost Path—only over water) distance presented after removing the Spitsbergen, Bjørnøya, Isle of May, and Canadian individuals. The diagonal line visualizes the result of the multiple regression on distance matrices (MRM) analysis (slope and y-intercept). The Mantel test between genetic and geographic distance ($R = 0.775$, $P = 0.012$, $n_{Colonies} = 7$) was significant and 60.08% of the variance in Slatkin's linearized F_{ST} was explained by geographic distance (regression coefficient of linear IBD model = 0.76×10^{-6}, $P = 0.006$, $n_{Colonies} = 7$). A two-dimensional kernel density estimation (kde2d) highlights dense groups of data points, thus substructure in the genomic landscape pattern. Analyses were conducted and results visualized in R using the *ecodist*, *marmap* and *MASS* packages. The dataset(s) needed to create this figure can be found at https://doi.org/10.6084/m9.figshare.14743323.

geographical separation during the winter season might limit the likelihood of immatures intermixing between the Isle of May and other colonies. Similarly, distinct overwintering distributions have been found to lead to increased genetic diversification in other philopatric seabird species[15,44,45], such as the thick-billed murre (*Uria lomvia*)[21] and black-browed albatross (*Thalassarche melanophris*)[48]. The presence of a Canadian cluster can also be largely explained by their winter distribution around Newfoundland[47,49]. There is, however, some fragmentary overlap in the overwintering distribution of the Canadian and Icelandic colonies off southwestern Greenland[47,49], suggesting that barriers to dispersal of immatures and gene flow in the western Atlantic may be further enforced by the large geographic distance. In contrast, the winter distribution from the colonies in Iceland, Norway, and the Faroe Islands overlaps off the coast of southern Greenland (Fig. S17)[47]. This shared overwintering area, combined with the tendency to return to the natal colony and immature visits to nearby (up to 100 s km) colonies during the summer, appears to drive a pattern of IBD among colonies (Fig. 3b). Indeed, IBD has previously been recognized as an important driver of genomic structure in seabirds, for instance in the little auk (*Alle alle*)[50] and band-rumped storm-petrel (*Oceanodroma castro*)[51]. While these illustrated mechanisms provide reasonable explanations for the observed dispersal barriers and population structure based on our current knowledge, validation requires additional evidence, specifically on the winter distribution of immature puffins and natal dispersal rates across colonies covering the entirety of the puffin's breeding range.

High Arctic puffins from Spitsbergen are genetically the most divergent group within our dataset harboring the highest genome-wide differentiation. They are also characterized by significantly lower levels of genetic diversity, greater inbreeding coefficients, and longer and more abundant RoHs compared to other colonies. These observations may either result from a historical bottleneck followed by isolation (e.g., founder effect), local adaptation to their extreme environment, or generally lower effective population sizes. Population abundance estimates of <10,000 breeding pairs on Spitsbergen compared to 500,000 in the West Atlantic, two million on Iceland and more than two million in the boreal East Atlantic potentially indicate a lower effective population size[28]. The High Arctic puffins exclusively inhabit harsh, cold-current environments year-round, as they likely stay in an area bounded by the East Greenland ice edge, a latitudinal border at 70° N, and the front between the Barents and Greenland Sea during winter (Fig. S17). They are also substantially larger than birds from lower latitudes[28,33,34], following Bergmann's[52] or James's[53] rule, as has been observed in other seabirds[54,55]. This matches the clinal size variation of puffins that closely tracks sea temperatures in their breeding areas[56]. Despite these distinctions, we find that the relatively small population of puffins on Bjørnøya (<1000 pairs[28]), midway between Spitsbergen and mainland Norway, represents an area of secondary contact between the puffins from the High Arctic and other puffin colonies. Based on D- and the *f*3-statistics, the most likely southern sources are Iceland, the Faroe Islands, Norway, or a combination thereof. Thus, the barriers to gene flow that keep the Spitsbergen colonies distinct do not prevent the formation of a hybrid colony where individuals from the High Arctic and the cluster composed of mainland Norwegian, Icelandic and Faroese colonies meet.

The distinct population structure in the nuclear data is not observed in the mitochondrial genomes, which reveal an abundance of rare alleles and lack of significant population differentiation. The mitogenomic variation suggests that puffins experienced a recent population expansion, possibly out of a refugium after the Last Glacial Maximum. Indeed, it has been shown that mitogenomic variation in seabirds is dominated by historical factors rather than representing contemporary gene flow[44], and a lack of mitogenomic population structure has been observed in many marine birds with high philopatry[50,57,58]. In contrast to the mitogenomes, the structure in the nuclear data therefore likely originated after the last glacial period and reflects the influence of relatively recent barriers to gene flow in a context of historical demography[15,44]. Such results are relevant for understanding the "seabird paradox", which contrasts the life-history traits of high philopatry and restricted dispersal in otherwise highly mobile species[59].

Our results have major implications for the conservation management of the Atlantic puffin. The genetic structure we identify in puffins disagrees with the suggestion of three subspecies (*F. a. naumanni, F. a. arctica, F. a. grabae*)[36]. Although the genetically distinct Spitsbergen cluster coincides with the classification of morphologically large puffins in the High Arctic (*F. a. naumanni*)[28], we observe gene flow from Spitsbergen into Bjørnøya, which has been considered *F. a. arctica*[28]. Furthermore, the geographic divide between *F. a. grabae* and *F. a. arctica* lies farther south than previously thought, with the Faroese puffins being genetically closer to *F. a. arctica* than to *F. a. grabae*. Nonetheless, *F. a. grabae* is currently represented by a single colony (Isle of May) in our study and the geographical extent of this genomic cluster needs to be refined by additional sampling, particularly in the western UK, Ireland, and France. Finally, puffins from the Western Atlantic region (e.g., colonies in Canada) form their own distinct genetic cluster that is not recognized within the current classification. Our results do not only warrant a revision of Salomonsen's taxonomic classification of three subspecies[36], but also highlight the need to acknowledge the four identified clusters as distinct units within the conservation management of puffins[11,13,14]. Although puffin colonies within clusters are not genetically distinct entities, ecological independence illustrated by contrasting population dynamics across relatively small spatial scales (e.g., western Norway[31]) suggests that higher resolution local management units based on ecological differences should be considered. Nonetheless, the genetically distinct clusters at the outer edges of the puffin's distribution with putative local adaptations that will not be easily replenished indicate that conservation of these distinct clusters must be a first priority. Finally, our sampling does not cover several outskirts of the puffin's distribution, such as the U.S., northern Canada, Greenland, Ireland, western UK, France or Russia, and we may therefore still underestimate the true biological and genetic complexity of this species.

In conclusion, our study shows that a complex interplay of barriers to gene flow drives a previously unrecognized population diversification in the iconic Atlantic puffin. So far, much of seabird population genetics research has been based on mitochondrial and microsatellite data[15,44], which have limited power to characterize contemporary factors that determine population structure and gene flow[20,60]. High-resolution nuclear data are therefore essential to help define evolutionary significant population units, disentangle convoluted ecological relationships, and are particularly important for seabird conservation, which aims to preserve genetic diversity considering profound global population declines[7,8], and the threat of global warming, which negatively impacts ecosystems worldwide[61].

Methods

Ethical statement. Feather and blood samples of puffins included in this study were collected and handled under the following permits.

1. Gåsøyane, Røst, Hornøya, Bjørnøya (Norway)—FOTS ID #15602 and #15603 from the Norwegian Food Safety Authority for SEATRACK and SEAPOP; Permit 2018/607 from Miljødirektoratet (Norwegian Environment Agency), dated 4 May 2018.

2. Gannet and Gull Island (Canada)—Canadian Wildlife Service Migratory Bird Banding Permit 10559 G, approved Animal Use Protocol (AUP) by Eastern Wildlife Animal Care Committee (17GR01, 18GR01), Newfoundland and Labrador Wilderness and Ecological Reserves Permit—Scientific Research (DOC/2017/02003), Canadian Wildlife Service Scientific Permit ST2785 (to M.L.M.), Canadian Wildlife Service Banding Permit 10694, and Acadia University Animal Care Committee Permits ACC 02-15 and 06-15 (to M.L.M.).
3. Isle of May (Scotland)—Scottish Natural Heritage licence 2014/MON/RP/156 and Ringing Permit A400 (to MPH).
4. Vestmannaeyjar, Papey, Breiðafjörður, Grímsey (Iceland)—Icelandic puffins were legally hunted during the hunting period of 1 July–15 August.
5. Faroe—Feathers came from predated birds collected in the field after the predator was finished with them.

Draft reference genome assembly. A de novo Atlantic puffin draft genome was generated from the blood of a female Atlantic puffin. Read data were sequenced on three Illumina HiSeqX lanes using the 10x Genomics Chromium technology and assembled with the Supernova assembler (v2.1.1)[62] after subsampling to 0.8 billion and 1 billion reads to maximize performance and remain within the computational capacity of the assembler. We refined the two assemblies through several steps, including merging of 'haplotigs', removal of contaminant sequences, misassembly correction, re-scaffolding using mapping coverage and linkage information, and gap filling (Supplementary Data 1a). The most complete and continuous 800 M and 1000 M assemblies together with the 3rd best assembly overall were selected for a second round of refinement (Supplementary Data 1b) resulting in a total of 72 draft assemblies. Of these, we kept the four most complete and continuous assemblies for additional gap filling and polishing, after which the most complete draft genome was selected for downstream analyses (Supplementary Data 1c). The puffin mitogenome was confidently identified by blasting (blastn) all scaffolds shorter than 25 kb against a custom-built database of 135 published mitogenomes of the order 'Charadriiformes' and annotated with the MITOS web server[63] (Fig. S1). The remaining nuclear scaffolds were ordered and concatenated into "pseudo-chromosomes" by mapping them to the razorbill genome (*Alca torda*—NCBI: bAlcTor1 primary, GCA_008658365.1) and applying 200 N's as padding between each scaffold. We combined unmapped scaffolds into an "unplaced" pseudo-chromosome. We assessed the order and placement of scaffolds by investigating synteny in coverage and length between the puffin and razorbill chromosomes (Table S1). Details on the draft reference genome assembly and refinement can be found in the Supplementary File.

DNA extraction and sequencing. Samples from a total of 72 puffins collected across 12 breeding colonies (Fig. 1a) were made available for the present study by SEAPOP (http://www.seapop.no/en), SEATRACK (http://www.seapop.no/en/seatrack) and ARCTOX (http://www.arctox.cnrs.fr/en/home—Canadian colonies). These samples had been collected between 2012 and 2018 and consisted of blood preserved in EtOH or lysis buffer, or feathers (Supplementary Data 2). We extracted DNA using the DNeasy Blood & Tissue kit (Qiagen) following the manufacturer's protocol for animal blood or the nail/hair/feathers protocol applying several modifications for improved lysis and DNA yield. Individuals that had no sexing data associated with them were sexed using PCR amplification of specific allosome loci and visualization via gel electrophoresis. Genomic libraries were built by the Norwegian Sequencing Centre and sequenced on an Illumina HiSeq4000. We processed sequencing reads in PALEOMIX v1.2.14[64] and split the resulting bam files into nuclear and mitochondrial bam files. Additional details on the DNA extraction, sexing, sequencing and mapping are listed in the Supplementary File.

Mitogenome analyses. Genotypes across the mitochondrial genome were jointly called with GATK v4.1.4[65] by using the *HaplotypeCaller*, *CombineGVCFs*, and *GenotypeGVCFs* tool. We filtered genotypes according to GATKs Best Practices[66] and set genotypes with a read depth <3 or a quality <15 as missing. Indels and non-biallelic SNPs were removed and only SNPs present in all individuals were kept for subsequent analyses. The SNP dataset was annotated (Supplementary Data 3) with snpEff[67] utilizing the annotation of the newly assembled mitogenome of the Atlantic puffin and converted into a mitogenome sequence alignment. To serve as an outgroup, we appended four other species of the family Alcidae, i.e., the Razorbill (*Alca torda*, NCBI: CM018102.1), the Crested Auklet (*Aethia cristatella*, NCBI: NC_045517.1), the Ancient Murrelet (*Synthliboramphus antiquus*, NCBI: NC_007978.1) and the Japanese Murrelet (*Synthliboramphus wumizusume*, NCBI: NC_029328.1), to the alignment. To construct a maximum-likelihood phylogenetic tree, we split the alignment into seven partitions, i.e., one partition for a concatenated alignment of each of the three codon positions of the protein-coding genes, one partition for the concatenated alignment of the rRNA regions, one partition for the concatenated alignment of the tRNAs, one partition for the alignment of the control region, and one partition for the concatenated alignment of the "intergenic" regions. The best-fitting evolutionary model for each partition was found by *ModelFinder*[68] and the tree was built with IQTree v1.6.12[69] using 1000 ultrafast bootstrap replicates. We used the resulting tree to draw a haplotype genealogy graph with Fitchi[70]. Using Arlequin v.3.5[71], we calculated haplotype (h), nucleotide diversity (π), and Tajima's D[72] for each colony, for each genomic cluster defined by the nuclear analysis, and globally. In addition, an Ewens–Watterson test[73], Chakraborty's test of population amalgamation[74], and Fu's F_s test[75] were conducted for each of those groups. To further identify population differentiation, the proportion of sequence variation (Φ_{ST}) was estimated for all pairs of populations and genomic clusters. Hierarchical AMOVA tests subsequently determined the significance of a priori subdivisions into colonies and genomic clusters. Calculation of Φ_{ST} and AMOVA tests were also conducted in Arlequin. Additional details on the mitochondrial analyses are given in the Supplementary File.

Nuclear genome clustering and phylogenetic analyses. The majority of population genomic analyses were based on nuclear genotype likelihoods as implemented in ANGSD v.0.931[76]. After assessing the quality of the mapped sequencing data in an ANGSD pre-run, we removed an individual from the Isle of May from the dataset. Genotype likelihoods for nuclear SNPs covered in all individuals were calculated and filtered in ANGSD. Accounting for linkage disequilibrium, we further pruned the dataset by only selecting the most central site within blocks of linked sites ($R^2 > 0.2$) as in Orlando and Librado[77]. Subsequently, all variants located on the Z-pseudo-chromosome and "unplaced scaffolds" were excluded from the analyses yielding a final genotype likelihood panel consisting of 1,093,765 sites. We investigated genomic population structure with a PCA of the genotype likelihood panel using PCAngsd v0.982[78]. Individual ancestry proportions were estimated using a maximum likelihood (ML) approach implemented in ngsAdmix v32[79], with the number of ancestral populations (K) set from 1 to 10 and conducting 50 replicate runs for each K. The runs were clustered after similarity for each K and ancestry proportions were averaged within the major cluster using Clumpak[80] with default settings. Additional "hierarchical" PCA and admixture analyses were conducted for genomic sub-cluster(s) using identical methods.

After adding the razorbill genome as an outgroup to the genotype likelihood panel by mapping unpublished, raw 10x Genomics sequencing data used for the assembly of the embargoed razorbill genome to the puffin draft assembly, we built a neighbor-joining (NJ) tree based on pairwise genetic distance matrices (p-distance) and a sample-based ML phylogenetic tree in FastMe v2.1.5[81] and Treemix v1.13[82], respectively. For both trees, 100 bootstrap replicates were generated. To infer patterns of population splitting and mixing, we produced population-based ML trees including up to ten migration edges. The optimal number of migrations was selected using a quantitative approach by evaluating the distribution of explained variance, the log likelihoods, the covariance with an increase in migration edges, and by applying the method of Evanno[41] and several different linear threshold models. The topology for m_0 and m_{BEST} was evaluated by generating 100 bootstrap replicates. Additional details on the cluster and phylogenetic analyses are given in the Supplementary File.

Genetic diversity, heterozygosity, and inbreeding. We calculated a set of neutrality tests and population statistics in ANGSD using colony-based one-dimensional (1D) folded site frequency spectra (SFS). For each population, genomic cluster, and globally, Tajima's D and nucleotide diversity (π) were computed utilizing the per-site θ estimates. Individual genome-wide heterozygosity was calculated in ANGSD using individual, folded, 1D SFS. We calculated heterozygosity by dividing the number of polymorphic sites by the number of total sites present in the SFS.

The proportion of RoH within each puffin genome was computed by calculating local estimates of heterozygosity in 100 kb sliding windows (50 kb slide) following the approach in Sánchez-Barreiro et al.[42]. We defined the 10% quantile of the average local heterozygosity across all samples as the cutoff for a "low heterozygosity region" (Fig. S18). RoHs were declared as all regions with at least two subsequent windows of low heterozygosity (below cutoff) and their final length was calculated as described in Sánchez-Barreiro et al.[42]. We calculated an individual inbreeding coefficient based on the RoH, F_{RoH}, as in Sánchez-Barreiro et al.[42] by computing the fraction of the entire genome falling into RoHs, with the entire genome being the total length of windows scanned. Additional details on these analyses can be found in the Supplementary File.

Patterns of gene flow and admixture. Assessing potential patterns of IBD within the breeding range of the puffin, the program EEMS[43] was used to model the association between genetic and geographic data by visualizing the existing population structure and highlighting regions of higher-than-average and lower-than-average historic gene flow. We calculated a pairwise genetic distance matrix in ANGSD by sampling the consensus base (*-doIBS 2 -makeMatrix 1*) at the sites included in the genotype likelihood set (see *Nuclear cluster and phylogenetic analyses*) for each sample. The matrix was fed into 10 independent runs of EEMS, each consisting of one MCMC chain of six million iterations with a two million iteration burn-in, 9999 thinning iterations, and 1000 underlying demes.

Supplementing the results of the EEMS analysis, we conducted a traditional IBD analysis by determining geographical and genetic distances between the 12 colonies and assessing the significance of the correlation between the two distance matrices with a Mantel test[83] and a multiple regression on distance matrix (MRM)[84] analysis. F_{ST} was used as a proxy for genetic distance and computed for each population pair in ANGSD by applying two-dimensional (2D), folded SFS. We converted pairwise F_{ST} values to Slatkin's linearized F_{ST}[85]. Least Cost Path distances (paths over water only) between colony coordinates (latitude/longitude)

were calculated using the R package *marmap*[86] and used as geographic distances. We performed the Mantel test (999 permutations) and MRM analysis with the R package *ecodist*[87]. All analyses for IBD were re-run on subsets of colonies by progressively removing the colony from the geographic and genetic distance matrices, whose removal led to the highest increase in the proportion of variance in genetic distance explained by geographic distance in the resulting regression model (Spitsbergen, Isle of May, Bjørnøya and Gannet Isl.).

A distance-based Redundancy Analysis (dbRDA)[88] was conducted to corroborate the results of the MRM analyses and Mantel tests and to estimate the relative contribution of IBD and IBE to the observed Atlantic puffin population structure. The dbRDA was run between the genetic distance matrix versus geographic and environmental parameters[88]. A global dbRDA was performed with all geographic and environmental variables, and for statistically significant global dbRDA models, the most significant variables (geographic or environmental) were selected via a stepwise regression[89]. Those served as input for a reduced dbRDA to calculate the marginal effect of each variable and for a partial dbRDA with variance partitioning to estimate the separate effects of IBD and IBE. Similar to the MRM analyses and Mantel tests, these analyses were repeated on subsets of colonies by progressively removing the colony from the geographic, environmental, and genetic distance matrices, whose removal led to the highest increase in variance explained in the resulting global dbRDA model. Methods and R code for the dbRDA were found at https://github.com/laurabenestan/db-RDA-and-db-MEM [90].

Additional assessments of gene flow and admixture were conducted by calculating *f*3-statistics and multi-population D-statistics (aka ABBA BABA test)[91]. We calculated *f*3-statistics in Treemix for each unique combination of ((A,B),C)) of the 12 puffin populations. D-statistics were calculated in ANGSD (-doAbbababa2) for each combination of ((A,B),C),Outgroup) using the 12 puffin colonies. The outgroup was generated in ANGSD using the 10xGenomics sequencing data of the razorbill mapped to the puffin reference genome (see *Nuclear cluster and phylogenetic analyses).*

Evaluating genome-wide patterns of genetic differentiation, pairwise F_{ST} values between the Norway/Iceland/Faroe cluster and the Spitsbergen, Isle of May, Canada colonies (three comparisons) were calculated in sliding windows of 50 kb with 12.5 kb steps across the 25 pseudo-chromsomes by applying 2D, folded SFS. The window size of 50 kb was chosen for sliding window analyses because LD decays to ca. 10% ($R < 0.025$) within this distance (Fig. S19). Additional details on the IBD, admixture, and sliding-window analyses are given in the Supplementary File.

Statistics and reproducibility. The research sample included 72 adult Atlantic puffins (*Fratercula arctica*) across 12 colonies located in Svalbard, northern mainland Norway, Iceland, the Faroe Islands, Scotland, and Canada. The sample included six individuals per colony (12 colonies), including an equal sex ratio (3 males and 3 females per colony). All statistical tests were conducted using publicly available programs and packages as described in the methodological sections above. Reproducibility can be accomplished by following the sample collection and laboratory methods outlined above and by following the author's GitHub (https://github.com/OKersten/PuffPopGen) using the specified parameters mentioned in in the code and methodological sections above.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw read data analyzed in the current study have been deposited in the European Nucleotide Archive (ENA, www.ebi.ac.uk/ena) under study accession number PRJEB40631 (see Supplementary Data 2 for individual sample accession numbers). Nuclear and mitochondrial scaffolds (GCA_905066775.1, CAJHIB010000001-CAJHIB010013329), as well as pseudo-chromosomes (GCA_905066775.2, CAJHIB020000001-CAJHIB020000027), have been uploaded to ENA (Project PRJEB40926, Sample SAMEA7482542).

Code availability

Full code used for the population genomic analyses is available on the first author's GitHub (https://github.com/OKersten/PuffPopGen) and on Zenodo under the https://doi.org/10.5281/zenodo.4899574[92]. This includes versions of any software used, if relevant, and any specific variables or parameters used to generate, test, and process the dataset of this study.

Received: 17 November 2020; Accepted: 1 July 2021;
Published online: 29 July 2021

References

1. Otero, X. L., De La Peña-Lastra, S., Pérez-Alberti, A., Ferreira, T. O. & Huerta-Diaz, M. A. Seabird colonies as important global drivers in the nitrogen and phosphorus cycles. *Nat. Commun.* **9**, 246 (2018).
2. Velarde, E., Anderson, D. W. & Ezcurra, E. Seabird clues to ecosystem health. *Science* **365**, 116–117 (2019).
3. Piatt, J. F., Sydeman, W. J. & Wiese, F. Introduction: a modern role for seabirds as indicators. *Mar. Ecol. Prog. Ser.* **352**, 199–204 (2007).
4. Boersma, P. D., Clark, J. A. & Hillgarth, N. Seabird conservation. In *Biology of Marine Birds* (eds. Schreiber, E. & Burger, J.) 559–579 (CRC Press Boca Raton, 2002).
5. Denlinger, L. & Wohl, K. Seabird harvest regimes in the circumpolar nations. *Conservation of Arctic Flora and Fauna (CAFF)*, (2001).
6. Merkel, F. & Barry, T. Seabird Harvest in the Arctic. *Conservation of Arctic Flora and Fauna (CAFF)*, (2008).
7. Croxall, J. P. et al. Seabird conservation status, threats and priority actions: a global assessment. *Bird. Conserv. Int.* **22**, 1–34 (2012).
8. Paleczny, M., Hammill, E., Karpouzi, V. & Pauly, D. Population trend of the world's monitored seabirds, 1950-2010. *PLoS ONE* **10**, e0129342 (2015).
9. Frederiksen, M. Seabirds in the North East Atlantic. Summary of status, trends and anthropogenic impact. *TemaNord* **587**, 21–24 (2010).
10. Chardine, J. & Mendenhall, V. Human Disturbance at Arctic Seabird Colonies. *Conservation of Arctic Flora and Fauna (CAFF)*, (1998).
11. Funk, W. C., McKay, J. K., Hohenlohe, P. A. & Allendorf, F. W. Harnessing genomics for delineating conservation units. *Trends Ecol. Evol.* **27**, 489–496 (2012).
12. Moritz, C. Defining 'Evolutionarily Significant Units' for conservation. *Trends Ecol. Evol.* **9**, 373–375 (1994).
13. Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **11**, 697 (2010).
14. Fraser, D. J. & Bernatchez, L. Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Mol. Ecol.* **10**, 2741–2752 (2001).
15. Friesen, V. L. Speciation in seabirds: why are there so many species… and why aren't there more? *J. Ornithol.* **156**, 27–39 (2015).
16. Taylor, R. S. et al. Sympatric population divergence within a highly pelagic seabird species complex (*Hydrobates* spp.). *J. Avian Biol.* **49**, 1–14 (2018).
17. Rexer-Huber, K. et al. Genomics detects population structure within and between ocean basins in a circumpolar seabird: the white-chinned petrel. *Mol. Ecol.* **28**, 4552–4572 (2019).
18. Clucas, G. V. et al. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol. Ecol.* **27**, 4680–4697 (2018).
19. Frugone, M. J. et al. More than the eye can see: Genomic insights into the drivers of genetic differentiation in Royal/Macaroni penguins across the Southern Ocean. *Mol. Phylogenet. Evol.* **139**, 106563 (2019).
20. Cristofari, R. et al. Unexpected population fragmentation in an endangered seabird: the case of the Peruvian diving-petrel. *Sci. Rep.* **9**, 2021 (2019).
21. Tigano, A., Shultz, A. J., Edwards, S. V., Robertson, G. J. & Friesen, V. L. Outlier analyses to test for local adaptation to breeding grounds in a migratory arctic seabird. *Ecol. Evol.* **7**, 2370–2381 (2017).
22. Lowry, D. B. et al. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol. Ecol. Resour.* **17**, 142–152 (2017).
23. Somvichian-Clausen, A. Behind the stunning photo of a puffin gorging on fish. *Natl Geographic* (2017).
24. Huijbens, E. H. & Einarsson, N. Feasting on Friends: Whales, Puffins, and Tourism in Iceland. In *Tourism Experiences and Animal Consumption* (ed. Kline, C.) 10–27 (Routledge, 2018).
25. Lund, K. A., Kjartansdóttir, K. & Loftsdóttir, K. 'Puffin love': performing and creating Arctic landscapes in Iceland through souvenirs. *Tour. Stud.* **18**, 142–158 (2018).
26. Hodgetts, L. M. Animal bones and human society in the late younger stone age of arctic Norway. (Durham University, 1999).
27. Dove, C. J. & Wickler, S. Identification of bird species used to make a Viking age feather pillow. *Arctic* **69**, 29–36 (2016).
28. Harris, M. P. & Wanless, S. The puffin (T & AD Poyser, Bloomsbury Publishing, 2011).
29. BirdLife International. Fratercula arctica. The IUCN Red List of Threatened Species *2017* (2017)
30. Anker-Nilssen, T. & Aarvak, T. The population ecology of puffins at Røst. Status after the breeding season 2001. *NINA Oppdragsmeld.* **736**, 1–40 (2002).
31. Anker-Nilssen, T. et al. Key-site monitoring in Norway 2019, including Svalbard and Jan Mayen. SEAPOP Short Report 1–2020 (2020).
32. Lilliendahl, K. et al. Recruitment failure of Atlantic puffins Fratercula arctica and sandeels Ammodytes marinus in Vestmannaeyjar Islands. *N. áttúrufræðingurinn* **83**, 65–79 (2013).
33. Walker, S. J. & Meijer, H. J. M. Size variation in mid-Holocene North Atlantic Puffins indicates a dynamic response to climate change. *PLoS ONE* **16**, e0246888 (2021).
34. Burnham, K. K., Burnham, J. L. & Johnson, J. A. Morphological measurements of Atlantic puffin (Fratercula arctica naumanni) in High-Arctic Greenland. *Polar Res.* **39**. https://doi.org/10.33265/polar.v39.5242 (2020).

35. Gaston, A. J. & Provencher, J. F. A specimen of the high arctic subspecies of Atlantic Puffin, Fratercula arctica naumanni, in Canada. *Can. Field-Nat.* **126**, 50–54 (2012).
36. Salomonsen, F. *The Atlantic Alcidae*. vol. 6 (Elanders boktryckeri aktiebolag, 1944).
37. Moen, S. M. Morphologic and genetic variation among breeding colonies of the Atlantic puffin (Fratercula arctica). *Auk* **108**, 755–763 (1991).
38. Harris, M. P. Measurements and weights of British Puffins. *Bird. Study* **26**, 179–186 (1979).
39. Kim, J. A., Kang, S.-G., Yang, J. W., Hur, W.-H. & Kil, H.-J. Complete mitochondrial genome of Aethia cristatella (Charadriiformes: Alcidae). *Mitochondrial DNA Part B* **5**, 31–32 (2020).
40. Eo, S. H. & An, J. The complete mitochondrial genome sequence of Japanese murrelet (Aves: Alcidae) and its phylogenetic position in Charadriiformes. *Mitochondrial DNA A DNA Mapp. Seq. Anal.* **27**, 4574–4575 (2016).
41. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
42. Sánchez-Barreiro, F. et al. Historical Population Declines Prompted Significant Genomic Erosion in the Northern and Southern White Rhinoceros (*Ceratotherium Simum*). *Molecular Ecology*. 1–15 https://doi.org/10.1111/mec.16043 (2021).
43. Petkova, D., Novembre, J. & Stephens, M. Visualizing spatial population structure with estimated effective migration surfaces. *Nat. Genet.* **48**, 94–100 (2016).
44. Lombal, A. J., O'dwyer, J. E., Friesen, V., Woehler, E. J. & Burridge, C. P. Identifying mechanisms of genetic differentiation among populations in vagile species: historical factors dominate genetic differentiation in seabirds. *Biol. Rev. Camb. Philos. Soc.* **95**, 625–651 (2020).
45. Friesen, V. L., Burg, T. M. & McCoy, K. D. Mechanisms of population differentiation in seabirds. *Mol. Ecol.* **16**, 1765–1785 (2007).
46. Breton, A. R., Diamond, A. W. & Kress, S. W. Encounter, survival, and movement probabilities from an Atlantic puffin (*Fratercula arctica*) metapopulation. *Ecol. Monogr.* **75**, 133–149 (2006).
47. Fayet, A. L. et al. Ocean-wide drivers of migration strategies and their influence on population breeding performance in a declining seabird. *Curr. Biol.* **27**, 3871–3878.e3 (2017).
48. Burg, T. M. & Croxall, J. P. Global relationships amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Mol. Ecol.* **10**, 2647–2660 (2001).
49. Lowther, P. E., Diamond, T., Kress, S. W., Robertson, G. J. & Gill, F. Atlantic Puffin (*Fratercula arctica*). The Birds of North America Online 18, (2002).
50. Wojczulanis-Jakubas, K. et al. Weak population genetic differentiation in the most numerous Arctic seabird, the little auk. *Polar Biol.* **37**, 621–630 (2014).
51. Smith, A. L., Monteiro, L., Hasegawa, O. & Friesen, V. L. Global phylogeography of the band-rumped storm-petrel (Oceanodroma castro; Procellariiformes: Hydrobatidae). *Mol. Phylogenet. Evol.* **43**, 755–773 (2007).
52. Bergmann, C. Über die Verhältnisse der Wärmeökonomie der Tiere zu ihrer Grösse. *Gottinger Stud.* **3**, 595–708 (1847).
53. James, F. C. Geographic size variation in birds and its relationship to climate. *Ecology* **51**, 365–390 (1970).
54. Yamamoto, T. et al. Geographical variation in body size of a pelagic seabird, the streaked shearwater Calonectris leucomelas. *J. Biogeogr.* **43**, 801–808 (2016).
55. Barrett, R. T., Anker-Nilssen, T. & Krasnov, Y. V. Can Norwegian and Russian razorbills (Alca torda) be identified by their measurements? *Mar. Ornithol.* **25**, 5–8 (1997).
56. Anker-Nilssen, T., Aarvak, T. & Bangjord, G. Mass mortality of Atlantic Puffins Fratercula arctica off Central Norway, spring 2002: causes and consequences. *Atl. Seab.* **5**, 57–72 (2003).
57. Pearce, R. L. et al. Mitochondrial DNA suggests high gene flow in ancient murrelets. *Condor* **104**, 84–91 (2002).
58. Thomas, J. E. et al. Demographic reconstruction from ancient DNA supports rapid extinction of the great auk. *eLife* **8**, e47509 (2019).
59. Milot, E., Weimerskirch, H. & Bernatchez, L. The seabird paradox: dispersal, genetic structure and population dynamics in a highly mobile, but philopatric albatross species. *Mol. Ecol.* **17**, 1658–1673 (2008).
60. Edwards, S. & Bensch, S. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Mol. Ecol.* **18**, 2930–2936 (2009).
61. IPCC. Global Warming of 1.5 °C—Summary for Policy Makers. (2018).
62. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
63. Bernt, M. et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
64. Schubert, M. et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**, 1056–1082 (2014).
65. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
66. Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1–33 (2013).
67. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
68. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
69. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
70. Matschiner, M. Fitchi: haplotype genealogy graphs based on the Fitch algorithm. *Bioinformatics* **32**, 1250–1252 (2016).
71. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
72. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
73. Watterson, G. A. Heterosis or neutrality? *Genetics* **85**, 789–814 (1977).
74. Chakraborty, R. & Mitochondrial, D. N. A. polymorphism reveals hidden heterogeneity within some Asian populations. *Am. J. Hum. Genet.* **47**, 87–94 (1990).
75. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**, 915–925 (1997).
76. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinforma.* **15**, 356 (2014).
77. Orlando, L. & Librado, P. Origin and evolution of deleterious mutations in horses. *Genes* **10**, 649 (2019).
78. Meisner, J. & Albrechtsen, A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731 (2018).
79. Skotte, L., Korneliussen, T. S. & Albrechtsen, A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**, 693–702 (2013).
80. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).
81. Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800 (2015).
82. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
83. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
84. Lichstein, J. W. Multiple regression on distance matrices: a multivariate spatial analysis tool. *Plant Ecol.* **188**, 117–131 (2007).
85. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995).
86. Pante, E., Simon-Bouhet, B. & Irisson, J.-O. *marmap—R package*. (2019).
87. Goslee, S. & Urban, D. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw., Artic.* **22**, 1–19 (2007).
88. Legendre, P. & Anderson, M. J. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.* **69**, 1–24 (1999).
89. Blanchet, F. G., Legendre, P. & Borcard, D. Modelling directional spatial processes in ecological data. *Ecol. Modell.* **215**, 325–336 (2008).
90. Benestan, L. M. et al. Population genomics and history of speciation reveal fishery management gaps in two related redfish species (*Sebastes mentella* and *Sebastes fasciatus*). *Evol. Appl.* **14**, 588–606 (2021).
91. Soraggi, S., Wiuf, C. & Albrechtsen, A. Powerful inference with the D-statistic on low-coverage whole-genome data. *G3* **8**, 551–566 (2018).
92. Kersten, O. Code for Population Genomics Analyses of Atlantic Puffin (Fratercula arctica) using Whole Genome Sequencing (Version v1.0). Zenodo. https://doi.org/10.5281/zenodo.4899575 (2021).

Acknowledgements

Financial support was provided by the Nansen Foundation and the Faculty of Mathematical and Natural Sciences, University of Oslo (UiO). We are grateful to the SEAPOP program (www.seapop.no/en, Norwegian Research Council grant number 192141), and the SEATRACK (http://www.seapop.no/en/seatrack) and ARCTOX (https://arctox.cnrs.

fr/en/home) projects for collecting and sharing of samples. In particular, we thank Dave Fifield and Greg Robertson for the provision of puffin samples from Gull Island, and Árni Ásgeirsson and Róbert Arnar Stefánsson for supplying samples from Breiðafjörður. The authors acknowledge support from the National Genomics Infrastructure in Stockholm funded by the Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, and the SNIC/Uppsala Multidisciplinary Center for Advanced Computational Science for assistance with massively parallel sequencing (reference genome) and access to the UPPMAX computational infrastructure. Special thanks goes to the Norwegian Sequencing Centre, UiO (https://www.sequencing.uio.no) for the genomic libraries and resequencing of samples analyzed in this study. The razorbill genome data was made available for this study by Tom Gilbert and the Vertebrate Genome Project. Computation was performed using the resources and assistance from SIGMA2. Albína Pálsdóttir shared scripts for ANGSD, and Emiliano Trucchii advised on the manuscript. Pictures of puffins used in the figures were taken by Annemarie Look.

Author contributions

S.B. and B.S. conceptualized the project. M.I. performed the DNA extraction for the reference genome. O.K. did all other laboratory work. K.S.J. advised on the sequencing strategy and co-supervised O.K. O.K. refined the reference genome assembly. O.K. carried out the population genomic analyses with input from S.B., B.S., and D.M.L. O.K. designed the figures with input from S.B., B.S., and D.M.L. T.A.N., H.S., and E.S.H. advised on colony selection and provided ecological context. T.A.N., H.S., Od.K., S.D., E.S.H., J.F., M. P.H., J.D., K.E., M.L.M., and M.G.F. provided samples. O.K. wrote the paper with S.B., B. S., and D.M.L. All authors read and revised the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-021-02415-4.

Correspondence and requests for materials should be addressed to O.K. or S.B.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Caitlin Karniski.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Material

Supplementary data to this article can be found online at:

https://www.nature.com/articles/s42003-021-02415-4

Direct download links are as follows:

Supplementary Information:
https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-021-02415-4/MediaObjects/42003_2021_2415_MOESM1_ESM.pdf

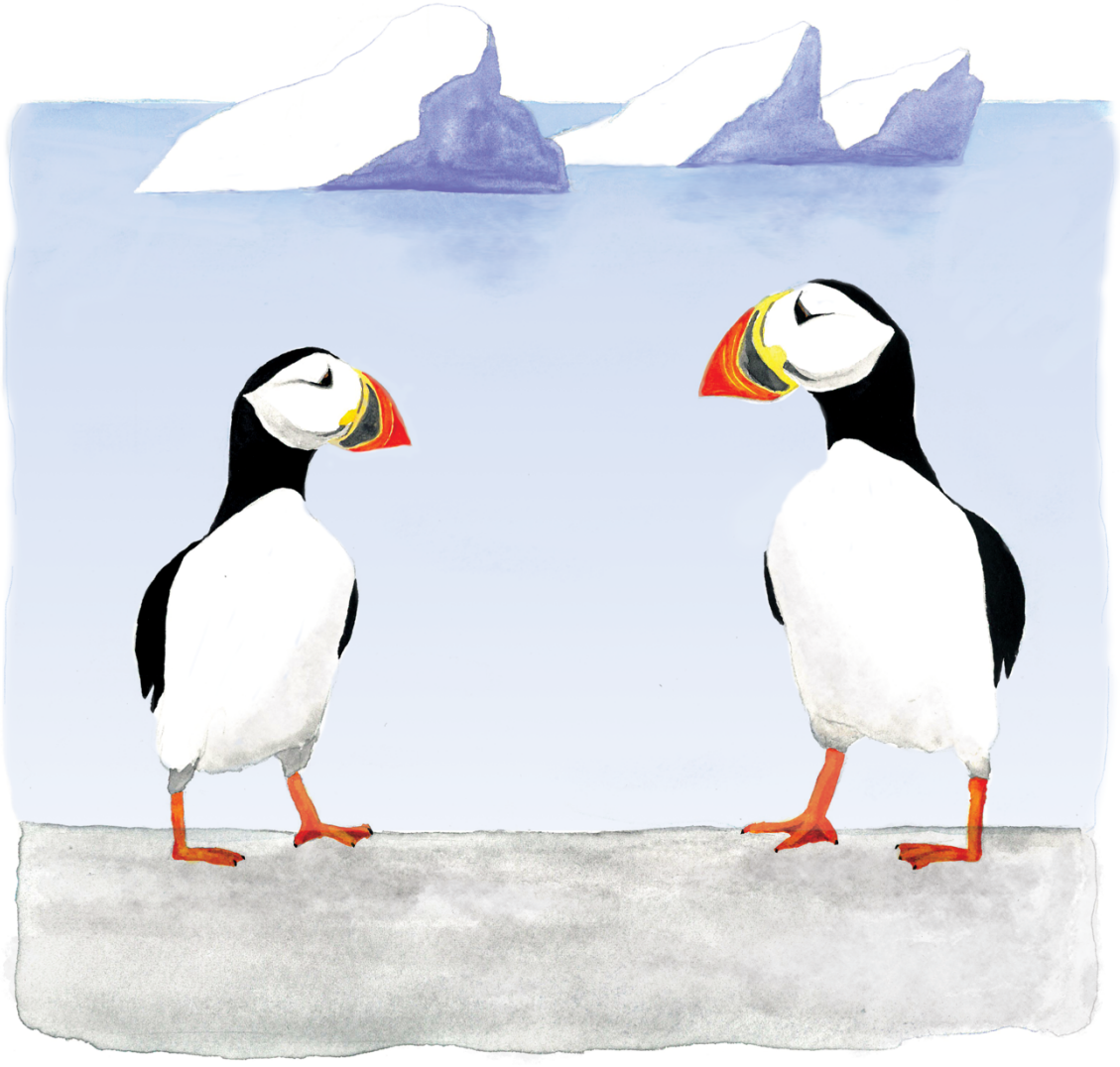
Reporting Summary:
https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-021-02415-4/MediaObjects/42003_2021_2415_MOESM2_ESM.pdf

Supplementary Data 1:
https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-021-02415-4/MediaObjects/42003_2021_2415_MOESM3_ESM.csv

Supplementary Data 2:
https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-021-02415-4/MediaObjects/42003_2021_2415_MOESM4_ESM.csv

Supplementary Data 3:
https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-021-02415-4/MediaObjects/42003_2021_2415_MOESM5_ESM.csv

Description of Additional Supplementary Files:
https://static-content.springer.com/esm/art%3A10.1038%2Fs42003-021-02415-4/MediaObjects/42003_2021_2415_MOESM6_ESM.pdf



Paper II

Sympatry of genetically distinct Atlantic Puffins (*Fratercula arctica*) in the High Arctic

IBIS
BOU international journal of avian science
Ibis (2022) doi: 10.1111/ibi.13153

Short Communication

Sympatry of genetically distinct Atlantic Puffins (*Fratercula arctica*) in the High Arctic

DEBORAH M. LEIGH,*,†,[1] OLIVER KERSTEN,*,†,[2] BASTIAAN STAR,[2] TYCHO ANKER-NILSSEN,[3] KURT BURNHAM,[4] JEFF JOHNSON,[5,6] JENNIFER PROVENCHER[7] & SANNE BOESSENKOOL*[2]

[1]*WSL Swiss Federal Research Institute, Zürcherstrasse 111, Birmensdorf, 8903, Switzerland*
[2]*Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Blindernveien 31, Oslo, 0371, Norway*
[3]*Norwegian Institute for Nature Research (NINA), Høgskoleringen 9, Trondheim, NO-7034, Norway*
[4]*High Arctic Institute, 603 10th Avenue, Orion, IL, 61273, USA*
[5]*Wolf Creek Operating Foundation, 1026 Soldier Creek Road, Wolf, WY, 82844, USA*
[6]*Department of Biological Sciences, University of North Texas, Denton, TX, 76201, USA*
[7]*National Wildlife Research Centre, Environment and Climate Change Canada, C/O Carleton University, 1125 Colonel By Drive, Ottawa, ON, K1A 0H3, Canada*

Across its range, the Atlantic Puffin *Fratercula arctica* is divided into four separate genetic clusters that correspond with geography and/or size differences. However, in the Western Atlantic High Arctic, there is a Puffin colony (Thule) that comprises two discrete size phenotypes. Using whole genome sequencing data of six Thule individuals from these two phenotypes, we found that Thule consists of three distinct genetic clusters, with no signs of recent interbreeding. Our results suggest the beginnings of a potential northward shift of boreal Atlantic Puffins in the West Atlantic, consistent with responses to a warming High Arctic climate.

*Corresponding authors.
Emails: deborah.leigh@wsl.ch; oliver.kersten@ibv.uio.no; sanne.boessenkool@ibv.uio.no
Twitters: @debbiemleigh; @OKerstenScience; @archaeogenomics

†*These authors contributed equally to this work.*

Keywords: climate change, genomics, Greenland, seabird.

The Arctic is undergoing an accelerated pace of warming and dramatic increases in human disturbance (Huntington *et al.* 2007, Serreze & Barry 2011). Ongoing northward range shifts of boreal species are increasing the likelihood of hybridization or lineage replacement of endemic Arctic populations (Kelly *et al.* 2010, Garcia-Elfring *et al.* 2017, Gallant *et al.* 2020). Although the logistical challenges intrinsic to the Arctic limit sample access, it is essential to expand genomic studies into the Arctic to help understand ongoing biotic change and taxonomic baselines, and to conserve Arctic biodiversity (Colella *et al.* 2020).

The Atlantic Puffin *Fratercula arctica* (hereafter 'Puffin', see Fig. S1) is an iconic seabird and of conservation concern (globally vulnerable, BirdLife International 2017). It is distributed across the North Atlantic from Spitsbergen and northern Greenland, to France and Maine (Harris & Wanless 2011; Fig. 1a). Whole genome resequencing has identified four separate Puffin genetic clusters that are partially consistent with subspecies delineations and latitudinal variation in body size (Harris & Wanless 2011). The smallest Puffins (*F. a. grabae*) form a single genetic cluster found in the UK and France. Mid-sized Puffins (*F. a. arctica*) are represented by two boreal genetic clusters along the North American Atlantic coast and in Iceland/Norway/Faroes, respectively. The largest Puffins (*F. a. naumanni*) form the most distinct genetic cluster and inhabit the High Arctic (e.g. Spitsbergen; Salomonsen 1944, Burnham *et al.* 2020a, Kersten *et al.* 2021). Finally, there is an *F. a. arctica/F. a. naumanni* hybrid population on the island of Bjørnøya (Fig. 1; Harris & Wanless 2011, Kersten *et al.* 2021).

On Dalrymple Rock Island (Igánaq: 76°28′21.65″N, 70°13′12.40″W; Greenland) near Thule Air Base, there is a small Puffin colony (hereafter 'Thule') that falls well within the expected High Arctic distribution of *F. a. naumanni* (66–79°) (Harris & Wanless 2011, Gaston & Provencher 2012, Burnham *et al.* 2020a). Unlike previously studied colonies, Thule consists of discrete large and mid-sized Puffin phenotypes. Large-sized Puffins are most common, with mid-sized Puffins representing fewer than 9% of individuals (the total Thule population size is 15–35 pairs) (Burnham *et al.* 2020a). Mid-sized individuals are similar in size to *F. a. arctica* and have been observed for multiple breeding seasons (Burnham *et al.* 2020a). Migratory monitoring data previously collected from both size phenotypes show an equally diverse non-breeding season distribution, with Thule Puffins using locations thousands of kilometres apart (Burnham *et al.* 2021). It is unclear if the size differentiation in Thule is the result of extreme size

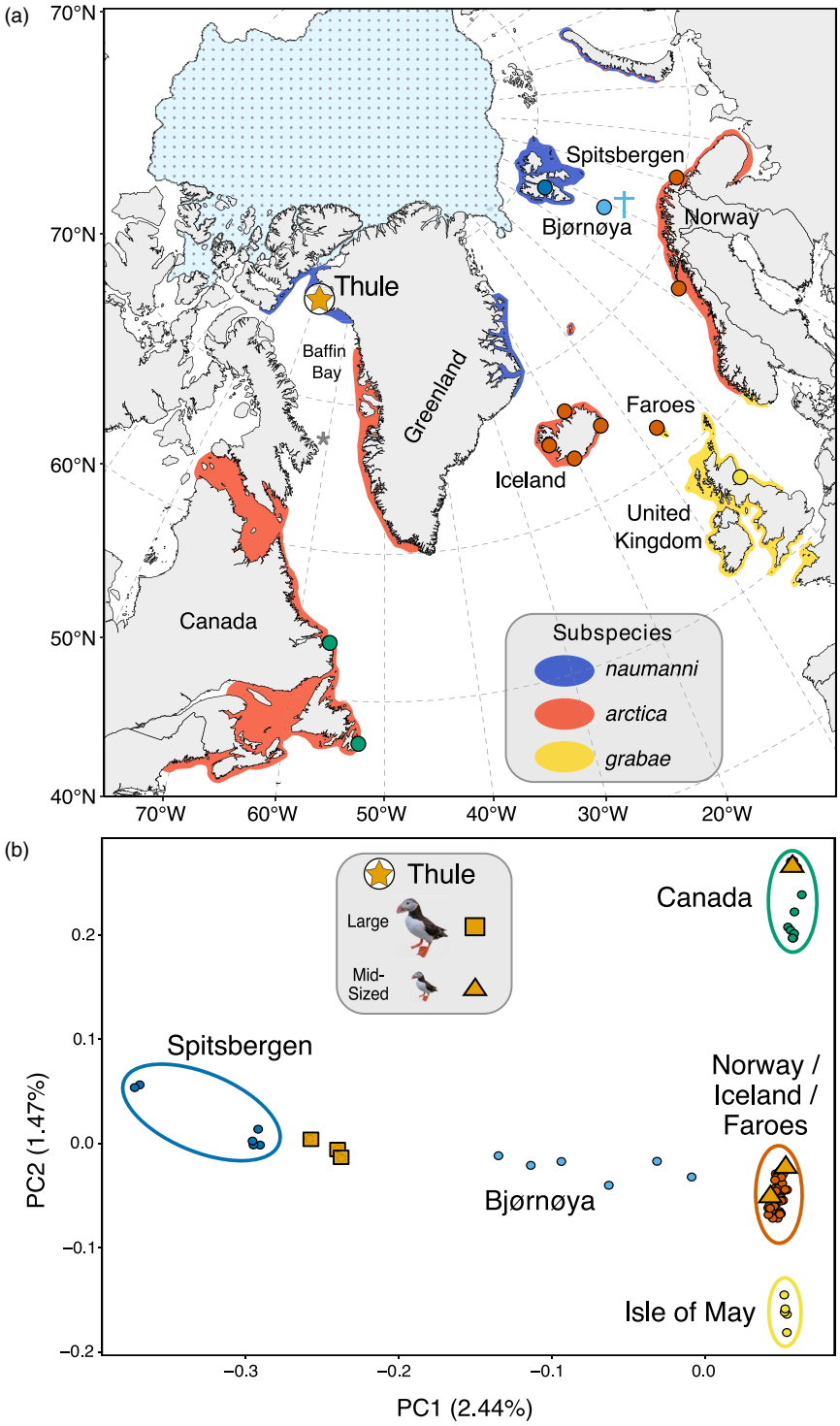


Figure 1. (a) Map presenting the 13 sites included in this study. Sites are coloured according to the genetic Atlantic Puffin clusters identified previously (Kersten *et al.* 2021) and shading highlights the range of the recognized subspecies. The cross depicts a confirmed hybrid zone. The asterisk indicates a large-bodied Puffin that was collected offshore at the Minarets during the breeding season. (b) Genetic structure (principal components analysis; PCA) based on genome-wide variation (n = 1 116 341 single nucleotide polymorphisms) for 77 individuals. Each circle represents a sample and colours indicate membership to a genetic cluster.

variation in Western Arctic *F. a. naumanni*, or if these mid-sized individuals are dispersed members of a different genetic cluster. Here we used whole genome sequencing to clarify the genomic relationship of Puffins in Thule. We discuss our results in light of ongoing boreal species shifts in response to climate change in the rapidly warming High Arctic.

METHODS

Blood from six adult Puffins from Thule (three mid-sized and three large-sized birds, see Fig. S1 and Data S1 and S2) was collected between 2012 and 2015 during the colony egg incubation period (July–August). Sampling was conducted following the guidelines established in Fair *et al.* (2010) and with permissions from the Greenland authorities. Size differences were visibly noticeable in the field, but to ensure a systematic classification, individuals were assigned a phenotype based on their wing length/beak size ratio cluster. The individuals sequenced were observed at Thule for two to three breeding seasons between 2012 and 2015, except for one large-sized male that was only observed during the last field season. Breeding status of sampled individuals was unknown because of site access constraints.

DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) or with a 5 M salt solution (Miller *et al.* 1988). Genomic libraries were built and sequenced (Illumina Hiseq4000; Illumina, San Diego, CA, USA) at the Norwegian Sequencing Centre (sequencing data have been deposited at the European Nucleotide Archive; Kersten 2022). Sequencing reads were mapped to the Atlantic Puffin assembly (European Nucleotide Archive Accession: CAJHIB020000000.2) using PALEOMIX v1.2.14 (Schubert *et al.* 2014), and analysed together with previously published genome data from 71 individuals representing 12 breeding colonies that acted as an integral reference for this study (Fig. 1a; Kersten *et al.* 2021, Kersten 2022). Further details on the methods and analyses can be found in Appendix S1.

Population structure was assessed using a genotype likelihood panel of 1 116 341 variant sites using ANGSD v.0.931 (Korneliussen *et al.* 2014; detailed in Appendix S1). A principal components analysis (PCA) was conducted with PCAngsd v0.982 (Meisner & Albrechtsen 2018) and individual ancestry proportions were estimated using ngsAdmix v32 (Skotte *et al.* 2013) and CLUMPAK (Kopelman *et al.* 2015), including a hierarchical approach, i.e. individuals from one cluster identified at $K = 2$ were removed followed by rerunning the analysis. Population and individual-based maximum likelihood phylogenetic trees with and without migration edges were built with Treemix v1.13 (Pickrell & Pritchard 2012) using the Razorbill *Alca torda* genome (GCA_008658365.1) as an outgroup. Individual pairwise genetic distance (p-distance) matrices were calculated with ngsDist v1.0.8 (Vieira *et al.* 2015).

Puffin colonies were divided into seven groups that included the four previously identified (Kersten *et al.* 2021) genetic clusters (Spitsbergen ($n = 6$), Iceland/Norway/Faroes ($n = 42$), Isle of May ($n = 5$), Canada ($n = 12$)) and the hybrid population on Bjørnøya ($n = 6$), as well as the two Thule size classes. These previous analyses by Kersten *et al.* (2021) detected varying levels of genetic diversity in the different clusters. To investigate genetic diversity of the Thule Puffins compared with their respective clusters, we analysed heterozygosity, runs of homozygosity (RoH) and individual inbreeding coefficients (F_{RoH}) using one- and two-dimensional site-frequency spectra (see Appendix S1).

Recent admixture was assessed by calculating *f*3-statistics in Treemix for each unique combination of ((A,B),C) of the seven Puffin groups. In addition, gene flow and admixture were investigated with a genome-wide ABBA–BABA *D*-statistics calculated in ANGSD comparing all possible triplets of the Puffin genetic groups and Thule morphologies with the Razorbill as outgroup and a significance threshold corrected for multiple testing (see Appendix S1).

RESULTS

The PCA revealed that the two size classes of Puffins at Thule were genetically distinct. We also observed genetic differentiation within the three mid-sized Thule Puffins (Fig. 1b). One individual grouped with the Western Atlantic cluster whereas the other two fell within the Iceland/Norway/Faroes cluster. In contrast, the three large individuals were all genetically similar to each other and most closely related to *F. a. naumanni* in Spitsbergen (Fig. 1b). Ancestry components estimated from the model-based clustering using $K = 2$–4 (Fig. 2), as supported by delta *K* (Evanno *et al.* 2005) and biological expectations (Fig. S2), as well as individual-based maximum likelihood phylogenetic trees (Fig. S3), confirmed these assignments of the Puffins from Thule. The best supported *K* value was 4 based on hierarchical analyses (Figs S4 and S5) and known biological and geographical differences between genetic clusters. Individual pairwise genetic distances mirrored the results visualized in the PCA and Admixture plot (Fig. S6).

Heterozygosity ($\chi^2 = 38.49$, $P = 8.99 \times 10^{-7}$, df = 6), inbreeding coefficients ($\chi^2 = 50.32$, $P = 4.06 \times 10^{-9}$, df = 6) and RoH lengths ($\chi^2 = 119.71$, $P = 1.90 \times 10^{-23}$, df = 6) were significantly different among the seven Puffin groups, and, for the Thule birds, values were largely consistent with their genetic clusters (Figs S7 and S8). The large Thule birds and *F. a. naumanni* had similar F_{RoH} that were significantly

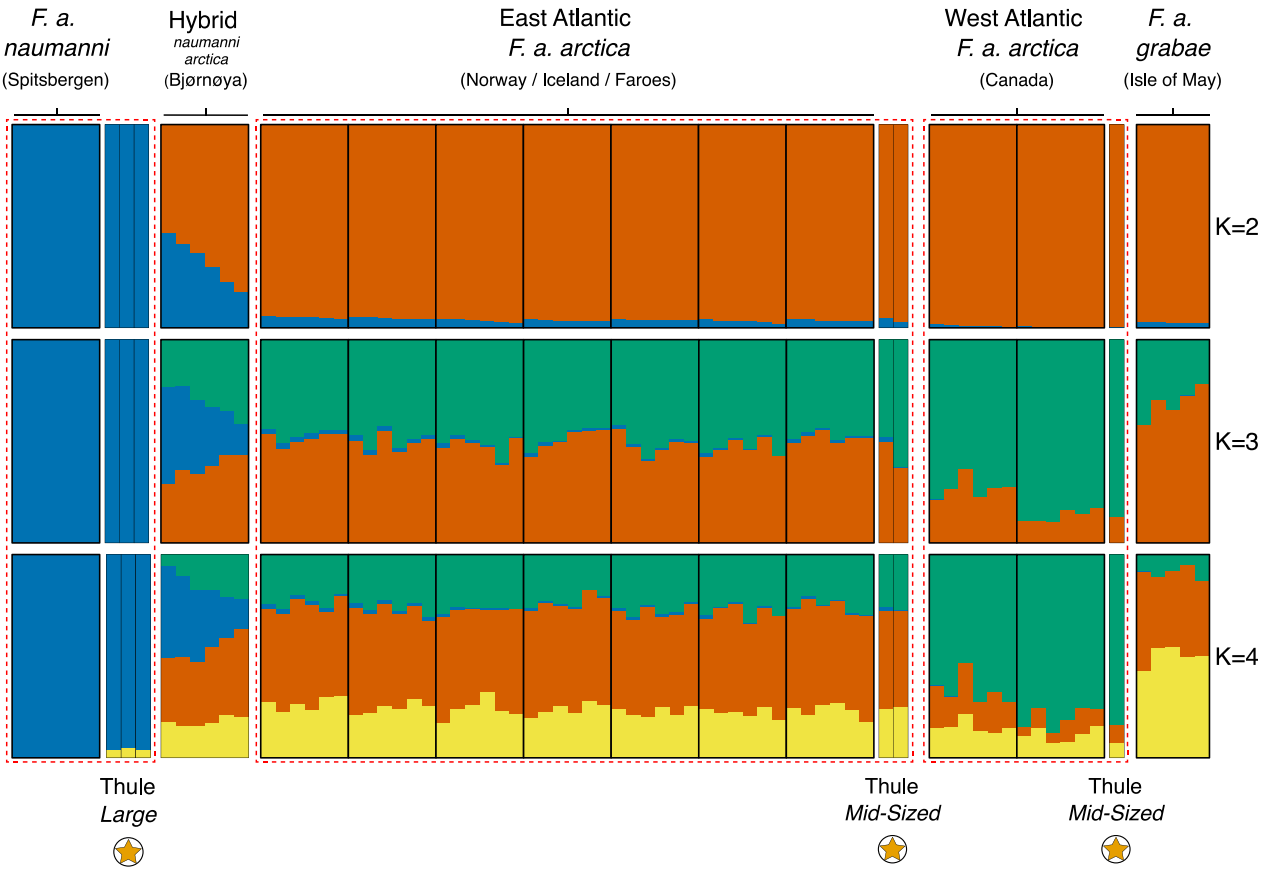


Figure 2. Genomic structure of 77 Atlantic Puffins across 13 colonies based on CLUMPAK-averaged admixture plots of the best *K* values. Colours indicate ancestry fraction to the different ancestral populations. Thule Puffins (labelled with a star symbol) have distinct ancestry components that are similar to those of three major genetic clusters (highlighted by dashed red line).

higher than the Norway/Iceland/Faroes genetic unit ($P < 0.05$ and $P < 0.0001$, respectively; Fig. S8). RoHs were significantly longer ($P < 0.05$) across the genome in large Thule Puffins relative to all other genetic units except *F. a. naumanni* (Fig. S7c), and the maximum size of homozygous tracts was similar in large Thule Puffins and *F. a. naumanni*.

No evidence for interbreeding was visible between the two Thule size phenotypes and genetic groups. Population-based maximum likelihood phylogenetic trees using up to two migration edges as determined by different threshold models (Figs S9 and S10) and *f*3-statistics (Table S1) did not show significant evidence for gene flow between any populations except from Spitsbergen into Bjørnøya, a known hybrid Puffin population (Kersten *et al.* 2021). The ABBA–BABA analysis showed significant *D*-statistics, and *Z* values were highest between large-sized Thule individuals and Spitsbergen (Table 1). Mid-sized Thule individuals showed weaker, albeit significant, signs of introgression with the Canadian genetic cluster. There was also an expected sign of introgression between Spitsbergen and Bjørnøya (Table 1). The admixture analysis showed a small (3.5–4.4%) admixed portion of the genome in the large Thule birds visible at $K = 4$, but not at $K = 2$ or 3 (Fig. 2).

DISCUSSION

Whole genome sequencing data analyses revealed that the Thule Puffin colony consists of individuals from multiple distinct genetic units, with no detectable recent interbreeding between size phenotypes. Remarkably, the mid-sized individuals found at Thule represented the two distinct West and East Atlantic *F. a. arctica* genetic clusters.

The large Thule Puffins were closely related to Puffins on Spitsbergen (*F. a. naumanni*) and showed similar levels of heterozygosity, inbreeding and RoH lengths. However, the populations at Thule and Spitsbergen

Table 1. ABBA–BABA analyses between Thule phenotypes and previously identified Atlantic Puffin genetic clusters (Kersten *et al.* 2021).

D	*Z*	*P* (adj)	nABBA	nBABA	nBlocks	H1	H2	H3	H4
Spitsbergen & Bjørnøya									
−0.0417	−17.32	0.000	3686.47	4007.01	19 152	Bjørnøya	Nor/Ice/Far	Spitsbergen	Razorbill
−0.0393	−15.20	0.000	3704.27	4007.01	19 446	Spitsbergen	Nor/Ice/Far	Bjørnøya	Razorbill
−0.0396	−14.25	0.000	3696.96	4001.84	18 526	Bjørnøya	Canada	Spitsbergen	Razorbill
−0.0386	−13.21	0.000	3704.27	4001.84	18 744	Spitsbergen	Canada	Bjørnøya	Razorbill
−0.0373	−11.32	0.000	3711.59	3999.58	17 995	Bjørnøya	Isle of May	Spitsbergen	Razorbill
−0.0373	−10.97	0.000	3711.77	3999.58	18 168	Spitsbergen	Isle of May	Bjørnøya	Razorbill
−0.0398	−10.61	0.000	3687.53	3993.49	17 690	Bjørnøya	Thule (mid-sized)	Spitsbergen	Razorbill
−0.0364	−9.49	0.000	3713.30	3993.49	17 781	Spitsbergen	Thule (mid-sized)	Bjørnøya	Razorbill
Spitsbergen & Thule (large)									
−0.1018	−35.17	0.000	3568.19	4377.16	17 759	Spitsbergen	Nor/Ice/Far	Thule (large)	Razorbill
−0.1014	−31.02	0.000	3566.21	4371.33	17 028	Spitsbergen	Canada	Thule (large)	Razorbill
0.1000	30.81	0.000	4377.16	3581.44	19 145	Nor/Ice/Far	Thule (large)	Spitsbergen	Razorbill
0.0980	27.88	0.000	4371.33	3591.38	18 368	Canada	Thule (large)	Spitsbergen	Razorbill
−0.1031	−27.15	0.000	3560.93	4379.88	16 415	Spitsbergen	Isle of May	Thule (large)	Razorbill
0.0954	24.12	0.000	4379.88	3616.92	17 534	Isle of May	Thule (large)	Spitsbergen	Razorbill
−0.0998	−23.51	0.000	3578.09	4371.15	16 093	Spitsbergen	Thule (mid-sized)	Thule (large)	Razorbill
−0.0981	−22.67	0.000	3590.41	4371.15	16 957	Thule (large)	Thule (mid-sized)	Spitsbergen	Razorbill
−0.0657	−18.79	0.000	3703.18	4224.30	16 622	Spitsbergen	Bjørnøya	Thule (large)	Razorbill
0.0596	15.98	0.000	4224.30	3748.99	17 816	Bjørnøya	Thule (large)	Spitsbergen	Razorbill
Bjørnøya & Thule (large)									
−0.0374	−14.18	0.000	3703.39	3991.44	17 732	Bjørnøya	Nor/Ice/Far	Thule (large)	Razorbill
−0.0370	−12.11	0.000	3704.24	3988.68	16 969	Bjørnøya	Canada	Thule (large)	Razorbill
−0.0387	−10.57	0.000	3698.75	3996.18	16 336	Bjørnøya	Isle of May	Thule (large)	Razorbill
0.0333	10.45	0.000	3991.44	3734.28	19 418	Nor/Ice/Far	Thule (large)	Bjørnøya	Razorbill
0.0326	9.42	0.000	3988.68	3736.52	18 572	Canada	Thule (large)	Bjørnøya	Razorbill
−0.0355	−8.52	0.000	3705.81	3978.37	16 036	Bjørnøya	Thule (mid-sized)	Thule (large)	Razorbill
0.0312	7.98	0.000	3996.18	3754.56	17 685	Isle of May	Thule (large)	Bjørnøya	Razorbill
−0.0304	−7.06	0.000	3743.59	3978.37	17 033	Thule (large)	Thule (mid-sized)	Bjørnøya	Razorbill
Canada & Thule (mid-sized)									
−0.0061	−3.01	0.006	3775.56	3822.24	17 920	Canada	Nor/Ice/Far	Thule (mid-sized)	Razorbill
0.0081	2.80	0.012	3822.24	3760.67	20 347	Nor/Ice/Far	Thule (mid-sized)	Canada	Razorbill
0.0112	2.77	0.013	3839.05	3753.65	18 024	Thule (large)	Thule (mid-sized)	Canada	Razorbill
−0.0097	−2.48	0.029	3765.61	3839.05	16 784	Canada	Thule (large)	Thule (mid-sized)	Razorbill
0.0081	2.46	0.029	3819.62	3758.02	16 979	Spitsbergen	Canada	Thule (mid-sized)	Razorbill
Others									
−0.0089	−2.47	0.029	3764.52	3832.06	18 081	Canada	Thule (large)	Isle of May	Razorbill
0.0105	2.34	0.039	3837.38	3757.49	16 402	Thule (large)	Thule (mid-sized)	Isle of May	Razorbill

Significant pairwise genome-wide comparisons are shown. Negative values signal introgression between H1 and H3, positive between H2 and H3.

were not panmictic, showing greater genetic differentiation than that between West and East Atlantic *F. a. arctica* (Fig. S6). Though our limited sample size probably impacts differentiation estimate accuracy, Thule and Spitsbergen are c. 5300 km apart (over water) and the observed patterns align with previous findings of isolation by distance within Puffin Evolutionary Significant Units (ESUs; Kersten *et al.* 2021). Moreover, available tracking data indicate no non-breeding season distribution overlap between birds from these High Arctic colonies (Fayet *et al.* 2017, Burnham *et al.* 2021, Kersten *et al.* 2021). Large Thule Puffins overwinter south of Iceland close to eastern Greenland whereas Spitsbergen Puffins overwinter north of Iceland (Burnham *et al.* 2021, Kersten *et al.* 2021). Non-overlapping overwintering grounds are recognized as a leading cause of population structure among seabirds (Puffins, Kersten *et al.* 2021; Black-browed Albatross *Thalassarche melanophris*, Burg & Croxall 2001; Thick-billed Murre *Uria lomvia*, Tigano *et al.* 2017). The western High Arctic (including Thule) has previously been speculated to represent a small, isolated, unique and vulnerable Puffin population (Gaston & Provencher 2012). Our results indicate that large Puffins at Thule are

genetically most similar to the Spitsbergen *F. a. naumanni*, but the two populations should be managed separately given the lack of non-breeding distribution overlap and observed genetic differentiation (Moritz 1994).

Mid-sized Thule Puffins clustered closely to either *F. a. arctica* genetic cluster. One female was closely related to Puffins from the boreal West Atlantic, whereas the other two Puffins (a male and a female) clustered with Puffins from Iceland/Norway/Faroes. The high genetic similarity to these genetic clusters and lack of detectable admixture with the large Thule birds suggests that these are dispersed individuals from southern natal colonies. Their overwintering areas support a southern origin (Burnham *et al.* 2021), corresponding to their respective genetic clusters and not the regions used by the larger Thule individuals (Fayet *et al.* 2017, Burnham *et al.* 2021, Kersten *et al.* 2021). Specifically, the female from the West Atlantic ESU (ID: 8408 in Burnham *et al.* 2021) overwintered in the Labrador Sea and North Atlantic, corresponding with colonies from Canada (Fayet *et al.* 2017). The two Puffins from the East Atlantic ESU overwintered near West Iceland (ID: 7363, male) and the Azores (ID: 8406, female; see Burnham *et al.* 2021), overlapping with Puffins from the Iceland/Norway/Faroes (Fayet *et al.* 2017).

Despite observing both phenotypes during the breeding season, no recent interbreeding was identified at Thule. Though detection of gene flow may be hampered by lack of genetic differentiation between closely related subspecies, contemporary introgression has previously been detected in Puffins at Bjørnøya (Kersten *et al.* 2021). Recent gene flow is expected to generate significant ABBA–BABA statistics (Barlow *et al.* 2018, Westbury *et al.* 2021). In Thule, no significant comparisons supported introgression between the size classes. Comparisons only supported our findings of genetic similarity between large Thule Puffins and Spitsbergen, as well as between mid-sized individuals and their boreal genetic clusters. Hence, there is no evidence that recent interbreeding has occurred between the morphologies at Thule. The potential cohabitation of distinct subspecies at Thule is a deviation from previously detected patterns of clear geographical boundaries and hybridization upon contact (Harris & Wanless 2011, Kersten *et al.* 2021). Historical records of phenotype variation at currently unsampled colonies in the East Arctic (Novaya Zemlya and Jan Mayen; Salomonsen 1944, Harris & Wanless 2011) suggest other Puffin colonies could also contain multiple subspecies; however, unlike Thule, this is probably accompanied by hybridization because there are also records of intermediate morphotypes (Salomonsen 1944, Harris & Wanless 2011). It can be speculated that a barrier to interbreeding at Thule may arise from sub-species' behavioural differences; different overwintering areas may lead to asynchronous colony arrival and mis-matched timing of pair bonding (Ketterson *et al.* 2015).

Sympatric distinct subspecies are unusual, especially in seabirds where new contact zones typically result in hybridization (Scopoli's Shearwater *Calonectris diomedea*, Munilla *et al.* 2016; gadfly petrels *Pterodroma* spp., Brown *et al.* 2010). The absence of evidence for hybridization at Thule is also unusual for Arctic species, where hybridization with low-latitude taxa upon contact is common and a key potential pathway for adaptation to climate change (Colella *et al.* 2020, Charles & Stehlik 2021). Importantly, hybridization expectations are clearly visible in the contact zone on Bjørnøya (this study, Kersten *et al.* 2021). Collectively, this suggests cohabitation at Thule may also be recent and interbreeding could arise in the future. Although we do not know the driving mechanisms, we hypothesize that climate warming may be pushing a northern range expansion of *F. a. arctica*, similar to those observed during the Little Ice Age (1620–1770 CE, Walker & Meijer 2021). Under this hypothesis, the mid-sized individuals at Thule may represent the very early stages of a range shift in boreal Puffins. Similar range shifts have already been detected in Western Atlantic Arctic populations of Thickbilled Murres *U. lomvia-* and Razorbills *A. torda* due to the extended habitable period in the Arctic (Patterson *et al.* 2021) and the northern shift of fish stocks (Gaston & Woo 2008). Additionally, an increase of boreal seabird species has been recorded in the East Atlantic Arctic (Descamps & Strøm 2021) and the first Atlantic records of Pacific species of *Fratercula* have also occurred at Thule, probably facilitated by recent Arctic sea-ice loss (Burnham *et al.* 2020b). The hypothesized northern range shifts of Puffins must now be confirmed with temporal samples and additional colonies (i.e. Arctic Canada). Nonetheless, it is clear that valuable insights about Arctic biodiversity can be gained even from a small number of individuals. Further studies are urgently needed across the Arctic to better understand the biodiversity present and the rapidly evolving responses to climate change.

The first authors (DML and OK) contributed equally and reserve the right to place themselves as first author on their CV. The Razorbill genome data were made available by Tom Gilbert and the Vertebrate Genome Project. Computation was performed using the resources from SIGMA2.

AUTHOR CONTRIBUTIONS

Deborah M. Leigh: Conceptualization (lead); data curation (supporting); formal analysis (supporting); funding acquisition (lead); investigation (equal); methodology (equal); resources (equal); validation (supporting); visualization (supporting); writing – original draft (lead); writing – review and editing (lead). **Oliver Kersten:** Data

curation (lead); formal analysis (lead); investigation (equal); methodology (lead); software (lead); validation (lead); visualization (lead); writing – original draft (equal); writing – review and editing (equal). **Bastiaan Star:** Data curation (supporting); formal analysis (supporting); funding acquisition (supporting); investigation (supporting); methodology (supporting); project administration (supporting); resources (supporting); supervision (supporting); validation (supporting); visualization (supporting); writing – original draft (equal); writing – review and editing (equal). **Tycho Anker-Nilssen:** Resources (supporting); writing – review and editing (supporting). **Kurt Burnham:** Investigation (lead); resources (supporting); writing – review and editing (supporting). **Jeff Johnson:** Resources (supporting); writing – review and editing (supporting). **Jennifer Provencher:** Conceptualization (lead); resources (supporting); writing – review and editing (supporting). **Sanne Boessenkool:** Formal analysis (equal); funding acquisition (lead); investigation (equal); methodology (equal); project administration (lead); resources (equal); supervision (lead); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

FUNDING

This work was supported by a LinnéSys Systematics Research Fund grant awarded to DML, and by the Faculty of Mathematical and Natural Sciences, University of Oslo. Sample acquisition was funded by The Offield Family Foundation, Wolf Creek Charitable Trust and Patagonia.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ETHICAL NOTE

All research in Greenland was conducted after ethical approval and issuance of permits by the Government of Greenland, Department of Fisheries, Hunting and Agriculture (High Arctic Institute permit numbers: Sags nr. 2012–065141, Dok. nr. 888887, Sags nr. 2013–083369, Dok. nr. 1204884, Sags nr. 2014–099682, Dok nr. 1594176, Sags nr. 2015–115204, Dok. nr. 1975643). Capture, handling and blood collection followed the Ornithological Council's *Guidelines to the use of wild birds in research* (Fair *et al.* 2010).

Data Availability Statement

Raw read data have been deposited in the European Nucleotide Archive (ENA, www.ebi.ac.uk/ena) under study accession number PRJEB40631 (see Data S1 for individual sample accession numbers) found at https://www.ebi.ac.uk/ena/browser/view/PRJEB40631. Full code used for the population genomic analyses is available on Zenodo under the DOI https://doi.org/10.5281/zenodo.5950720. This includes versions of any software used, if relevant, and any specific variables or parameters used to generate, test and process the dataset of this study.

REFERENCES

Barlow, A., Cahill, J.A., Hartmann, S., Theunert, C., Xenikoudakis, G., Fortes, G.G., Paijmans, J.L.A., Rabeder, G., Frischauf, C., Grandal-d'Anglade, A., García-Vázquez, A., Murtskhvaladze, M., Saarma, U., Anijalg, P., Skrbinšek, T., Bertorelle, G., Gasparian, B., Bar-Oz, G., Pinhasi, R., Slatkin, M., Dalén, L., Shapiro, B. & Hofreiter, M. 2018. Partial genomic survival of cave bears in living brown bears. *Nat. Ecol. Evol.* **2**: 1563–1570.

BirdLife International. 2017. *Fratercula arctica. The IUCN Red List of Threatened Species 2017.* Available at: https://doi.org/10.2305/IUCN.UK.2017-3.RLTS.T22694927A117606911.en

Brown, R.M., Nichols, R.A., Faulkes, C.G., Jones, C.G., Bugoni, L., Tatayah, V., Gottelli, D. & Jordan, W.C. 2010. Range expansion and hybridization in Round Island petrels (*Pterodroma* spp.): evidence from microsatellite genotypes. *Mol. Ecol.* **19**: 3157–3170.

Burg, T.M. & Croxall, J.P. 2001. Global relationships amongst black-browed and grey-headed albatrosses: analysis of population structure using mitochondrial DNA and microsatellites. *Mol. Ecol.* **10**: 2647–2660.

Burnham, K.K., Burnham, J.L., Johnson, J.A., Konkel, B.W., Stephens, J. & Badgett, H. 2020a. First record of horned puffin in the North Atlantic and tufted puffin in High Arctic Greenland. *Polar Res.* **39**: 4458.

Burnham, K.K., Burnham, J.L. & Johnson, J.A. 2020b. Morphological measurements of Atlantic puffin (*Fratercula arctica naumanni*) in High-Arctic Greenland. *Polar Res.* **39**: 3614.

Burnham, K.K., Burnham, J.L., Johnson, J.A. & Huffman, A. 2021. Migratory movements of Atlantic Puffins *Fratercula arctica naumanni* from high Arctic Greenland. *PLoS One* **16**: e0252055.

Charles, K.M. & Stehlik, I. 2021. Assisted species migration and hybridization to conserve cold-adapted plants under climate change. *Conserv. Biol.* **35**: 559–566.

Colella, J.P., Talbot, S.L., Brochmann, C., Taylor, E.B., Hoberg, E.P. & Cook, J.A. 2020. Conservation genomics in a changing Arctic. *Trends Ecol. Evol.* **35**: 149–162.

Descamps, S. & Strøm, H. 2021. As the Arctic becomes boreal: ongoing shifts in a high-Arctic seabird community. *Ecology* **102**: e03485.

Evanno, G., Regnaut, S. & Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**: 2611–2620.

Fair, J.M., Paul, E. & Jones, J. (eds) 2010. *Guidelines to the Use of Wild Birds in Research*, 3rd edn. Washington, DC: Ornithological Council.

Fayet, A.L., Freeman, R., Anker-Nilssen, T., Diamond, A., Erikstad, K.E., Fifield, D., Fitzsimmons, M.G., Hansen, E.S., Harris, M.P., Jessopp, M., Kouwenberg, A.L., Kress, S., Mowat, S., Perrins, C.M., Petersen, A., Petersen, I.K., Reiertsen, T.K., Robertson, G.J., Shannon, P., Sigurðsson, I.A., Shoji, A., Wanless, S. & Guilford, T. 2017. Ocean-wide drivers of migration strategies and their influence on population breeding performance in a declining seabird. *Curr. Biol.* **27**: 3871–3878.

Gallant, D., Lecomte, N. & Berteaux, D. 2020. Disentangling the relative influences of global drivers of change in biodiversity: a study of the twentieth-century red fox expansion into the Canadian Arctic. *J. Anim. Ecol.* **89**: 565–576.

Garcia-Elfring, A., Barrett, R.D.H., Combs, M., Davies, T.J., Munshi-South, J. & Millien, V. 2017. Admixture on the northern front: population genomics of range expansion in the white-footed mouse (*Peromyscus leucopus*) and secondary contact with the deer mouse (*Peromyscus maniculatus*). *Heredity* **119**: 447–458.

Gaston, A.J. & Provencher, J.F. 2012. A specimen of the High Arctic subspecies of Atlantic Puffin, *Fratercula arctica naumanni*, in Canada. *Can. Field Nat.* **126**: 50–54.

Gaston, A.J. & Woo, K. 2008. Razorbills (*Alca torda*) follow subarctic prey into the Canadian Arctic: colonization results from climate change? *Ornithology* **125**: 939–942.

Harris, M.P. & Wanless, S. 2011. *The Puffin*. London: T & AD Poyser, Bloomsbury Publishing.

Huntington, H.P., Boyle, M., Flowers, G.E., Weatherly, J.W., Hamilton, L.C., Hinzman, L., Gerlach, C., Zulueta, R., Nicolson, C. & Overpeck, J. 2007. The influence of human activity in the Arctic on climate and climate impacts. *Clim. Change* **82**: 77–92.

Kelly, B.P., Whiteley, A. & Tallmon, D. 2010. The Arctic melting pot. *Nature* **468**: 891.

Kersten. 2022. *Project: PRJEB40631; ENA – European Nucleotide Archive*. Available at: https://www.ebi.ac.uk/ena/browser/view/PRJEB40631

Kersten, O., Star, B., Leigh, D.M., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps, S., Erikstad, K.E., Fitzsimmons, M.G., Fort, J., Hansen, E.S., Harris, M.P., Irestedt, M., Kleven, O., Mallory, M.L., Jakobsen, K.S. & Boessenkool, S. 2021. Complex population structure of the Atlantic puffin revealed by whole genome analyses. *Commun. Biol.* **4**: 922.

Ketterson, E.D., Fudickar, A.M., Atwell, J.W. & Greives, T.J. 2015. Seasonal timing and population divergence: when to breed, when to migrate. *Curr. Opin. Behav. Sci.* **6**: 50–58.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A. & Mayrose, I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**: 1179–1191.

Korneliussen, T.S., Albrechtsen, A. & Nielsen, R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**: 356.

Meisner, J. & Albrechtsen, A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**: 719–731.

Miller, S.A., Dykes, D.D. & Polesky, H.F. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* **16**: 1215.

Moritz, C. 1994. Defining 'evolutionarily significant units' for conservation. *Trends Ecol. Evol.* **9**: 373–375.

Munilla, I., Genovart, M., Paiva, V.H. & Velando, A. 2016. Colony Foundation in an oceanic seabird. *PLoS One* **11**: e0147222.

Patterson, A., Gilchrist, H.G., Gaston, A. & Elliott, K.H. 2021. Northwest range shifts and shorter wintering period of an Arctic seabird in response to four decades of changing ocean climate. *Mar. Ecol. Prog. Ser.* **679**: 163–179.

Pickrell, J.K. & Pritchard, J.K. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**: e1002967.

Salomonsen, F. 1944. *The Atlantic Alcidae - The Seasonal and Geographical Variation of the Auks Inhabiting the Atlantic Ocean and the Adjacent Waters*. Göteborg: Elanders Boktryckeri Aktiebolag.

Schubert, M., Ermini, L., Sarkissian, C.D., Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D., Fernández, R., Kircher, M., McCue, M., Willerslev, E. & Orlando, L. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat. Protoc.* **9**: 1056–1082.

Serreze, M.C. & Barry, R.G. 2011. Processes and impacts of Arctic amplification: a research synthesis. *Global Planet. Change* **77**: 85–96.

Skotte, L., Korneliussen, T.S. & Albrechtsen, A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**: 693–702.

Tigano, A., Shultz, A.J., Edwards, S.V., Robertson, G.J. & Friesen, V.L. 2017. Outlier analyses to test for local adaptation to breeding grounds in a migratory arctic seabird. *Ecol. Evol.* **7**: 2370–2381.

Vieira, F.G., Lassalle, F., Korneliussen, T.S. & Fumagalli, M. 2015. Improving the estimation of genetic distances from next-generation sequencing data. *Biol. J. Linn. Soc.* **117**: 139–149.

Walker, S.J. & Meijer, H.J.M. 2021. Size variation in mid-Holocene North Atlantic Puffins indicates a dynamic response to climate change. *PLoS One* **16**: e0246888.

Westbury, M.V., Thompson, K.F., Louis, M., Cabrera, A.A., Skovrind, M., Castruita, J.A.S., Constantine, R., Stevens, J.R. & Lorenzen, E.D. 2021. Ocean-wide genomic variation in Gray's beaked whales, *Mesoplodon grayi*. *R. Soc. Open Sci.* **8**: 201788.

Received 27 June 2022;
Revision 3 September 2022;
revision accepted 10 October 2022.
Associate Editor: Matt Wood.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Size comparison using wing length and beak size between adult Atlantic Puffins measured at the Thule colony during the breeding season in 2010–16.

Figure S2. Delta *K* as a function of the number of ancestral clusters (*K*) as calculated by the method of Evanno for *K* = 1–9.

Figure S3. Individual-based Treemix analysis of 77 Atlantic Puffins.

Figure S4. Delta *K* as a function of the number of ancestral clusters (*K*) as calculated by the method of Evanno for *K* = 1–9 after removing Spitsbergen, the large Thule morphs and Bjørnøya individuals.

Figure S5. Hierarchical genomic structure of 62 Puffins based on CLUMPAK-averaged admixture plots of the best *K* values.

Figure S6. Heatmap of genetic distances between 77 Atlantic Puffin individuals.

Figure S7. Estimates of individual genome-wide heterozygosity, individual inbreeding coefficients and length distribution of runs of homozygosity tracts longer than 500 kb for Puffins from each genomic cluster.

Figure S8. Genome-wide heterozygosity and inbreeding compared between Puffins of the Thule colony and colonies of the previously identified population genomic clusters.

Figure S9. Estimation of the optimal number of migration edges (m) for a Treemix-generated population-based maximum likelihood tree using optM.

Figure S10. Population-based Treemix analyses of 13 Atlantic Puffin colonies applying up to two migrations.

Table S1. Significant recent admixture signal between genomic Atlantic Puffin clusters as revealed by *f*3-statistics.

Appendix S1. Detailed methodological description of the whole genome analysis of six Atlantic Puffin individuals from a Western Atlantic High Arctic colony.

Data S1. Summary information on all analysed samples.

Data S2. Biometrics of all analysed specimens.

Supplementary Material

Supplementary data to this article can be found online at:

https://onlinelibrary.wiley.com/doi/full/10.1111/ibi.13153

Direct download links are as follows:

Appendix S1:
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fibi.13153&file=ibi13153-sup-0003-AppendixS1.docx

Supplementary Figures:
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fibi.13153&file=ibi13153-sup-0001-FigureS1-S10.docx

Supplementary Tables:
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fibi.13153&file=ibi13153-sup-0002-TableS1.docx

Supplementary Data Table 1:
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fibi.13153&file=ibi13153-sup-0004-DataS1.csv

Supplementary Data Table 2:
https://onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1111%2Fibi.13153&file=ibi13153-sup-0005-DataS2.csv

Paper III

Hybridization of Atlantic puffins in the Arctic coincides with 20th-century climate change

Paper IV

The Genomic Basis of Differentiation in the Atlantic Puffin