A 'JAUNT' ACROSS THE POND:

INVESTIGATING PATTERNS OF GENETIC DIFFFERENTIATION AND TRANS-ATLANTIC MIGRATION IN A TEMPERATE SEABIRD (*MORUS BASSANUS*)

by

Nate J. Clark

A thesis submitted to the Department of Biology

in conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

(December 2017)

Abstract

Our understanding of the mechanisms driving population divergence is dependent on our ability to effectively detect and interpret patterns of genetic differentiation. In natural populations, the evolutionary processes influencing genetic differentiation can be difficult to ascertain. Seabirds are highly mobile, yet often display population differentiation, suggesting intrinsic barriers to gene flow are important in driving their diversification.

The northern gannet (*Morus bassanus*), a piscivorous, colonial seabird, is widely distributed across the North Atlantic Ocean. Despite strong dispersal abilities, decades of banding studies indicate that the North Atlantic Ocean acts as an effective barrier for migration in northern gannets. However, recent telemetry revealed that some northern gannets migrate across the ocean each year, suggesting trans-Atlantic gene flow is possible.

To investigate population differentiation in northern gannets, I analyzed genomic patterns of diversity among eight North American and European colonies. Genetic structure was characterized using indices of genetic differentiation, principal component analysis, a Bayesian clustering method, and discriminant analysis of principal components. I found significant population genetic structure across the breeding range, corresponding primarily to the division between North American and European gannet populations. My analyses did not find significant differentiation among colonies on either side of the Atlantic.

I employed Bayesian approaches to estimate contemporary gene flow and divergence between North American and European gannet populations. I found low but

significant rates of migration from Europe into North America, but no migration from North America into Europe. My analyses suggest that North American and European populations diverged recently, potentially occupying separate refugia during the Last Glacial Maximum. Finally, I used outlier analysis and random forest feature selection to investigate fine-scale genetic structure, and identified loci that allowed population assignment of 85% of sampled northern gannets.

Low genetic differentiation in northern gannets may be due to shared variation following recent range expansion or to admixture resulting from overlap in nonbreeding distribution. My results suggest that the North Atlantic Ocean acts as a semi-permeable barrier to gene flow in northern gannets. The informative genetic markers I identified may allow colony-specific impact assessment and the development of targeted management strategies in northern gannets.

Co-Authorship

This thesis has been prepared in the traditional format as defined by the Queen's University School of Graduate Studies. Dr. Vicki Friesen was critical in research design, project development, and manuscript preparation, and financially supported much of this research. Dr. Tim Birt assisted with the conceptual design and a significant proportion of the laboratory work, and Dr. Gregory Robertson helped fund the project and provided input on the ecological aspects and conservation implications of this research. Drs. Bill Montevecchi, Steve Votier, Stefan Garthe and David Grémillet provided many of the samples that made this research possible.

Acknowledgements

This thesis would never have been completed without the support and friendship of the people I am so lucky to have around me. Foremost, I would like to thank my supervisor Vicki Friesen. Vicki was unerringly patient and generous as a supervisor during this project. She has provided invaluable assistance with every aspect of my thesis, and has always encouraged and inspired me to grow both as a biologist and as a person. She has been a calm voice when I am stressed, a stern voice when I am behind, a sure voice when I am uncertain and a kind voice always.

My thanks to Tim Birt, our resident lab guru, who devoted a significant amount of time and effort in the lab to make sure this project would succeed. I especially want to thank Tim for his patience and encouragement, and for always having a joke ready for when I needed it the most. Thanks also to Peter Davies and Rob Colautti for taking the time and energy to serve on my committee.

Many individuals and organizations have contributed to the gannet project over the years, providing the samples, expertise and funding that made my research possible: Steve Votier, David Grémillet, Stefan Garthe, Bill Montevecchi, David Cairns, Bryan Nelson, Robert Barrett, Sir Hugh Hamilton Dalrympel, l'Institut de biologie integrative et des systèmes (IBIS) de L'Université Laval, Environment and Climate Change Canada, Genome Quebec, Compute Canada and Queen's University.

I would like to thank everyone in the Friesen Lab for accepting and supporting me throughout my Master's thesis. Without their help, I can confidently say I would never have reached this point. I would like to extend special thanks to Hannah 'Hahna' Driver,

Brody Crosby, Lily Colston-Nepali, Russell Turner, Katie Birchard, and Ryan Franckowiak for their friendship, guidance, and encouragement. You have made this degree a truly enjoyable experience.

Finally, I would like to thank my family and friends for their unwavering support throughout this process. Thank you, Mom, Dad, Jason and Cynthia, for never letting me quit no matter how hard I tried. You have all been there to catch me when I fall, and I will never forget it.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ABC: Approximate Bayesian computation

AMOVA: Analysis of molecular variance

ANOVA: Analysis of variance

BIC: Bayesian information criterion

bp: Base pairs

DAPC: Discriminant analysis of principal components

ddRAD-seq: Double digest restriction site-associated DNA sequencing

H_E: Expected heterozygosity

H_O: Observed heterozygosity

IBIS: l'Institut de biologie integrative et des systèmes

K: Number of genetic clusters

kya: Thousand years ago

LGM: The Last Glacial Maximum

LD: Linkage disequilibrium

m: Estimate of contemporary migration rate

MCMC: Markov chain Monte Carlo

mtDNA: Mitochondrial DNA

mya: Million years ago

n: Sample size

Ne: Effective population size

p: Statistical probability value

P: Average frequency of the most common allele

PCA: Principal components analysis

PCR: Polymerase chain reaction

r^2: Coefficient of determination

RAD-seq: Restriction site-associated DNA sequencing

SNP: Single nucleotide polymorphism

t: Estimated time of divergence

α: Probability of rejecting the null hypothesis when it is true

π: Nucleotide diversity

%Poly: The percentage of sites that aree polymorphic in a population

Glossary

Batch effect - Sub-groups of measurements that have qualitatively different behaviour across conditions and are unrelated to the biological variables in a study (Leek *et al.*, 2010).

Breeding site fidelity or philopatry – Fidelity to the site of last breeding (Greenwood, 1980).

Decision tree – A statistical model mapping the consequences of a series of choices. Choices are made at nodes and consequences are mapped using branches from those nodes. As more decisions are made, the model grows.

Demographic independence – A state in which population dynamics are driven more by local death and birth rates than by immigration or emigration (Hanski, 1998).

Extrinsic barrier (to gene flow) – An environmental feature constraining the movement of genes between two groups of organisms.

Generation time – Many definitions are commonly used. Defined here as the weighted mean age of parents for a new-born cohort when the population has reached near stable age distribution (Gaillard *et al.*, 2005).

High-grading/upward grading bias – The overestimation of assignment success that results from developing and testing a panel of genetic markers on the same dataset (Waples, 2010).

Intrinsic barrier (to gene flow) – Any non-geographic factor constraining the contemporary movement of genes between two groups of organisms (e.g. behavioural incompatibilities).

Last Glacial Maximum (LGM) – The most recent interval in Earth history when global ice sheets reached their maximum integrated volume, beginning 26.4 to 33 kya and ending 19 to 20 kya in the Northern Hemisphere (Clark *et al.*, 2009).

Machine-learning – A basic form of artificial intelligence whereby a program can use pattern recognition on a dataset to inform future analysis without explicitly being programmed to do so.

Natal site fidelity or philopatry – Fidelity to the site of birth beyond the age of independence (Greenwood, 1980).

Quaternary Period – A geological subdivision of the Cenozoic Era, beginning 2.58 mya and continuing to present day (Head & Gibbard, 2015).

Refugia (refugium s.) **–** Regions in which isolated or relict populations of organisms can persist during periods of unfavorable conditions.

Site fidelity – The tendency to frequent the same location.

Chapter 1: Introduction and Literature Review

1.1 General introduction

Contemporary patterns of genetic diversity represent a 'snapshot' into the long and often complicated evolutionary history of a species. Population genetic structure is determined by the historical and contemporary interplay between the evolutionary forces of selection, mutation, gene flow and genetic drift (Wright, 1931, 1951). Detecting and interpreting patterns of genetic differentiation can therefore provide insight into these fundamental processes, not only revealing the mechanisms driving population divergence in the present, but also reflecting a species' evolutionary history and informing conservation efforts in the future. In natural populations, however, the factors promoting or preventing genetic differentiation can be multifaceted, difficult to characterize and poorly understood.

Under an allopatric model of divergence, **extrinsic barriers**[1] to gene flow facilitate the differentiation of populations through drift and selection. Seabirds are generally characterized by a high dispersal ability, undertaking some of the longest recorded animal migrations (e.g. Egevang *et al.*, 2010), and thus encountering few apparent extrinsic barriers to gene flow. In many species of seabirds, however, populations can display striking levels of genetic, morphological and behavioural differentiation (reviewed in Friesen *et al.* 2007; Friesen 2015). The apparent conflict between high vagility and low genetically effective dispersal in seabirds has long been recognized and has been termed the seabird 'paradox' or 'enigma' when discussing the

[1] Definitions for words in bold are found in the glossary.

importance of extrinsic barriers to gene flow in population divergence (Steeves *et al.*, 2005; Friesen *et al.*, 2007; Milot *et al.*, 2008). This 'paradox' suggests that **intrinsic barriers** to gene flow must play important roles in the contemporary evolution of seabirds. Seabirds therefore represent a useful study system for investigating the non-physical or cryptic mechanisms driving population differentiation across a wide range of organisms (Friesen *et al.*, 2007).

This thesis is a study of population differentiation in a widely distributed, colonial species of seabird, the northern gannet (*Morus bassanus;* hereafter "gannet"). In this general introduction, I aim to (i) review the role of glaciation and historical association in shaping patterns of genetic diversity in temperate and polar species, (ii) discuss the importance of site fidelity in driving and maintaining differentiation in contemporary populations, (iii) examine the utility of population assignment tests in the study and conservation of species characterized by low overall genetic differentiation and (iv) review what is currently known about gannet dispersal and population genetics.

1.2 Patterns of genetic variation in the Northern Hemisphere

For species inhabiting high latitudes, contemporary patterns of genetic diversity may reflect recolonization history and the presence of past extrinsic barriers, more than recent demography or adaptation (Hewitt, 2004). During the **Quaternary Period**, the expansion of the Cordilleran, Laurentide and Scandinavian ice sheets rendered the majority of Northern Europe and North America uninhabitable for many organisms, restricting those species' ranges to ice-free **refugia** (Hewitt, 2000). Under the Expansion-Contraction biogeographic model, temperate species were constrained to lower latitudes

during periods of cooling, and expanded to higher latitudes following glacial retreat and postglacial warming (Provan & Bennett, 2008). Following a range expansion, if the newly established populations have not undergone bottlenecks, genetic structure is expected to be low until sufficient time has passed for drift and selection to erase shared ancestral variation (Slatkin, 1985, 1987; Birky *et al.*, 1989). Understanding the mechanisms driving population differentiation in species inhabiting the Northern Hemisphere is therefore complicated by the repeated range expansions and contractions that many of these taxa may have experienced (Hewitt 2000; Alberdi *et al.* 2015; Vargas-Rodriguez *et al.* 2015).

Many temperate and polar seabird species have colonies occurring in areas that were either covered or isolated by ice during the **Last Glacial Maximum** (LGM). As most species of seabirds do not fly over ice, gene flow between populations inhabiting distinct refugia would have been restricted, and many contemporary colonies could only have been recolonized following glacial retreat (Morris-Pocock *et al.*, 2008; Pereira & Baker, 2008). Consequently, climate zone appears to be a strong predictor of the extent of population structure in seabirds (Friesen, 2015). In the Northern Hemisphere many high latitude seabirds display weak genetic differentiation, which has often been attributed to recolonization from a single, shared refugium during the Pleistocene epoch (e.g. Milot *et al.* 2008; Bicknell *et al.* 2012; Yannic *et al.* 2016). Population genetic structure in other species may reflect differentiation by isolation in separate refugia or hierarchal recolonization history (Sternkopf *et al.*, 2010). These patterns of genetic diversity, however, are often similar to what is expected under several contemporary evolutionary and demographic scenarios (e.g. local adaptation; genetic connectivity). Indeed, these

patterns may even be maintained today by mechanisms resulting from the isolation or connectivity experienced by populations during past climatic cycles (Friesen *et al.*, 2007). Uncovering and distinguishing between historical and contemporary demographic and selective processes has therefore proven difficult, particularly in highly vagile organisms, like seabirds.

1.3 Site fidelity

In contemporary populations, intrinsic reproductive barriers may prevent the distributions of populations from overlapping, regardless of their geographic proximity. The isolation resulting from these barriers will reduce gene flow between populations and may influence population genetic structure. **Site fidelity** is expected to be selectively advantageous if the quality of a site is predictable in time (Switzer, 1993). The 'quality' of a site may be determined by the availability of resources at the site (Switzer, 1993), or by the intensity of competition for those resources (Wilson & Yoshimura, 1994). Importantly, the strength of site fidelity may vary between populations or individuals within a population, potentially influencing ecological and evolutionary dynamics between groups. In seabirds, site fidelity takes several forms (see below), with many species exhibiting more than one type (e.g. Fifield *et al.* 2014). Its prevalence suggests that site fidelity may represent a common intrinsic barrier to gene flow in seabirds and may play an important role in the evolution of seabird diversity.

Banding and telemetry studies indicate that seabirds generally display strong **natal and breeding site philopatry** (but see Coulson 2016). By promoting isolation during the breeding season, philopatry is expected to reduce gene flow between colonies.

However, not all species with reportedly strong philopatry exhibit population genetic structure (e.g. common murres, *Uria aalge*; Tigano *et al.* 2015). Indeed, in their review of the mechanisms driving speciation in seabirds, Friesen *et al.* (2015) found that across many seabird taxa, philopatry alone may be a sufficient barrier to gene flow but most often acts in combination with other barriers to drive differentiation (e.g. Rayner *et al.* 2011).

For many species of seabirds, nonbreeding distribution appears to be a strong predictor of population differentiation (reviewed in Friesen *et al.* 2007; Friesen 2015; Munro & Burg 2017). If populations are isolated by philopatry during the breeding season, and individuals remain near their breeding colony or travel to distinct wintering grounds during the nonbreeding season (i.e. have winter-site fidelity), the opportunity for gene flow between colonies is limited. Conversely, overlap of populations during the nonbreeding season may result in gene flow, which could counteract geographical and behavioural isolation during the breeding months.

Accordingly, species with populations that exhibit divergent migratory behaviours or occupy distinct wintering grounds commonly have stronger genetic structure than species with individuals that simply disperse or that migrate to shared nonbreeding areas (Friesen *et al.*, 2007; Friesen, 2015). Although some studies have associated the use of different nonbreeding sites with asynchronies in breeding phenology (Rayner *et al.*, 2011), for most species exactly how connectivity during the nonbreeding season influences population genetic structure remains unclear.

1.4 Population assignment

The genetic assignment of individuals to their population of origin is essential to our understanding and characterization of contemporary ecological and evolutionary processes (Manel *et al.*, 2005). Dispersal between populations can influence population growth, gene flow, and, ultimately, species persistence (Lowe & Allendorf, 2010). Assignment tests calculate the explicit probability that an individual's multilocus genotype originated in each of a set of predefined populations (Wilson *et al.*, 2003; Piry *et al.*, 2004). Genetic assignments are commonly employed to assess population connectivity, as they allow indirect estimates of dispersal at scales not feasible through direct measurement alone (Waser & Strobeck, 1998). Amongst its many uses, population assignment has recently been employed to detect hybridization, assess population boundaries, characterize source-sink dynamics, estimate migration rates, identify cryptic species, improve fisheries stock assessment and aid with wildlife forensics (Wasser *et al.*, 2015; Leys *et al.*, 2016; Pritchard *et al.*, 2016; Cure *et al.*, 2017).

By providing estimates of population genetic connectivity, assignment tests can facilitate effective species conservation and management. Accurate genetic assignment allows managers to determine which populations are impacted by disturbance and to establish the effect of local stressors on population decline (Milano *et al.*, 2014; Ardren & Bernall, 2017; Domínguez *et al.*, 2017; Ng *et al.*, 2017; Storfer *et al.*, 2017). Assignment methods are particularly important for the management and conservation of migratory species, where populations may overlap in nonbreeding range and may be differentially impacted by stressors along alternative migratory routes (Webster *et al.*, 2002; Ruegg *et al.*, 2014; Baerwald *et al.*, 2016).

The efficacy of population assignment tests is often dependent on the number of individuals being assigned, the number of markers employed, and the degree of differentiation between the populations being compared. Accurate assignment using traditional molecular methods has therefore proven difficult when overall population differentiation is low (Cornuet & Luikart, 1996; Berry *et al.*, 2004).

The advent of high-throughput sequencing represented a revolution in molecular evolution and ecology. Recently developed techniques, like restriction site-associated DNA sequencing (RAD-seq), allow rapid and robust discovery and genotyping of thousands of markers across the genome. Comparisons across such a high number of loci increase the resolution at which we can examine evolutionary processes and characterize demography. By selecting a subset of informative markers from a larger pool, population assignment should be possible even for species with low overall genetic differentiation (Paetkau *et al.*, 2004).

1.4.1 Outlier detection for population assignment

Theoretical expectations regarding the effects of selection on allelic frequencies can be used to identify the loci important for accurate population assignment. Genome scans for selection attempt to identify putatively adaptive variation by screening for unusual differences in allele frequencies between populations. The nature of the loci, the measure of differentiation (F_{ST}, Tajima's D, d_{xy} etc.) and the groups used in the comparison are dependent on the specific questions being addressed (Wolf & Ellegren, 2016). Loci that exhibit extreme differences in allele frequency are often referred to as outliers. Divergent selection, resulting from heterogeneous environmental conditions, is

predicted to drive local adaptation and maintain adaptive variation between populations (Felsenstein, 1976; Hedrick *et al.*, 1976). Identifying outliers potentially resulting from strong divergent selection may therefore be useful in assigning individuals to their population of origin.

A major assumption of many outlier detection methods is that most genetic variation is neutral, and that outliers, by deviating from neutral expectations, have experienced relatively strong selection (Wolf & Ellegren, 2016). However, this assumption is easily violated, as various other evolutionary and demographic processes (e.g. population bottlenecks) can produce genomic signatures like those generated by selection (Savolainen *et al.*, 2013).

Although not necessarily under selection, outlier loci are by definition, those markers that exhibit the most extreme differences in allele frequency between populations. This implies that, even without functional annotation or validation, outliers can still be useful for population assignment, allowing researchers to examine subtle genetic structure in otherwise homogeneous populations (Allendorf *et al.*, 2010). Indeed, this approach has been successfully employed to investigate complex demography, uncover fine-scale genetic structure, and improve population assignment across a wide range of non-model species (Ogden *et al.*, 2013; Fraïsse *et al.*, 2014; Barley *et al.*, 2015; Benestan *et al.*, 2015; Candy *et al.*, 2015; Oliveira *et al.*, 2015; Cahill & Levinton, 2016; Foote *et al.*, 2016; Tigano *et al.*, 2017).

1.4.2 Random forest feature selection for population assignment

An emerging alternative to differentiation-based approaches is to rank the importance of a locus specifically for the purposes of population assignment. Random forest is a **machine-learning** method that constructs **decision trees** by assessing the importance of a subset of features at each individual node. When random forest is applied to population assignment, each genetic marker (e.g. SNP) is given an importance value in relation to the mean decrease in accuracy (MDA) associated with its removal from a series of assignment tests. This MDA value can then be used in place of traditional metrics like F_{ST} to select those markers that allow accurate classification. By iteratively constructing decision trees, random forest algorithms consider loci in various combinations rather than assessing the impact of each individual locus. Random forest algorithms have been successfully employed to assign individuals to their population of origin across several taxa (Stephan *et al.*, 2015; Sylvester *et al.*, 2017; Jacobs *et al., in press*) but remain a relatively underutilized alternative to F_{ST}-based assignment.

1.5 Study system – the northern gannet

The northern gannet is a temperate, colonial seabird with breeding colonies on islands distributed across the North Atlantic Ocean (Figure 1a; Nelson 2002). It is one of three extant species of gannets (order Pelecaniformes; family Sulidae) and is the largest breeding seabird in the North Atlantic (Nelson 1978). The world population is approximately 1,069,000 individuals (835,000 in Europe; 234,000 in North America), with population sizes varying greatly among colonies (Chardine *et al.*, 2013; Murray *et al.*, 2015). Gannets begin breeding between the ages of five and seven and pair

monogamously, with successful pairs producing one offspring each breeding season (Nelson 1978). Incubation time for gannets last approximately 6 weeks and is typically followed by 12 – 13 weeks of biparental chick-rearing (Nelson 2002). Gannets are medium range foragers, often feeding within a few hundred kilometers of the breeding colony (Hamer *et al.*, 2007; Montevecchi *et al.*, 2009). They employ flexible foraging tactics, exploiting a broad range of pelagic fish and competing with other species for discards from commercial fishing vessels (Hamer *et al.*, 2000; Montevecchi *et al.*, 2009; Votier *et al.*, 2010). Northern gannets will forage using a variety of different techniques, including surface feeding, vertical plunge-diving and underwater pursuit (Garthe *et al.*, 2000). Gannets are well-studied, widely distributed, abundant apex predators, marking them as a possible indicator species for marine ecosystem health (Montevecchi *et al.*, 2012a; Lescroel *et al.*, 2016).

1.5.1 Northern gannet migration

Like many northern seabirds, gannets migrate south during the winter (Figure 1b). On average they travel between 3500-4000 km per migration, and wintering sites appear to be selected based on food availability (Fort *et al.*, 2012; Fifield *et al.*, 2014). Individual gannets display strong breeding and natal philopatry (Nelson, 2002), and recent telemetry has revealed remarkable winter site-fidelity as well (Fifield *et al.*, 2014). Although there is evidence of population-specific migration strategies, tracking has revealed a high degree in overlap in the nonbreeding range among North American and among European colonies of gannets (Fort *et al.*, 2012; Fifield *et al.*, 2014).

Until recently, decades of mark-recapture studies suggested that the North Atlantic acts as an effective barrier to gene flow between North American and European colonies of gannets. By 2008, only nine gannets banded in North America had been recovered in eastern Atlantic waters (Gaston *et al.*, 2008). Despite substantially greater banding effort, only one European gannet has ever been found in western Atlantic waters (62,328 European vs. 13,946 North American bandings; Fort *et al.*, 2012; Stauss *et al.*, 2012).

Recently, a geolocation study revealed three adult North American gannets migrating to the west coast of Europe and Africa during the nonbreeding season (7% of tracked birds; Fifield *et al.*, 2014). The recorded trans-Atlantic crossings were as rapid as 5 days and most of the gannets did not rest while crossing the ocean. A significant number of European gannets winter along the west coast of Europe and Africa (Kubetzki *et al.*, 2009; Fort *et al.*, 2012). This suggests that some North American gannets may be wintering with European gannets and opens the possibility of gene flow across the Atlantic. Although gannets are not known to form pair-bonds on wintering-grounds and all three gannets returned to North America to breed, Fifield *et al.* (2014) suggested that the gannet's group migration strategy may result in immature gannets following adults to North American colonies. Again, despite significantly greater tracking efforts (106 vs. 46 individuals), no European gannets have ever been tracked migrating to western Atlantic waters during the nonbreeding season (Kubetzki *et al.* 2009; Fort *et al.*, 2012; Stauss *et al.*, 2012). Together, mark-recapture and tracking studies suggest limited opportunity for gene flow between North American and European populations of gannets.

1.5.2 Previous genetic analysis in the northern gannet

Friesen *et al.* (*unpublished*) were the first to assess the extent of differentiation among populations of gannets using mitochondrial DNA (mtDNA) control region sequences. They reported that North American and European populations of gannets were significantly differentiated, with no evidence of differentiation among either North American or European colonies. They also found no evidence of female-mediated gene flow between Europe and North America, estimating that the mitochondrial lineages of these two populations may have diverged between 22 and 86 thousand years ago (kya). These divergence dates suggest that North American and European populations of gannets may have survived the last Pleistocene glaciation in separate refugia, consistent with results for other species of North Atlantic seabirds (e.g. common murres, Morris-Pocock *et al.*, 2008).

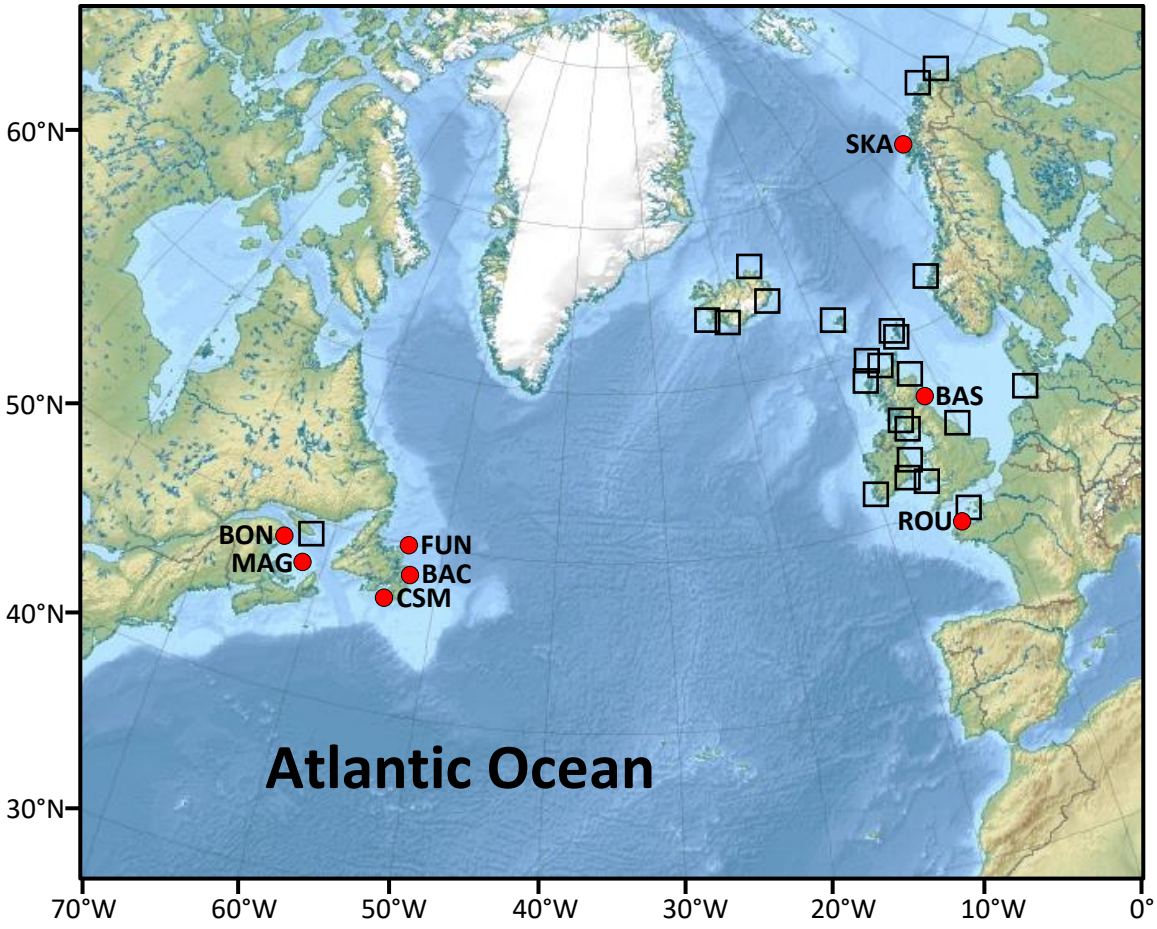
1.6 Research objectives

Using high-throughput sequencing I analyzed patterns of genome-wide variation in gannets from eight colonies across the species' breeding range. This thesis had three broad objectives: 1) quantify genetic variation within and among colonies of northern gannets, 2) investigate the roles of contemporary gene flow and historical association in driving patterns of genetic differentiation across the northern gannet's range, and 3) develop a panel of informative genetic markers that will allow the assignment of individual gannets to their colony of origin.

Given results of mark-recapture and tracking studies, as well as previous genetic analysis, I hypothesize that European and North American colonies of gannets are

genetically differentiated, and that little gene flow occurs between these populations. Evidence from mtDNA further suggests that these populations diverged before the end of the LGM, possibly inhabiting separate refugia. Given evidence that gannet populations diverged recently, have a high degree of overlap in nonbreeding range and are not differentiated at mtDNA markers, I hypothesize that, despite reportedly high philopatry, little differentiation exists either among North American or among European colonies.

a)


60°N
50°N
40°N
30°N
70°W
60°W
50°W
40°W
30°W
20°W
10°W
0°
SKA
BAS
ROU
BON
MAG
FUN
BAC
CSM
Atlantic Ocean

b)

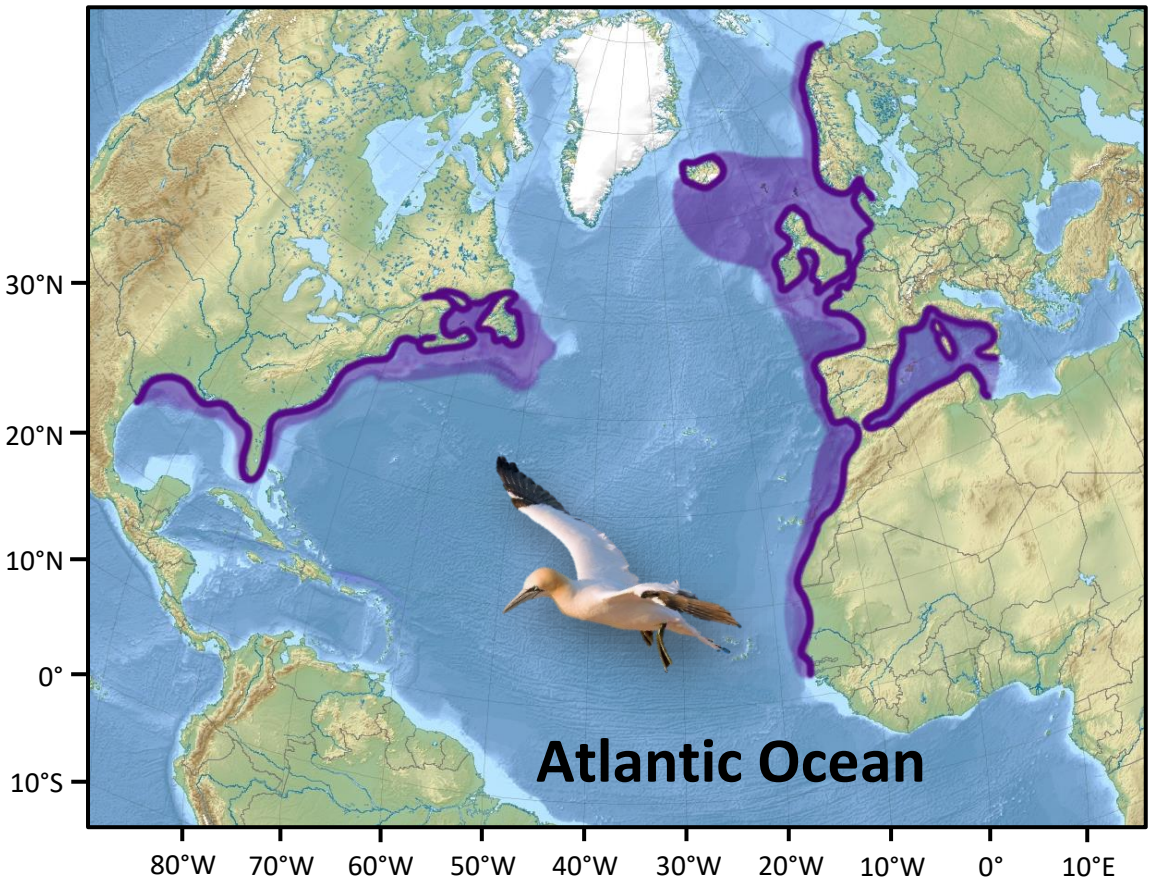


Figure 1. **a)** Map showing sampling locations for northern gannets. Red circles correspond to the individual colonies sampled for this study. Black squares represent breeding localities not sampled in this study. See Table 1 for sample sizes. **b)** Map showing approximate non-breeding distribution of northern gannets (Kubetzki *et al.*, 2009; Fort *et al.*, 2012; Fifield *et al.*, 2014; Garthe *et al.*, 2016). BAC, Baccalieu I.; Bon, Bonaventure I.; CSM, Cape St. Mary's; FUN, Funk I.; MAG, Magdalen I.; BAS, Bass Rock; ROU, Rouzic I.; SKA, Skarveklakken. Image: Andreas Trepte.

Chapter 2: Materials and methods

2.1 Sampling

Blood samples were collected from 117 gannets from eight colonies throughout the species' breeding range (Figure 1a and Table 1). All samples are archived at Queen's University, Kingston, Ontario. DNA was extracted following a standard proteinase-K/phenol-cholorform procedure and purified using ethanol precipitation (Sambrook *et al.* 1989). DNA concentrations were determined using a Qubit® 3.0 Fluorometer and Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Massachusetts, USA). Each sample was normalized to 20 ng/µL for use in library preparation.

2.2 Library preparation

Double-digest restriction site-associated DNA sequencing (ddRAD-seq; Peterson *et al.*, 2012) libraries were prepared at l'Institut de biologie integrative et des systèmes (IBIS) de L'Université Laval following the protocols outlined in Poland *et al.* (2012). Briefly, two restriction enzymes (*Sbf*I and *Msp*I) were used to digest 200 ng of normalized genomic DNA. Digested samples were then ligated to adaptors barcoded with four to eight base pairs (bp) for individual assignment. Uniquely barcoded samples were pooled and fragments 200 to 500 bp in length were selected using a BluePippin (Sage Science, Inc., Beverly, MA, USA). Size-selected fragments were then amplified via a high-fidelity PCR for 18 cycles. Libraries were sequenced in two lanes of Illumina HiSeq2500 with 100 bp, single-end sequencing at the McGill University-Génome Québec Innovation Centre in Quebec, Canada. As my two libraries were prepared and

sequenced at different times, inclusion of duplicate individuals in both lanes allowed removal of any **batch effects** associated with sequencing date. Upon receiving the sequenced reads, I checked the quality of the raw data using the program FASTQC (Andrews, 2010).

2.3 SNP and genotype calling

Raw reads were demultiplexed, trimmed to a length of 90 bp to remove low quality bases at the 3' end, and filtered for uncalled bases, low average quality score (Phred score < 10) and correct barcode sequence, using the program *process_radtags* from the STACKS pipeline version 1.44 (Catchen, 2013). Single nucleotide polymorphisms (SNPs) were called and catalogued using the built-in STACKS pipeline *denovo_map*. Briefly, *denovo_map* aligned identical reads into putative alleles. Alleles were then compared to identify putative loci and SNPs were called at each locus. Loci were first identified in each sample and then assembled into a catalog for each population. Using the *populations* module of STACKS, loci were processed and filtered, retaining only those SNPs that had a sequencing depth greater than 9X, were present in at least 70% of individuals in each population and were genotyped in five of eight populations. To avoid biases in estimates of genetic connectivity and to minimize detection of false positives in subsequent outlier analyses, SNPs with a global minor allele frequency of less than 5% were removed (Roesti *et al.*, 2012). To minimize linkage disequilibrium (LD), only one random SNP was retained for each locus. All SNPs missing in more than 70% of individuals and all individuals missing data for more than 30% of loci were subsequently removed from the dataset using VCFtools version 0.1.14

(Danecek *et al.*, 2011). Pairwise LD was calculated using VCFtools and one SNP from each pair with a coefficient of determination (i.e. r^2) above 0.5 was excluded from downstream analysis (Carlson *et al.*, 2004; Kawakami *et al.*, 2014; Benestan *et al.*, 2015).

I performed sensitivity analysis in STACKS by adjusting the value of a single parameter and holding all other parameters constant. The aim of the analysis was to find the value of a parameter that allowed the highest level of stringency while simultaneously maintaining a high number of informative polymorphic loci (Table 2). Specifically, I determined how adjusting the value used for certain parameters influenced: 1) the average number of SNPs retained per individual, 2) the average sequencing depth per individual, 3) the average percentage of missing data per individual and 4) the pairwise F_{ST} values between my colonies.

2.4 Summary statistics

I calculated global and population summary statistics (observed [H_O] and expected heterozygosity [H_E], major allele frequency [P], percentage of polymorphic sites [% Poly], Wright's inbreeding coefficients F_{IS} and F_{ST}, and nucleotide diversity [π]) using the *populations* module in STACKS, BAYESASS 3.0.4 (Wilson *et al.*, 2003) and Arlequin version 3.5 (Excoffier & Lischer, 2010). All statistical calculations were bootstrapped for a minimum of 10,000 iterations to test for significance. I performed an analysis of molecular variance (AMOVA) in Arlequin to test for significant global differentiation. I conducted an Analysis of variance (ANOVA) on all remaining statistics to determine whether average values differed significantly across populations. All

resulting p-values were adjusted using Benjamini-Hochberg correction for multiple comparisons (Benjamini & Hochberg, 1995).

2.5 Clustering analyses

I assessed the most probable number of genetic clusters in my dataset using three clustering methods. I first performed principal component analysis (PCA) using the R package *adegenet* (Jombart & Ahmed, 2011). PCA aims to summarise the global variability within a dataset and does not rely on *a priori* assignment of individuals to populations. Note, however, that a PCA includes both within and between individual genetic diversity and is therefore not a direct measure of population structure but rather overall genetic structure.

I next used the program STRUCTURE version 2.3.4 (Pritchard *et al.*, 2000) distributing runs across multiple cores using the *ParallelStructure* R package (Besnier & Glover, 2013). The program uses a Bayesian clustering method to find groupings that will minimize deviations from Hardy-Weinberg and linkage equilibrium, and determines the probability of assignment for each individual's multilocus genotype to those clusters. STRUCTURE was run using the admixture model without location as *a priori* knowledge and with correlated allele frequencies. I used a burn-in of 1,000,000 iterations and a run length of 10,000,000 steps to allow the Markov chain Monte Carlo algorithm to stabilize. This was repeated 10 times for each value of K (the number of potential clusters) ranging from one to eight. I assessed the most probable number of K using the *ad-hoc* statistic ΔK (Evanno *et al.*, 2005) calculated using STRUCTURE HARVESTER (Earl & vonHoldt, 2012) and DISTRUCT 1.1 (Rosenberg, 2004).

Finally, I performed discriminant analysis of principal components (DAPC; Jombart & Ahmed, 2011) also in the R package *adegenet.* I began my analysis by conducting a K-means clustering of principal components to produce clusters of individuals that minimize within-group genetic variation, while simultaneously maximizing between-group variation. Using this approach, each value of K will produce a model with an associated likelihood. A Bayesian information criterion (BIC) can then be used to assess the best supported number of clusters without prior population information. I then performed a DAPC using the number of clusters indicated by the BIC.

2.6 Outlier detection

I performed outlier detection using the FDist2 (Beaumont & Nichols, 1996) program implemented in LOSITAN (Antao *et al.*, 2008). LOSITAN identifies outliers by comparing observed F_{ST} values to a simulated distribution of expected F_{ST} values, calculated as a function of expected heterozygosity (H_E) under an island model of migration. Loci with observed F_{ST} values outside of this distribution are labelled as putatively under directional or purifying selection. Each population pairwise comparison was conducted four times, with every simulation running for 1,000,000 iterations. Mean neutral F_{ST} was estimated for each run by iteratively identifying and removing loci putatively under selection. Each pairwise comparison used a subsample size of eight individuals. Once the four runs were completed, I selected only those loci that were reported as outliers and had a probability of 1.0 of deviating from neutral expectations across multiple independent runs. To eliminate bias associated with missing data, all SNPs used in this analysis were present in all individuals (n = 898 SNPs).

I ranked outlier loci by F_{ST} value and assessed assignment success at three arbitrary rankings ($F_{ST} \geq 0.5$, 0.1 and 0.2). Assignment tests were performed using a panel of SNPs taken from all population pairwise comparisons, as well as from North American and European populations independently.

I employed a randomization approach to determine whether selecting loci with extreme F_{ST} values for population assignment is a circular method and to test whether my selected panels of outliers were informative of fine-scale population structure among my sampled gannet populations (Benestan *et al.*, 2015; Campagna *et al.*, 2015; Tigano *et al.*, 2017). I randomly reassigned individuals to eight equal populations and, following the protocol outlined above, performed outlier analysis on the resulting dataset. I then tested population assignment in GeneClass 2.0 using the randomized SNP panels, comparing the assignment success rates between the randomized and actual datasets. If the pattern of differentiation we detect using this method is driven by a random process, I would expect to have high assignment success for the random populations, as well as the original dataset.

I further improved population assignment success by performing a jackknife resampling procedure on the selected outlier SNPs. By iteratively removing SNPs from the selected panel, I could assess the individual impact of each marker on overall assignment success. If removing a SNP positively impacted assignment success, I removed it from the dataset. Resampling occurred only after the outlier panel was compared to the randomized panel to ensure that both datasets were treated equally up until that point in the analysis.

2.7 Random forest feature selection

Regularized random forest analysis was run using R package *RRF* (Deng & Runger, 2012). Regularized random forest is a stringent application of random forest, introducing a penalty coefficient (λ) to each feature (SNP) at a decision node. In *RRF*, the penalty coefficient (default λ = 0.8) is defined by the user. The coefficient will influence the size of the selected feature panel and should ensure that each feature is highly informative. For a SNP to be selected for the assignment panel, that marker must be more informative than the other SNPs previously selected in the panel, as well as the SNPs being compared at the node, despite the applied penalty. For each pairwise comparison I ran 15 independent runs, generating 15 lists of SNPs ranked by MDA.

Ideally, an assignment test would use a subset of samples for marker selection and a second, independent subset to conduct the assignment analysis (Anderson, 2010). Dividing a dataset into two independent groups effectively reduces **high-grade or upward grading bias** (Waples, 2010). However, when sample sizes are relatively low, it may not be feasible to have two independent subsets of samples. As such, Anderson (2010) suggested that a subset of samples (training-set) is used to select the markers and the complete dataset is used to test assignment power. Accordingly, for each pairwise comparison I randomly selected 50% of individuals from each population for use as a training-set. Each run generated 8,000,000 decision trees and allowed twice the default number of features to be considered at each node (*mtry* = 2(√(number of features))). All other parameters were set to default values. Only those SNPs with positive MDA values across multiple independent runs were selected for subsequent population assignment. To eliminate bias associated with missing data, all SNPs used in this analysis were present in

all individuals (n = 898 SNPs). To compare the performance of this method to outlier analysis under low overall levels of differentiation, I performed pairwise comparisons among North American and among European gannet colonies.

2.8 Population assignment tests

Assignment tests were conducted using the program GeneClass 2.0 (Piry *et al.*, 2004). A Bayesian probability approach and Monte Carlo re-sampling computation (100,000 simulated individuals; Paetkau *et al.*, 2004) were used to assess the probability that each multilocus genotype originated from a specific colony (alpha = 0.01). Additionally, STRUCTURE and DAPC analyses were used to assess the assignment power of the selected SNP panels, using location as *a priori* knowledge. STRUCTURE was run assuming no admixture and correlated allele frequencies. To ensure that my DAPC was discriminating and stable, I assessed the optimal number of principal components to retain using the α-score calculated in *adegenet*.

2.9 Estimations of contemporary gene flow

Contemporary migration rates between European versus North American populations of gannets were estimated using a Bayesian approach implemented in a modified version of BAYESASS 3.0.4 that could process thousands of SNPs at once (Wilson *et al.*, 2003). BAYESASS uses genotypic disequilibrium to identify individuals with recent migrant ancestry (first- and second-generation) in each sample. Using the frequency of these individuals, the program then implements a Bayesian framework with

a Markov chain Monte Carlo (MCMC) to estimate contemporary migration rates (m). For each run, I performed 9,000,000 MCMC iterations with a burnin of 900,000. Delta values were adjusted to optimize terminal acceptance rates (30% - 60%; Wilson *et al.*, 2003). Mixing and convergence were assessed using the program TRACER v1.6 (Rambaut et al. 2014). Three independent runs, each using a different seed, were performed to check for consistent convergence in estimates. The estimates produced using different seeds were very similar and were not averaged across runs (Appendix A3).

2.10 Estimations of population divergence

To obtain estimates of divergence time for the North American and European populations of gannets, I used approximate Bayesian computation (ABC), implemented in the program DIYABC (Cornuet *et al.*, 2008). The ABC approach enables the testing of complex population genetic models by simulating predefined scenarios and comparing the simulated summary statistics to those of the observed dataset (Beaumont *et al.*, 2002).

Recently, Cabrera and Palsbøll (2017) demonstrated that DIYABC is more effective at accurately resolving simple genetic models than complex ones. Accordingly, I compared three simple scenarios, all proposing that European and North American gannets originated in one population, and diverged into two populations at some point in the past (Appendix A4). Scenarios differed only in the proposed effect of divergence on effective population size. Priors for the simulation were initially uniform and broad to account for the lack of available estimates of effective population sizes and divergence time (Ne = 10 – 1,000,000; t = 10 – 1,000,000). Each scenario was simulated 1,000,000

times and were compared by estimating posterior probabilities using logistic regression (Cornuet *et al.*, 2014).

DIYABC does not explicitly address contemporary gene flow, which if unaccounted for, may result in underestimates of divergence time. To avoid including recent migrants in the simulation, first-generation migrants detected by GeneClass 2.0 were removed from the analysis (n = 1).

By default, DIYABC outputs divergence times in generations. Unfortunately, to my knowledge no formal estimates of **generation time** exist for the northern gannet. Age of first breeding for gannets is commonly reported at five years (Nelson, 1978), however breeding success and survivorship are typically lower in juveniles (Nelson 2002; Wanless *et al.*, 2006), suggesting that cohort replacement will take longer. The mean generation time of a population is defined as

$$\bar{T} = \sum_{i=1}^{\infty} i l_i f_i \lambda^{-i}$$

where

$$l_i = \prod_{j=1}^{i} s_j$$

is the probability of surviving from birth to age i, f_i the annual fecundity at age i, and λ is the annual growth rate of the population (Leslie, 1966; Gaillard *et al.*, 2005; Niel & Lebreton, 2005). Nelson (1978) reported that inexperienced or first-time breeders typically have an annual breeding success of 0.53 chicks per annum (pa) and that the average life span of a gannet is approximately 16.2 years. Estimates of age-specific survival, adult breeding success (0.72 chicks pa) and global population growth rate (2%

pa) were taken from Wanless *et al.*, (2006). If it is assumed that fecundity and survival remain constant beyond the sixth year (Dillingham, 2010), gannets have a mean generation time of 11.4 years. This estimate was used to analyze DIYABC divergence times.

Table 1. Northern gannet colonies sampled for this study, initial and (final) sample sizes (n), geographic coordinates and population abbreviations.

Colony	***N***	**Latitude**	**Longitude**	**Abbreviation**
Baccalieu Island	9 (9)	48.1263° N	52.7990° W	BAC
Bonaventure Island	16 (10)	48.4943° N	64.1608° W	BON
Cape St. Mary's	9 (9)	46.8232° N	54.1922° W	CSM
Funk Island	10 (9)	49.7569° N	53.1811° W	FUN
Magdalen Island	20 (20)	47.3877° N	61.9012 ° W	MAG
Bass Rock	10 (8)	56.0773 ° N	2.6408 ° W	BAS
Rouzic Island	22 (21)	48.8992° N	3.4373° W	ROU
Skarveklakken	21 (19)	69.1586° N	15.6710° E	SKA

Table 2. Parameters adjusted for sensitivity analysis in the pipeline STACKS (Catchen *et al.*, 2013). The program used in the analysis, the parameter adjusted for the analysis, the function of the specified parameter, the range of values analyzed, and the final value selected for the parameter.

Program	**Parameter**	**Function**	**Range**	**Value selected**
process_radtags	-s	Discard reads with an average Phred score below specified value.	10 - 30	10
	--barcode_dist_1	The allowable distance between barcodes to allow for barcode rescue.	1-3	1
denovo_map	-m	The minimum number of identical reads required to create a stack (allele) when processing an individual.	2-8	4
	-M	The maximum number of mismatches allowed between loci when processing an individual.	2-8	4
	-n	The maximum number of mismatches allowed between loci when building the catalog.	2-8	6
populations	-r	The minimum proportion of individuals in a population a locus must be present in to process the locus.	0.6 – 0.9	0.7
	-p	The minimum number of populations a locus must be present in to process the locus.	5-8	5
	-m	The minimum stack depth required to process a locus.	6-10	9

Chapter 3: Results

3.1 Sequence quality and filtering

I obtained a total of 260,193,827 raw single-end reads, representing an average of approximately 2,000,000 reads per individual. Quality of reads was high, with an average phred score per read of 36. The catalog built by STACKS consisted of 46,613 loci. After filtering in *populations,* 2752 SNPs were found in at least five of the eight populations and 70% of individuals. Filtering to remove markers missing more than 30% of data further reduced the number of retained loci to 2190 SNPs. VCFtools detected a further 140 pairs of SNPs (involving 83 unique markers) with an r^2 value above 0.5. Several of these markers were in LD with the same SNP, necessitating the removal of an additional 45 SNPs. After applying all filtering steps, 105 gannet individuals (Table 1) and 2155 SNPs were retained across all populations.

3.2 Batch effect removal

I detected a batch effect, in that individuals sequenced in different lanes grouped together in my clustering analyses regardless of geographic origin (Appendix A1). Although the cause is not known with certainty, a possible explanation is slight differences in fragment size selection during library preparation (Davey *et al.*, 2011; Andrews *et al.*, 2016). Because several individuals were included in both batches as duplicates (i.e. individuals sequenced in both lanes), I was able to develop a list of loci with alleles unique to either lane. I removed these loci from the full dataset using a blacklist (-B) in the *populations* module in STACKS. I then selected an equal number of individuals from colonies sequenced in both lanes and performed an AMOVA to ensure

that sequencing lane accounted for a limited proportion of the variance among my samples (Appendix A2).

3.3 Summary statistics

Most summary statistics had similar values among the eight gannet colonies (Table 3). Across all populations the average proportion of polymorphic sites was 93%, the frequency of the most common allele (P) was 0.79, and the nucleotide diversity (π) was 0.30. Average F_{IS} estimates for Europe ($F_{IS} = 0.021 \pm 0.007$) and North America ($F_{IS} = 0.0082 \pm 0.006$) were significantly different ($p < 0.001$), however both estimates were relatively low. The average F_{IS} for Funk Island, Newfoundland was significantly lower than all other colonies ($F_{IS} = -0.0082$, $p < 0.05$). The Norwegian colony on Skarveklakken had the highest F_{IS} value ($F_{IS} = 0.038$) and differed significantly from every colony except Rouzic Island, France ($F_{IS} = 0.026$, $p = 0.058$). Similarly, Rouzic Island's F_{IS} estimate was significantly greater than all other colonies except for Bonaventure Island, Quebec ($F_{IS} = 0.014$, $p = 0.089$).

Average observed heterozygosity was significantly higher for the colony on Funk Island than for any other colony ($H_O = 0.32$, $p < 0.05$). Additionally, the Magdalen Islands had significantly higher observed heterozygosity ($H_O = 0.30$) than the European colonies of Rouzic ($H_O = 0.29$, $p < 0.05$) and Skarveklakken ($H_O = 0.29$, $p < 0.01$). Both Rouzic Island and Skarveklakken were established within the last 60 and 40 years respectively (Siorat & Rocamora, 1995; Barrett, 2008), suggesting that the high F_{IS} and low observed heterozygosity at these sites may be indicative of recent founder effects.

The AMOVA implemented in Arlequin indicated that the global F_{ST} across the gannet's range did differ significantly from zero (global $F_{ST} = 0.017$, $p << 0.001$). Pairwise comparisons either among North American or among European colonies revealed no significant differentiation. In contrast, all but two pairwise comparison between European and North American populations of gannets were significant ($p << 0.001$; Table 4). The North American colony at Funk Island did not differentiate significantly from the European colonies of Bass Rock ($F_{ST} = 0.016$, $p = 0.083$) or Skarveklakken ($F_{ST} = 0.0089$, $p = 0.0012$).

3.4 Clustering analyses

Results from PCA, STRUCTURE and DAPC were highly concordant. In my PCA, the orientation of clusters along the primary principal component reflects the geographical divide between European and North American colonies of gannets (Figure 2). Similarly, STRUCTURE grouped individuals into two genetic clusters largely representing one European and one North American population (Figure 4a). Accordingly, ΔK showed a peak at $K = 2$ (Figure 3a). The K-means clustering algorithm implemented in *adegenet* also indicated that $K = 2$ was the most probable number of genetic clusters in the data (Figure 3b). A histogram of assignment probabilities produced by DAPC revealed similar genetic structure to that produced by PCA and STRUCTURE (Figure 4b).

3.5 Outlier analysis

LOSITAN identified 221 outlier loci across all pairwise comparisons. The number of outliers varied greatly between comparisons, ranging from 4 to 50 SNPs. Perhaps unsurprisingly, more outliers were detected in pairwise comparisons between European and North American colonies (average = 28 SNPs; range = 17 – 42 SNPs), than when comparing amongst just North American (average = 21 SNPs; range = 10 – 50 SNPs) or just European colonies (average = 9 SNPs; range = 4 – 15 SNPs). The comparison between the European colonies on Skarveklakken and Rouzic Island produced the fewest number of outliers (4 SNPs). Interestingly, the Newfoundland colony of Funk Island had the highest number of outliers when compared to the North American colonies on Cape St. Mary's (50 SNPs) and Baccalieu Island (44 SNPs).

When only North American populations were compared, 151 SNPs were detected as outliers. Conversely, when only European populations were compared, only 25 unique outliers were detected overall. Jackknife resampling further reduced the selected panel to 111 SNPs in North America, 20 SNPs in Europe, and 173 SNPs when comparing amongst all colonies.

3.6 Population assignment success

Assignment of individuals to Europe versus North America was 99% successful using the full dataset (2155 SNPs) in GeneClass 2.0, with only a single individual from Funk Island (Newfoundland) being "misassigned" to Europe. This same individual was labelled as a putative first-generation migrant in GeneClass 2.0 and had a membership probability of over 95% to the European cluster in STRUCTURE. Assignment of gannets

to individual colonies was much lower using the full dataset, with only 31% of individuals assigning to the colony where they were sampled.

3.6.1 Outlier assignment success

Population assignment was 77% using an F_{ST} cutoff of 0.1, and a panel of 221 outlier loci. Importantly, aside from two North American individuals with high membership probabilities to Europe in the STRUCTURE analysis, all misassignments were to colonies in their respective regions. When assigning individuals amongst North American colonies, assignment success was relatively high at 81%. Most misassignments amongst the North American colonies were to the colony on the Magdalen Islands. Assignment success was lower amongst European colonies, dropping to 73%, with most misassignments occurring between Skarveklakken and Rouzic Island.

Both STRUCTURE and DAPC showed greater genetic division between colonies using the subset of selected outlier SNPs than when using the full dataset. When assigning North American gannets, STRUCTURE indicated the most probable number of populations to be five (Figure 5a) and produced clusters largely representing the five North American colonies (Figure 6a). DAPC could clearly distinguish between four genetic clusters, but had a high proportion of misassignments between the colonies on Bonaventure Island and Magdalen Islands (Figure 6b). When assigning European gannets, STRUCTURE found best support for two genetic clusters (Figure 5b), grouping the Skarveklakken and Rouzic Island colonies together into a single cluster (Figure 7a). DAPC produced similar results (Figure 7b).

LOSITAN detected a total of 239 SNPs in the randomized dataset. Subsequent PCA, STRUCURE analysis and DAPC all failed to detect any genetic structure using these outlier loci (Figure 8). Furthermore, assignment success in GeneClass 2.0, using these random outliers, was only 14%. These results suggest that the structure detected with outliers derived from the original dataset is not the result of a random process or the analytical method I employed.

Once compared to the randomized dataset, subsequent jackknifing increased the assignment success of outlier SNP panels to 90% in North America, 82% in Europe and 85% across all colonies.

3.6.2 Random forest assignment success

Pairwise comparisons among North American and among European gannet colonies produced panels of 107 and 34 SNPs respectively. Assignment success using the random forest selected panel was lower than when using outlier loci. Amongst North American colonies 70% of individuals were correctly assigned. Amongst European colonies assignment success was only 62%.

3.7 Estimates of contemporary gene flow

Estimates of contemporary migration produced by BAYESASS 3.0.4 suggest ongoing, unidirectional gene flow between European and North American colonies of gannets (Figure 9). Confidence intervals (95%) for the proportion of individuals with European ancestry in North American colonies ranged from between 2 and 8%. Conversely, estimates of migration into Europe from North America were as high as 3%

but overlapped with zero, indicating no significant gene flow. All three independent runs, using different seeds, produced very similar migration estimates and reached convergence (Appendix A3).

Relatively low self-recruitment rates (range: 0.77 to 0.91) and nonconvergence suggested that my data lacked the resolution or requisite level of differentiation necessary to estimate migration rates between individual colonies (Appendix A5).

3.8 Estimates of population divergence

The results of the ABC analysis in DIYABC were very similar across the three scenarios. All three simulations produced posterior distributions with median estimates of population divergence between 33 and 25 kya. The most highly supported scenario (Appendix A6) accounted for a change in the effective population sizes of the European and North American ancestral groups following range expansion (Figure 10). This scenario produced a population divergence date of approximately 25 kya.

Table 3. Summary genetic statistics for eight gannet colonies calculated across all retained variable sites. Statistics include the average number of individuals genotyped per locus in each population (N), the total number of variable sites in each population (Sites), the percentage of those sites that were polymorphic in each population (% Poly), the average frequency of the most common allele (P), the expected heterozygosity for each population (H_E), the observed heterozygosity for each population (H_O), the average nucleotide diversity across all loci (π) and the average Wright's inbreeding coefficient (F_{IS}). See Table 1 for colony abbreviations.

	N	**Sites**	**% Poly**	**P (%)**	**H_E**	**H_O**	**Π**	**F_{IS}**
BAC	8.73	2108	88.57	79.29	0.285	0.298	0.302	0.012
BON	9.58	2151	90.80	79.04	0.288	0.299	0.304	0.014
CSM	8.86	2135	90.30	79.38	0.284	0.298	0.301	0.011
FUN	8.77	2125	91.20	78.83	0.292	0.315	0.310	-0.008
MAG	19.44	2155	97.82	78.58	0.299	0.304	0.307	0.011
BAS	7.80	2116	87.29	79.66	0.280	0.296	0.299	0.009
ROU	20.02	2096	97.04	78.92	0.293	0.294	0.301	0.026
SKA	18.76	2154	97.49	79.00	0.294	0.290	0.302	0.038

Table 4. Average pairwise F_{ST} values among the eight populations analyzed in this study. See Table 1 for colony abbreviations.

	BON	CSM	FUN	MAG	BAS	ROU	SKA
BAC	-0.004	0.005	0.003	-0.003	0.021**	0.020**	0.019**
BON		0.003	0.000	-0.002	0.021**	0.019**	0.018**
CSM			0.002	0.002	0.015**	0.018**	0.017**
FUN				-0.001	0.004	0.008**	0.009*
MAG					0.016**	0.017**	0.018**
BAS						0.000	-0.001
ROU							-0.001

*Significantly greater than 0 at $\alpha = 0.05$ after Benjamini-Hochberg corrections for multiple comparisons.

**Significantly greater than 0 at $\alpha = 0.001$ after Benjamini-Hochberg corrections for multiple comparisons.

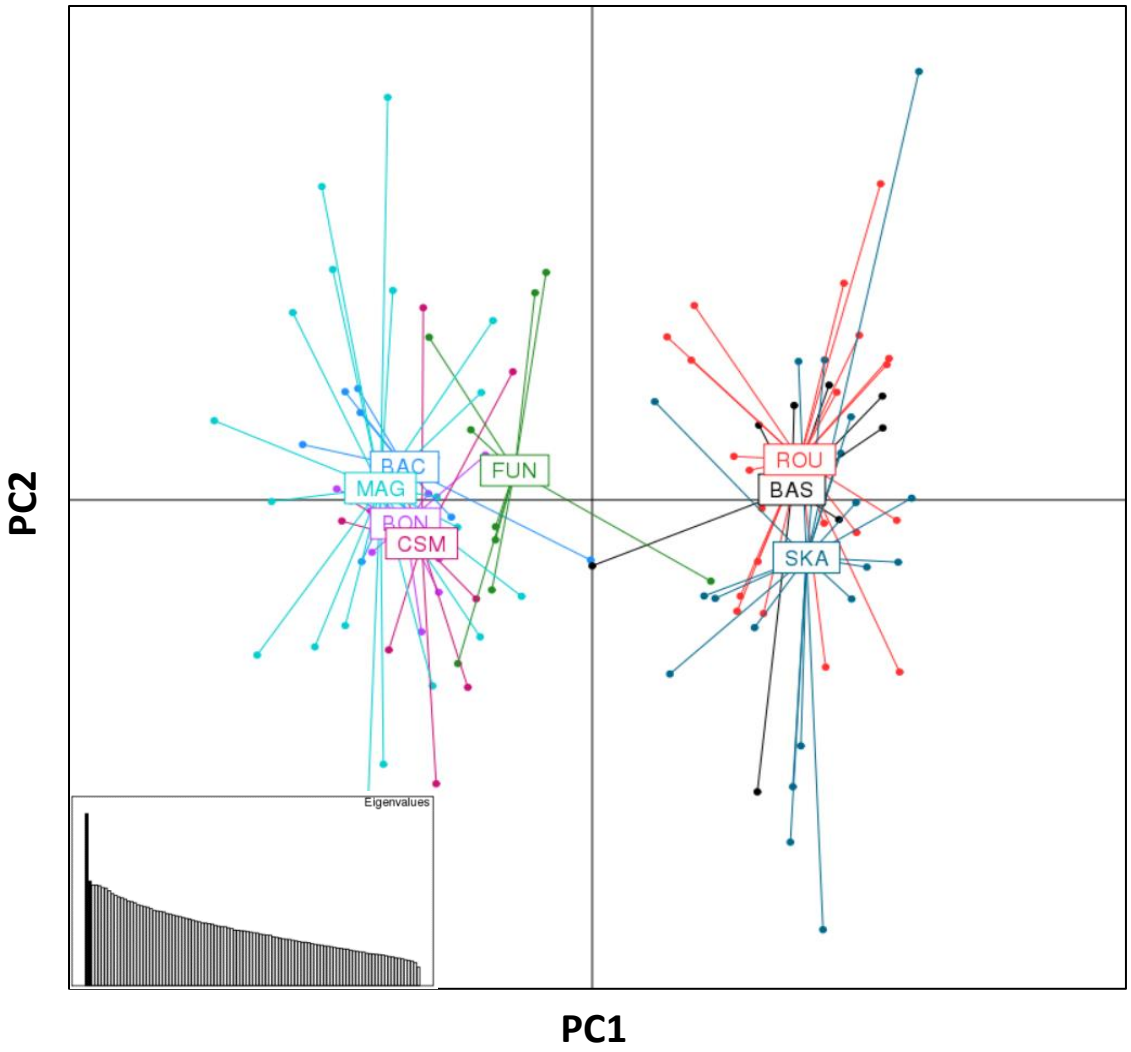


Figure 2. Principal component analysis biplot for all loci (n = 2155) across eight northern gannet colonies. See Table 1 for colony abbreviations.

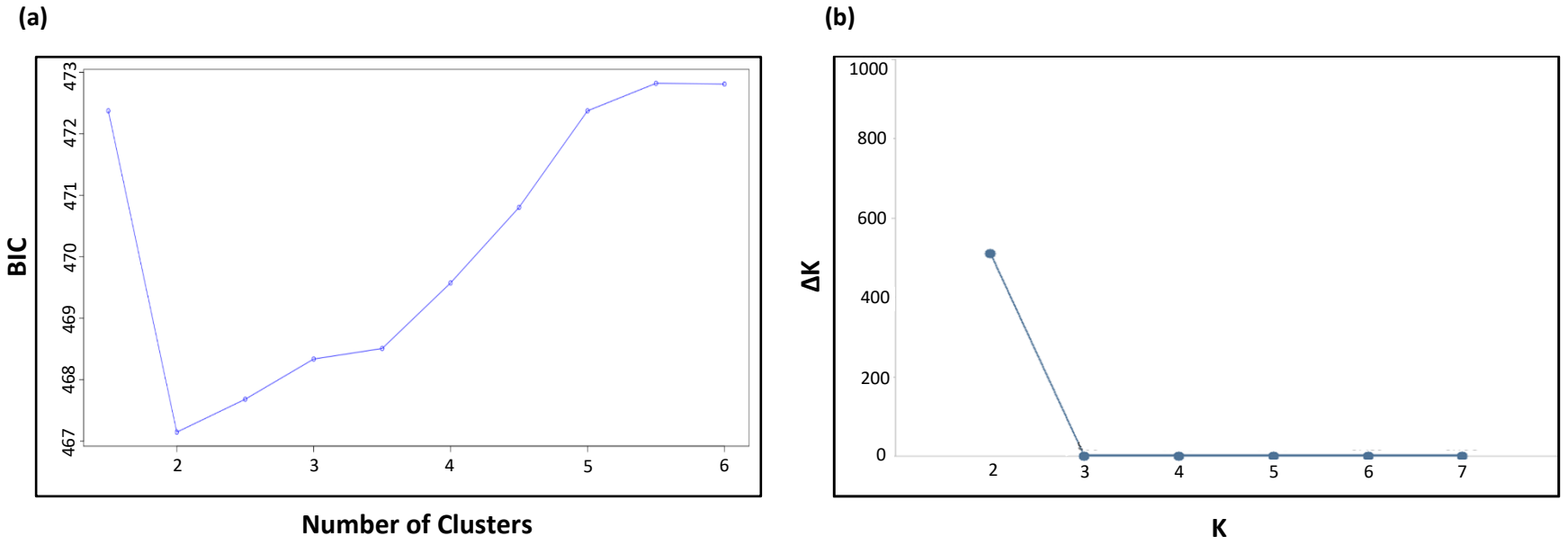


Figure 3. The most highly supported number of genetic populations (K) as indicated by (a) STRUCTURE and (b) *adegenet*. The highest ΔK values and lowest Bayesian information criterion (BIC) correspond to the most highly supported number of genetic clusters.

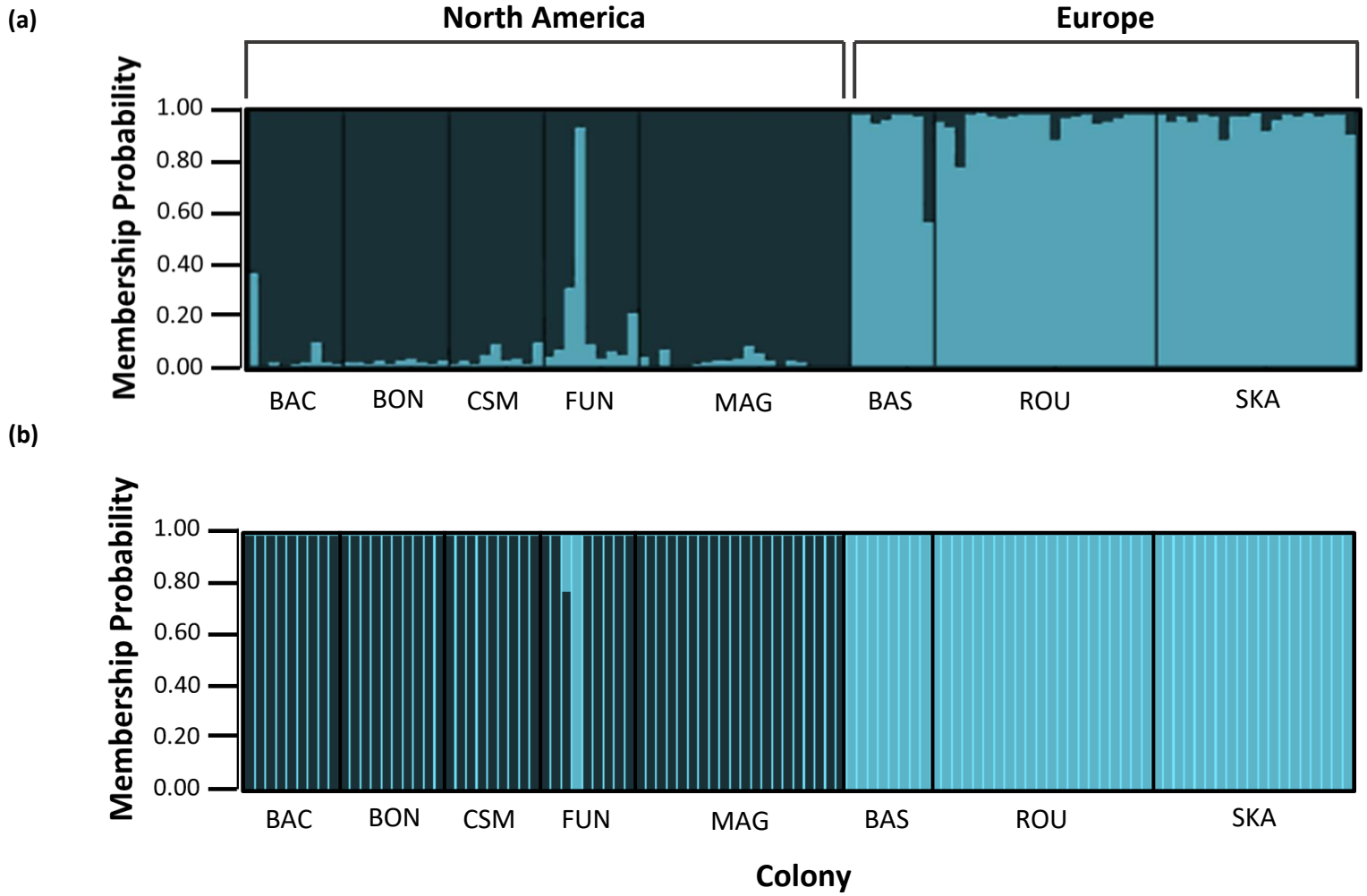


Figure 4. Probabilities of assignment of individual gannets to two genetic populations as determined by (a) STRUCTURE using a model of admixture, no prior population information and correlated allele frequencies, and (b) discriminant analysis of principal components. See Table 1 for colony abbreviations.

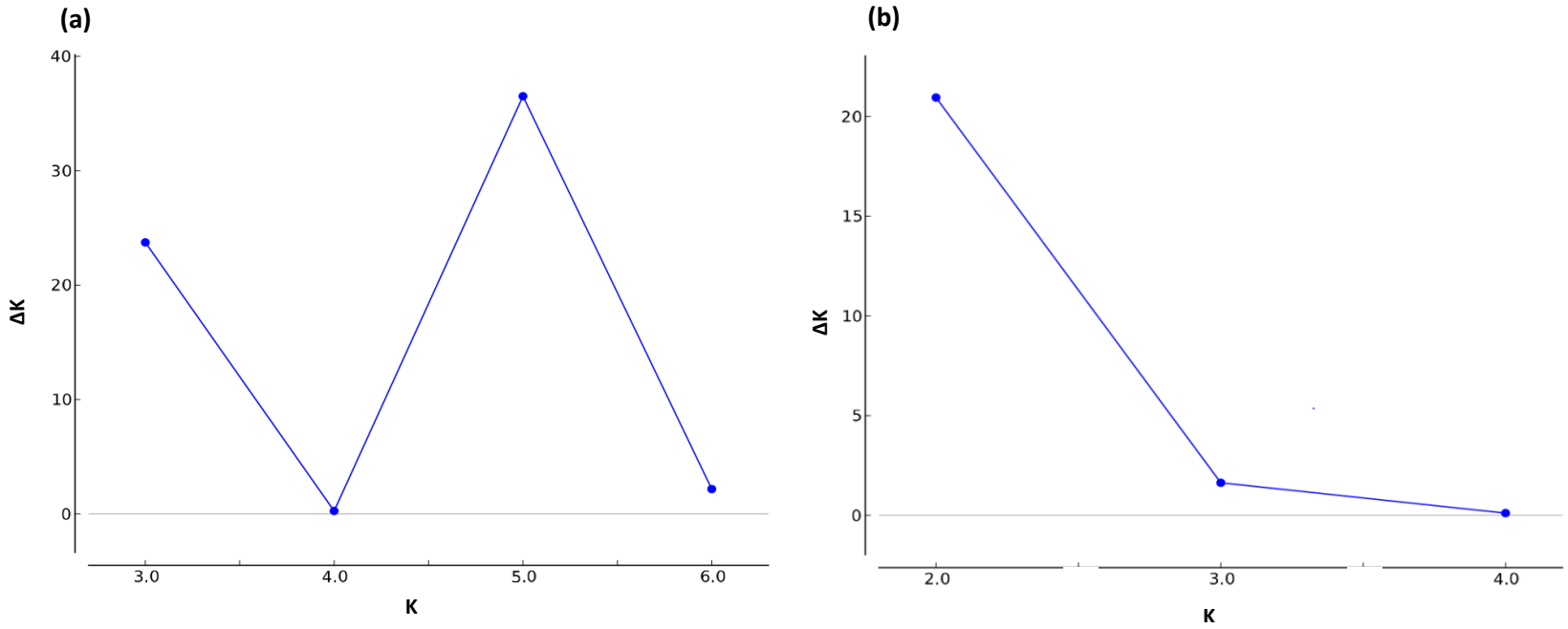


Figure 5. The most highly supported number of genetic populations (K) as indicated by STRUCTURE using the outlier panel of SNPs produced for (a) North American and (b) European gannets only.

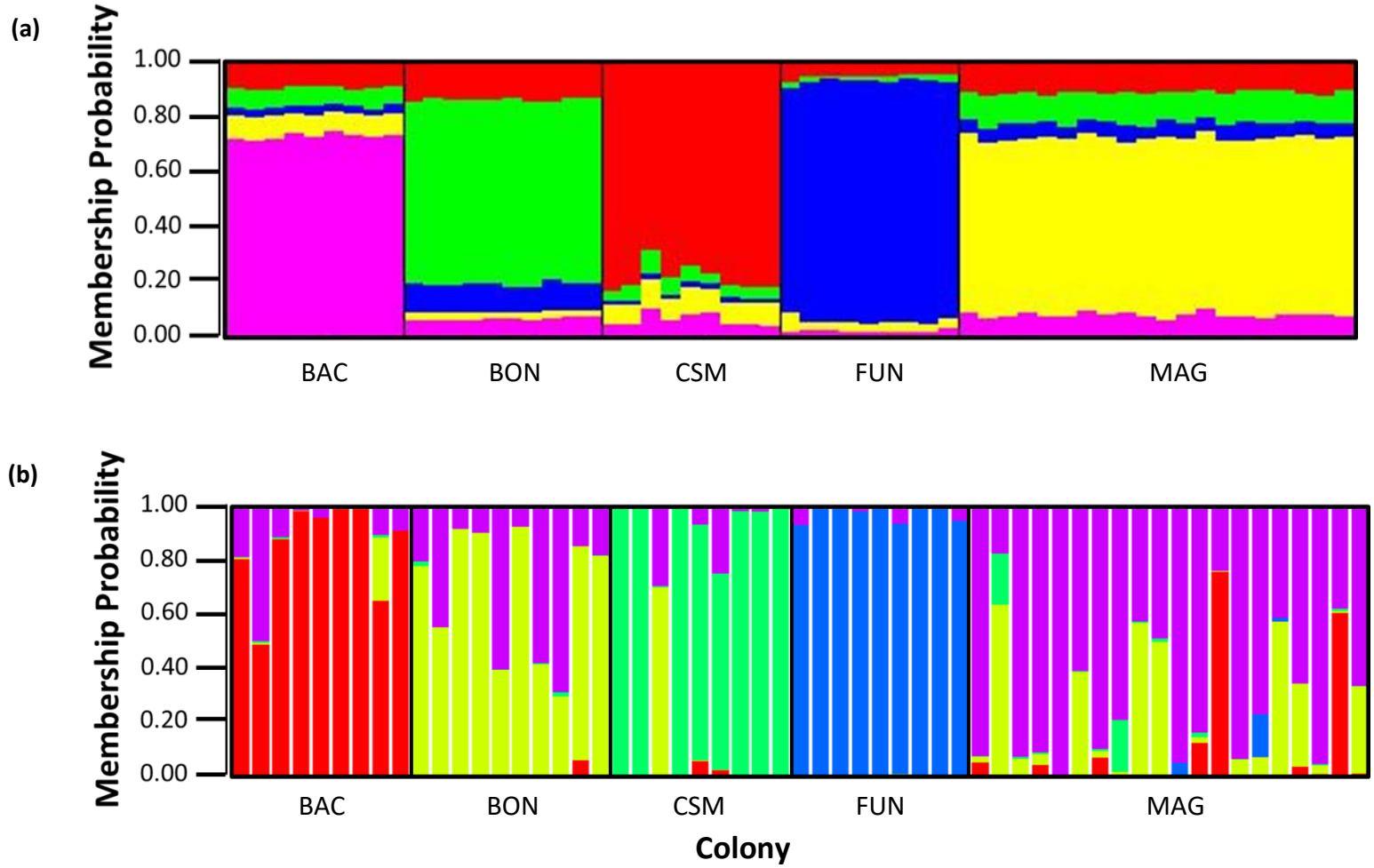


Figure 6. Probabilities of assignment of North American gannets to five genetic populations with the selected panel of outlier markers (151 SNPs) using (a) STRUCTURE with a model of no admixture, and correlated allele frequencies, and (b) discriminant analysis of principal components. Both analyses used prior population information. See Table 1 for colony abbreviations.

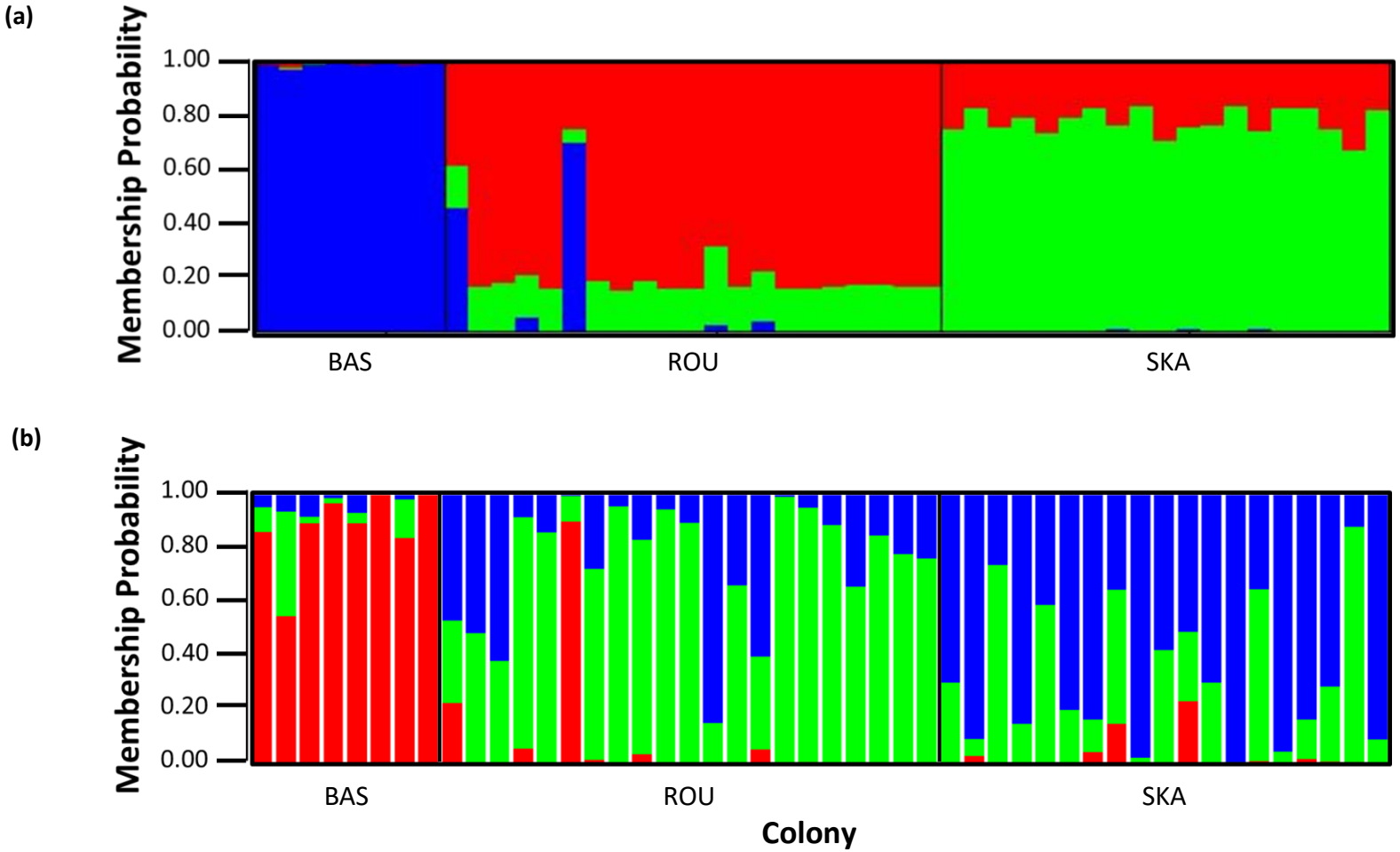


Figure 7. Probabilities of assignment of European gannets to three genetic populations with the selected panel of outlier markers (25 SNPs) using (a) STRUCTURE with a model of no admixture, and correlated allele frequencies, and (b) discriminant analysis of principal components. Both analyses used prior population information. See Table 1 for colony abbreviations.

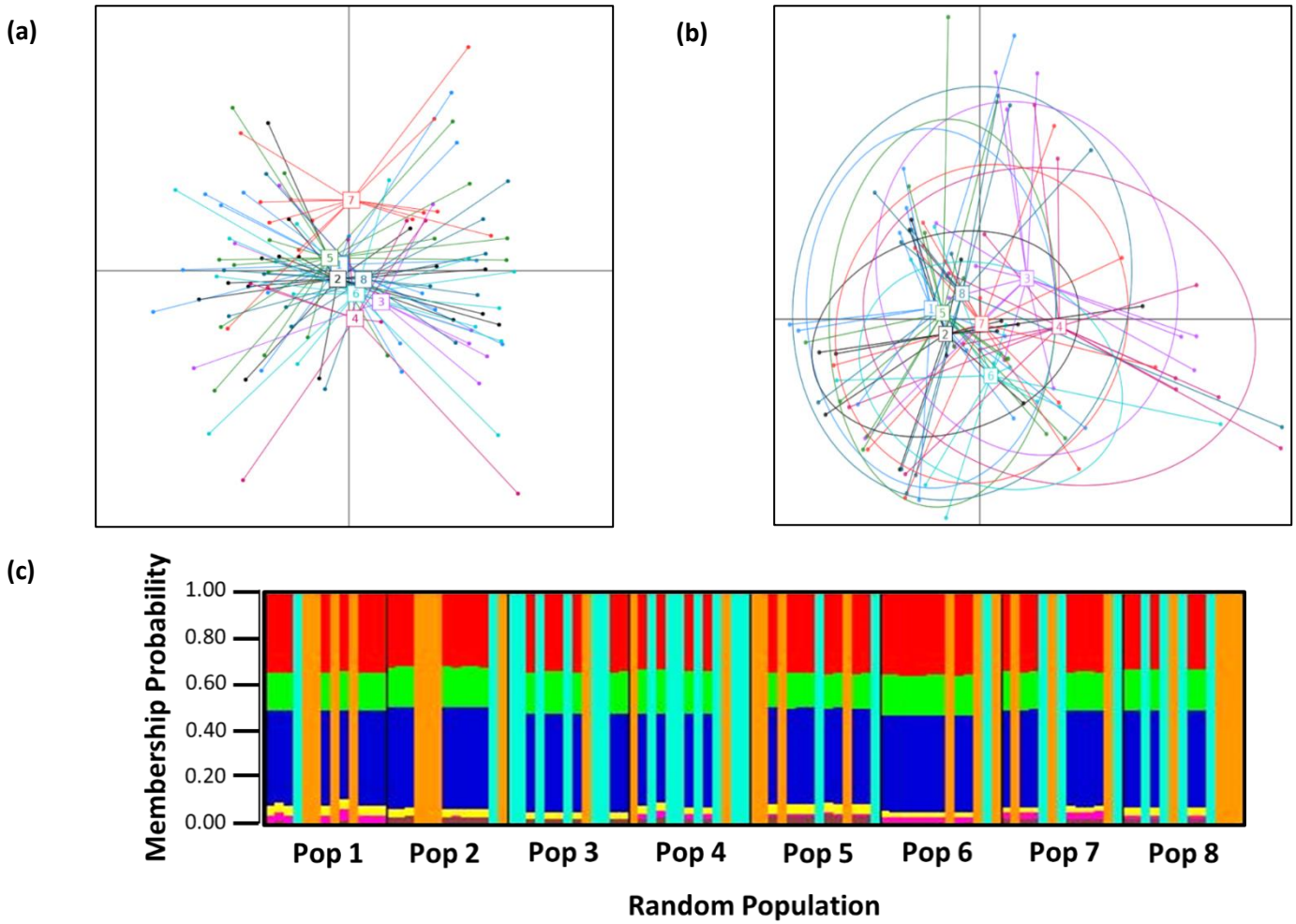


Figure 8. Results of clustering analysis using only outliers detected from the randomized dataset (239 SNPS). (a) PCA biplot, (b) DAPC plot and (c) STRUCTURE bar plot.

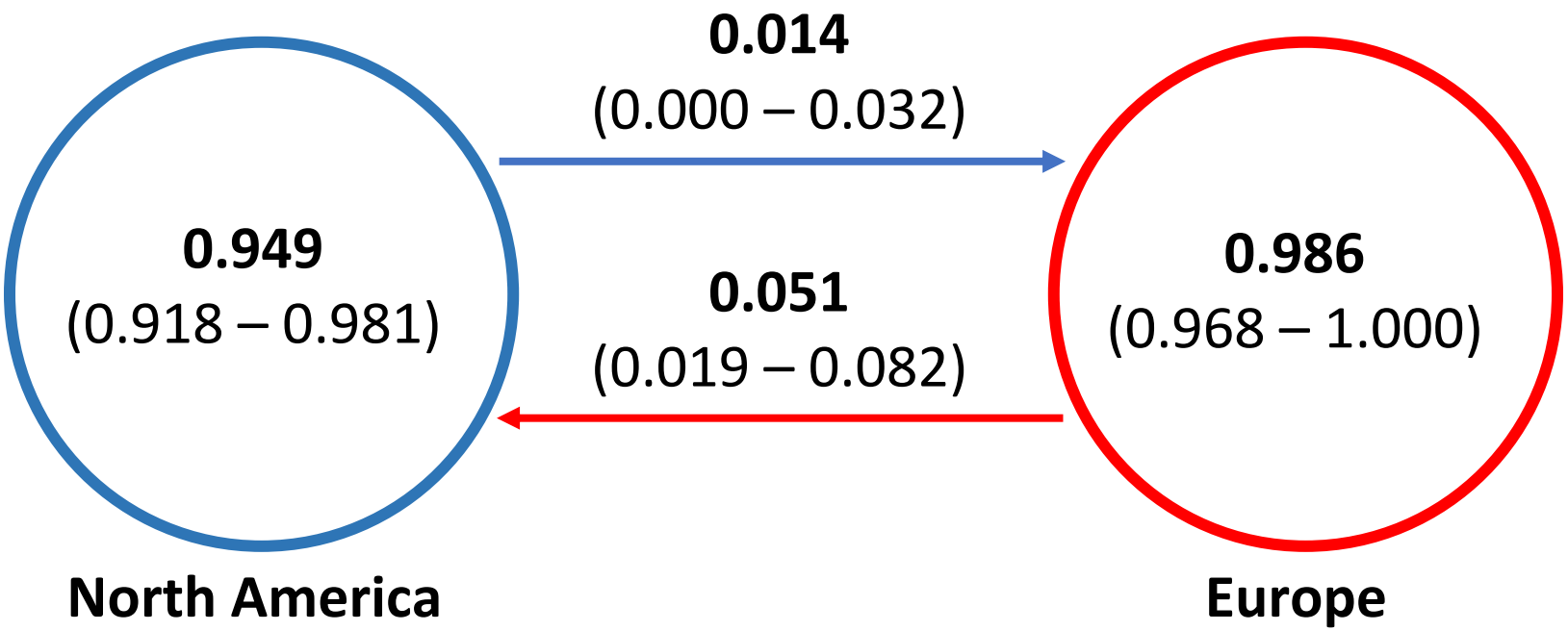


Figure 9. Contemporary migration rate estimates (m; BAYESASS) between the European and North American colonies of northern gannets using 2155 SNPs. Migration estimates are given in bold above the arrows. Self-recruitment rates are given in bold inside the circles. Upper and lower 95% confidence limits are given in parentheses.

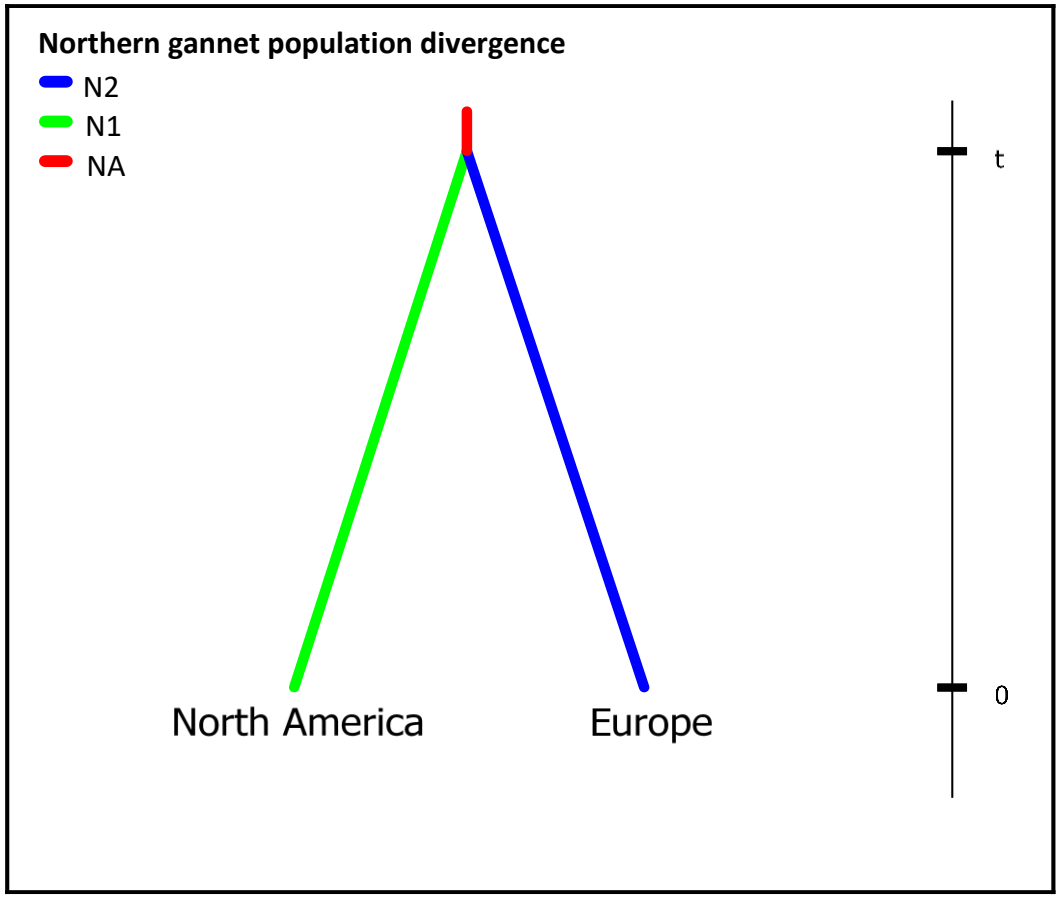


Figure 10. Most highly supported scenario from simulated data sets in DIYABC. Figure shows a simple population divergence, whereby one ancestral population of gannets diverged into the contemporary North American and European populations at some point in the past (t). NA represents the effective population size before the divergence event and N1 and N2 represent the present-day effective populations sizes of gannets in North America and Europe respectively.

Chapter 4: Discussion

Clustering analyses indicated significant population genetic structure across the northern gannet breeding range. The genetic structure I detected corresponds primarily to the division between North American and European populations of gannets. Results from ABC suggested that these two populations diverged during the LGM between 33 and 25 kya. Despite significant genetic differentiation between the two regions, I found unequal proportions of migrant ancestry within European and North American colonies.

When analyzing all 2155 SNPs, I found no evidence of genetic differentiation among colonies within North America or Europe. I did, however, detect fine-scale genetic structure between all colonies using outlier loci (173 SNPs), allowing high population assignment success. Evidence and possible evolutionary mechanisms for these patterns of differentiation are detailed below.

4.1 Trans-Atlantic genetic differentiation and gene flow

My analyses indicate that northern gannets have significant population genetic structure. Clustering analyses found greatest support for two genetic groups, corresponding largely to the European and North American populations of gannets. Assignment to these two clusters was 99% using all 2155 SNPs, with only one North American gannet being missassigned to Europe. Moreover, most pairwise F_{ST} estimates between European and North American gannet colonies were significantly different from zero. Friesen *et al.* (unpublished) also reported that North American and European populations of gannets were significantly differentiated at the mtDNA control region. My

results, in combination with previous mtDNA analysis, therefore provide strong support for the existence of two genetic populations across the gannet's breeding range.

I detected uneven proportions of individuals with migrant ancestry within the North American and European gannet populations. Confidence intervals for estimates of migration from North America into Europe overlapped with zero (0 - 3%). Conversely, gene flow from Europe into North America was significantly greater than zero (2 - 8%), suggesting unidirectional dispersal across the Atlantic. Interestingly, I found that a single population (Funk Island) in North America did not differentiate significantly from two of the three European gannet colonies. This same colony also displayed evidence of recent immigration from Europe in my clustering analysis and population assignment, possibly harboring a first-generation migrant. Furthermore, the Funk Island colony had the lowest average F_{IS} estimate and higher numbers of outliers when compared to North American colonies than European colonies, potentially indicative of admixture with Europe.

Large tracts of island-free, open ocean may act as cryptic barriers to gene flow in some species of seabirds. For example, Steeves *et al.* (2003) found that, despite reports of recent dispersal, colonies of brown boobies (*Sula leucogaster*) separated by the Eastern Pacific Basin were genetically differentiated. They suggested that the ocean basin acts as a dynamic barrier to gene flow in the brown booby, promoting population divergence while still allowing episodic migration. My results indicate that the North Atlantic Basin may play a similar role in gannets, potentially representing a cryptic or semi-permeable barrier to gene flow. Simulations conducted by Hastings (1993) and Waples & Gaggiotti (2006) revealed that **demographic independence** often occurs when migration rates are below 10% ($m < 0.1$) . Both of my migration rate estimates were below this threshold,

suggesting that demographic independence is being maintained between these populations. That both mtDNA (Friesen *et al., unpublished*) and nuclear DNA show significant genetic differentiation, despite the evidence of gene flow and recent population divergence (discussed below), further suggests that the North Atlantic, although evidently semi-permeable, acts as an effective barrier to gene flow in gannets.

4.1.1 Juvenile recruitment

Tracking and banding studies have revealed that juvenile gannets will spend the majority of their first few years travelling large distances and visiting breeding sites, often settling for short periods just outside established colonies (Nelson, 2002; Votier *et al.*, 2011). Theoretically, these immature birds are assessing the quality of the sites they visit, and prospector attendance appears to be important for recruitment in several other species of seabirds (Becker & Bradley, 2007; Dittmann *et al.*, 2007; Campioni *et al.*, 2017). Although the majority of gannet juveniles ultimately return to their natal colony, many have been found breeding at distant colonies (Barrett, 1996; Veron & Lawlor, 2009).

In a metanalysis modelling population growth rates, Moss *et al.*(2002) found that small gannet colonies typically grow faster than larger colonies. Often these small colonies are growing faster than is expected through self-recruitment alone. Moreover, Moss *et al.* found that larger colonies rear more offspring than they can sustainably recruit. Together with evidence of strong breeding philopatry (Nelson, 2002), Moss *et al.* suggested that gene flow between northern gannet colonies is mediated by immature individuals.

Gene flow across the North Atlantic Ocean might also be mediated by dispersal of juvenile gannets. Social or conspecific attraction may play a vital role in the recruitment process for many seabirds, and is commonly employed in restoration projects to establish or repopulate colonies (e.g. Australasian gannets, *Morus serrator*; Sawyer & Fogle, 2013). The degree of migratory connectivity between seabird populations appears to be an important correlate of genetic structure (Friesen *et al.*, 2007). The relationship between nonbreeding distribution and genetic differentiation indicates that interactions at the wintering ground, or along the migratory route, may result in recruitment between populations. The three gannets tracked crossing the North Atlantic all appeared to have overwintered off the west coast of Africa, where most European gannets spend the winter months (Kubetzki *et al.*, 2009). Fifield *et al.* (2014) suggested that conspecific attraction and the gannets' group migration strategy may result in the recruitment of juvenile gannets across the Atlantic. Furthermore, nearly all trans-Atlantic band recoveries have been of immature gannets (Gaston *et al.*, 2008). If estimates of high breeding philopatry (95%; Nelson, 2002) and winter-site fidelity (Fifield *et al.*, 2014) in adult gannets are correct, then nonbreeding or juvenile gannets may be the most likely trans-Atlantic recruits.

Despite impressive mark-recapture efforts in birds, for most species very little has been revealed about migratory connectivity and dispersal through band recovery alone (Webster *et al.*, 2002). In gannets, banding efforts have typically been biased towards a select number of large colonies and band recovery rates have traditionally been very low (Gaston *et al.*, 2008). Moreover, relatively conservative tracking efforts have already begun to reveal foraging and migratory behaviours that remained unreported through

decades of band recoveries in both adult (Wakefield *et al.*, 2013; Fifield *et al.*, 2014) and immature gannets (Votier *et al.*, 2011; Montevecchi *et al.*, 2012b). However, to my knowledge, only 46 adult and 18 immature North American gannets, and 106 adult and 15 immature European gannets have been tracked throughout the nonbreeding season (Grémillet *et al.*, 2006; Montevecchi *et al.*, 2012b).

When considering these numbers relative to a world population size of over 1,000,000 individual gannets, our current inability to accurately predict patterns of dispersal through mark-recapture and tracking alone may be unsurprising. My study highlights the importance of an integrative approach, incorporating demographic and genetic information, as a means of understanding and interpreting dispersal patterns. Although I cannot determine whether gene flow across the Atlantic is mediated through immature or adult dispersal, I can provide the first evidence confirming its presence.

4.1.2 Vagrancy

An alternative explanation for the ongoing gene flow detected between European and North American populations of gannets is the recruitment of vagrant individuals. Gannets are sometimes encountered in regions far outside of their species' normal range, presumably forced to those areas by strong winds or inclement conditions (Boev, 2009; Hazevoet, 2014). Gannets wintering off the coast of Africa would be exposed to the powerful northeasterly trade winds and could be blown into the Atlantic during a storm. The most oceanic colony of gannets in North America is on Funk Island. Should a vagrant individual from Europe settle in the first colony it encounters, it may help to explain the presence of a first-generation migrant and the European admixture found on

Funk Island. However, several lines of evidence suggest that vagrancy may be an unlikely explanation.

Firstly, tracking studies have revealed that an individual gannet is able to cross the Atlantic ocean twice in a single breeding season (Fifield *et al.*, 2014). Although the North Atlantic undoubtedly represents a formidable barrier for dispersal, the capacity for trans-Atlantic navigation suggests that a vagrant individual could return to its colony of origin. Secondly, my analyses found evidence of migrant ancestry in several individuals and across multiple populations, indicating that the rate of trans-Atlantic vagrancy would have to be quite high. Finally, unless vagrancy is very common, the percentage chance that a first-generation vagrant was sampled at random from amongst thousands of North American gannets for this study is low (0.005%; assuming Funk Island population size of 20,000 individuals; Chardine *et al.*, 2013).

4.2 Northern gannet population divergence

Divergence dates from my ABC analysis, suggest that the North American and European populations of gannets diverged during the LGM, possibly in separate refugia. Using mtDNA, Friesen *et al.* (*unpublished*) predicted that North American and European populations of gannets diverged during the late-Pleistocene, between 22 and 86 kya. My ABC analysis, using over 2100 independent genetic markers, also indicated that European and North American populations diverged towards the end of the Pleistocene epoch ($t \approx 25$ kya). Strong linkage and extremely low recombination suggests that, from a population genetic perspective, mtDNA effectively acts as a single genetic marker (Avise, 1991). Single-locus estimates of divergence time are expected to predate

population divergence (Edwards & Beerli, 2000; Arbogast *et al.*, 2002; Zhang & Hewitt, 2003; Weir *et al.*, 2016). I suggest that the use of multiple nuclear loci to estimate divergence date may explain why my approximation overlaps with the lower dates produced using mtDNA.

Fossil records indicate that, during the late-Pleistocene, the Mediterranean Sea acted as an important refugium for several species of temperate seabirds (Tyrberg, 1999; Zotier *et al.*, 1999). Species assemblages from the region suggest conditions approximating those at the British Isles today (Tyrberg, 1999), where the majority of northern gannets now reside, and numerous fossil recoveries suggest that gannets were once widely distributed across the Mediterranean (Tyrberg, 1999) reaching as far as the islands of Crete to the east (Weesie, 1987). The earliest dated fossils indicate that gannets may have occupied the sea as early as 30 to 40 kya, potentially remaining in the area until the late-Holocene (approximately 2.7 kya; Oros Sršen *et al.*, 2017).

Patterns of genetic diversity in some seabirds have been linked to late-Pleistocene range expansion (Moum & Arnason, 2001; Liebers & Helbig, 2002; Morris-Pocock *et al.*, 2008). Furthermore, phylogeographic patterns suggest a southern refugium in the Mediterranean region for a wide range of taxa (e.g. mammals, Barbosa *et al.*, 2017; insects; Kühne *et al.*, 2017; aquatic invertebrates, Laggis *et al.*, 2017; plants, Lee *et al.*, 2017; amphibians, Wielstra *et al.*, 2017), including several species of seabirds (European storm-petrels, *Hydrobates pelagicus*, Cagnon *et al.*, 2004; Scopoli's shearwater, *Calonectris diomedea*, Gómez-Díaz *et al.*, 2006; European shag, *Phalacrocorax aristotelis*, Thanou *et al.*, 2016). Fossil records and phylogeographic studies therefore support the suggestion that northern gannets may have occupied a Mediterranean

refugium during the LGM, however, my results suggest that gannets may have persisted elsewhere as well.

In the Northern Hemisphere, cryptic refugia likely played an important role in the persistence of many temperate and polar species (Stewart *et al.*, 2010). Patterns of genetic variation suggest that some seabirds may have persisted during the LGM in a cryptic refugium in the Northeast Atlantic (Liebers *et al.*, 2004; Tigano *et al.*, 2015). Across several species of seabirds, a phylogeographic boundary dating to the Pleistocene, has been identified between the Atlantic Ocean and the Mediterranean sea (Cagnon *et al.*, 2004; Gómez-Díaz *et al.*, 2006; Thanou *et al.*, 2016). I suggest that European and North American gannets were isolated from each other during the LGM, potentially inhabiting the Mediterranean region as well as a cryptic refugium elsewhere in the Atlantic.

4.3 Population genetic structure among North American and European gannet colonies

When analyzing my full dataset (2155 SNPs), I found no significant population differentiation among North American or European colonies of gannets. Unless newly established colonies have experienced bottlenecks, population differentiation is expected to remain low following range expansion until enough time has passed for genetic drift and selection to erase shared standing genetic variation (Slatkin, 1985, 1987; Birky *et al.*, 1989). The gannet's abundance and breeding distribution are very well documented, owing to over a century of consistent census data (Wanless *et al.*, 2006). Most gannet colonies were founded within the last 100 years and the global population is predicted to be increasing at a rate of approximately 2% pa (Wanless *et al.*, 2006). Beginning in 1913,

nearly decadal censuses of European colonies indicate that the gannet breeding range has expanded northward by 4° latitude and southward by 3° latitude during the 20th century alone (Moss *et al.*, 2002). This information strongly suggests a recent, and apparently ongoing, range expansion. The shared genetic variation that usually accompanies such a rapid and recent expansion may explain the low overall population differentiation I detected (Slatkin, 1987).

The weak genetic structure among European and among North American colonies may also reflect contemporary gene flow. Gene flow is expected to counteract genetic drift and local adaptation, thereby reducing population differentiation (Kawecki & Ebert, 2004; but see Tigano & Friesen, 2016). Nonbreeding range appears to be an important predictor of population differentiation and genetic structure in seabirds (Friesen *et al.*, 2007). Telemetry indicates that there is a high degree of overlap in the nonbreeding distribution of northern gannets on both sides of the Atlantic (Fort *et al.*, 2012; Fifield *et al.*, 2014). Theory predicts that very few migrants per generation are required to genetically homogenize two populations (Wright, 1931). Should overlap in nonbreeding range result in the recruitment of individuals between colonies, the movement of even a few gannets per generation could result in the pattern of differentiation I have reported here.

Unfortunately, low self-recruitment rates and nonconvergence in BAYEASS meant that I did not have the genetic differentiation necessary to estimate colony-specific dispersal rates. Although some researchers have suggested that a shared wintering distribution may result in juvenile recruitment between colonies (e.g., Tigano *et al.*, 2017), empirical evidence demonstrating a direct link between demographic connectivity

on the nonbreeding ground and genetic connectivity at the breeding colony is still lacking for most seabird species. While I cannot report direct evidence of gene flow among European and North American colonies, my study adds to the growing body of evidence suggesting that seabird populations with shared nonbreeding distributions typically display weak genetic structure (Friesen *et al.*, 2007; Friesen, 2015).

4.4 Conservation implications

Effective conservation management requires the identification and separate management of distinct populations within a species (Crandall *et al.*, 2000; Funk *et al.*, 2012). Genetic characterization refers to the process by which genetically distinct units are identified, with the purpose of conserving a species' genetic variation (deVicente *et al.*, 2005). Estimates of F_{ST} and STRUCTURE analyses both suggest the presence of two genetically distinct units across the northern gannet range. I therefore suggest that these populations likely represent two separate management units. However, evidence of recent divergence, low overall differentiation and ongoing gene flow indicate that European and North American populations of gannets constitute a single evolutionary significant unit (Moritz, 1994).

I established that, despite low overall genetic differentiation, population assignment is possible in the northern gannet using a subset of informative genetic markers. I found that assignment using outlier loci was much higher than when using random forest selected markers. I further demonstrated that the pattern of differentiation I detected with outlier loci was not replicable when individuals were randomly assigned to populations (Campagna *et al.*, 2015; Tigano *et al.*, 2017).

When analyzing divergence between six species of capuchino seedeater (genus *Sporophila*), Campagna *et al.* (2015) reported that the genetic structure they found using outlier loci was uninformative because the same pattern could be detected using a random dataset. Alternatively, using a methodology very similar to the one employed here, Tigano *et al.* (2017) found that the differentiation they detected in thick-billed murres (*Uria lomvia*) was most likely indicative of true genetic structure. Together these studies suggest I have taken a robust approach to demonstrating the informative nature of the selected outlier SNP panels.

In 2010, an explosion occurred on the *Deepwater Horizon* oil-drilling rig, resulting in the largest marine oil spill in history. The northern gannet was amongst the most heavily affected seabird species, with the majority of mortality being amongst juveniles (Montevecchi *et al.*, 2012a). Contrary to previous reports from band recoveries, geolocator tracking has since revealed that over 25% of the North American population of gannets winters in the Gulf of Mexico, with concentrations in those areas most impacted by the spill (Montevecchi *et al.*, 2012a; b; Fifield *et al.*, 2014). Although perhaps an extreme example, the *Deepwater Horizon* disaster reflects the sudden, and potentially long-lasting, impact human disturbance can have on northern gannets and other marine species.

Individual population assignment using the genetic markers I have identified may have several applications to the conservation and management of northern gannets. Firstly, if used at a site of disturbance, these markers will help managers to determine whether specific colonies are differentially impacted by stressors. These sites can be restricted to breeding grounds, along migratory routes or at nonbreeding site.

Importantly, the loci I have identified allow the detection of first-generation migrants, whilst simultaneously allowing colony-specific assignment. Large-scale application of these panels may allow the characterization of local and trans-Atlantic gene flow in contemporary populations. With estimates of genetic connectivity, in combination with estimates of demographic connectivity taken from band recovery or tracking, we can begin to understand how localized disturbance impacts gannet evolutionary and ecological dynamics on a metapopulation scale.

Seabirds have been proposed as important sentinels of marine ecosystem health (Montevecchi *et al.*, 2012b; Lescroel *et al.*, 2016). Tracking and monitoring of gannets may help assess the impacts of oil pollution, fisheries exploitation, and wind and wave energy facilities on both sides of the Atlantic Ocean (Montevecchi *et al.*, 2012a; Garthe *et al.*, 2017; Haney *et al.*, 2017). Large population sizes, diverse diet, seasonal movement, and an ocean-wide distribution suggest that gannets are likely very important components of many marine ecosystems and may therefore prove useful as biomonitoring species for marine ecosystem health across the Atlantic (Lescroel *et al.* 2016). The panels of SNPs produced from my thesis will increase our ability to monitor and track the movements of gannets between colonies and at sea.

4.5 Future directions

For the purposes of population assignment, the adaptive roles of the outlier loci I detected do not need to be known. However, with the development of a reference genome, functional annotation and landscape genomic analysis may be used to understand local adaptation in the gannet. Furthermore, alignment to a reference genome

will greatly improve the number of markers retained overall and may therefore increase the resolution at which fine-scale genetic structure can be detected, potentially resulting in greater overall assignment success.

A major gap in my sampling design is the population of gannets on Iceland. Garthe *et al.* (2016) found that Icelandic gannets migrate to the western coasts of Europe and Africa, and suggested that the population may be European in origin. Although band recoveries suggest little migration with North America, genetic analysis can reveal whether Iceland serves as a stepping-stone for gene flow across the Atlantic. More comprehensive sampling, potentially including Icelandic colonies, would improve understanding of genetic connectivity and recolonization history in the gannet.

Summary

Population genetic structure is determined by the historical and contemporary interplay between evolutionary forces. Detecting and interpreting patterns of genetic differentiation can therefore not only reveal the mechanisms driving population divergence in the present, but can also reflect a species' evolutionary history and inform conservation efforts in the future. To investigate population differentiation in the northern gannet, I analyzed genomic patterns of diversity among eight North American and European colonies. My estimates of trans-Atlantic migration suggest that the North Atlantic Ocean is acting as a semi-permeable barrier to gene flow, allowing low rates of unidirectional migration from European colonies into North American colonies of gannets. Despite the ongoing gene flow, I found significant population differentiation, corresponding primarily to the division between North American and European colonies of gannets. Additionally, I found limited genetic differentiation among either European or North American colonies of gannets. My analyses suggested that the two populations diverged recently, potentially occupying separate refugia during the Last Glacial Maximum. My research suggests that North American and European populations of gannets should be managed as separate management units. Lastly, I identified genetic markers that allowed accurate colony assignment. These markers may allow colony-specific impact assessment and the development of targeted management strategies in the northern gannet.

Literature Cited

Alberdi, A., Gilbert, M.T.P., Razgour, O., Aizpurua, O., Aihartza, J. & Garin, I. (2015) Contrasting population-level responses to Pleistocene climatic oscillations in an alpine bat revealed by complete mitochondrial genomes and evolutionary history inference. *Journal of Biogeography*, **42**, 1689–1700.

Allendorf, F.W., Hohenlohe, P. & Luikart, G. (2010) Genomics and the future of conservation genetics. *Nature Reviews Genetics*, **11**, 697–709.

Anderson, E.C. (2010) Assessing the power of informative subsets of loci for population assignment: Standard methods are upwardly biased. *Molecular Ecology Resources*, **10**, 701–710.

Andrews S. (2010) FastQC: a quality control tool for high throughput sequence data. (http://www.bioinformatics.babraham.ac.uk/projects/fastqc)

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G. & Hohenlohe, P.A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

Antao, T., Lopes, A., Lopes, R.J., Beja-Pereira, A. & Luikart, G. (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Arbogast, B.S., Edwards, S. V., Wakeley, J., Beerli, P. & Slowinski, J.B. (2002) Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology, Evolution, and Systematics*,. **33**, 707–

740.

Ardren, W.R. & Bernall, S.R. (2017) Dams impact westslope cutthroat trout metapopulation structure and hybridization dynamics. *Conservation Genetics*, **18** 297–312.

Avise, J.C. (1991) Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Review of Genetics*, **25**, 45–69.

Baerwald, M.R., Meek, M.H., Stephens, M.R., Nagarajan, R.P., Goodbla, A.M., Tomalty, K.M., Thorgaard, G.H., May, B.N. & Nichols, K.M. (2016) Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout. *Molecular Ecology*, **25**, 1785–1800.

Barbosa, S., Paupério, J., Herman, J.S., Ferreira, C.M., Pita, R., Vale-Gonçalves, H.M., Cabral, J.A., Garrido-García, J.A., Soriguer, R.C., Beja, P., Mira, A., Alves, P.C. & Searle, J. B. (2017) Endemic species may have complex histories: within-refugium phylogeography of an endangered Iberian vole. *Molecular Ecology*, **26**, 951–967.

Barley, A.J., Monnahan, P.J., Thomson, R.C., Grismer, L.L. & Brown, R.M. (2015) Sun skink landscape genomics: Assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology*, **24**, 1696–1712.

Barrett, R.T. (2008) Recent establishments and extinctions of Northern Gannet Morus bassanus colonies in North Norway, 1995-2008. *Ornis Norvegica*, **31**, 172–182.

Barrett, R.T. (1996)The status of the North Atlantic Gannet Morus bassanus after 50 years in Norway. *Seabird*, **18**, 30–37.

Beaumont, M.A. & Nichols, R.A. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Beaumont, M.A., Zhang, W. & Balding, D.J. (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Becker, P.H. & Bradley, S.J. (2007) The role of intrinsic factors for the recruitment process in long-lived birds. *Journal of Ornithology*, **148**, 377–384.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R. & Bernatchez, L. (2015) RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, **24**, 3299–3315.

Benjamini, Y. & Hochberg, Y. (1995) Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society*, **57**, 289–300.

Berry, O., Tocher, M. & Stephen, S. (2004) Can assignment tests measure dispersal? *Molecular Ecology*, **13**, 551–561.

Besnier, F. & Glover, K.A. (2013) ParallelStructure: A R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. *PLoS One*, **8**, 1–5.

Bicknell, A.W.J., Knight, M.E., Bilton, D., Reid, J.B., Burke, T. & Votier, S.C. (2012) Population genetic structure and long-distance dispersal among seabird populations: Implications for colony persistence. *Molecular Ecology*, **21**, 2863–2876.

Birky, C.W., Fuerst, P. & Maruyama, T. (1989) Organelle gene diversity under migration, mutation, and drift: Equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, **121**, 613–627.

Boev, Z. (2009) Status of the gannet *Morus bassanus* in the Black Sea region (E Bulgaria). *Acrocephalus*, **30**, 2007–2010.

Cabrera, A.A. & Palsbøll, P.J. (2017) Inferring past demographic changes from contemporary genetic data: A simulation-based evaluation of the ABC methods implemented in DIYABC. *Molecular Ecology Resources*, 1–17.

Cagnon, C., Lauga, B., Hemery, G. & Mouches, C. (2004) Phylogeographic differentiation of storm petrels (*Hydrobates pelagicus*) based on cytochrome b mitochondrial DNA variation. *Marine Biology*, **145**, 1257–1264.

Cahill, A.E. & Levinton, J.S. (2016) Genetic differentiation and reduced genetic diversity at the northern range edge of two species with different dispersal modes. *Molecular Ecology*, **25**, 515–526.

Campagna, L., Gronau, I., Silveira, L.F., Siepel, A. & Lovette, I.J. (2015) Distinguishing noise from signal in patterns of genomic divergence in a highly polymorphic avian radiation. *Molecular Ecology*, **24**, 4238–4251.

Campioni, L., Granadeiro, J.P. & Catry, P. (2017) Albatrosses prospect before choosing a home: intrinsic and extrinsic sources of variability in visit rates. *Animal Behaviour*, **128**, 85–93.

Candy, J.R., Campbell, N.R., Grinnell, M.H., Beacham, T.D., Larson, W.A. & Narum, S.R. (2015) Population differentiation determined from putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. *Molecular Ecology Resources*, **15**, 1421–1434.

Carlson, C.S., Eberle, M.A., Rieder, M.J., Yi, Q., Kruglyak, L. & Nickerson, D.A. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *American Journal of Human Genetics*, **74**, 106–120.

Catchen, J.M. (2013) STACKS: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Chardine, J.W., Rail, J.F. & Wilhelm, S. (2013) Population dynamics of northern gannets in North America, 1984-2009. *Journal of Field Ornithology*, **84**, 187–192.

Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W. & McCabe, A.M. (2009) The Last Glacial Maximum. *Science*, **325**, 710–714.

Cornuet, J.M. & Luikart, G. (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Cornuet, J.M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J. M. & Estoup, A. (2014) DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* **30**, 1187 - 1189.

Cornuet, J.M., Santos, F., Beaumont, M.A., Robert, C.P., Marin, J.M., Balding, D.J., Guillemaud, T. & Estoup, A. (2008) Inferring population history with DIYABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.

Coulson, J.C. (2016) A Review of Philopatry in Seabirds and Comparisons with Other Waterbird Species. *Waterbirds*, **39**, 229–240.

Crandall, K.A., Bininda-emonds, O.R., Mace, G.M. & Wayne, R.K. (2000) Consideing evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, **15**, 290–295.

Cure, K., Thomas, L., Hobbs, J.-P.A., Fairclough, D.V. & Kennington, W.J. (2017) Genomic signatures of local adaptation reveal source-sink dynamics in a high gene flow fish species. *Scientific Reports*, **7**, 8618. Springer US.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G.M., Gabor T.S., Stephen T., McVean, G. & Durbin, R. (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. & Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-

generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Deng, H. & Runger, G. (2012) Feature selection via regularized trees. *Proceeding of the International Joint Conference on Neural Networks*, 10–15.

Dillingham, P.W. (2010) Generation time and the maximum growth rate for populations with age-specific fecundities and unknown juvenile survival. *Ecoloigcal Modelling*, **221**, 895–899.

Dittmann, T., Ezard, T.H. & Becker, P.H. (2007) Prospectors' colony attendance is sex-specific and increases future recruitment chances in a seabird. *Behavioural Processes*, **76**, 198–205.

Domínguez, M., Tiedemann, R., Reboreda, J.C., Segura, L., Tittarelli, F. & Mahler, B. (2017) Genetic structure reveals management units for the yellow cardinal (*Gubernatrix cristata*), endangered by habitat loss and illegal trapping. *Conservation Genetics*, **18**, 1131–1140.

Earl, D.A. & vonHoldt, B.M. (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Edwards, S.V. & Beerli, P. (2000) Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*, **54**, 1839.

Egevang, C., Stenhouse, I.J., Phillips, R.A., Petersen, A., Fox, J.W. & Silk, J.R. (2010) Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration.

Proceeding of the National Academy of Sciences, **107**, 2078–2081.

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

Excoffier, L. & Lischer, H.E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Felsenstein, J. (1976) The theoretical population genetics of variable selection and migration. *Annual Review of Genetics*, **10**, 253–280.

Fifield, D.A., Montevecchi, W.A., Garthe, S., Gregory, J., Kubetzki, U. & Rail, J. (2014) Migratory tactics and wintering areas of northern gannets *(Morus Bassanus)* breeding in North America. *Ornithological Monographs*, **79**.

Foote, A., Vijay, N., Avila-Arcos, M., Baird, R., Durban, J., Morin, P., Fumagalli, M., Gibbs, R., Hanson, B., Korneliussen, T., Martin, M., Robertson, K., Sousa, V., Vieira, F., Vinar, T., Wade, P., Worley, K., Excoffier, L., Gilbert, T. & Wolf, J. (2016) Genome-culture coevolution promotes rapid divergence in the killer whale. *Nature Communications,* 1–48.

Fort, J., Pettex, E., Tremblay, Y., Lorentsen, S.H., Garthe, S., Votier, S., Pons, J.B., Siorat, F., Furness, R.W., Grecian, W.J., Bearhop, S., Montevecchi, W.A. & Grémillet, D. (2012) Meta-population evidence of oriented chain migration in northern gannets (*Morus bassanus*). *Frontiers in Ecology and the Environment*, **10**, 237–242.

Fraïsse, C., Roux, C., Welch, J.J. & Bierne, N. (2014) Gene-flow in a mosaic hybrid zone: Is local introgression adaptive? *Genetics*, **197**, 939–951.

Friesen, V.L. (2015) Speciation in seabirds: why are there so many species…and why aren't there more? *Journal of Ornithology*, **156**, 27–39.

Friesen, V.L., Burg, T.M. & McCoy, K.D. (2007) Mechanisms of population differentiation in seabirds: Invited review. *Molecular Ecology*, **16**, 1765–1785.

Funk, W.C., McKay, J.K., Hohenlohe, P.A. & Allendorf, F.W. (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, **27**, 489–496.

Gaillard, J.M., Yoccoz, N.G., Lebreton, J.D., Bonenfant, C., Devillard, S., Loison, A., Pontier, D. & Allaine, D. (2005) Generation time: A reliable metric to measure life-history variation among mammalian populations. *American Naturalist*, **166**, 119–123.

Garthe, S., Benvenuti, S. & Montevecchi, W.A. (2000) Pursuit plunging by northern gannets (Sula bassana) feeding on capelin (Mallotus villosus). *Proceedings of the Royal Society B: Biological Science*, **267**, 1717–1722.

Garthe, S., Hallgrimsson, G.T., Montevecchi, W.A., Fifield, D. & Furness, R.W. (2016) East or west? Migration routes and wintering sites of northern gannets *Morus bassanus* from south-eastern Iceland. *Marine Biology*, **163**.

Garthe, S., Peschko, V., Kubetzki, U. & Corman, A.M. (2017) Seabirds as samplers of the marine environment - a case study of northern gannets. *Ocean Sciences*, **13**,

337–347.

Gaston, A., Brewer, D. & Diamond, A.W. (2008) Canadian Atlas of Bird Banding-Volume 2.

Gómez-Díaz, E., González-Solís, J., Peinado, M.A. & Page, R.D. (2006) Phylogeography of the Calonectris shearwaters using molecular and morphometric data. *Molecular Phylogenetics and Evolution*, **41** 322–332.

Greenwood, P.J. (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, **28**, 1140–1162.

Grémillet, D., Pichegru, L., Siorat, F. & Georges, J.Y. (2006) Conservation implications of the apparent mismatch between population dynamics and foraging effort in French northern gannets from the English Channel. *Marine Ecology Progress Series*, **319**, 15–25.

Hamer, K.C., Humphreys, E.M., Garthe, S., Hennicke, J., Peters, G., Grémillet, D., Phillips, R. A., Harris, M. P. & Wanless, S. (2007) Annual variation in diets, feeding locations and foraging behaviour of gannets in the North Sea: Flexibility, consistency and constraint. *Marine Ecology Progress Series*, **338**, 295–305.

Hamer, K.C., Phillips, R.A., Wanless, S., Harris, M.P. & Wood, A.G. (2000) Foraging ranges, diets and feeding locations of gannets Morus bassanus in the North Sea: Evidence from satellite telemetry. *Marine Ecology Progress Series*, **200**, 257–264.

Haney, J.C., Jodice, P.G.R., Montevecchi, W.A. & Evers, D.C. (2017) Challenges to oil spill assessment for seabirds in the deep ocean. *Archives of Environmental*

Contamination and Toxicology, **73**, 33–39.

Hanski, I. (1998) Metapopulation dynamics. *Nature*, **396**, 41–49.

Hastings, A. (1993) Complex interaction between dispersal and dnamics: lessons from coupled logistic equations. *Ecology*, **74**, 1362–1372.

Hazevoet, C.J. (2014) Eighth report on birds from the Cape Verde Islands, including records of nine taxa new to the archipelago. *Zoologia Caboverdiana*, **5**, 29–56.

Head, M.J. & Gibbard, P.L. (2015) Formal subdivision of the Quaternary System/Period: Past, present, and future. *Quaternary International*, **383**, 4–35.

Hedrick, P.W., Ginevan, M.E. & Ewing, E.P. (1976) Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology, Evolution, and Systematics*, **7**, 1–32.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt, G. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B Biological Sciences*, **359**, 183–195.

Jacobs, A., De Noia, M., Praebel, K., Kanstad-Hanssen, Ø., Paterno, M., Jackson, D., McGinnity, P., Sturm, A., Elmer, K.R. & Llewellyn, M. S. (2017) Genetic fingerprinting of salmon louse (*Lepeophtheirus salmonis*) populations in the North-East Atlantic using a random forest classification approach. *bioRxiv*, 179218.

Jombart, T. & Ahmed, I. (2011) adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Kawakami, T., Backström, N., Burri, R., Husby, A., Olason, P., Rice, A.M., Ålund, M., Qvarnström, A. & Ellegren, H. (2014) Estimation of linkage disequilibrium and interspecific gene flow in Ficedula flycatchers by a newly developed 50k single-nucleotide polymorphism array. *Molecular Ecology Resources*, **14**, 1248–1260.

Kawecki, T.J. & Ebert, D. (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.

Kubetzki, U., Garthe, S., Fifield, D., Mendel, B. & Furness, R.W. (2009) Individual migratory schedules and wintering areas of northern gannets. *Marine Ecology Progress Series*, **391**, 257–265.

Kühne, G., Kosuch, J., Hochkirch, A. & Schmitt, T. (2017) Extra-Mediterranean glacial refugia in a Mediterranean faunal element : the phylogeography of the chalk- hill blue *Polyommatus coridon*. *Scientific Reports*, 1–11.

Laggis, A., Baxevanis, A.D., Charalampidou, A., Maniatsi, S., Triantafyllidis, A. & Abatzopoulos, T.J. (2017) Microevolution of the noble crayfish (*Astacus astacus*) in the Southern Balkan Peninsula. *BMC Evolutionary Biology*, **17**, 122.

Lee, C.R., Svardal, H., Farlow, A., Exposito-Alonso, M., Ding, W., Novikova, P., Alonso-Blanco, C., Weigel, D. & Nordborg, M. (2017) On the post-glacial spread of human commensal *Arabidopsis thaliana*. *Nature Communications*, **8**, 14458.

Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E., Geman, D., Baggerly, K. & Irizarry, R.A. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, **11**,733–739.

Lescroel, A., Mathevet, R., Peron, C., Authier, M., Provost, P., Takahashi, A. & Gremillet, D. (2016) Seeing the ocean through the eyes of seabirds: A new path for marine conservation? *Marine Policy*, **68**, 212–220.

Leslie, P.H. (1966) The intrinsic rate of increase and the overlap of successive generations in a population of guillemots (*Uria aalge*). *Journal of Animal Ecology*, **35**, 291–301.

Leys, M., Keller, I., Räsänen, K., Gattolliat, J.L. & Robinson, C.T. (2016) Distribution and population genetic variation of cryptic species of the Alpine mayfly Baetis alpinus (Ephemeroptera: Baetidae) in the Central Alps. *BMC Evolutionary Biology*, **16**, 77.

Liebers, D., de Knijff, P. & Helbig, A.J. (2004) The herring gull complex is not a ring species. *Transactions of the Royal Society B: Biological Sciences*, **271**, 893–901.

Liebers, D. & Helbig, A.J. (2002) Phylogeography and colonization history of Lesser Black-backed Gulls (*Larus fuscus*) as revealed by mtDNA sequences. *Journal of Evolutionary Biology*, **15**, 1021–1033.

Lowe, W.H. & Allendorf, F.W. (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–3051.

Manel, S., Gaggiotti, O.E. & Waples, R.S. (2005) Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, **20**, 136–142.

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G.R.,

Espiñeira, M., Fiorentino, F., Garofalo, G., Geffen, A.J., Hansen, J.H., Helyar, S.J., Nielsen, E.E., Ogden, R., Patarnello, T., Stagioni, M., Tinti, F. & Bargelloni, L. (2014) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, **23**, 118–135.

Milot, E., Weimerskirch, H. & Bernatchez, L. (2008) The seabird paradox: Dispersal, genetic structure and population dynamics in a highly mobile, but philopatric albatross species. *Molecular Ecology*, **17**, 1658–1673.

Montevecchi, W.A., Benvenuti, S., Garthe, S., Davoren, G.K. & Fifield, D. (2009) Flexible foraging tactics by a large opportunistic seabird preying on forage-and large pelagic fishes. *Marine Ecology Progress Series*, **385**, 295–306.

Montevecchi, W.A., Hedd, A., McFarlane Tranquilla, L., Fifield, D.A., Burke, C.M., Regular, P.M., Davoren, G.K., Garthe, S., Robertson, G.J. & Phillips, R.A. (2012a) Tracking seabirds to identify ecologically important and high risk marine areas in the western North Atlantic. *Biological Conservation*, **156**, 62–71.

Montevecchi, W., Fifield, D., Burke, C., Garthe, S., Hedd, A., Rail, J.F. & Robertson, G.J. (2012b) Tracking long-distance migration to assess marine pollution impact. *Biology Letters*, **8**, 218–21.

Moritz, C. (1994) Defining "evolutionarily significant units" for conservation. *Trends in Ecology and Evolution*, **9**, 373–375.

Morris-Pocock, J.A., Taylor, S.A., Birt, T.P., Damus, M., Piatt, J.F., Warheit, K.I. & Friesen, V.L. (2008) Population genetic structure in Atlantic and Pacific Ocean common murres (*Uria aalge*): Natural replicate tests of post-Pleistocene evolution.

Molecular Ecology, **17**, 4859–4873.

Moss, R., Wanless, S. & Harris, M.P. (2002) How small northern gannet colonies grow faster than big ones. *Waterbirds*, **25**, 442–448.

Moum, T. & Arnason, E. (2001) Genetic diversity and population history of two related seabird species based on mitochondrial DNA control. *Molecular Ecology*, **10**, 2463–2478.

Munro, K.J. & Burg, T.M. (2017) A review of historical and contemporary processes affecting population genetic structure of Southern Ocean seabirds. *Emu Austral Ornithology*, **117**, 4–18.

Murray, S., Harris, M.P. & Wanless, S. (2015) The status of the gannet in Scotland in 2013-14. *Scottish Birds*, **35**, 3–18.

Nelson, J.B. (1978). The Gannet. Buteo Books. Vermillion, South Dakota.

Nelson, J.B. (2002) The atlantic gannet. Fenix Books Ltd, Norfolk

Ng, E.Y., Garg, K.M., Low, G.W., Chattopadhyay, B., Oh, R.R., Lee, J.G. & Rheindt, F.E. (2017) Conservation genomics identifies impact of trade in a threatened songbird. *Biological Conservation*, **214**, 101–108.

Niel, C. & Lebreton, J.D. (2005) Using demographic invariants to detect overharvested bird populations from incomplete data. *Conservation Biology*, **19**, 826–835.

Ogden, R., Gharbi, K., Mugue, N., Martinsohn, J., Senn, H., Davey, J.W., Pourkazemi, M., McEwing, R., Eland, C., Vidotto, M., Sergeev, A. & Congiu, L. (2013) Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing.

Molecular Ecology, **22**, 3112–3123.

Oliveira, R., Randi, E., Mattucci, F., Kurushima, J.D., Lyons, L. & Alves, P.C. (2015) Toward a genome-wide approach for detecting hybrids: informative SNPs to detect introgression between domestic cats and European wildcats (*Felis silvestris*). *Heredity*, **115**, 195–205.

Oros Sršen, A., Kralj, J. & Šešelj, L. (2017) The late-Holocene avifaunal assemblage from the island of Palagruža (Croatia): The earliest record of the Northern Gannet in the Adriatic Sea. *The Holocene*, **27**, 1540–1549.

Paetkau, D., Slade, R., Burden, M. & Estoup, A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Pereira, S.L. & Baker, A.J. (2008) DNA evidence for a Paleocene origin of the Alcidae (Aves: Charadriiformes) in the Pacific and multiple dispersals across northern oceans. *Molecular Phylogenetics and Evolution*, **46**, 430–445.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, **7**.

Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L. & Estoup, A. (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536–539.

Poland, J.A., Brown, P.J., Sorrells, M.E. & Jannink, J.L. (2012) Development of high-

density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, **7**.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Pritchard, V.L., Erkinaro, J., Kent, M.P., Niemelä, E., Orell, P., Lien, S. & Primmer, C.R. (2016) Single nucleotide polymorphisms to discriminate different classes of hybrid between wild Atlantic salmon and aquaculture escapees. *Evolutionary Applications*, **9**, 1017–1031.

Provan, J. & Bennett, K.D. (2008) Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology and Evolution*, **23**, 564–571.

Rambaut, A., Suchard, M.A., Xie, D. and Drummond, A.J. (2014) Tracer v1.6. (http://beast.bio.ed.ac.uk/Tracer).

Rayner, M.J., Hauber, M.E., Steeves, T.E., Lawrence, H.A., Thompson, D.R., Sagar, P.M., Bury, S.J., Landers, T.J., Phillips, R.A., Ranjard, L. & Shaffer, S.A. (2011) Contemporary and historical separation of transequatorial migration between genetically distinct seabird populations. *Nature Communications*, **2**, 332.

Roesti, M., Salzburger, W. & Berner, D. (2012) Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evolutionary Biology*, **12**, 94.

Rosenberg, N.A. (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Ruegg, K.C., Anderson, E.C., Paxton, K.L., Apkenas, V., Lao, S., Siegel, R.B., Desante,

D.F., Moore, F. & Smith, T.B. (2014) Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology*, **23**, 5726–5739.

Savolainen, O., Lascoux, M. & Merilä, J. (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics*, **14**, 807–20.

Sawyer, S.L. & Fogle, S.R. (2013) Establishment of a new breeding colony of Australasian gannets (*Morus serrator*) at Young Nick's Head Peninsula. *Notornis*, **60**, 180–182.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) Molecular cloning: a laboratory manual (No. Ed. 2). *Cold spring harbor laboratory press*.

Siorat, F. & Rocamora, G. (1995). Changes in numbers and distribution of the Northern Gannet (*Morus bassanus*) on Rouzic Island, (Reserve naturelle des Sept-Iles, Bretagne), France 1939-1994. *Colonial Waterbirds*, **18**, 172–178.

Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Slatkin, M. (1985) Gene flow in natural populations. *Annual Review of Ecology, Evolution and Systematics*, **16**, 393-430.

Stauss, C., Bearhop, S., Bodey, T.W., Garthe, S., Gunn, C., Grecian, W.J., Inger, R., Knight, M.E., Newton, J., Patrick, S.C., Phillips, R.A., Waggitt, J.J. & Votier, S.C. (2012) Sex-specific foraging behaviour in northern gannets *Morus bassanus*: Incidence and implications. *Marine Ecology Progress Series*, **457**, 151–162.

Steeves, T.E., Anderson, D.J. & Friesen, V.L. (2005) A role for nonphysical barriers to

gene flow in the diversification of a highly vagile seabird, the masked booby (*Sula dactylatra*). *Molecular Ecology*, **14**, 3877–3887.

Steeves, T.E., Anderson, D.J., McNally, H., Kim, M.H. & Friesen, V.L. (2003) Phylogeography of Sula: the role of physical barriers to gene flow in the diversification of tropical seabirds. *Journal of Avian Biology*, **34**, 217–223.

Stephan, J., Stegle, O. & Beyer, A. (2015) A random forest approach to capture genetic effects in the presence of population structure. *Nature Communications*, **6**, 7432.

Sternkopf, V., Liebers-Helbig, D., Ritz, M.S., Zhang, J., Helbig, A.J. & de Knijff, P. (2010) Introgressive hybridization and the evolutionary history of the herring gull complex revealed by mitochondrial and nuclear DNA. *BMC Evolutionary Biology*, **10**, 348.

Stewart, J.R., Lister, A.M., Barnes, I. & Dalen, L. (2010) Refugia revisited: individualistic responses of species in space and time. *Transactions of the Royal Society B: Biological Sciences*, **277**, 661–671.

Storfer, A., Epstein, B., Jones, M., Micheletti, S., Spear, S.F., Lachish, S. & Fox, S. (2017) Landscape genetics of the Tasmanian devil: implications for spread of an infectious cancer. *Conservation Genetics*, **18**, 1–11.

Switzer, P.V. (1993) Site fidelity in predictable and unpredictable habitats. *Evolutionary Ecology*, **7**, 533–555.

Sylvester, E.V., Bentzen, P., Bradbury, I.R., Clément, M., Pearce, J., Horne, J. & Beiko, R.G (2017) Applications of random forest feature selection for fine-scale genetic

population assignment. *Evolutionary Applications*, 1–13.

Thanou, E., Sponza, S., Nelson, E.J., Perry, A., Wanless, S., Daunt, F. & Cavers, S. 2016. Genetic structure in the European endemic seabird, *Phalacrocorax aristotelis* , was shaped by a complex interaction of historical and contemporary, physical and non-physical drivers. *Molecular Ecology*, doi: 10.1111/mec.13996.

Tigano, A., Damus, M., Birt, T.P., Morris-Pocock, J.A., Artukhin, Y.B. & Friesen, V.L. (2015) The arctic: Glacial refugium or area of secondary contact? Inference from the population genetic structure of the thick-billed murre (*Uria lomvia*), with Implications for management. *Journal of Heredity*, **106**, 238–246.

Tigano, A. & Friesen, V.L. (2016) Genomics of local adaptation with gene flow. *Molecular Ecology*, **25**, 2144–2164

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**, 2370–2381.

Tyrberg, T. (1999) Seabirds and late Pleistocene marine environments in the northeast Atlantic and the Mediterranean. *Smithsonian Contributions to Zoology*, **89**, 139–157.

Vargas-Rodriguez, Y.L., Platt, W.J., Urbatsch, L.E. & Foltz, D.W. (2015) Large scale patterns of genetic variation and differentiation in sugar maple from tropical Central America to temperate North America. *BMC Evolutionary Biology*, **15**, 257.

Veron, P.K. & Lawlor, M.P. (2009) The dispersal and migration of the Northern Gannet

Morus bassanus from Channel Islands breeding colonies. *Seabird*, **22**, 37–47.

de Vicente, M.C., Guzman, F.A., Engels, J. & Rao, V.A. (2006) Genetic characterization and its use in decision-making for the conservation of crop germplasm. *The role of biotechnology in exploring and protecting agricultural genetic resources*, 2129-2138.

Votier, S.C., Bearhop, S., Witt, M.J., Inger, R., Thompson, D. & Newton, J. (2010) Individual responses of seabirds to commercial fisheries revealed using GPS tracking, stable isotopes and vessel monitoring systems. *Journal of Applied Ecology*, **47**, 487–497.

Votier, S.C., Grecian, W.J., Patrick, S. & Newton, J. (2011) Inter-colony movements, at-sea behaviour and foraging in an immature seabird: Results from GPS-PPT tracking, radio-tracking and stable isotope analysis. *Marine Biology*, **158**, 355–362.

Wakefield, E.D., Bodey, T.W., Bearhop, S., Blackburn, J., Colhoun, K., Davies, R.,Dwyer, R.G., Green, J.A., Gremillet, D., Jackson, A.L., Jessopp, M.J., Kane, A.,Langston, R.H., Lescroel, A., Murray, S., Le Nuz, M., Patrick, S.C.,Peron, C., Soanes, L.M., Wanless, S., Votier, S.C. & Hamer, K.C. (2013) Space partitioning without territoriality in gannets. *Science*, **341**, 68–70.

Wanless, S., Frederiksen, M., Harris, M.P. & Freeman, S.N. (2006) Survival of Gannets *Morus bassanus* in Britain and Ireland, 1959-2002. *Bird Study*, **53**, 79–85.

Waples, R.S. (2010) High-grading bias: Subtle problems with assessing power of selected subsets of loci for population assignment. *Molecular Ecology*, **19**, 2599–2601.

Waples, R.S. & Gaggiotti, O. (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

Waser, P.M. & Strobeck, C. (1998) Genetic signatures of interpopulation dispersal. *Trends in Ecology and Evolution*, **13**, 43–44.

Wasser, S.K., Brown, L., Mailand, C., Mondol, S., Clark, W., Laurie, C. & Weir, B.S. (2015) Genetic assignment of large seizures of elephant ivory reveals Africa ' s major poaching hotspots, **349**, 84–88.

Webster, M.S., Marra, P.P., Haig, S.M., Bensch, S. & Holmes, R.T. (2002) Links between worlds: Unraveling migratory connectivity. *Trends in Ecology and Evolution*, **17**, 76-83

Weesie, P.D.M. (1987) The Quaternary avifauna of Crete, Greece. *Palaeovertebrata*, **18**, 1–94.

Weir, J.T., Haddrath, O., Robertson, H.A., Colbourne, R.M. & Baker, A.J. (2016) Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences*, **113**, E5580–E5587.

Wielstra, B., ZieliŃski, P. & Babik, W. (2017) The Carpathians hosted extra-Mediterranean refugia-within-refugia during the Pleistocene Ice Age: genomic evidence from two newt genera. *Biological Journal of the Linnean Society*, **14**, 352–360.

Wilson, D.S. & Yoshimura, J. (1994) On the Coexistence of Specialists and Generalists.

American Naturalist, **144**, 692–707.

Wilson, G.A. & Rannala, B. (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177–91.

Wolf, J.B. & Ellegren, H. (2016) Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, doi: 10.1038/nrg.2016.133.

Wright, S. (1931) Evolution in mendelian populations. *Genetics*, **16**, 97–159.

Wright, S. (1951) The Genetical Structure of Populations. *Annals of Eugenics*, **15**, 322–354.

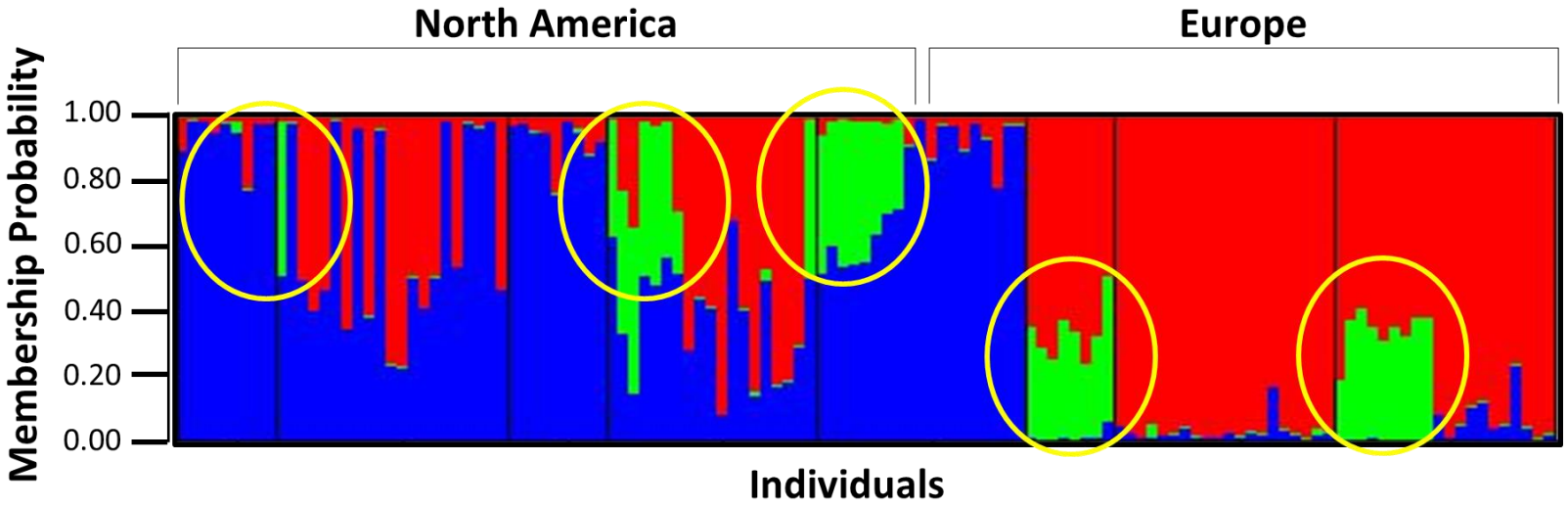
Yannic, G., Yearsley, J.M., Sermier, R., Dufresnes, C., Gilg, O., Aebischer, A., Gavrilo, M.V., Strøm, H., Mallory, M.L., Guy Morrison, R.I., Gilchrist, H.G. & Broquet, T. (2016) High connectivity in a long-lived high-Arctic seabird, the ivory gull *Pagophila eburnea*. *Polar Biology*, **39**, 221–236.

Zhang, D.-X. & Hewitt, G.M. (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology*, **12**, 563–584.

Zotier, R., Bretagnolle, V. & Thibault, J.C. (1999) Biogeography of the marine birds of a confined sea, the Mediterranean. *Journal of Biogeography*, **26**, 297–313.

Appendix

Appendix A1. A bar plot produced in STRUCTURE, displaying the batch effect detected in the northern gannet sequencing data. The effect (circles) was associated with sequencing lane. Several individuals were included in both batches as duplicates. These duplicate individuals were used to identify and remove loci unique to either sequencing lane in STACKS (Catchen, 2013). Once these loci were filtered from the dataset, a subsequent AMOVA found no significant difference between a subset of individuals sequenced across the two lanes, indicating that the effect was removed.



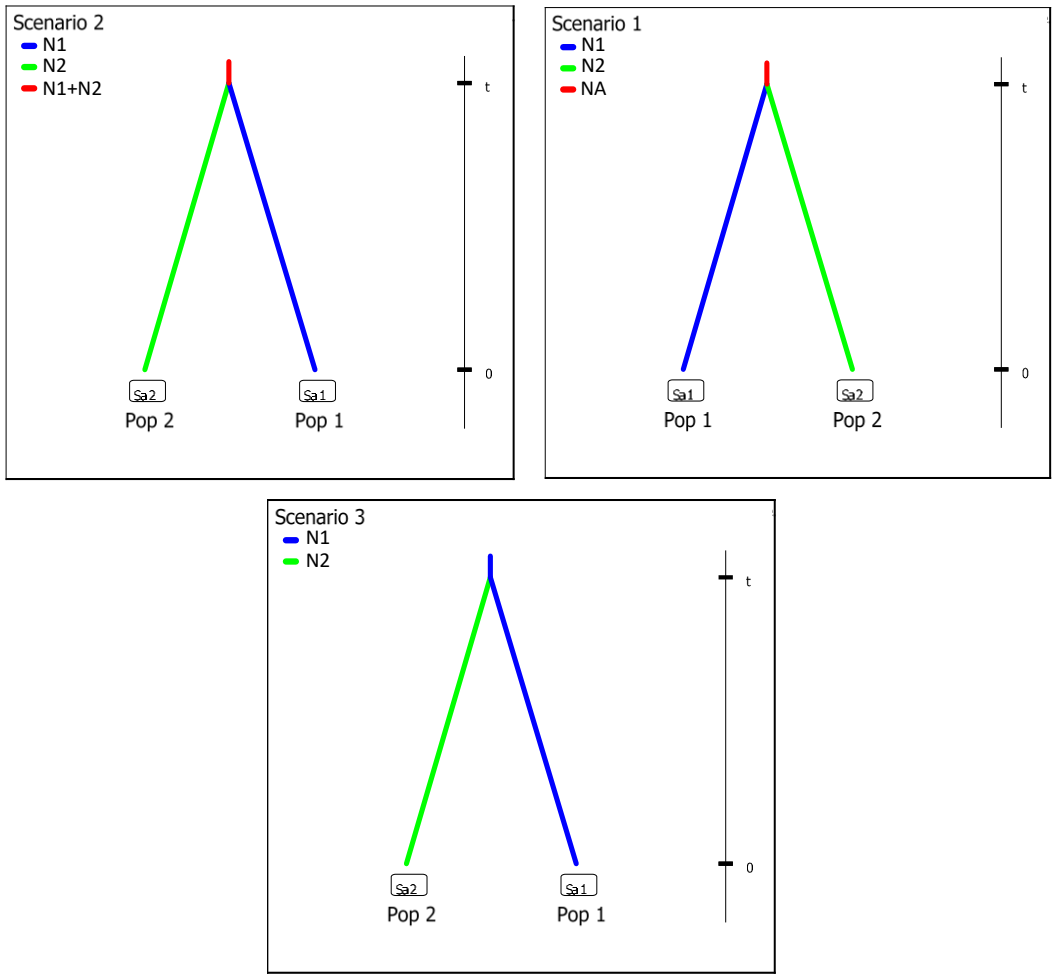
Appendix A2. Results of the locus-by-locus analysis of molecular variance (AMOVA) used to confirm the removal of the detected batch effect. Samples were grouped by sequencing lane.

Source of variation	Sum of squares	Variance components	Percentage of variation	p-value	Fixation indices
Among lanes	205.420	-0.452	-0.246	0.95186	-0.00246
Among colonies within lanes	858.021	3.125	1.702	0.00000	0.01698
Within colonies	12309.034	180.906	98.540		
Total	13372.476	183.579	100	0.00000	0.01456

Appendix A3. Proportion of individuals that are recent migrants (first- or second-generation) as inferred with BAYESASS (Wilson *et al.*, 2003). Columns indicate the source of migrants and rows the receiving population. All runs used a burn-in of 900,000, and 9,000,000 iterations. Self-recruitment rates are given in bold. Upper and lower 95% confidence limits are given in parentheses.

Run	Seed	Proportion of recent migrants (m)		
			North America	Europe
1	10	North America	**0.949** (0.918 – 0.981)	0.051 (0.019 – 0.082)
		Europe	0.014 (0.000 – 0.032)	**0.986** (0.968 – 1.000)
2	68	North America	**0.949** (0.918 – 0.981)	0.051 (0.019 – 0.082)
		Europe	0.014 (0.000 – 0.032)	**0.986** (0.968 – 1.000)
3	90	North America	**0.948** (0.916 – 0.981)	0.052 (0.020 – 0.085)
		Europe	0.014 (0.000 – 0.032)	**0.986** (0.967 – 1.000)

Appendix A4. The three scenarios used to simulate data sets in DIYABC. Figures show a simple divergence, whereby one population of gannets diverged into the contemporary North American and European populations at some point in the past (t). NA represents the effective population size before the divergence event and N1 and N2 represent the present-day effective population sizes of gannets in North America and Europe respectively.



Appendix A5. Proportion of individuals that are recent migrants (first- or second-generation) as inferred with BAYESASS (Wilson *et al.*, 2003). Columns indicate the source of migrants and rows the receiving population. Run used a burn-in of 900,000 and 9,000,000 iterations. Self-recruitment rates are given in bold. Standard deviations are given in parentheses.

	BAC	BAS	BON	CSM	FUN	MAG	ROU	SKA
BAC	**0.810**	0.018	0.018	0.018	0.020	0.069	0.021	0.025
	(0.066)	(0.019)	(0.016)	(0.018)	(0.020)	(0.058)	(0.018)	(0.022)
BAS	0.020	**0.774**	0.020	0.022	0.021	0.028	0.033	0.082
	(0.020)	(0.072)	(0.019)	(0.020)	(0.021)	(0.025)	(0.025)	(0.054)
BON	0.019	0.020	**0.765**	0.019	0.020	0.118	0.020	0.019
	(0.017)	(0.018)	(0.079)	(0.017)	(0.020)	(0.079)	(0.017)	(0.017)
CSM	0.018	0.021	0.018	**0.768**	0.017	0.096	0.021	0.041
	(0.017)	(0.018)	(0.017)	(0.076)	(0.017)	(0.062)	(0.021)	(0.029)
FUN	0.020	0.019	0.020	0.020	**0.765**	0.099	0.020	0.038
	(0.018)	(0.016)	(0.018)	(0.019)	(0.069)	(0.056)	(0.018)	(0.029)
MAG	0.014	0.013	0.012	0.012	0.024	**0.890**	0.012	0.023
	(0.013)	(0.013)	(0.012)	(0.012)	(0.016)	(0.030)	(0.011)	(0.015)
ROU	0.010	0.011	0.011	0.012	0.014	0.011	**0.912**	0.020
	(0.010)	(0.011)	(0.010)	(0.011)	(0.012)	(0.011)	(0.031)	(0.020)
SKA	0.010	0.013	0.015	0.012	0.014	0.013	0.031	**0.892**
	(0.010)	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)	(0.020)	(0.031)

Appendix A6. Type I and Type II error rates and posterior probabilities for each scenario simulated in DIYABC.

True Scenario	Type I error rate	Type II error rate	Posterior probability
Scenario 1	0.23	0.30	0.2548 [0.2290,0.2805]
Scenario 2	0.18	0.26	0.5407 [0.5146,0.5669]
Scenario 3	0.32	0.28	0.2045 [0.1827,0.2263]