UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO BIOLOGIA ANIMAL

Population genomics of the endangered Bermuda petrel

Rita Oliveira Afonso

Mestrado em Biologia Evolutiva e do Desenvolvimento

Dissertação orientada por:
Doutora Mónica Silva e Doutor Francisco Pina Martins

2021

Para a Beatriz

Agradecimentos/Acknowledgments

Em primeiro lugar, quero agradecer à Mónica Silva, por toda a orientação ao longo deste ano, principalmente pela enorme disponibilidade e prioridade que deu a este trabalho, e também por ter criado sempre um ambiente descontraído, mas sério, ao longo de todo o processo. Por todos os conselhos honestos, chamadas de atenção e pelo respeito que sempre senti. Agradeço especialmente a confiança desde o início e por me ter "deixado" errar, sozinha, o que me fez evoluir bastante. Sem dúvida, esta tese também é fruto da sua dedicação e do conhecimento que me transmitiu.

Agradeço ao professor Francisco, por toda a ajuda, sugestões e ideias incríveis surgidas na hora, o que sempre me impressionou. Por me ter impulsionado no mundo bioinformático e por ter me ter mostrado que as linhas de comandos assustadoras, afinal são simples e até bonitas. Pelo método suave e paciente com que sempre ensinou e pelo excelente sentido de humor, que muito aprecio. Agradeço pelo entusiasmo com meus pequenos progressos, até por vezes, maior que o meu e pelas palavras de encorajamento, que me fizeram mais confiante.

Quero também agradecer ao Vítor, pelos conselhos nas análises de demografia e pela simpatia, sempre. À Letizia pela partilha de conhecimentos e dados sobre a ecologia da espécie. Ao Jeremy Madeiros, pelo esforço na recolha e entrega dos dados, valiosos, e também por toda a dedicação inspiradora à Freira da Bermuda. Agradeço também à Matilde, pela amplificação do fragmento do gene COI da espécie e também pela ajuda no laboratório. À Fátima, pela companhia e partilha, no laboratório, que tornaram aqueles dias mais calorosos, mesmo com toda a situação pandémica. Por fim, agradeço à Vera, por todos os conselhos desde o início e pelo interesse sincero em ajudar, e também pelas reuniões das "freiras", que me ajudaram, muitas vezes, a orientar.

Quero deixar um agradecimento à comunidade anónima *online*, especialmente na área da bioinformática, que me "safou" várias vezes e que é realmente inspiradora, neste mundo estranho. Também à Alexandra pela generosidade e coragem, que tanto me facilitou a vida.

Agradeço aos meus colegas de mestrado, em especial ao Salo, à Mafs e à Sofia, pela amizade e pela partilha de felicidade, de frustrações e de parvoíces.

A todos os meus amigos, que me acompanharam nesta jornada e que tanto adoro. Ao Ricardo, o meu Balsito, um agradecimento maior pelo carinho e força que me deu e dá.

Agradeço também à minha família, à "bó" Margarida, tios e primos e também aos que já não estão connosco, pelo apoio que me fez chegar até aqui. À Isilda e Paulo pela comidinha boa e por me terem feito sentir sempre em casa. Um agradecimento especial à Paula, Tozé e Carlos, família de coração, por tanto que contribuíram para a minha educação e felicidade e pelo afeto que me dão até hoje. Aos meus pais, que permitiram que eu estudasse até aqui e me deram a oportunidade de escolha e pelo apoio e valores que me passaram ao longo da vida.

Ao Simão, meu mano e padrinho, agradeço a falta de paciência, que me fez aprender mais, mas também o apoio incondicional em tudo. Agradeço os kiwis partidos e as loiças que não cobrou nos momentos mais aflitos da tese.

Ao André, meu amor, pelo carinho e compreensão, principalmente neste último ano. Por ser o meu maior apoio e por me trazer momentos de felicidade profunda. Agradeço também a partilha da paixão por pardela, que foi essencial nestes últimos tempos.

Em último lugar, agradeço à Beatriz, minha companheira de vida, de todas as horas, pela cumplicidade intensa, inata e sem explicação, que moldou a minha vida. Embora agora de maneira diferente, agradeço que nunca me tenha deixado sozinha, mesmo quando a casa estava vazia. Para ti dedico todo o meu esforço e trabalho, já que foi em ti que teve origem.

Sem jeito para mais palavras, bem-haja a todos que fizeram parte do meu caminho!

Rita

Resumo

A freira da Bermuda (*Pterodroma cahow*) é uma espécie endémica das Ilhas das Bermudas e pertence à Ordem dos Procelariiformes, um dos grupos de aves marinhas mais ameaçadas do planeta. Demonstra características e comportamentos comuns desta Ordem de espécies, tal como a postura de um único ovo por época reprodutora, o alto investimento parental por parte dos dois membros do casal e filopatria, característica que designa a tendência de retorno dos indivíduos ao local onde foram criados, para se reproduzirem. Esta espécie passa a maior parte do seu tempo de vida em alto mar no norte do Oceano Atlântico e apenas regressa a terra na época reprodutora, onde ocorre a escolha de parceiro, que se mantém normalmente para a vida, assim como a escolha do ninho, geralmente construídos através de escavações no solo. Registos históricos indicam que a freira da Bermuda terá sofrido um enorme decréscimo populacional aquando da chegada dos humanos às ilhas, no século XVII. A desflorestação e a introdução de espécies predadoras invasoras terão causado esse decréscimo, levando ao desaparecimento de avistamentos da espécie em apenas 12 anos. Assim, a crença de que a espécie se tivera extinguido permaneceu durante 300 anos, cerca de 15 gerações. Surpreendentemente, em 1951, novos avistamentos terão levado à descoberta de cerca de 14 casais num ilhéu de difícil acesso. Consequentemente, foi implantado um extenso plano de conservação que dura até aos dias de hoje e que tem resultado numa extraordinária recuperação, existindo cerca de 150 casais a nidificar nas Bermudas, atualmente. No entanto, esta espécie ainda enfrenta várias ameaças, nomeadamente ambientais. Além disso, teme-se que o decréscimo populacional que a freira da Bermuda sofreu tenha causado um efeito de gargalo que tem consequências genéticas negativas, resultando numa forte deriva genética e consanguinidade, que são as principais preocupações em espécies endémicas insulares, principalmente aquelas que passaram por um decréscimo populacional relativamente recente.

Assim, este trabalho tem como objetivo aferir certas características genéticas da freira da Bermuda para estudar as consequências da sua história demográfica na sua variabilidade genética. Este trabalho utiliza dados genéticos diversos: ddRAD-seq, uma tecnologia de sequenciação de 3ª geração capaz de caracterizar milhares de polimorfismos de nucleótido único (SNPs) do genoma nuclear neutral da espécie; fragmento do gene COI do genoma mitocondrial; e dois genes do sistema imunitário, pertencentes à família dos recetores do tipo Toll (TLRs), descritos como estando sob seleção em algumas espécies de aves nas quais foram caracterizados. Utilizando estes dados, os principais objetivos deste trabalho são: a) inferir o impacto relativo da deriva genética e da seleção em genes com impacto potencial na fitness individual; b) caracterizar a diversidade genética neutral da espécie; c) inferir os níveis de consanguinidade da população, assim como as suas consequências na escolha de parceiro (acasalamento preferencial) e no sucesso de eclosão do casal; d) inferir a história demográfica e confirmar a presença de um efeito de gargalo genético.

Dois genes do tipo Toll foram sequenciados com sucesso, no entanto o TLR5 mostrou resultados inconsistentes da clonagem, que podem ser devidos à duplicação do gene, comum nesta família de genes, ou com erros de sequenciação ou clonagem. Já o TLR4 revelou baixa diversidade na região que codifica o péptido de ligação que reconhecem o patogénio, a zona que se esperaria estar sob seleção balanceadora, sugerindo assim que a deriva genética poderá ter afetado este gene apesar da pressão seletiva.

Esta espécie de freiras nidifica em cinco ilhéus diferentes, próximos um dos outros se considerarmos a sua capacidade dispersiva. Os dados genómicos revelaram não haver estrutura populacional e como tal, pode ser considerada uma espécie panmítica. Este resultado era esperado, já que a população atual derivou de poucos casais e também porque a filopatria encontrada nesta espécie não é ilhéu-específica.

A diversidade genética neutral nuclear é já há muito considerada no campo da genómica populacional, pela sua importância na compreensão da história demográfica das populações, assim como na área da genética da conservação para a inferência do risco de extinção das espécies, ainda que a sua importância seja alvo de controvérsia. A diversidade genómica nuclear neutral de *Pterodroma cahow* é semelhante à de outras espécies do género *Pterodroma*. No entanto, o ADN mitocondrial indicou um resultado diferente, tendo a freira da Bermuda a diversidade mais baixa entre sete espécies de freiras, algumas das quais igualmente ameaçadas. Visto que o ADN mitocondrial é haploide e herdado por via materna, tem apenas ¼ do efetivo populacional em relação ao ADN nuclear. Por essa razão, reflete a história demográfica da população mais recente, enquanto o ADN nuclear reflete diversidade genética mais ancestral. Resultados díspares entre a diversidade dos genomas mitocondrial e nuclear podem assim indicar um decréscimo populacional recente.

A consanguinidade é um dos principais problemas de populações pequenas porque podem causar efeitos na *fitness* dos indivíduos, efeito esse a que usualmente se chama depressão de consanguinidade. Dadas as características demográficas da freira da Bermuda e a sua história demográfica, esta temática é altamente pertinente. O cálculo do nível de consanguinidade individual foi feito através de uma matriz de afinidade genómica (F_{GRM}). Os resultados revelaram que a média da consanguinidade da população é baixa. No entanto, alguns indivíduos revelaram valores preocupantes, equivalentes ao nível de consanguinidade de filhos de irmãos e de filhos de primos. A taxa de sucesso de eclosão dos ovos está descrita nesta espécie de freira como uma das mais baixas de entre os Procellariiformes. Visto que a fertilidade está diretamente relacionada com a *fitness* dos indivíduos, testou-se se a taxa de eclosão, medida de fertilidade, é influenciada pela consanguinidade. Para isto, usámos duas medidas em dois modelos lineares generalizados mistos (GLMM) diferentes. Primeiro, usámos a heterozigotia multi-*locus* individual (*MLH*), uma medida inversamente correlacionada com o F_{GRM}, que revelou não existir uma correlação heterozigotia-*fitness* (HFC), sugerindo que uma possível diminuição da heterozigotia devido à consanguinidade não aparece afetar a taxa de eclosão. Por outro lado, testou-se se o grau de parentesco entre os dois membros do casal influencia o sucesso de eclosão. O grau de parentesco entre indivíduos obteve-se através de um estimador genómico de parentesco, que se revelou fiável e consistente com os dados da genealogia existentes para esta população obtidos com os dados de campo. Os resultados sugerem que o grau de parentesco partilhado pelo casal não influencia a taxa de eclosão da sua descendência.

Como a maioria dos Procellariiformes, a Freira da Bermuda é socialmente monogâmica e a escolha do parceiro é, geralmente, para a vida. Uma situação de stress, como é a de depressão de consanguinidade, pode levar à mudança de comportamento reprodutor, como por exemplo a prevenção de reprodução entre indivíduos aparentados. Os dados genómicos usados neste trabalho revelaram, através de um *bootstrap* não paramétrico, que os indivíduos estão a escolher o seu parceiro ao acaso relativamente ao grau de parentesco. Este resultado pode demonstrar a falta de oportunidade de escolha, ou, por outro lado, a incapacidade dos indivíduos de reconhecerem os seus familiares. Ainda assim, a existência de paternidades fora do casal (EPPs) deve ser estudada nesta espécie, para melhor compreensão do impacto da consanguinidade no seu comportamento reprodutor.

A inferência demográfica foi realizada por dois métodos distintos, ambos baseados no espectro de frequência dos nucleótidos (SFS) dos dados genómicos, através dos programas informáticos Fastsimcoal e Stairwayplot. Este último sugere dois eventos demográficos mais evidentes. O mais antigo evidencia um decréscimo ligeiro e prolongado a começar há cerca de 100 mil anos, coincidente com os períodos glaciares e interglaciares do Pleistoceno, conhecidos por provocar decréscimos populacionais noutras espécies costeiras devido às oscilações do nível da água do mar. Além disso, esta análise demográfica sugere um decréscimo acentuado no efetivo populacional num período mais recente da

história da espécie, há 2 mil anos. Embora não coincida com a chegada dos humanos às Bermudas, pode ainda estar relacionado, visto que os dados baseados no SFS podem não ser suficientemente precisos para eventos demográficos recentes. Já a análise do Fastsimcoal revelou que o modelo com efeito de gargalo mais recente é o que mais se adequa aos dados, corroborando assim a hipótese da Freira da Bermuda ter sofrido um decréscimo populacional severo num passado evolutivo recente.

Em resumo, tanto as análises de inferência demográfica, como a diversidade genética nuclear e mitocondrial sugerem a ocorrência de um efeito de gargalo genético e consequente perda de diversidade genética. No entanto, o pequeno tamanho populacional não parece estar a causar depressão de consanguinidade relativamente à taxa de eclosão. Além disso, não foi encontrado um comportamento de prevenção de consanguinidade através da escolha de parceiro. Por outro lado, foi encontrada baixa diversidade num gene candidato a estar sob seleção, o que pode sugerir um efeito poderoso da deriva genética relativo à pressão seletiva a que estes genes podem estar sujeitos.

Palavras-chave: Freira da Bermuda, Espécies endémicas insulares, Depressão de consanguinidade, ddRAD, genes TLR

Abstract

The Bermuda petrel is an island endemic species belonging to the Procellariiformes, one of the most endangered orders of seabirds. Spending most of their lifetime in the open Atlantic Ocean, Bermuda petrel breed exclusively in Bermuda Islands, where females lay a single egg per season that demands high biparental investment. Historical records tell the tale of a significant decline of population size at the time of human settlement in Bermuda, throwing the species to near-extinction. Even so, the population was able to recover aided by an ongoing conservation plan. However, it still faces several threats, and genetic consequences resulted from that drastic decline are expected, being inbreeding and genetic drift the main concerns.

This study aimed to study neutral and functional genetic diversity levels and inbreeding levels, as well as its effects on individual fitness and on mating choice. Also, it aimed to look for a genetic signature of the recent bottleneck suggested by historical records. For this, we used a high-throughput sequencing technique, ddRAD-seq, to survey thousands of nuclear SNPs, one mitochondrial gene (COI), and two candidate genes to selection (TLRs).

The results obtained revealed that Bermuda petrel has suffered a recent genetic bottleneck and showed low mtDNA diversity when compared to other petrel species. Conversely, nuclear genomic diversity was not relatively lower than in other endangered petrels. Inbreeding levels were not high overall, although some individuals showed to be highly inbred. However, individual inbreeding and relatedness between couples is not affecting their hatching success. Additionally, individual's mating choice is not influenced by kinship. Finally, one of the two candidate genes to selection was dismissed since the results revealed unexpected signs of gene duplication, while the other revealed low genetic diversity in the pathogen biding coding region, hinting for overcome of genetic drift relative to selection.

Keywords: Bermuda petrel, Island Endemic Species, Inbreeding Depression, ddRAD, TLR genes

Table of Contents

List of Figures

List of Tables

Abbreviatures

AIC – Akaike information criterion

COI – Cytochrome Oxidase I

DBR – Degenerated base region

DP – Depth of coverage

EPPs – Extra-pair paternities

FUBAR – Fast unconstrained Bayesian approximation

GARD – Genetic Algorithm Recombination Detection

GLMM – Generalized linear mixed model

HFC – Heterozygosity- fitness correlation

HTS – Hight throughput sequencing

HWE – Hardy Weinberg equilibrium

IBD – Identical by descent

ID – Identity disequilibrium

LD – Linkage disequilibrium

LPS – Lipopolysaccharides

LRR – Leucine-rich repeat

MAF – Minor Allele Frequency

MEME – Mixed-effect model of evolution

MHC – Major Histocompatibility complex

MLH – Multi-locus heterozygosity

mtDNA – mitochondrial DNA

PAMPs – Pathogen associated molecular patterns

PCA – Principal component analysis

PCR – Polymerase chain reaction

RADseq – Restriction site associated DNA sequencing

ddRADseq – Double digestive restriction site associated DNA sequencing

ROH – Runs of Homozygosity

SFS – Site frequency spectrum

SNP – Single nucleotide polymorphism

TLR – Toll-like receptor

VCF – Variant Call format

Chapter 1 - **General Introduction**

1.1 –– *Pterodroma cahow*

The Bermuda petrel (*Pterodroma cahow*) is an endemic seabird species to the islands of Bermuda located in the western North Atlantic Ocean (Figure 1.1). It belongs to the Procellariiformes, which is one of the most endangered order of seabirds that face several threats (Dias et al., 2019). It is a Gadfly Petrel, a group of seabirds that live mainly confined to a single ocean basin. It is a medium-sized seabird and does not present sexual dimorphism, although males are usually larger and heavier. They are burrow nesters which makes them particularly vulnerable to habitat degradation and land invasive predators.

Figure 1.1 Spatial localization of Bermuda Islands including two expansions. The red areas on the second amplification are the current breeding areas of Bermuda petrel, located among five islands labelled by their initials. Map source: Esri

NS - Nonsuch; GI - Green Island; HR - Horn Rock; IP – Inner Pear; LR – Long Rock.

1.1.1 -Life cycle

Like all Gadfly Petrels, the Bermuda petrel, also known as Cahow, is pelagic, spending most of its life in the open ocean, rarely returning to land, except for breeding. They can fly effortlessly for long periods, soaring most of the time, due to their long and narrow wings (Figure 1.2), sometimes visiting areas of thousands of kilometres away from their breeding areas (J. L. Madeiros, 2005; Ventura et al., 2020). This species is highly philopatric, juveniles returning to the place where they were raised, when they reach about five years of age. Once they arrive to the Bermuda Islands, in mid-late October, they engage in courtship activity, where they perform long, unequal moans paced with extended yelps and shrieks during night-time, giving rise to the name Cahow and to the myth of Bermuda being an "Isle of Devils" or "Isla de Demonios" as once believed by the 15th and 16th century explorers (Gehrman, 2012). Lifespan of petrels is relatively high. It is believed that a Cahow can live up to 30 years and most individuals reach 20 years of age (J. L. Madeiros, 2005). Such lifespan affects its breeding ecology, similar to all other petrel species (Bried et al., 2003; Hamer, 2001)

Petrels have a long breeding season with significant parental investment of both the female and male (Hamer, 2001). Cahows need to balance their current energy budget and the requirements of future reproduction in order to maximize their reproductive output along their mature life. The female lays a single egg clutch per year, meaning only one chance to reproduce per year. Once fertilization occurs, the female departs again to the sea, where it engages in a massive feeding period in order to develop her single, large egg. The mate chosen is generally for life, as well as the nest burrow. Couples tend to break only in the case of the death of one of the mates, where the surviving adult finds another match. Incubation is shared by the couple and lasts about 50 days. Chick feeding is also shared by the pair. The chick stays in the nest for 3 months (Figure 1.3) and as soon it can fly, it leaves the nest and remains in the open North Atlantic Ocean until it matures, when ~ five years old (J. L. Madeiros, 2005).

Figure 1.2 - Two Bermuda petrels flying in open ocean. Source: Richard Crossley

Figure 1.3 - Bermuda petrel chick ~20 days after the hatching

1.1.2 -History and status

Sub-fossil evidence, as well as historical records suggest that Bermuda petrel was abundant, possibly with millions of nesting birds, by the beginning of the 16th century, the time when the islands were first discovered by Spanish explorers. A usual routine of the mariners at that time was leaving pigs in discovered islands. The first settlers arrived at Bermuda in 1609 and by that time, the population of Bermuda petrel was already decimated by the hogs, only remaining 10 to 50 percent of the birds (Gehrman, 2012). Adding to that, the settlement of the English brought other threats to the species survival, such as the introduction of exotic predator species, including rats, cats, and dogs. Deforestation and widespread burning also menaced Bermuda petrel's habitat and availability of nests. The disturbance was so severe that only 12 years after, in 1621, the Bermuda petrel had disappeared. It was thought extinct for three centuries. Ultimately, back in 1951, a bird was found dead after hitting a lighthouse and it was identified as a Bermuda petrel (Del Rey, 1964). Promptly, a group of conservationists planned a journey to the most remote islets of Bermuda. Astonishingly, around 14 pairs were found nesting in rock cavities, which made this a "Lazarus species" (Keith & Burgman, 2004). Years later, in 1962, an ambitious recovery programme started and lasts until the present day. The goals of the programme include the eradication of exotic predators and habitat ecological restoration, recreating the native forest of Bermuda in one of its islands, Nonsuch Island (Figure 1.1). It also includes building of artificial concrete burrows that provide extra nesting opportunities. Recently, a translocation plan started, where near-fledged chicks are taken from native islands and placed in new nests on Nonsuch island. The plan has already shown success with most grown chicks coming back to the island when mature (Carlile et al., 2012). In fact, breeding pairs increased from 18 in 1962 to 56 in 2000 and to 156 in 2021. Despite this population growth, the species is still considered Endangered by the IUCN

(International Union for Conservation of Nature and Natural Resources) and still faces several threats, such as introduced predators, increased intensity and frequency of hurricanes and sea level rise, due to Climate Change, that result in habitat loss.

The conservation plan of Bermuda petrel has led to the gathering of detailed data about the species ecology and breeding success along the years for most of the population, rarely available for wild species (J. Madeiros, 2018). Such data revealed a low hatching success rate (~67%), being one of the lowest among Gadfly petrels (Figure 1.4). This information and the historical drastic decline of the population leads to increasing concern for the species future.

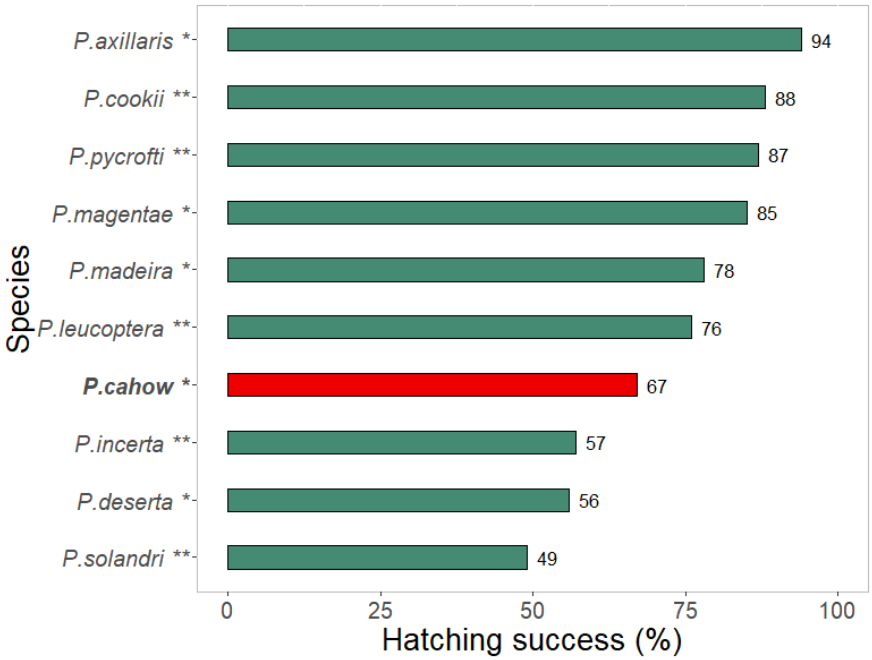


Figure 1.4 - Hatching success of ten Gadfly petrel species, all belonging to *Pterodroma* genus. Bermuda petrel is highlighted in red and bold. The asterisks after species name are relative to the species census size. Only one indicate a small census size (<1500), while two asterisks indicate larger census size (>1500).

Pterodroma axillaris (Gummer et al., 2015); *Pterodroma cookii* (Imber et al., 2003); *Pterodroma pycrofti* (Bartle, 1968); *Pterodroma magentae* (Imber et al., 2003); *Pterodroma madeira* (Menezes et al., 2010); *Pterodroma leucoptera* (O'Dwyer et al., 2006); *Pterodroma cahow* (present thesis); *Pterodroma.incerta* (Cuthbert, 2004); *Pterodroma deserta* (Menezes et al., 2010); *Pterodroma solandri* (Bester et al., 2007).

1.2 -Genetic diversity and demography

Genetic variation in a wild population can vary over generations by the influence of four combined evolutionary forces: mutation, gene flow, selection, and genetic drift. Such forces are responsible for the fluctuations of allele frequencies in a population over time. Changes in allele frequency between generations occur randomly (Hamilton, 2009). The alleles that a certain generation passes to their offspring are always a subsample of the alleles available in that parental generation, which causes a stochastic process designated by genetic drift. The impact of this process is higher in smaller populations with a reduced genetic pool, generating higher fluctuations in allele frequencies, which increases the likelihood of fixation (frequency of 1) or loss of a particular allele. Thus, genetic drift reduces genetic diversity (, 2009b). Single island endemic species do not undergo migration, thus do not benefit from the income of diversity generated by the admixture of two different genetic pools. Finally, natural selection is the process through which organism adapt differently to a certain environment due to genetic variation. Individuals with well adapted genotypes will be more fit, thus producing higher number of offspring, which will increase the beneficial alleles in the population over time. Selection can be positive, increasing the frequency of the beneficial allele; purifying, removing the maladapted allele; and balancing, which maintains the different alleles at intermediate frequencies. Both positive and

purifying selection decrease genetic diversity, while balancing selection can maintain or increase the variability of the population's genetic pool (Hamilton, 2009).

To understand how neutral genetic diversity is originated and maintained in natural populations has always been an important goal in the field of population genetics. Neutral regions of the genome are not related to fitness, thus are not under selective pressures (Kimura, 1983). The interaction between genetic drift and mutation is responsible for the neutral genetic diversity segregating in an island population. Hence, neutral genetic diversity is proportional to the mutation rate (μ) and the population size (Teixeira & Huber, 2021). The relation between an historical period of intense genetic drift, caused by a reduced effective population size (N_e) and low neutral genetic diversity is well established (García-Dorado & Caballero, 2021). However, its association with reduced potential of the population's survival and viability can be controversial. Reduced neutral genetic diversity has often been associated with maladapted individuals showing poor health which reduce the viability of the population in a certain environment (Evans & Sheldon, 2008; Újvári et al., 2002). However, Teixeira & Huber, (2021) doubt such causal relation, arguing about its weak empirical evidence, pointing out that IUCN red list, an international list of endangered species, is as a poor predictor of neutral genetic diversity of the listed species, and even questioning its relevance in conservation genetics.

Severe demographic events on a population's history have a very important effect on its present genetic diversity. Many seabird species share a history of accentuated decline in some point of their history, that was either caused by an extreme climatic event (e.g. Pleistocene climatic oscillations) (Silva et al., 2016) or by anthropogenic intrusion (e.g. habitat destruction, introduced predators) (Paleczny et al., 2015). In these cases, a considerable amount of genetic diversity is lost directly by the decrease in population size. Additionally, genetic drift has a higher impact on smaller populations leading to further loss of genetic diversity. Genetic bottlenecks can have long-lasting effects and destroy part of the population's adaptative potential that may never be recovered. Hence, learning about the presence and extent of such events is very important to understand the evolutionary consequences they may have, as well as to comprehend the conservation status of the population, which will help its conservation management (Bortoluzzi et al., 2020; Sanderson et al., 2018).

Whilst consequences of a genetic bottleneck are well-described, research on changes in genetic diversity though the recovery process are sparse (Cammen et al., 2018; Keller et al., 2001). Demographic events such as immigration or an improved birth rate can increase population size and eventually genetic diversity, although its rate of increase remains uncertain. This uncertainty has recently expanded the interest on this subject, combined with the growing number of success cases that have arisen after decades of conservation efforts (Cammen et al., 2018). Island species are thought to be at disadvantage regarding the ability to recover from bottlenecks. Geographical distance and lack of potential for dispersal make migration very rare or impossible (Helmus & Behm, 2020). Some species might have the means to disperse, like seabirds who are able to fly for thousands of kilometres (Ventura et al., 2020), but philopatry influence them to breed in the same local. Adding to that, several island endemics comprehend a single population. All these factors contribute to make single island endemics among the most endangered in the wild and sometimes extinction is a likely scenario (Wood et al., 2017). Nonetheless, some success cases show the importance of well implemented conservation strategies, also revealing an impressive endurance of apparent fragile populations (Jones & Kress, 2012).

1.3 -Causes and consequences of inbreeding depression

Following a bottleneck, mating choice gets stricter due to the decrease of the number of individuals available for mating. The possibility of inbreeding is a known consequence of a severe

reduction in population size and has consequences of its own. The term inbreeding applies when there is mating between relatives with at least one common ancestor. The resulted progeny shows a genome-wide increase in homozygosity, which is a consequence of most of alleles being identical-by-descent (IBD). When two mating individuals are closely related, the probability that their offspring will have two identical alleles due to the transmission of a common ancestor allele from both parents are much higher than with random mating in an outbred population. These offspring often reveal reduced fitness. The main impact of inbreeding is the reduction of mean phenotypic value, caused by the changes in genotype frequencies, as well as reduced fitness effect associated with the same changes. Detrimental fitness effects might be associated with reduced fertility, and the ability of competition with adjoining species or increased vulnerability to diseases and environmental stress (Charlesworth & Charlesworth, 1999; Hamilton, 2009).

Reduction of the average phenotype value and fitness in a population due to inbreeding is often referred to as inbreeding depression (Hamilton, 2009). This is a widely studied subject since the work of Charles Darwin, considered very important and concerning many species. Two classical hypotheses for explaining the positive correlation between the inbreeding coefficient and fitness, and thus inbreeding depression, are described, both involving dominance effects in a locus (Charlesworth & Charlesworth, 1999). The first is called the dominance hypothesis and is due to an increase in homozygosity which will rise the expression of recessive alleles that are deleterious and that are not expressed in heterozygosity of that locus. The other hypothesis, often called the overdominance hypothesis, suggests that inbreeding depression results from the decrease in heterozygotes. It assumes heterosis, when heterozygotes show higher fitness than each homozygote. Both hypotheses are not exclusive and can be affecting the population at the same time (Charlesworth & Charlesworth, 1999; Hamilton, 2009). A potentially useful application to explore with genomics of highly endangered species is to test whether there is a sign of inbreeding, if there is variance in inbreeding coefficients within populations, and whether variation in the heterozygosity of individuals is correlated with fitness traits. If a significant correlation is found, it may be interpreted as a sign of inbreeding depression. Inbreeding depression has been suggested to lead to heterozygosity-fitness correlations (HFC) when there are significant correlations between fitness-related traits and individual heterozygosity. HFCs can be caused by a single locus (‘local effect’) or by genome-wide heterozygosity (‘general effect’) (Velando et al., 2015). Variance in heterozygosity across the genome, also known as identity disequilibrium (ID) occurs when individual heterozygosity varies more between individuals than expected, which is due to variance in inbreeding coefficients among individuals. The presence of ID can indicate inbreeding depression and can be used to infer if that is the cause of ‘general effect’ HFCs.

The rise in homozygosity of deleterious alleles can also increase the selection against them, and may be an advantageous method for purging a population’s genetic load and help the recovery of fitness (Boakes et al., 2007; Hedrick & Garcia-Dorado, 2016). Empirical studies in captive populations showed that purging indeed increases fitness and it may be beneficial to subject species to deliberate inbreeding. Recent studies have also suggested that purging has occurred in wild populations, providing an advantage for these species (Grossen et al., 2020; Robinson et al., 2018). However, it is not always straightforward to detect purging both in wild and captive populations due to concurring adaptative processes and lack of experimental power, respectively. It is also described that purging is most effective when inbreeding increases slowly or in low magnitude across generations. This may also increase the failure to detect it, although it may not imply that is not happening or that it is not important for the species recovery (Boakes et al., 2007).

The mating system of a population can help mitigate the effects of inbreeding depression. Assortative mating is defined as non-random mating in which the resulting genotype frequencies differ

from those in a random mating population. Negative assortative mating, also called disassortative mating occurs when individuals tend to choose their mates based on unlike genotype or phenotype (Hamilton, 2009), (e.g. Dargent et al., 2019; Maisonneuve et al., 2020). The most well-known example is that mammals occasionally mate disassortatively regarding the major histocompatibility complex (MHC) (Penn & Potts, 1998; Wedekind & Furi, 1997). This has been seen as a strategy to fight inbreeding depression. On the other hand, positive assortative mating may occur when individuals choose like-mates, which has been described to occur in some populations as a way to avoid outbreeding depression (Kokko & Ots, 2006). In some cases, inbreeding can be advantageous, such as contributing to the maintenance of locally adapted genes (Pertoldi et al., 2007) or the increase of fitness of an individual in case it contributes to the reproductive success of kin. Such phenomena have often been called "inbreeding paradox" (Kokko & Ots, 2006).

1.4 - Drift and selection in immune-system genes

Evolution of immune system genes has been of particular interest in the field of evolutionary biology due to its complexity and direct role in species survival (Velová et al., 2018). There is usually high, intraspecific polymorphism in these genes due to balancing selection mediated by environment specific pathogens (Grueber et al., 2014). However, drastic demographic events can also have an impact on variation and evolution of immune genes. A severe genetic bottleneck can shape not only neutral but also functional genomic regions (Gilroy, Phillips, et al., 2017; Grueber et al., 2013). In this case, genetic drift can outweigh the effects of diversifying selection depending on population's response after the bottleneck, leading to the loss of genetic diversity. Still, detection of a drift-selection equilibrium in recent bottleneck populations can be challenging since signs of balancing selection can persist for many generations after a bottleneck, not reflecting necessarily current selection pressures (Gilroy, Phillips, et al., 2017).

Research on variation at functional genes, combined with neutral genomic regions assessments, can bring insightful information on the effects of demographic events, such as a population bottleneck, as well as its impact on the adaptative ability of endangered species. Thus, such knowledge can be of interest in both evolutionary biology and conservation genetics (Grueber et al., 2014; Nandakumar & Ishtiaq, 2020).

1.5 - Aims

In order to assess the impact of the recent demographic bottleneck on the population genetics of Bermuda petrel, the present study focuses on the following goals:

1) Evaluate the genetic diversity of two immune-system genes and understand the dynamic between genetic drift and selection.

2) Evaluate nuclear genome-wide neutral diversity and single-gene mitochondrial diversity and compare it to sister taxa.

3) Estimate inbreeding levels and its influence on a single fitness trait, as well as on mating behaviour.

4) Confirm a recent genetic bottleneck based on nuclear genome-wide data.

Chapter 2 - **Toll-like receptor genes**

2.1 -Introduction

An important goal of conservation genomics is to determine if the viability of small, endangered populations is compromised by the loss of adaptive variation due to genetic drift. Over the years, an intense focus has been given to the study of the vertebrate's immune systems, namely to the interaction of the forces governing the evolution of functionally relevant genes, such as those of the Major Histocompatibility Complex (MHC) (Gilroy, Phillips, et al., 2017). However, the complexity of these gene families, with many recently duplicated genes, represents a challenge in their study in non-model organisms, which has led to the recognition of the importance of other less complex gene families involved in the immune response such as the Toll-like receptor (TLR) genes. These have been, lately frequently used for functional variation and selection studies (Gilroy, Phillips, et al., 2017; Nandakumar & Ishtiaq, 2020; Podlaszczuk et al., 2020).

TLRs is a family of genes that encode transmembrane sensors capable of recognizing and binding pathogen-associated molecular patterns (PAMPs), present in bacteria, fungi, viruses, etc. Such recognition occurs in the extracellular domain of the receptor, which triggers a cascade of intracellular immune responses (Alcaide & Edwards, 2011). Studies using protein crystallography have revealed that variation in the binding region of these receptors can compromise pathogen recognition and lead to variation in the immune response between individuals (Levy et al., 2020). Recent studies have shown that TLR genes have evolved by positive or balancing selection in several avian species. Velová et al., (2018) reported signatures of positive selection in the majority of the avian TLR genes across more than 60 species, with most sites under selection encoding the extracellular domain of the receptors. Also, balancing selection has been suggested to explain the maintenance of variation in TLR genes despite lower levels of neutral genetic variation in bottlenecked avian species (Gilroy, Phillips, et al., 2017; Grueber et al., 2012). This family of genes has been well characterized in birds and it comprises ten genes, four of which have orthologs in other vertebrate groups (Alcaide & Edwards, 2011). All TLR genes are candidates for being under selection and for helping understand the effect of demographic events in functional variation of populations, namely how balancing selection and drift interact to maintain adaptive variation in small populations. Also, these genes provide additional insights to the study of endangered species, due to their role in species fitness and adaptability in the face of the emergence of new pathogens (Gilroy, Phillips, et al., 2017).

For this study, it was our aim to amplify and sequence the complete coding region of the genes TLR4 (three exons) and TLR5 (single exon) in the Bermuda petrel population. The two genes present different ligand specificities. TLR4 allows recognition of lipopolysaccharides (LPS) present in gram-negative bacteria's cell wall, while TLR5 detects flagellins, proteins present in flagellated bacteria (Alcaide & Edwards, 2011). Both have been shown to be under positive selection in several avian species (Velová et al., 2018) and recently in the Gentoo penguin (Levy et al., 2020), a seabird species in the sister group to the Procellariiformes. Both genes comprise a leucine-rich repeat (LRR) region, which is described as the most polymorphic region compared to the remaining regions, more conserved (Levy et al., 2020). The goal of surveying these TLR genes is to contrast levels of adaptive variation with those from neutral loci and better comprehend the effects of a demographic bottleneck on the maintenance of functional genomic variation and its impact on fitness. A deeper understanding of the role of both drift and selection on these genes can be accomplished as a tool to better comprehend the impacts of the demographic history and future sustainability of the endangered Bermuda petrel.

2.2 -Material and Methods

2.2.1 -Sample collection

The population of Bermuda petrel was sampled in 5 of the 6 islands where they breed in Castle Harbour, Bermuda, in 2019 and 2020. Overall, blood samples were collected for 104 individuals and kept at 4ºC in the field and frozen at -20º C upon arrival to the lab. All the sampling process was conducted previously to the start of this master dissertation.

2.2.2 -DNA Extraction

Genomic DNA of 104 samples from Bermuda Petrel individuals was extracted from blood samples using the QIAGEN DNeasy Blood and Tissue Kit. The quality of the extracted DNA was verified though electrophoresis on a 1% agarose gel, to confirm the presence of high-molecular weight DNA. DNA concentration was measured using both NanoDrop and Qubit technologies (Thermofisher Scientific). All DNA samples were stored at -20ºC.

2.2.3 -Toll-like receptors (TLR) Genotyping

Analysis of TLR genes, which include amplification and polymorphism and selection analysis of the DNA sequences, comprised randomly chosen 29 individuals out of the 104 sampled for Bermuda petrel species.

2.2.3.1- Primer optimization and development

The amplification of TLR genes in the Bermuda petrel comprised the use of primers already designed for close species and the design of Cahow-specific primers. For TLR4 we were able to use primers from Levy et al., (2020) to amplify exons 1 and 2, using a single primer pair (Figure 2.1, Table S 1). Regarding the third exon of TLR4, Levy et al., (2020) primer pairs did not work. Therefore, we used a set of primers for all known avian TLRs, designed by Alcaide & Edwards, (2011). One of the primer pairs that target an internal segment of the third exon of TLR4 (avTLR4) was tested in *P. cahow* and amplified successfully. The amplified segment was used as a reference to design species-specific internal primers (cT4-3EF and cT4-3SR) to obtain the remaining regions of the gene. Combined with Levy et al., (2020) primers, located in the exon's flanking regions (Figure 2.1), it was possible to amplify all three exons. Regarding TLR5, primers from Levy et al., (2020) targeting fragments from both ends of the gene's single exon (A and C) worked well for our study species (Figure 2.1). Internal primers (cT5BF and cT5BR) were designed based on *P. cahow* sequences of A and C fragments to amplify the remaining coding region between those fragments. Primer design for both genes was performed using Primer3Plus v2.4.2 (Untergasser et al., 2012). For each desired region, two primers were selected considering the location relative to the region of interest, GC content (%) and melting temperature (Tm). Primer sequences are fully detailed in Table S 1.

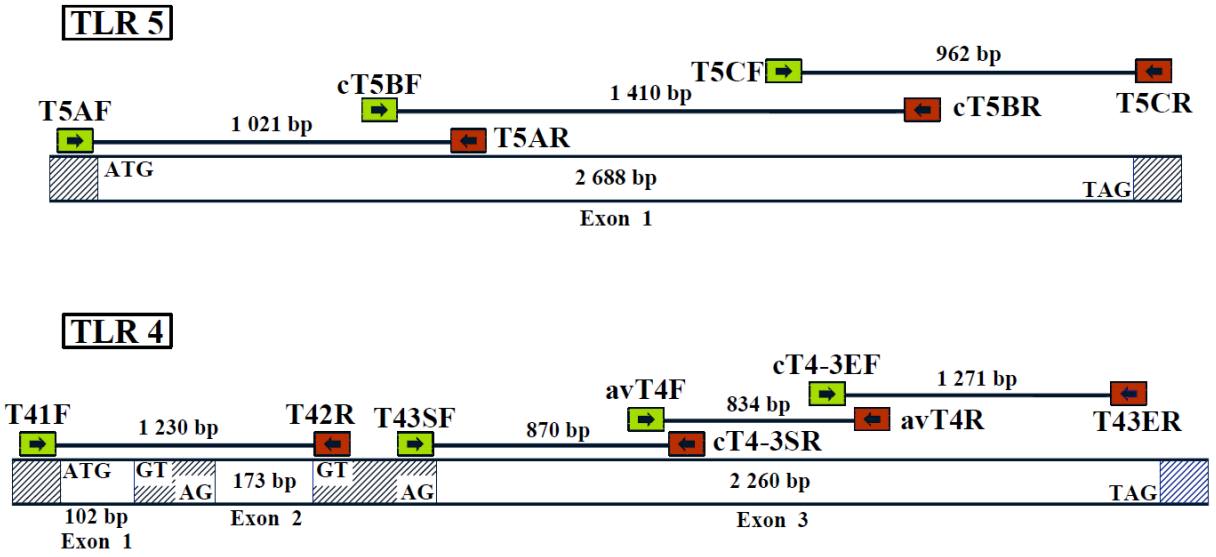


Figure 2.1 – Amplification scheme of TLR5 and TLR4. Coloured boxes represent the forward and reverse primers. PCR fragments and exons sizes are described in base pairs (bp), although some might not be proportional.

2.2.3.2- PCR and sequencing

Polymerase chain reactions (PCR) were carried out in a final volume of 15 µl and contained 2 µl of dNTPs (266 µM), 1.5 µl of reaction buffer (1 X), 8.05 µl of nuclease free water, 0.25 µl of Taq polymerase (0.083 U/µl), 0.3 µl of each primer (0.2 µM), 0.6 µl of $MgCl_2$ (2mM) and 2 µl of the DNA template. PCRs were performed in BioRad Thermocycler under the following conditions: first denaturation step of 3 min at 94ºC; 35 cycles, each one comprising a 94ºC step for 40s, primer-specific annealing temperature for 40s and an extension step at 72ºC for 80s; and a final extension step of 72ºC during 10 min. PCR products were visualized in 1% agarose gels stained with SYBR Safe, using a Ladder to verify the expected size of the DNA fragment. The successfully amplified products were purified using an ExoSap purification protocol. Purified PCR products were sequenced in both directions in outsourcing (Macrogen-Inc).

Raw sequences were edited and analysed in Sequencher v. 5.4.6 (Gene Codes Corporation) for the coding region of both genes. SNPs and heterozygote sites were visually identified, applying the IUPAC (International Union of Pure and Applied Chemistry) code. Alignment was performed by ClustalW Multiple alignment tool implemented in Bioedit 7.2.6.1 (Hall, 1999).

2.2.3.3- Genetic diversity

Allelic phasing in both TLR4 and TLR5 genes for those individuals with more than two SNPs was conducted using the command line version of the Bayesian PHASE algorithm (Stephens et al., 2001) with 1 000 burn-in iterations, a thinning interval of 10 and 10 000 iterations. All fragments that showed poor resolved phased alleles (<0.7) were cloned in outsourcing (NZYtech). Standard polymorphism statistics were estimated in DnaSP v.6 (Rozas et al., 2003) including the number of segregating sites (S), number of haplotypes (H), haplotype and nucleotide diversity (*Hd,* π) and number of synonymous and non-synonymous sites (Syn, Non Syn).

2.2.3.4- Selection analysis

Inference of selection comprised two different approaches. The first approach was based on allele frequency spectrum of the population and was computed in DnaSP v.6 (Rozas et al., 2003): Tajima's *D*. This statistic tests for significant deviances from a neutral expectation. Tajima's *D* is

calculated as the difference between the mean number of pairwise differences and the number of segregating sites between sequences (Hamilton, 2009).

The other approach in based on the rates of synonymous and non-synonymous substitutions (*dS*, *dN*) and how they vary along a phylogenetic tree. To assess the phylogenetic relationships of TLR genes between Bermuda petrel and other avian species, the 9 closest taxa with sequences on GenBank were extracted. Additionally, ten individuals of Zino's Petrel *P.madeira*, a closely related gadfly petrel, were screened for both genes following the same method used for *P.cahow* . Finally, *Gallus gallus* and *Calidris pugnax* TLR sequences were also extracted and used as outgroups. For species with more than one sequence, a consensus was used for both the alignment and phylogenetic analysis. Bioedit (Hall, 1999) was used for alignment and a maximum-likelihood tree was constructed using MEGA 10.1.1 (Kumar et al., 2016) with 500 bootstrap replicates and under GTR+G (Generalized time reversible with gamma distribution) substitution model as proposed by jModelTest 2.1.3 (Darriba et al., 2012) as being the most appropriate. Two codon-based selection models were run using the command line version of the HyPhy v. 2.5 package available on Datamonkey 2.0 (http://www.datamonkey .org/; Weaver et al., 2018) : Mixed-effect model of evolution (MEME), used for detection of episodical selection under a proportion of branches of the phylogeny; and fast unconstrained Bayesian approximation (FUBAR), a Bayesian approach to detect pervasive selection pressure along the full phylogeny. Posterior probability was set to 0.9 and significance value to 0.1. Prior to these tests, the Genetic Algorithm Recombination Detection (GARD) analysis was performed in Datamonkey 2.0 (http://www.datamonkey .org/; Weaver et al., 2018) to look for possible recombination breakpoints in TLR4. This is important since the selection tests chosen consider recombination effects (which can generate false positives) for predicting positively selected sites (Murrell et al., 2012, Murrel et al., 2013).

2.3 –– Results

The results of TLR5 analysis are shown separately followed by TLR4 results of polymorphism and detection of selection assessments.

2.3.1 -TLR5

Cloning results suggested that TLR5 gene may be duplicated in *P. cahow* or that an error occurred during the cloning process. At least four clones were sequenced per PCR product in 2 individuals. All polymorphic sites found with Sanger sequencing were recorded in the different clones. However, for two of the fragments that comprise the exon, more than two alleles were found in the four clones sequenced for each individual. Additionally, Sanger sequencing revealed that all sampled individuals showed two adjacent stop codons in the end of the gene. This may suggest that in case of duplication, one copy can be one-codon longer than the other. Thus, no further analyses were done for TLR5.

2.3.2 -TLR4 polymorphism

We successfully obtained 2408 bp of TLR4 for 29 individuals, including the three exons, although we were not able to genotype the complete coding region, lacking 124 bp from the end of the third exon. No frameshift mutations or premature stop codons were found. Overall, the Cahow's TLR4 showed 10 variable sites, of which 3 were non-synonymous changes forming 11 haplotypes, which resulted in high haplotype diversity (*Hd* = 0.826) (Table 2.1). The Leucine-rich repeat (LRR) is 1842bp long in TLR4, corresponding to ~80% of the entire coding region of TLR4 and including sections of

both exons 2 and 3. Polymorphism of this region in *P. cahow* did not differ from the entire coding region sequenced, implying that the polymorphism found is widely spread through the gene (Table 2.1).

Table 2.1 Polymorphism measures characterizing TLR4 entire coding region and only LRR domain in Bermuda petrel

Locus	**Fragment size (bp)**	**N**	**S**	**Syn**	**Non Syn**	**H**	***Hd***	***π***
TLR4	2408	29	10	7	3	11	0.826	0.0014
TLR4 - LRR	1842	29	8	6	2	9	0.813	0.0015

Abbreviations: N, number of individuals sampled; S, number of segregating sites; Syn, number of synonymous changes; Non Syn, number of non-synonymous changes; H, number of haplotypes; *Hd*, haplotype diversity; *π*, nucleotide diversity; LRR, leucine rich repeat.

2.3.3 - Tests of selection of TLR4

Tajima's D was positive for TLR4, although not significantly different from 0 ($D = 1.56$, $p > 0.1$).

A phylogenetic tree was constructed for 12 species (Figure S 1), on which the codon-based selection methods were based. The GARD analysis detected 6 recombination breakpoints along the TLR4 alignment, and a partitioned dataset was used as input for the MEME and FUBAR analyses. MEME was able to detect 12 episodical codon sites under diversifying selection in specific branches of the phylogeny. All the sites except for two (53, 121, 187, 302, 380, 398, 405, 424, 522, 533) were located in the leucine-rich repeat (LRR) region (extracellular domain), responsible for pathogen recognition. The remaining two selected codons were in positions 6 and 747 and were in the signal peptide, located in the N-terminus of the TLR4 protein and in the Toll/Interleukin-1R (TIR) (intercellular domain), respectively. MEME results suggest that five codon sites (398, 405, 522, 747) were episodically positively selected in a number of branches preceding divergence of *P. cahow*. The FUBAR model identified 5 positively selected sites along the entire phylogeny. All the sites detected by FUBAR were also detected in MEME and all belong to the LRR domain of TLR4 (Table 2.2). Ten of the codon sites found in these analyses has been previously found to be positively selected in birds (Table 2.2) (Velová et al., 2018).

Table 2.2 Codons of TLR4 that showed evidence of selection among twelve species of avian species, using two different selection detection methods: MEME and FUBAR.

Locus	**MEME**	**FUBAR**
TLR4	6*, 53, 121, **187***, **302***, 380*, **398***, **405***, 424, **522***, 533, 747	**187***, **302***, **398***, **405***, **522***

Codons in bold were detected by both methods. Sites with asterisk means the site was already described to be under positive selection pressure in Velová et al., (2018). Sites underlined for MEME methods are the ones detected from a number of branches preceding *P. cahow* divergence. Codon number correspond to *G. gallus* coding sequence.

2.4 -Discussion

Multigene families evolution include gene duplications over time that form several paralogous genes (Hamilton, 2009). New mutations in duplicated genes allow these paralogous to evolve different purposes and thus provide higher diversity in their functions, which is beneficial for populations. Immune system genes are commonly affected by duplication due to the impact in species survival, since the different gene copies allow the recognition of a larger array of pathogens. Still, in the duplication process there might be mutations or insertions/deletions (indels) that result in loss of functionality of the genes, originating pseudogenes. This pseudogenization can occur due to premature stop codons or frameshift mutations, and it usually goes hand in hand with duplication events (Velová et al., 2018). Gene duplications, as well as pseudogenization events are widely described in TLR genes of mammals and birds (Alcaide & Edwards, 2011; Velová et al., 2018). Pseudogenization of TLR5 is described for several lineages in the bird phylogeny, such as Passeriformes, and purifying selection towards TLR5's loss of function in certain lineages has been suggested (Velová et al., 2018). Nevertheless, gene duplication of TLR5 has not been described in avian lineages, despite its presence in fishes (Khan et al., 2019). Recent studies in seabird species have not reported gene duplication or pseudogenization in TLR5 (Levy et al., 2020; Podlaszczuk et al., 2020) and several studies have described positive selection on this gene (Grueber et al., 2014; Khan et al., 2019; Levy et al., 2020; Nandakumar & Ishtiaq, 2020, Velová et al., 2018). Despite not being deeply studied in Procellariiformes (Alcaide & Edwards, 2011), TLR5 was considered a good candidate to study diversity and selection in functional DNA in Bermuda petrel. Hints of duplication revealed by the cloning results for this species were not surprising due to the record of duplications in other TLR genes. However, this result was unexpected since the duplication of TLR5 has never been described for any species of bird. Our results suggest that Sanger sequencing is not sufficient to detect duplicated genes because the patterns of polymorphism can be interpreted as high polymorphism of a single gene. Thus, suspicions rise about putative duplications of TLR5 on other avian lineages only analysed by Sanger sequencing (Grueber et al., 2014; Nandakumar & Ishtiaq, 2020; Zapata et al., 2020). Errors during PCR reactions and cloning are also a possible explanation for the pattern found in the cloning results: more than two alleles for a single individual. During the PCRs that precede the cloning process, disruption of phase of heterozygous polymorphisms can occur due to misleading recombination, although this is more likely for high polymorphic fragments, which will create chimeric alleles (Meyerhans et al., 1990). Both duplication and cloning errors are plausible explanations for the cloning results of TLR5 found in the Bermuda petrel, however, information for only two individuals cannot fully confirm either one of them. Further analyses are needed, such as extended sequencing of further flanking regions of the exon and cloning for higher number of individuals.

Genetic diversity measures assessed for Bermuda petrel's TLR4 gene revealed 10 polymorphic sites defining 11 haplotypes among 29 individuals. This level of TLR4 variation is intermediate relatively to that reported for species that share a similar demography to the Bermuda petrel. Recent studies of recently bottlenecked bird species have shown very low or no variation in TLR4, which was suggested to be the effect of drift caused by small population size in those species (Gilroy, van Oosterhout, et al., 2017; Grueber et al., 2012; Podlaszczuk et al., 2020; Zapata et al., 2020). However, other bird bottlenecked species revealed high levels of TLR4 variation, which the authors suggest might be due to the long-lasting effects of ancient balancing selection, not weakened by genetic drift (Grueber et al., 2015; Nandakumar & Ishtiaq, 2020). Bermuda petrel shows intermediate levels of TLR4 diversity compared to those species, which might be due to a weaker effect of drift, stronger effect of diversifying selection or both. Comparisons with other species are often done but should be considered with caution. Several factors can influence variation in functional genes, such as the severity and duration of the

bottleneck, as well as different selective pressures on TLR4, since pathogen pools differ according to habitat characteristics.

The Leucine-rich repeat (LRR) is an extracellular domain of TLR4 responsible for recognizing and binding to pathogen's derived peptides and has been described as the most variable region of the gene. The sites encoding this region have been shown to be under balancing selection which maintains variation that is important for pathogen recognition (Grueber et al., 2014). Thus, genetic variation was assessed from this region alone in Bermuda petrel and only small differences were noted when compared to the whole coding region (see 2.3.2 -TLR4 polymorphism). This pattern suggests that the LRR region may not be under balancing selection in Bermuda petrels, or that TLR4 variation is not adaptive in this species. Adding to this suggestion, neutral polymorphism as obtained by the RAD loci sequencing in Bermuda petrel revealed relatively higher RAD nucleotide diversity (π_{RAD}= 0.0020, see 3.3.2 -Genetic diversity and population structure) than that found in TLR4 (π_{TLR4}=0.0014, Table 2.1), which would not be expected if the gene was under selection.

Codon-based selection analyses on TLR4 inferred the presence of positive selection in particular codon sites along the TLR4 avian phylogeny, but no selection was detected along the branch leading to *P. cahow*, since it diverged from the common ancestor with *P. madeira* (see 2.3.3 -Tests of selection of TLR4). Additionally, we found 45 codon sites with evidence of purifying selection in the TLR4 alignment, which is expected in genes that highly impact fitness, in which some codon positions are conserved due to functional constrains in the protein (Nandakumar & Ishtiaq, 2020). Purifying selection can be the main force acting in most of the open reading frame of the gene, although positive selection acting on specific sites may be important to maintain the receptor variability in the population. Tajima's *D* test was applied to TLR4 in Bermuda petrel and was found to be not significantly different from zero, suggesting that null hypothesis of the gene evolving under neutrality cannot be rejected (Hamilton, 2009).

Overall, neither genetic variation nor the selection analysis show evidence of balancing selection acting on the TLR4 gene in the Bermuda petrel. Although this gene has been shown to be under balancing selection on other avian species, namely in penguins, the sister clade of the Procellariiformes, the lack of evidence in *P. cahow* can be due to low variation in pathogen species present in their currently restricted breeding region that should have led to population's adaptation to the local environment. Also, the relatively recent genetic bottleneck, as suggested in historical records for Bermuda petrel, which is also suggested by the demographic analysis using ddRAD seq data in the present thesis (see 3.3.6 -Demographic history) may have left the population more exposed to the effects of drift. Consequently, drift may be outweighing the effect of balancing selection, leading to a lack of polymorphism in this gene. Finally, absence of strong pre-bottleneck balancing selection should also be considered as a different explanatory hypothesis, since it has been shown that a strong drift pressure is needed to erase signatures of ancestral balancing selection in current genomes (Gilroy, Phillips, et al., 2017).

Results found in this chapter offer important insights regarding the evolution of two TLR genes reported for the first time in an endangered procellariiform species. Besides, this study exposed a putative TLR4 duplication, for the first time in birds. However, this research presents several limitations and further studies are needed to better understand the role of TLRs variation in the fitness of Bermuda petrels, as well as the overall impact of the historical genetic bottleneck in adaptive diversity in this species. For this, more TLR genes should be studied, complemented with functional analyses, including expression patterns. Complementary surveys of neutral nuclear variation, based on other than RAD seq loci might be valuable to better estimate relative genetic diversity in TLR genes. Finally, pathogen

prevalence in Bermuda petrel habitat, both at sea and in land, should be surveyed to understand how much, if any, selective pressure must be upon the population.

Chapter 3 - **Population Genomics**

3.1 - Introduction

High throughput sequencing (HTS) technology has had a very large impact in the field of evolutionary biology. Sequencing of thousands of markers across the genome increased the power of parameter estimation which improved the degree of reliability to study evolutionary processes in wild populations (Stapley et al., 2010). HTS includes several different techniques from reduced-representation methods to whole-genome sequencing. Nowadays, reduced-representation methods are relatively inexpensive and easily provide thousands of markers for hundreds of individuals. Restriction site associated DNA sequencing (RADseq) relies on restriction enzymes that cut the DNA into fragments, following the ligation of adaptors that serve as barcodes that allow individual identification. A polymerase chain reaction (PCR) and Ilumina sequencing results in thousands of reads for several individuals in the same run (Davey & Blaxter, 2010). Such reads are further analysed by bioinformatics tools for aligning and demultiplexing, which allows for the detection of a pool of single nucleotide polymorphisms (SNPs) in the sampled individuals. RADseq has a derivative technique called double-digestive restriction site associated DNA sequencing (ddRADseq), which differs from the original by using two different restriction enzymes. This technique enables the extraction of similar location and size reads for the different individuals increasing DNA sequence coverage obtained among individuals, although it reduces the number of SNPs obtained when compared with the original RADseq technique (Cumer et al., 2018).

During the last few years, several studies have compared the performance of genomic markers resulting from reduced-representation methods with "classic" genetic markers such as microsatellites. SNPs are biallelic, thus are less informative than microsatellites, which are highly variable among different individuals. However, the number of SNPs obtained outweighs its low variation disadvantage (Attard et al., 2018). In fact, HTS has showed a great improvement compared to microsatellites, concerning, for example, the accuracy of relatedness estimation (Lemopoulos et al., 2019; Thrasher et al., 2018), parentage analysis (Flanagan & Jones, 2019), or inbreeding coefficient estimation and estimates of genome-wide heterozygosity (Hoffman et al., 2014; Huisman et al., 2016). This improved power for parameter estimation is very relevant to the study case of Bermuda petrel since its known history and endangered status brings concern on inbreeding levels in the population and its putative impacts on fitness and consequently on species' potential sustainability.

The main aim of this chapter is to analyse ddRAD seq data from the Bermuda petrel population to: a) estimate genome-wide neutral variation and compare it to other endangered taxa; b) estimate inbreeding by measuring the mean multi-locus heterozygosity (*MLH*) of the sampled population; c) infer the presence of inbreeding depression by detecting significant heterozygosity-fitness correlations (HFCs), using the hatching success as a fitness trait; d) estimate pairwise relatedness with genomic data and compare it to known pedigree relations from the field data, as well as to test its influence on the couple's hatching success; e) test the hypothesis of random mate choice by comparing differences between the real couples relatedness distributions and those simulated under random mating; f) infer the demography history of Bermuda petrel using two coalescence based-methods based on the site frequency spectrum (SFS).

3.2 - Material and Methods

The aims of this chapter concern a genomic approach on a considerable proportion of the Bermuda petrel population, which included several procedures and analyses, as further described.

3.2.1 -ddRAD sequencing

The high throughput sequencing method used was double-digest restriction-site associated DNA sequencing (ddRAD seq) and was applied to 104 individuals.

3.2.1.1- Library preparation

The DNA extracts from the 104 Bermuda petrel samples were sent to Queens University in Canada for library preparation using ddRAD seq. DNA samples were digested with *SbfI* and *MluCI* restriction enzymes and ligated with custom adapters including a degenerated base region (DBR) adaptor, a unique individual inline barcode and a restriction overhang produced by the restriction enzymes. Following digestion and ligation, the DNA fragments were size-selected and amplified through PCR (Polymerase Chain Reaction). The DNA concentration of each sample was quantified and pooled in equimolar proportions for sequencing across one lane of Ilumina HiSeq.

3.2.1.2- Bioinformatics pipeline

Raw reads resulting from the ddRAD protocol were parsed to remove putative duplicates resulting from the PCR reactions steps conducted during library preparation. For this purpose, we used the ParseFastQ.py pyhon script (https://github.com/Eljensen/ParseDBR_ddRAD, commit: 8fc0edc), which detects duplicates based on the fragment's DBR sequence. Duplicate-free raw data was then inspected with FastQC v0.11.9 (www.bioinformatics.babraham.ac.uk/projects/fastqc/) for quality controls.

3.2.1.2.1 -*de novo* assembly

De novo assembly of the ddRADseq data was conducted in Stacks v2.55 (Catchen et al., 2013). A *de novo* assembly approach is used when a reference genome for the studied species, or for a closely related species is not available. The first step of the Stacks pipeline is demultiplexing of the raw reads. For this, we used *process_radtags* to discard reads with any uncalled base (-*c*) or low-quality scores (-*q*), as well as to sort sequence reads by individual's barcode.

Stacks comprises six different successive modules in the complete assembly pipeline, which allow SNP calling. The first two programs (*ustacks* and *cstacks*) employ three major parameters that manage the *de novo* assembly of loci and determines the resulting number of RAD loci and SNPs. Therefore, the optimization of these parameters to each dataset is very important as it affects the accuracy of the extracted biological patterns. Optimization of these parameters (*m, M and n*) was processed following Paris et al., (2017). The minimum number of raw reads required to form a putative allele (or stack) (*m*) and the maximum number of mismatches accepted between alleles to consider them the same locus (*M*) are the two parameters that determine the loci within each sample created in *ustacks*. After the building of loci for every individual, Stacks builds a *catalog* in *cstacks*, containing all the assembled loci of the population. The *n* parameter corresponds to the maximum number of mismatches allowed between sequences of different individuals to be considered the same locus (Paris et al., 2017). Standard metrics, such as mean coverage, number of assembled loci, number of polymorphic loci and the number of SNPs, were computed for each parameter, ranging from 2 to 6. As suggested by Paris et al., (2017), such metrics were calculated using a subsample of the data (30 individuals) and only on loci that were present in at least 80% of the sample size (*r80 loci*). These loci are unlikely to have high levels

of sequencing error due to its presence in a high proportion of the population. The results were plotted using GNU R v. 4.0.5 software (Supplementary Fig. S1). This process was done as automated as possible using a shell script written purposely for this analysis (*stacks_param_optimization.sh*) and an R script (*param_optimization_plots.R*) combined (https://github.com/ritafonso/stacks_parameter_optimizer, commits: c7e3a00, c7e3a00). All parameters were set to 3. Detailed choice process of each parameter is fully described in Supplementary Method S1 – Detailed description of parameter choice for Stacks assembly.

The building of the *catalog* employed all 104 individuals sequenced, since all showed a coverage depth > 15%. The assembled loci were then matched against the *catalog* using *sstacks*. The following module, *tsv2bam*, was run with the defaults. *gstacks* incorporated the paired-end reads and removed the remaining single-end reads. The *populations* module was further employed to apply filters of interest. Loci present in less than 60% of the individuals (*-r* 0.6) were filtered. Variable sites with more than 60% observed heterozygosity (*--max-obs-het* 0.60) were purged as they may correspond to pooled paralogs. Minor Allele Frequency (MAF) values are very sensitive to the demography of the population and need to be examined carefully. To avoid considering erroneous genetic signal, a minimum MAF threshold filter is recommended for HTS dataset, such as RADseq (Linck & Battey, 2019). A subsampled dataset of 25 individuals was tested under two different minimum MAF (5% and 10%) and Principal Component Analysis (PCA) was computed for both values using an accessible R script *snp_pca_static.R* (https://github.com/CoBiG2/RAD_Tools, commit: a69e9bd). MAF filter is described to impact genetic structure analysis, which may result in misleading results (Linck & Battey, 2019). Since PCA of both values revealed similar results regarding population structure (Figure S 4), a minimum MAF threshold of 5% (*--min-maf* 0.05) was set in *populations* to process a nucleotide site at a locus, as recommended for population genetic studies (Linck & Battey, 2019). Finally, several different output formats were generated conveniently for further analyses.

3.2.1.2.2 - SNP filtering

The Stacks final assembly generated a Variant Call Format (VCF) file containing all the information of the SNPs called and filtered throughout the assembly. In order to improve the quality and reliability of our SNP's dataset, further filtering was applied to the assembly's resulting VCF file using VCFTools v0.1.16 (Danecek et al., 2011). Such step is important to reduce ambiguities caused by lab procedures or bioinformatics analyses. Mean depth of coverage per-locus (DP) across all individuals was calculated (*--site-mean-depth*) and plotted (Figure S 5). A minimum DP filter is important to remove false positives, while maximum DP cut-offs remove likely paralogous or repetitive regions that erroneously clustered together. Our data was filtered at minimum DP of 10X (*--min-meanDP* 10), which corresponds to approximately 0.25X of the median and maximum DP (*--max-meanDP* 80) was set to 80X, corresponding to around 2X the median (Figure S 5). Variable sites comprehending 20% or higher missing data were also eliminated (*--max-missing* 0.8). Deviations from Hardy Weinberg Equilibrium (HWE) were estimated (*--hardy*) and SNPs that were out of HWE at P-value < 0.05 were removed (*--hwe* 0.05).

The final SNPs dataset resulted from the filters described above was called *"all_snp"* dataset. Since certain analysis can be impacted by linkage disequilibrium (LD), a single SNP was selected for every locus of the *"all_snp"* dataset (Table 3.1). Homologous SNPs in the same locus have a high probability to be in linkage due to physical distance. The SNP located closest to the centre of the locus was the one chosen as sequencing reads exhibit poorer quality scores on their edges. Filtering was employed using the *vcf_parser.py* python script (https://github.com/CoBiG2/RAD_Tools, commit: 0e96c2c), which resulted in the *"one_snp"* dataset (Table 3.1).

3.2.2 - Genetic diversity and population structure analyses

Genomic diversity measures are better understood when compared between closely related populations or species. Thus, a comparative approach brings much value to the analysis and its interpretation. A *de novo* Stacks assembly was done on three closely related, and also highly endangered gadfly petrels: Bermuda petrel, Zino's petrel *Pterodroma madeira* and Desertas petrel, *Pterodroma deserta*. The ddRAD data of the latter species were obtained for 28 and 20 individuals respectively. In order to balance the assembly procedure and not bias the results, the sample size for each species was settled to 20 individuals. Madeira's samples were randomly subsampled to 20 individuals, while the Bermuda petrel implied a more complex procedure due to its sample size being around five times larger. A bash script was produced including a bootstrap procedure of 100 runs, where the Bermuda sample size was randomly subsampled to 20 and subsequently joined with the other two species for the Stacks *de novo* assembly, for every run (all_freiras_assembly_script.sh in https://github.com/ritafonso/misc-scripts, commit: 8763017). Thus, 100 de novo assemblies were run, and the results represent the average of those runs. Parameter optimization method was done as described above for the Bermuda petrel assembly and the parameters chosen were (*m*=3, *M*=3, *n*=3), although customized scripts were created for this assembly (param_optimization_combined_assembly.sh and plot_params_opt_combined_assembly.sh in https://github.com/ritafonso/misc-scripts, commit: 8763017).

Diversity indices were then calculated for each species, without any bias from sample size or parameter choice. Nucleotide diversity was estimated for every variable site (π_{SNP}) and for RAD loci of the Stacks's *catalog* (π_{RAD}). The mean per-individual observed and expected heterozygosities (*Ho* and *He*) were estimated. All measures were obtained using the *populations* module of Stacks. SNP filters described for the *"all_snp"* dataset (Table 3.1) were also applied to this dataset.

We also surveyed one mitochondrial locus. The mitochondrial genome is haploid and maternally inherited, causing it to have a quarter of the population effective population size of the nuclear genome. This causes mitochondrial loci to have higher rates of genetic drift, which adding to a generally higher substitution rate, results in faster responses of mitochondrial loci to demographic events than nuclear loci (Hamilton, 2009). Nucleotide and haplotype diversity for the three petrel species described above were calculated for a fragment of the cytochrome oxidase I (COI) gene for 29 individuals. COI sequences of species close to the Bermuda petrel were obtained from NCBI GenBank followed by a random subsample of 29 individuals.

Population genetic structure was inferred through a Principal Component Analysis (PCA) of the filtered SNPs information from the *"one_snp"* dataset. Computation of the PCA was performed using the R package SNPRelate v. 1.24.0 (Zheng et al., 2012) as implemented in the R script *snp_pca_static.R* (https://github.com/CoBiG2/RAD_Tools, commit: a69e9bd). The script was partially modified, using the ggplot2 v. 3.3.3 package (Wickham H, 2016) and the pairs function to plot the results (https://github.com/CoBiG2/RAD_Tools/pull/3). A PCA including the three petrel species was also computed for validation of genetic structure between them, using the same script.

3.2.3 - Inbreeding inference

Individual inbreeding (*f*) is described as the proportion of the genome that is Identical-by-descent (IBD). The *f* resulted from a relationship of full-sibling or parent-offspring is approximately 0.25, while for half-siblings or grandparent-grandchildren relations, *f* of the offspring is approximately 0.125. Relatedness is the expected proportion of alleles between individuals that are IBD (Hamilton, 2009). High throughput sequencing (HTS) methodologies that allow the use of mapped molecular

markers, such as a whole-genome-sequencing approach or RAD-seq when a reference genome is available, has shown the highest power to estimate *f*, by identifying homozygous IBD chromosome segments, often called runs of homozygosity (ROH) (Kardos et al., 2016). Nevertheless, unmapped molecular techniques, such as ddRAD-seq, are also very useful to estimate *f* through individual heterozygosity and have been shown to have higher power than other markers, such as microsatellites (Hoffman et al., 2014; Huisman et al., 2016). Individual genome-wide heterozygosity is described as a good estimator of *f*, based on the idea that offspring of more closely related parents may show lower heterozygosity due to larger IBD segments across the genome. The real proportion of heterozygous loci across the genome, which is a function of *f* and the heterozygosity of non-inbred individuals (Crow and Kimura, 1970), can be approximated with our ddRAD data set by the multi-locus heterozygosity (*MLH*), which represents the proportion of genotyped loci that are heterozygous. *f* can also be measured using the diagonal elements of the genomic relatedness matrix (F_{GRM}) (Kardos et al., 2016). *MLH* and F_{GRM} were calculated for every individual using the *inbreedR* package v. 0.3.2 (Stoffel et al., 2016) in GNU R. A Pearson correlation between the two marker-based estimators was estimated, and results of the distribution of *MLH* and F_{GRM} across the population were plotted using *ggplot2* v. 3.3.3 in GNU R.

3.2.3.1- Inbreeding avoidance through assortative mating

Genomic data combined with breeding system information provided by the extensive field work in Bermuda allowed the gathering of genomic data for 31 couples. This allowed to test for the presence of assortative mating in the population regarding the degree of relatedness of the chosen mates. The mean kinship value between real couples was compared against the frequency distribution of kinships generated from 1000 simulations of the same number (i.e., n=31) of randomly selected pairs. These simulated couples were randomly created with a non-reposition system incorporated in a python script (*distributor.py*, https://github.com/ritafonso/misc-scripts, commit: 8763017). The significance level used of this non-parametric procedure was α= 0.05. We used our PCA results to look for a signal of population sub-structuring among our *P. cahow* samples which might also result in a pattern of non-random mating. If no sub-structuring is found in our dataset, then a significant deviation from random mating in our test may be the result of inbreeding or outbreeding avoidance.

3.2.3.2- Testing for signatures of inbreeding: Heterozygosity-fitness correlations

The correlation of heterozygosity across loci in a genome, also called identity disequilibrium (ID) was estimated in our dataset by calculating the parameter g_2 (David et al., 2007), using the *inbreedR* package v. 0.3.2 in R. The probability that g_2 differs from 0 was computed with 1 000 permutations to estimate the p-value and to estimate a 95% confidence interval.

According to the general effect hypothesis, a significant ID is related to heterozygosity-fitness correlations (HFC) (Brommer et al., 2015). The impact of the mean genome-wide heterozygosity of the couple in their hatching success across years for which this fitness trait was available was tested, using hatching success as the response variable, *MLH* as fixed effects and the nest-ID as a random effect. Information on the yearly hatching success was available for 31 couples (Figure S 6), thus we used 62 individuals from the genomic dataset out of the total 104 individuals. Heterozygosity-fitness correlation analysis were conducted in GNU R using a generalized linear mixed model (GLMM) in the *lme4* package v. 1.1-27.1 (Bates et al., 2015).

3.2.4 -Relatedness analysis

The R package *related* v. 1.0 (Pew et al., 2015) is based on the algorithm by (Wang, 2011) as implemented in the COANCESTRY and it was used to estimate pairwise relatedness of all individuals in the dataset. The function *compareestimators* was applied to test the performance of different

relatedness estimators and choose the best for our dataset. This function creates simulated pairs of individuals of known relationship degrees (unrelated (UN), parent-offspring (PO), full-sibling (FS) and half-sibling (HS)) from the real allele frequency file provided by the user. Four different non-likelihood-based estimators are then applied to the simulated data and the results were plotted also using GNU R. Finally, the function uses the Pearson correlation coefficient between the real and simulated data to assess the performance of each estimator. Stricter filters were applied to the "one_snp" dataset, resulting in *"strict"* dataset (Table 3.1). The VCF file was converted to geneland format using PGDSpider v. 2.1.1.5 (Lischer & Excoffier, 2012). Random subsets of 500, 1000, 2000, 3000 and 4000 were created by the R script *subsample_geneland_script.R* (https://github.com/ritafonso/geneland_subsample, commit: 1620cb9) and were tested in *compareestimators* to measure the impact of the number of SNPs used. Higher numbers of SNPs would hit the software's limits, and thus were untestable. Each test was applied to 1000 simulated pairs of known relationship. The results of *compareestimators* function revealed that the precision of estimation of each relationship degree increased with the number of SNPs, but the effect disappeared after 3000 SNPs. Such relationship was similar for every estimator and the correlation between the four estimators was very high (>0.995). The Wang, (2002) estimator was thus chosen and applied to the 4000 SNPs subsample using the *coancestry* function.

Pairs where relationships are known from field data were crossed with genetic relatedness estimation and a density plot was created with *ggplot2* (Wickham H, 2016) in R to check how well the chosen estimator distinguishes the different known relationships present in the dataset.

3.2.4.1- Effect of Pair Relatedness on hatching success

Particularly in a context of inbreeding, relatedness between mates can impact fertility and the production and offspring (Trask et al., 2021). Such impact was studied in the Bermuda petrel by looking at the hatching success of the couples across time taking into consideration their degree of relatedness. Hatching success for each couple was obtained from field data, and it was defined as the proportion of successful hatchings per number of years that the couple attempted to mate. Hatching was considered successful if the chick hatched and survived the next days after hatching, since chick survival in this period is still highly dependent on egg quality. All other scenarios were considered failures, including egg unviability to embryo death. To test for a possible correlation between each couple's degree of relatedness and its hatching success, a generalized linear mixed-effect model (GLMM) was employed using the R package *lme4* v. 1.1-27.1. Pairwise relatedness was set as a fixed effect, with hatching success as the dependence variable and nest-ID as a random effect. This analysis was applied to the *"one snp"* dataset.

3.2.5 - Inference of demographic history

To test whether the Bermuda petrel underwent a bottleneck, as suggested by historical records, the demographic history of the Bermuda petrel was investigated by two different approaches, both based on the site frequency spectrum (SFS). For these analyses, a new dataset was used without SNPs or individual's missing data. The filters of MAF and maximum heterozygosity were also not applied to the dataset and all SNPs per RAD loci were included. The SFS was calculated using easySFS (https://github.com/isaacovercast/easySFS, commit: b866269). This tool contains a projecting function (*--preview*) that counts the number of segregating sites for all the possible number of individuals in our dataset. This allows maximizing the number of SNPs, which is critical to calculate the SFS (Gutenkunst et al., 2009), even though it may reduce the number of individuals in the dataset and thus the number of coalescences used to estimate variation in N_e. We used the full dataset of 104 individuals with 27 062 SNPs.

To perform the demographic inference analysis on the Bermuda petrel we used Stairway plot v.2 (Liu & Fu, 2020), a software that fits a multi-epoch demographic model to the data. For this, we tested four different numbers of random breakpoints (51, 103, 154, 206) based on the sample size. As recommended, 67% of sites were used for training data as well as 200 bootstraps to estimate 95% confident intervals. We used a mutation rate of 2.89×10^{-9} per site per generation based on an estimate for the northern fulmar (Nadachowska-Brzyska et al., 2015).

Generation time was estimated using the following formula:

$$T = A + p/(1-p)$$

where T represents generation time, and A and p represent age of maturity and adult survival rate, respectively (see Gangloff et al. 2013). It has been estimated in other *Pterodroma* that the adult survival rate is 0.93 (Brooke et al., 2010). Although the actual adult survival rate is not known for *P. cahow*, it is known that it has higher survival rates than other seabird species due to the ongoing conservation plan that mitigates some of the threats to their survival. The average age of maturity in *P. cahow* has been defined as 5 years old (J. L. Madeiros, 2005). Those values result in a generation time of around 18 years old, but it was increased to 20 years old due to the external care that Bermuda petrel undergo under the current Conservation Plan (see Gangloff et al. 2013). The remaining parameters in Stairway plot were set to default.

The second demography inference method used is implemented under a faster continuous-time sequential Markovian coalescent approximation in fastsimcoal2 (Excoffier et al., 2021). The observed SFS included the number of monomorphic sites, obtained in Stack's module *populations*. We compared the fit of three different models that we considered most likely representing the demographic history of *P. cahow*: "Recent Bot", "Null" and "Old Bot". The "Null" model assumes a constant population size over time (Figure 3.1). The "Recent Bot" model tests the likelihood of a genetic bottleneck from 3 to 30 generations ago, while "Old Bot", an older bottleneck, from 30 to 200 000 generations ago (Figure 3.1). For each model, 100 independent runs were performed with 30 cycles and 100 000 coalescent simulations to estimate an expected SFS. The best run for each model was found through the highest likelihood, as implemented in *fsc-selectbestrun.sh* script (https://github.com/speciationgenomics/scripts, commit: 48a9fb2). Likelihoods of the models are obtained by the difference between the observed likelihood (MaxObs), which is the maximum possible likelihood with a perfect fit of the expected SFS, and the estimated likelihood (MaxEst), the maximum likelihood estimated for each model. The fit of each model was analysed by comparing the observed and expected SFS, using the *utilFscOutput.r* R script (https://github.com/vsousa/EG_cE3c/blob/master/CustomScripts/Fastsimcoal_ProcessOutput/Scripts_AnalyseFsc, commit: c3a8871). The most suitable model to our dataset was assessed by the Akaike information criterion (AIC), which incorporates the numbers of parameters of the model, using *calculateAIC.sh* script (https://github.com/speciationgenomics/scripts, commit: 48a9fb2).

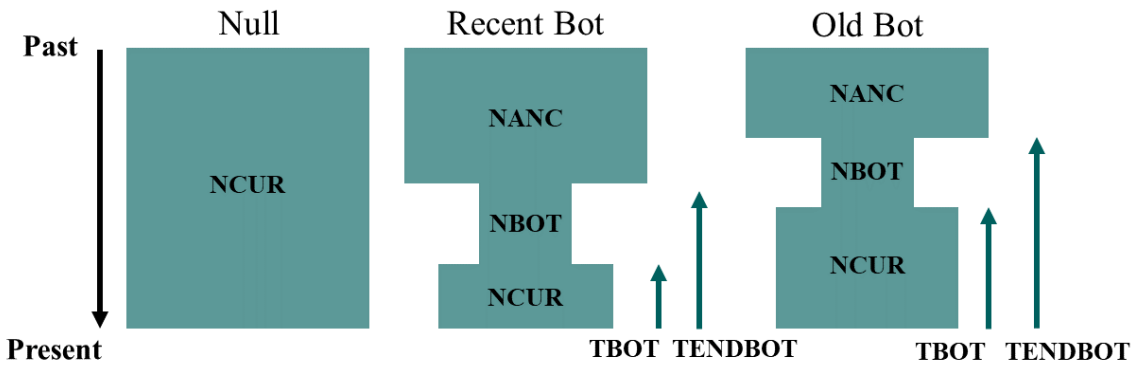


Figure 3.1 Historical demographic models analysed using fastsimcoal2. Model names are on top of each scheme. NCUR represents the estimation for current N_e, NANC, N_e before the bottleneck and NBOT, the N_e during the bottleneck. TBOT defines the time (in generations) passed since the end of the bottleneck, while TENDBOT identifies the beginning. Results of parameter estimation and AIC of all models are given in Table 3.5.

3.3 - Results

The results regarding ddRAD seq method and analyses concerning genetic diversity, inbreeding and demography are further described separately.

3.3.1 - RAD sequencing

The RAD library constructed produced 403 387 152 raw reads. Demultiplexing and quality filtering resulted in 399 838 435 reads. Further filtering on SNP extraction resulted in three different datasets, comprehending different number of variable sites (Table 3.1).

Table 3.1 -SNP filtering steps and final datasets.

Filtering step	Nr of Retained SNPs	Notes	
Stacks final assembly	126 681	-	
individuals per locus > 0.6	86 048	Loci with information for less than 60% of the individuals were removed	*populations* – Stacks module [1]
Maximum observed heterozygosity = 0.6	85 560	SNPs with more than 60% observed heterozygosity are removed.	
MAF>0.05	56 536	SNPs with MAF lower than 0.05 were removed	
Missing data > 0.8	49100	SNPs with greater than 20% missing data were removed	VCFTools [2]
min DP = 10 and max DP=80	39716	Only SNPs with depth of coverage higher than 10 and lower than 80 were kept	
HWE deviation (p-value > 0.05)	33 501	Loci that showed significant ($p < 0.05$) departure from HWE were removed.	
Centred SNP filter	12 484	Only the SNP closest to the centre of the locus was kept	*vcf_parser.py* [3]
Missing data > 1	7 167	All SNPs with missing data were excluded	VCFTools [2]
min DP = 20 and max DP=60	6 814	Only SNPs with depth of coverage higher than 20 and lower than 60 were kept	
Final "all_snps" dataset	33 501	-	
Final "one_snp" dataset	12 484	-	
Final "strict" dataset	6 814	-	

1 - (Catchen et al., 2013) ; 2 - (Danecek et al., 2011) ; 3 – (https://github.com/CoBiG2/RAD_Tools, commit: 0e96c2c).

3.3.2 - Genetic diversity and population structure

Two measures of nuclear nucleotide diversity (π_{SNP} and π_{RAD}) were used to compare genetic diversity between the Bermuda petrel and the two northeast Atlantic, related petrel species. Both measures showed similar levels of diversity between species (Table 3.2), although *P. madeira* showed higher diversity (π_{SNP} =0.202, π_{RAD} = 0.200 x 10^{-2}), while *P. deserta* (π_{SNP} =0.181, π_{RAD} = 0.176 x 10^{-2}) showed the lowest, and *P. cahow* revealed intermediate diversity (π_{SNP} =0.193, π_{RAD} = 0.197 x 10^{-2}). This trend was also confirmed by both the estimated observed (*Ho*) and expected heterozygosity (*He*).

Table 3.2 - Genetic diversity indices concerning ddRAD seq analysis

	Variable sites			All sites		
Species	π_{SNP}	*Ho*	*He*	π_{RAD} (x 10^2)	*Ho* (x 10^2)	*He* (x 10^2)
Pterodroma cahow	0.193	0.180	0.187	0.197	0.183	0.191
Pterodroma deserta	0.181	0.180	0.175	0.176	0.173	0.171
Pterodroma madeira	0.202	0.182	0.196	0.200	0.180	0.194

The same pattern was revealed by the mitochondrial COI fragment, although it showed higher differences between the three species (Table 3.3). The same diversity measures were also applied to other species in the *Pterodroma* genus for further comparison. *Pterodroma madeira* revealed greater nucleotide and haplotype diversity (π_{COI}= 0.00344, *Hd*=0.773), as well as a higher number of haplotypes and appears to be the most diverse of the seven *Pterodroma* species included. On the other hand, *P. cahow* showed the lowest diversity (π_{COI}= 0.00027, *Hd*=0.192), with only 2 different haplotypes in 29 individuals genotyped (Table 3.3).

Table 3.3 - Nucleotide and haplotype diversity for the COI mitochondrial gene of seven *Pterodroma* species. N represents the number of sampled individuals and H the number of haplotypes.

Scientific name	**Common name**	**N**	**Fragment size (bp)**	**Polymorphic sites / 100bp**	**Nucleotide diversity (π x 10^2)**	**H**	**Haplotype diversity (*Hd*)**
P. cahow	Bermuda petrel	29	722	0.139	0.027	2	0.192
P. deserta	Desertas petrel	29	732	0.41	0.079	3	0.490
P. madeira	Zino's petrel	29	732	2.32	0.344	7	0.773
P. feae	Cape Verde petrel	29	732	0.683	0.118	6	0.658
P. cookii	Cook's petrel	29	677	0.443	0.106	4	0.584
P. macroptera	Great-winged petrel	29	648	0.463	0.032	4	0.200
P. magentae	Magenta petrel	10	648	0.309	0.086	3	0.378

DNA sequences used for this analysis were obtained in the following references: *P. cahow* (present thesis); *P. deserta*, *P. madeira* and *P. feae* (Gangloff et al., 2013); *P. cookii* (Rayner et al., 2010); *P. macroptera* (Lawrence et al., 2016); *P. magentae* (Tizard et al., 2019).

Population structure analysis in *P. cahow* as revealed by a PCA showed no evidence of sub-structure between the islands where individuals currently breed.

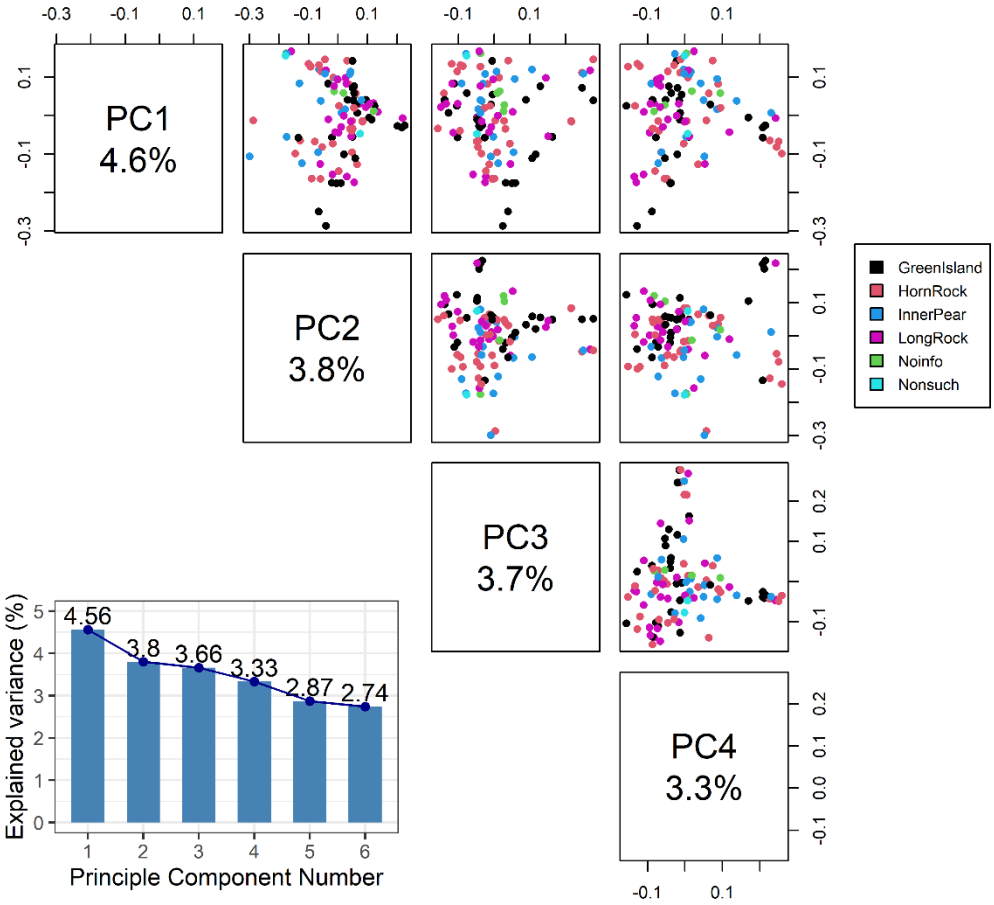


Figure 3.2 - Principal Component Analysis (PCA) of genome-wide markers (ddRAD seq data) of Bermuda petrel population across the five islands where they were sampled. The barplot on the bottom left corresponds to the first six eigenvectors and their respective percentage of variance explained.

3.3.3 - Inbreeding

Estimates of the inbreeding coefficient as calculated by the genomic relatedness matrix (F_{GRM}) ranged from -0.093 to 0.24, with a mean value of -0.017 (Figure 3.3). F_{GRM} can take negative values, because the current population is used as reference, implying that an individual with a negative F_{GRM} is outbred compared to the mean of the population (Huisman et al., 2016). Three individuals show F_{GRM} >0.15, representing a high level of inbreeding, while five others show values between 0.05 and 0.10, estimated for offspring of a cousin's couple, for example, which also represents a considerable high inbreeding level (Figure 3.3). This variation suggests that it is possible to find HFC in the Bermuda petrel population.

Multi -locus heterozygosity (*MLH*) estimates ranged from 0.21 to 0.33, with a mean of 0.294. Three individuals revealed extremely low heterozygosity values, below 0.25 (Figure 3.4). *MLH* and F_{GRM} were highly correlated (Pearson's coefficient r=-0.93, p =2.2e-16). Thus, *MLH* estimates were used for further enquiries related to heterozygosity fitness correlations analysis.

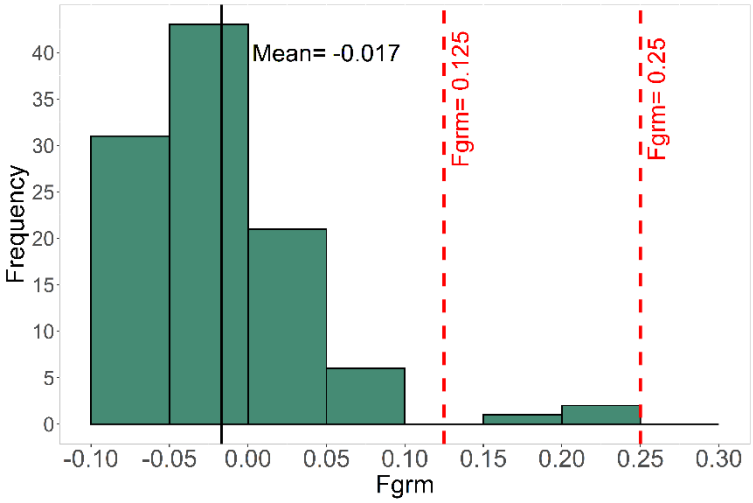


Figure 3.3 F_{GRM} distribution of the 104 sampled individuals. The black vertical line is relative to the mean value. The red vertical lines correspond to reference values of F_{GRM}. F_{GRM}=0.125 is equivalent to offspring of two cousins, and F_{GRM}=0.25 is equivalent to offspring of two siblings.

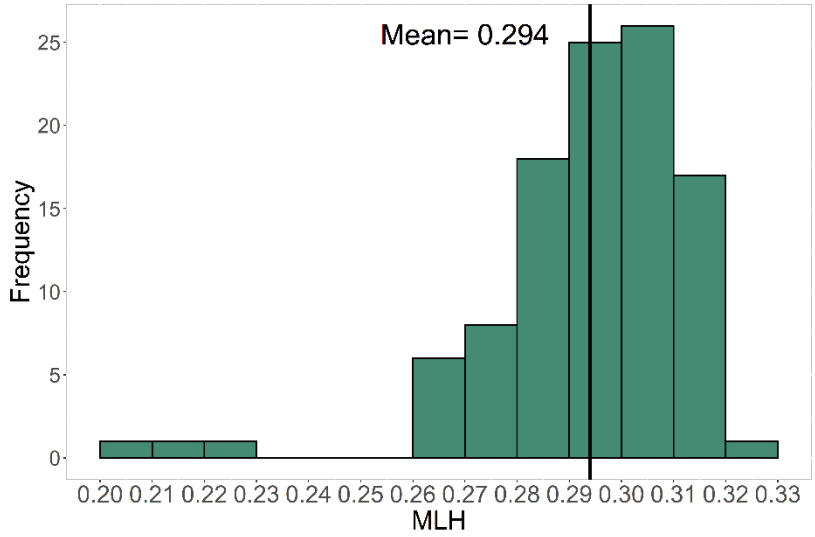


Figure 3.4 Multi locus heterozygosity (*MLH*) distribution of the sampled population. The black vertical line represents the mean value.

3.3.3.1- Heterozygosity fitness correlations

Identity disequilibrium (ID) was found in the ddRAD seq dataset of Bermuda petrel, as the parameter *g2* was significantly different from zero (0.0046, $p < 0.001$). This result indicates that heterozygosity is significantly correlated across loci among the individuals and that the heterozygosity in the ddRAD seq data of 62 individuals appears to be representative of genome wide heterozygosity. This also suggests that, although not very strong, there is a sign of inbreeding in the Bermuda petrel population and that a putative presence of HFC may indeed result from inbreeding depression.

There was no significant effect of the *MLH* on the hatching success, a clear fitness-related trait (Table 3.4). Therefore, in contrast to our expectations, we found no evidence that individual heterozygosity is associated with the likelihood of its egg hatching successfully, at least with this set of markers and for the 62 sampled individuals.

3.3.4 - Relatedness

Pairwise relatedness was estimated for the sampled individuals and, overall, the density plots showed a strong overlap with the values expected from the known pairwise relationships from the field data (Figure 3.5). Lower precision was found for full siblings (FS). Exceptions were found, most of them probably resulted errors during field data collection.

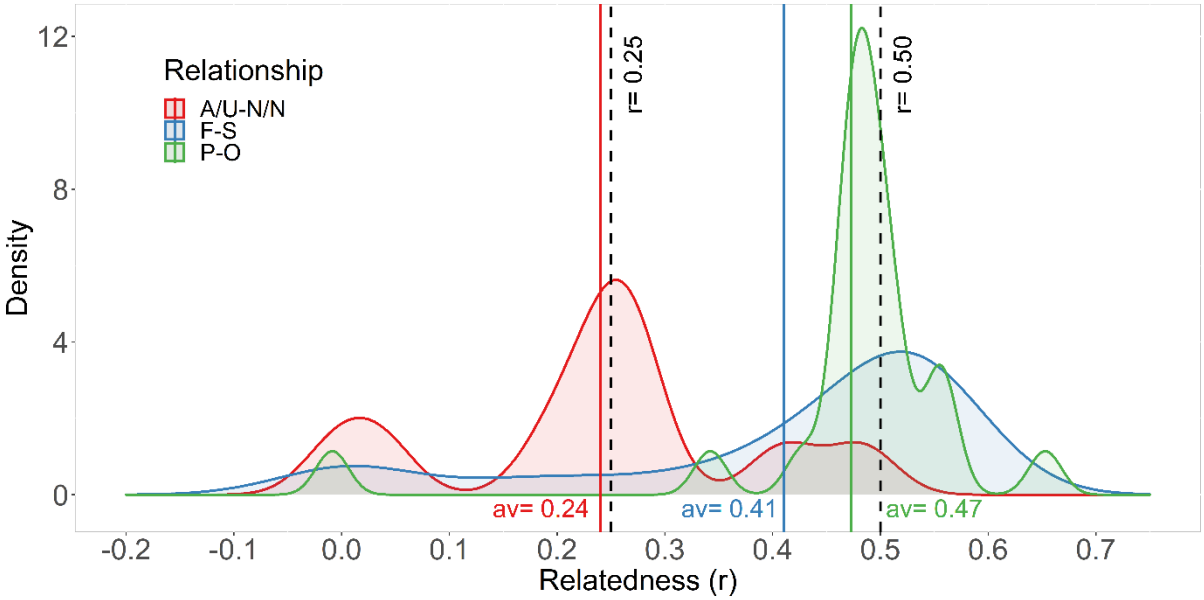


Figure 3.5 Relatedness estimates based on ddRAD seq data grouped by known relationships from the field data

P-O (Parent-offspring); F-S (Full-sibling); A/U-N/N (Aunt/Uncle – Niece/Nephew)

Mean pairwise relatedness between the 31 couples sampled from the Bermuda petrel population was 0.0197 (sd = 0.131). Three couples showed outlier values of relatedness: two had values of 0.177 and 0.258, which may represent pairings between cousins and half-siblings, respectively. The other couple showed relatedness of 0.557, which field data revealed being a full-sibling couple.

There was no significant relation between the relatedness of the couples and their hatching success along the years as revealed by GLMM analysis (Table 3.4). Hence, hatching success of the couple is not influenced by their degree of genetic relatedness as revealed by ddRAD seq data.

Table 3.4 – Results of the generalized linear mixed models (GLMM). The dependence variable is presented first and the fixed effect after, while the random effect is in parenthesis.

Model	**Estimate ± SE**	***p***
hatching success ~ *MLH* + (1|nest id)	1.380 ± 6.568	0.834
hatching success ~ relatedness + (1|nest id)	-0.543 ± 1.531	0.723

MLH – multi locus heterozygosity; SE – standard error; *p* – p-value.

3.3.5 - Assortative mating

The mean value of pairwise relatedness of the 31 couples sampled did not significantly differ from the relatedness of the simulated couples generated by the non-parametric bootstrap (Figure 3.6). Since these 31 couples are almost a quarter of the total breeding population of Bermuda petrel, data do not support assortative mating regarding relatedness being present in the Bermuda petrel. In other words, the results suggest that in this species, individuals do not choose mates that are more or less genetically related to themselves than if mate choice occurred at random.

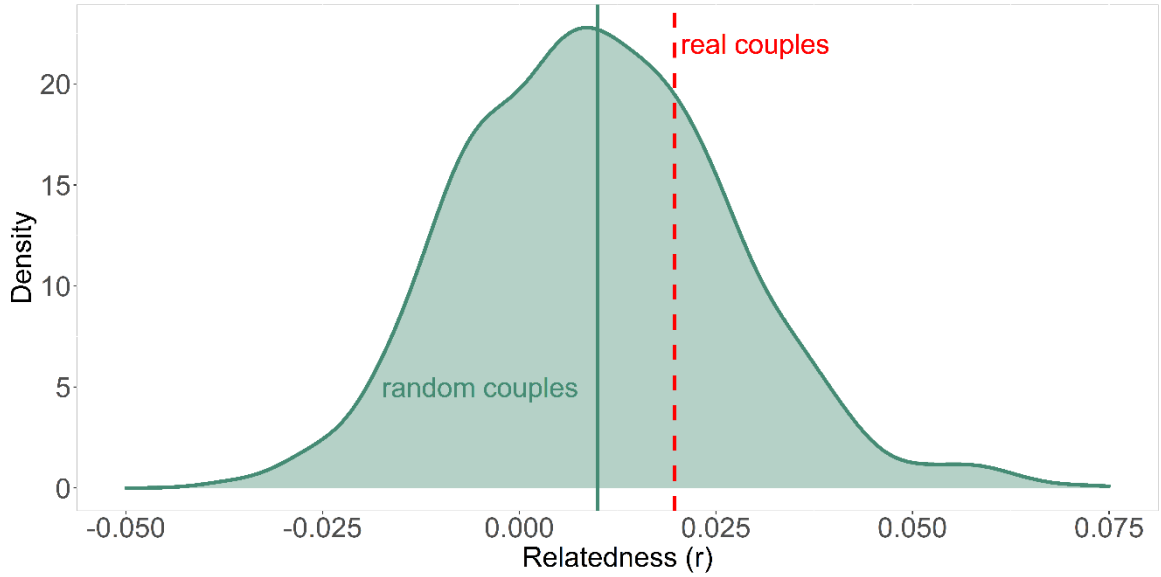


Figure 3.6 Distribution of random generated couples. The vertical solid line represents the mean value. The red line represents the mean value of pairwise relatedness for the 31 couples sampled for Bermuda petrel population.

3.3.6 - Demographic history

Data filtering applied to the dataset to estimate the SFS, resulted in a 27 062 SNPs present in all 104 individuals. The results of the Stairway Plot analysis are shown in Figure 3.7. It revealed a gentle and long-lasting decrease in the effective population size which took the population from ~150 000 individuals about 100 000 years ago to ~4 000 individuals, 2 000 year ago, representing a reduction of ~98%. Following this long-lasting decrease, Stairway Plot suggests a considerable population bottleneck (from 4 000 to 600 individuals) that lasted around 1 200 years followed by an apparent increase in the following 300 years (from 600 to 4 000 individuals). The apparent expansion occurring from the recovery of the bottleneck ceased around 500 years ago and since then the population size has been stable.

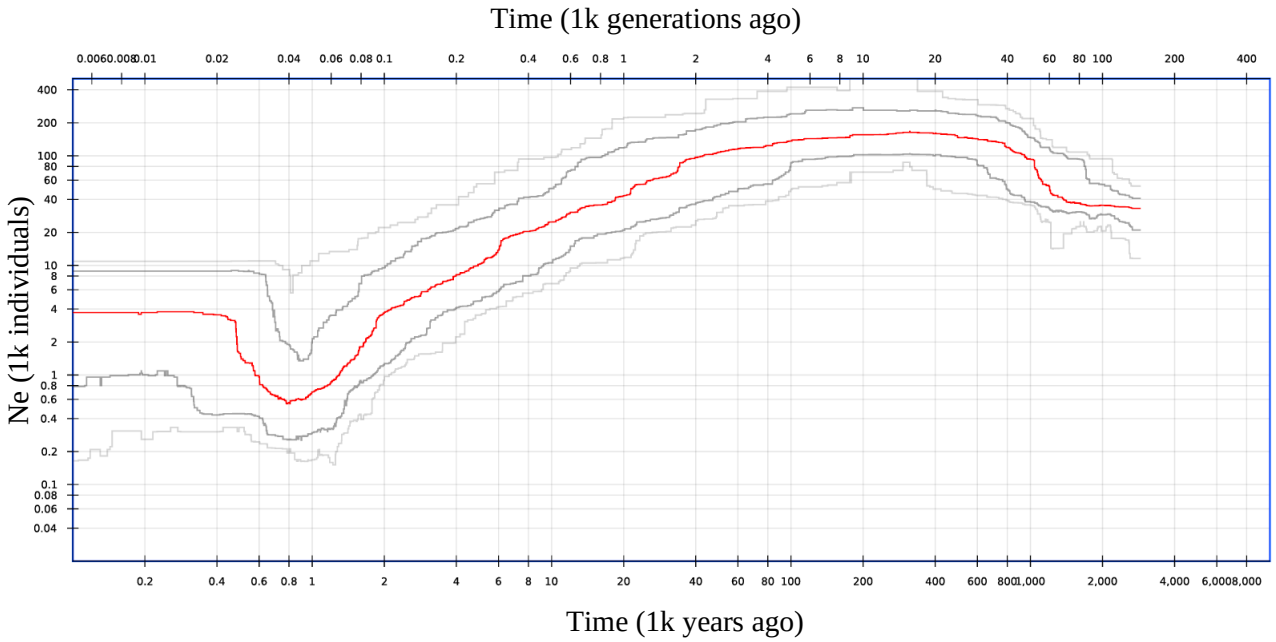


Figure 3.7 - Stairway Plot analysis results. Bermuda petrel's generation time was set as 20 years old. All axes are presented in logarithmic scale.

Regarding the three demographic models tested by fastsimcoal2, the "Recent bot" model showed the highest likelihood, followed by the "Old bot", showing a slightly lower likelihood, and finally the "Null" model (Table 3.5). The model of the "Recent bot" suggests a ~99.8 % population's reduction between 13 and 3 generations ago, followed by a ~90% expansion. Generally, the model suggests that the current *Ne* is ~98.3% lower than before the bottleneck. Despite this model showing the best fit for our data, it does not exactly fit the observed SFS (Figure S 7). Parameter estimates, likelihoods, and AIC of each model are described in Table 3.5.

Table 3.5 – Description of parameter estimates of ach model and respective likelihood and Akaike criteria correction (AIC)

Model	NANC	TENDBOT	NBOT	TBOT	NCUR	MaxEst – MaxObs (Likelihoods)	AIC
Null	-	-	-	-	770	9864.11	209811
Rec Bot	57089	13	99	3	961	300.23	205441
Old Bot	144434	2684	61	2664	733	1249.22	249478

NANC – Effective population size before the bottleneck; TENDBOT – Time since the start of the bottleneck; NBOT – Effective population size during the bottleneck; TBOT – Time since the end of the bottleneck; NCUR – Effective population size after the bottleneck.

3.4 - Discussion

This project aimed to understand the possible genetic consequences of a known historical bottleneck in the Bermuda petrel population. For this, mtDNA and nuclear SNPs were used to infer genetic diversity and heterozygosity, and to estimate the inbreeding level and the putative effects of an expected increase in individual homozigosity. The demographic history was also inferred with genomic data analyses to look for a signature of this bottleneck. Both mtDNA and nuclear SNPs were successfully obtained for a large fraction of the population and ddRAD data was shown to be representative of genome-wide heterozygosity levels. Even though no signs of inbreeding depression were found for hatching success, low genetic diversity and significant inbreeding levels found for some individuals raise concern for this species' survival potential.

3.4.1 - Genetic diversity pattern in Bermuda petrel

The nuclear neutral genetic diversity of Bermuda petrel was assessed on a combined analysis with two other northeast Atlantic petrel species: the Portuguese Desertas and Zino's petrel. These species are both single island endemic and are considered Vulnerable and Endangered by the IUCN, respectively, because of their small size, habitat loss and other threats to their survival (Menezes et al., 2010). Nuclear genome-wide heterozygosity and nucleotide diversity were similar between the three petrel species (Table 3.2). However, mtDNA results suggested a different scenario, showing Zino's petrel with the highest diversity, followed by Desertas and finally Bermuda's. Zino's petrel highest diversity may be explained because of its ancestral position as suggested by its more basal position in a *Pterodroma* species tree whereas Desertas petrel is a more recent, derived taxon (Gangloff et al., 2013). In fact, Bermuda petrels showed the lowest nucleotide and haplotype diversity of COI gene among seven *Pterodroma* species, most of which have undergone a recent bottleneck and all of them being endangered by similar threats. Conflicting results between nuclear and mitochondrial diversity are well described (e.g. Silva et al., 2015) and a potential explanation relies on the fact that nuclear DNA has a higher effective population size and consequently longer coalescence times than mitochondrial, leading to a slower rate of loss of ancestral genetic variation. This is due to mitochondrial DNA being only maternally inherited and haploid. Thus, the relatively lower level of mtDNA diversity of Bermuda petrel compared to pattern of the nuclear genome may be due to the recent loss of genetic diversity still not being represented in the nuclear genome since mtDNA tracks demographic events much faster. These results lend further support to the recency of the genetic bottleneck in the Bermuda petrel. Additionally, similar COI diversity levels were found in Magenta and Grey-winged petrel (Table 3.3), two southern hemisphere petrels which have suffered very similar recent bottlenecks.

3.4.2 - Inbreeding level and impact in the Bermuda petrel

Inbreeding levels are expected to be higher in small populations, especially those that went through a genetic bottleneck, such as the case of Bermuda petrel. Thus, these species are at particular concern regarding inbreeding depression and its impact on their vulnerability, as already shown in several avian species (Blanco & Morinha, 2021; Townsend & Jamieson, 2013).

Analysis of the Bermuda petrel showed overall low levels of inbreeding for most individuals. Nevertheless, for a few individuals inbreeding was high, which should not be neglected since such values are not expected in a large population under random mating. In addition, multi-locus heterozygosity (*MLH*), as estimated from our large panel of RAD loci, was shown to be related to inbreeding, as expected because inbreeding will tend to reduce the individual's heterozygosity (Hoffman et al., 2014), and was employed as its surrogate. Variation found in the observed *MLH* distribution was

also illustrated by the identity disequilibrium (ID) estimates for the Bermuda petrel population. The parameter g_2 was positive, although not very high (see 3.3.3.1- Heterozygosity fitness correlations) when compared with other small bird populations (Foster et al., 2021; Velando et al., 2015) suggesting less inbreeding level variation. The fact that all the individuals sampled for the present thesis are adults should be taken into consideration when interpreting identity disequilibrium since this metric decreases with the age due to selective pressure against the most inbred individuals (Huisman et al., 2016). Despite this fact, g_2 was still significantly different from 0, which suggests that the heterozygosity assessed through the ddRADseq assembly is representative of the genome-wide heterozygosity of Bermuda petrel population. Furthermore, it also suggests, that genome-wide heterozygosity correlates with individual inbreeding and that heterozygosity-fitness correlations (HFCs) are likely to exist in the population which should be caused by a genome-wide effect (Szulkin et al., 2010).

Assessment of HFCs in small populations like Bermuda petrel's is key to better understand the effects of inbreeding in the different fitness components, particularly whether there are traits under inbreeding depression. The Bermuda petrel has been shown to have among the lowest levels of hatching success for *Pterodroma* species. In fact, it has been shown in other avian species that inbreeding can have a strong effect on fertility and hatchability, as well as in other fitness traits (Hemmings et al., 2012; Trask et al., 2021). High levels of inbreeding can lead to genetic incompatibility between mates that end in hatching failure, being early embryo death described as the most likely explanation in case of inbreeding depression (Hemmings et al., 2012). However, and in contrast to our expectations, present results suggest that the pair's average multi-locus heterozygosity (*MLH*) does not have a significant impact on its ability to successfully hatching an egg (see 3.3.3.1- Heterozygosity fitness correlations).

Other possible causes of the low hatching success reported for Bermuda petrel may be related to maternal condition and environmental factors (Hemmings et al., 2012), since our results exclude inbreeding depression as a likely explanation. Although our results failed to detect an HFC regarding hatching success, the identity disequilibrium found in our samples still advocates that inbreeding might have an impact in fitness. Hence, looking at other fitness traits, as well as sampling chicks and juveniles is needed to better understand the impact of inbreeding and to test the presence of inbreeding depression more accurately in Bermuda petrel population.

It has been shown that thousands of SNPs obtained from high throughput sequencing can have high power for relatedness estimation, despite putative genotyping errors (Thrasher et al., 2018). We confirmed such power by the close match between relationships based on field observations and the relatedness estimation for these pairs based on genomic data, bar a few exceptions which can be explained by errors in the field data or during genotyping.

Similar to the analysis testing the influence of individual inbreeding on hatching success, pair relatedness estimation of 31 couples was assessed for its effect on their hatching success. The results indicate that the degree of kinship between Bermuda petrel's couples does not influence their hatching success. Although not expected, this result is not exclusive in seabirds since no relation was found between genetic relatedness, based on microsatellites, and breeding success in the Monteiro's Storm-petrel (Nava et al., 2017). One explanation for our results is that our sample size, particularly of closely related pairs, is not large enough to allow the detection of a relation between the pair's relatedness and its hatching success. Another alternative explanation is the presence of extra-pair paternities (EPPs). Although poorly explored in *Pterodroma* genus, EPPs have been described among Procellariiformes, even if not expected (Bried et al., 2021). Certain circumstances may lead to EPPs in social monogamous species as it may increase the offspring genetic diversity. In fact, EPPs have been described as a potential way of avoiding inbreeding (Brouwer & Griffith, 2019). Although the occurrence of EPPs has never been studied in Bermuda petrels, this might be an alternative strategy for inbreeding avoidance in this

species. Also, this may explain the lack of evidence of genetic relatedness influence on the couple's hatching success. However, it is also a possibility that relatedness between the couple does not influence hatching, as discussed by Mays et al., (2008).

Single island endemics and highly philopatric species, such as the Bermuda petrel, are limited regarding dispersal strategies, therefore may rely on mating behaviours to avoid inbreeding. Assortative mating is one of the known strategies of inbreeding avoidance, in a sense of preventing mating between relatives, and it is described that procellariifom species can differentiate kin from non-kin individuals through olfactory cues (Bonadonna & Sanz-Aguilar, 2012). Additionally, in social monogamous species with necessary biparental care, non-random mating choice should be essential and expected since pairs mate for life (Nava et al., 2017), and a bad choice could impact life-time reproductive success. Thus, the Bermuda petrel is a good candidate to study assortative mating regarding kinship degree. Results of the present study show that the individuals of this population are choosing mates randomly regarding neutral genomic relatedness. This result may be explained by the limited pool of available mates or the inability to detect relatives. Also, other characteristics can surpass genetic relatedness when it comes to mate choice, such as certain phenotypic traits, behaviours, or relatedness of specific functional genes. Hence, we cannot conclude that assortative mating is not a strategy to mitigate inbreeding depression effects in Bermuda petrel only based on neutral genomic variation and degree of kinship. Functional regions of the genome that highly impact fitness, such as those related to the immune function, should also be accounted to fully understand the relation between inbreeding and mating behaviour.

The exposure of deleterious recessive alleles due to increase of homozygosity under inbreeding can also lead to a natural process of purging of the detrimental genotypes as they become exposed to natural selection. This can be an active process against the effects of inbreeding depression in wild populations, but its effectiveness depends on the rate of increase of the inbreeding level (Hedrick & Garcia-Dorado, 2016). The lack of evidence of inbreeding depression in Bermuda petrel regarding hatching success, a key fitness trait in birds (Blanco & Morinha, 2021) may be due to a past event of severe inbreeding depression, followed by the purging of detrimental alleles as already found in other island bottlenecked species (Laws & Jamieson, 2011).

All in all, the results of the present study show a lack of strong inbreeding depression indicators in Bermuda petrel. Nevertheless, we only looked at a single fitness trait and further research should consider additional traits, as well as the impact of alternative inbreeding avoidance strategies so that the extent of inbreeding depression in this species is thoroughly studied.

3.4.3 - Demographic inference

We used demographic inference analyses to look for a genomic signature of the historical bottleneck that has been described for the Bermuda petrel, as well as to better understand the overall demographic history of this species.

We used two different methods for estimating its demography based on the site frequency spectrum (SFS) derived from our RAD loci. The Stairway Plot results showed a clear prolonged population size decrease starting in the last 100 000 years ago. The beginning of this population decline seems to have occurred in the late Pleistocene, which was characterized by severe sea-level fluctuations due to global glacial periods. It is described that Bermuda Islands endured a severe habitat loss during such fluctuations and a massive sea-level rise occurred between 90 000 and 130 000 years ago during an interglacial period (Harmon et al., 1978). Habitat loss during times of high sea levels may have caused a permanent threat to Bermuda petrel resulting in a slow population decline, which could explain the effective population size (N_e) decline around that time.

Following this slow, steady decline of effective population size possibly related with the glaciation cycles of the Pleistocene, Stairway Plot suggests a much sharper N_e decline between 800 and 2000 y ago. This is earlier than one might expect from a potential population bottleneck driven by the arrival of the first humans to Bermuda, which happened 400 to 500 years ago (Gehrman, 2012). However, this discrepancy in terms of the timing of events might be explained by the fact that current methods of demographic inference based on the SFS derived from RADseq data may not provide extremely accurate estimates of the timing of demographic events (Santiago et al., 2020). This is related to the difficulty of having coalescent events in a relatively short time-frame, particularly for species with high generation times. If we consider that the human arrival to Bermuda occurred in the last 20 to 25 generations of Bermuda petrels, we are attempting to infer the demography of the species in a contemporary time scale, during which there might have not been enough lineage coalescences for proper inference. After this drastic decline, the Stairway Plot analysis suggests a steep N_e increase, which however may not represent an actual population expansion. Since during a genetic bottleneck the coalescence rate is very high, when it eventually ceases, coalescence between lineages is almost null. This mimics a population expansion and a high N_e, even though it is an incorrect signal generated by the drastic change in the coalescence rate following the bottleneck (Hamilton 2009).

The second approach for demographic inference, using fastsimcoal2, suggested the recent bottleneck model as the most likely, against a model with constant population size and a model with an older bottleneck (>30 generations ago). This result reinforces the recent genetic bottleneck hypothesis. However, the likelihood of the models was relatively low, and the fitting of the expected SFS was poor. Further approaches to increase the likelihood of the models should be applied, by considering demographic models better suited for inbred populations (Blischak et al., 2020).

Chapter 4 - **Final Remarks**

The present study aimed to find the genomic signature of a recent genetic bottleneck in Bermuda petrel, an island endemic species, as well as to comprehend its impacts on neutral and adaptive genetic diversity and mating system in order to address the future sustainability of the species. Such objectives are critical since this is a unique small population facing several current threats. Demographic inference analysis, as well as genetic diversity patterns in both neutral nuclear genome and mtDNA support the existence of a relatively recent genetic bottleneck, although we could not associate it exclusively to human arrival to Bermuda. Such genetic bottleneck does not seem to be highly impacting Bermuda petrel regarding inbreeding levels and the fitness component related to hatching success. Still, low genetic diversity in a candidate gene to selection, TLR4, suggests a strong impact of genetic drift. Additionally, mating behaviour does not seem to be a strategy of inbreeding avoidance. Nonetheless, the negative effects of inbreeding on species survival are well described because of inbreeding depression and reduced adaptive potential. The bottleneck experienced by the Bermuda petrel due to human arrival may be too recent (~ 20 generations), and consequently its effects may still not be felt at the present, but it will likely impact future generations. Furthermore, this study revealed the first sign of gene duplication in TLR5 described in birds, although more analysis should be done to confirm it.

Results of this study revealed for the first time genomic information for a significant proportion of the highly endangered Bermuda petrel which is relevant to its conservation strategy. Kinship estimates can complement existing pedigree data and detect possible extra-pair paternities (EPPs). Inbreeding and relatedness degree between individuals can also be accounted for in translocation actions, for example by taking to Nonsuch Island only individuals that are below a certain threshold of relatedness. Although we found no evidence of inbreeding depression, confirmation of a recent genetic bottleneck creates concern on the species sustainability and ability to adapt to present and future challenges. Future studies should focus in acquiring a deeper knowledge of HFCs, by obtaining more information on further fitness traits, as well as assess “local effect” correlations in the Bermuda petrel population. This should provide such information to understand the population’s evolutionary potential and properly plan conservation strategies that should mitigate its vulnerability and slowly decrease the need of human intervention.

References

Alcaide, M., & Edwards, S. V. (2011). Molecular evolution of the toll-like receptor multigene family in birds. *Molecular Biology and Evolution*, *28*(5), 1703–1715. https://doi.org/10.1093/molbev/msq351

Attard, C. R. M., Beheregaray, L. B., & Möller, L. M. (2018). Genotyping-by-sequencing for estimating relatedness in nonmodel organisms: Avoiding the trap of precise bias. *Molecular Ecology Resources*, *18*(3), 381–390. https://doi.org/10.1111/1755-0998.12739

Bartle, J. A. (1968). Observations on the breeding habits of Pycroft's petrel. *Notornis*, *15*(70:99).

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bester, A. J., Priddel, D., Klomp, N. I., Carlile, N., & O'Neill, L. E. (2007). Reproductive success of the Providence Petrel *Pterodroma* solandri on Lord Howe Island, Australia. *Marine Ornithology*, *35*(1), 21–28.

Blanco, G., & Morinha, F. (2021). Genetic signatures of population bottlenecks, relatedness, and inbreeding highlight recent and novel conservation concerns in the Egyptian vulture. *PeerJ*, *9*. https://doi.org/10.7717/peerj.11139

Blischak, P. D., Barker, M. S., Gutenkunst, R. N., & Falush, D. (2020). Inferring the Demographic History of Inbred Species from Genome-Wide SNP Frequency Data. *Molecular Biology and Evolution*, *37*(7), 2124–2136. https://doi.org/10.1093/molbev/msaa042

Boakes, E. H., Wang, J., & Amos, W. (2007). An investigation of inbreeding depression and purging in captive pedigreed populations. *Heredity*, *98*(3), 172–182. https://doi.org/10.1038/sj.hdy.6800923

Bonadonna, F., & Sanz-Aguilar, A. (2012). Kin recognition and inbreeding avoidance in wild birds: The first evidence for individual kin-related odour recognition. *Animal Behaviour*, *84*(3), 509–513. https://doi.org/10.1016/j.anbehav.2012.06.014

Bortoluzzi, C., Bosse, M., Derks, M. F. L., Crooijmans, R. P. M. A., Groenen, M. A. M., & Megens, H. J. (2020). The type of bottleneck matters: Insights into the deleterious variation landscape of small managed populations. *Evolutionary Applications*, *13*(2), 330–341. https://doi.org/10.1111/eva.12872

Bried, J., Andris, M., Dubois, M. P., & Jarne, P. (2021). Decreased selectivity during mate choice in a small-sized population of a long-lived seabird. *Journal of Avian Biology*, 1–13. https://doi.org/10.1111/jav.02837

Bried, J., Pontier, D., & Jouventin, P. (2003). Mate fidelity in monogamous birds: A re-examination of the Procellariiformes. *Animal Behaviour*, *65*(1), 235–246. https://doi.org/10.1006/anbe.2002.2045

Brommer, J. E., Kekkonen, J., & Wikström, M. (2015). Using heterozygosity-fitness correlations to study inbreeding depression in an isolated population of white-tailed deer founded by few individuals. *Ecology and Evolution*, *5*(2), 357–367. https://doi.org/10.1002/ece3.1362

Brooke, M. de L., O'Connell, T. C., Wingate, D., Madeiros, J., Hilton, G. M., & Ratcliffe, N. (2010). Potential for rat predation to cause decline of the globally threatened Henderson petrel *Pterodroma atrata*: Evidence from the field, stable isotopes and population modelling. *Endangered Species Research*, *11*(1), 47–59. https://doi.org/10.3354/esr00249

Brouwer, L., & Griffith, S. C. (2019). Extra-pair paternity in birds. *Molecular Ecology*, *28*(22), 4864–4882. https://doi.org/10.1111/mec.15259

Cammen, K. M., Vincze, S., Heller, A. S., McLeod, B. A., Wood, S. A., Bowen, W. D., Hammill, M. O., Puryear, W. B., Runstadler, J., Wenzel, F. W., Kinnison, M., & Frasier, T. R. (2018). Genetic diversity from pre-bottleneck to recovery in two sympatric pinniped species in the Northwest Atlantic. *Conservation Genetics*, *19*(3), 555–569. https://doi.org/10.1007/s10592-017-1032-9

Carlile, N., Priddel, D., & Madeiros, J. (2012). Establishment of a new, secure colony of Endangered Bermuda Petrel *Pterodroma cahow* by translocation of near-fledged nestlings. *Bird Conservation International*, *22*(1), 46–58. https://doi.org/10.1017/S0959270911000372

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. https://doi.org/10.1111/mec.12354.Stacks

Charlesworth, B., & Charlesworth, D. (1999). The genetic basis of inbreeding depression. *Genetical Research*, *74*(3), 329–340. https://doi.org/10.1017/S0016672399004152

Corporation, G. C. (n.d.). *Sequencher® version 5.4.6*. http://www.genecodes.com

Crow, J.F. and Kimura, M. (1970). *An introduction in Population Genetics Theory*. Harper and Row.

Cumer, T., Pouchon, C., Boyer, F., Yannic, G., Rioux, D., Bonin, A., & Capblancq, T. (2018). Double-digest RAD-sequencing: do wet and dry protocol parameters impact biological results? *BioRxiv*, *33*(0), 445122. https://doi.org/10.1101/445122

Cuthbert, R. (2004). Breeding biology of the Atlantic Petrel, *Pterodroma incerta* , and a population estimate of this and other burrowing petrels on Gough Island, South Atlantic Ocean . *Emu - Austral Ornithology*, *104*(3), 221–228. https://doi.org/10.1071/mu03037

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Dargent, F., Chen, L., Fussmann, G. F., Ghalambor, C. K., & Hendry, A. P. (2019). Female preference for novel males constrains the contemporary evolution of assortative mating in guppies. *Behavioral Ecology*, *30*(3), 646–657. https://doi.org/10.1093/beheco/ary202

Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods*, *9*, 772.

Davey, J. L., & Blaxter, M. W. (2010). RADseq: Next-generation population genetics. *Briefings in Functional Genomics*, *9*(5–6), 416–423. https://doi.org/10.1093/bfgp/elq031

David P, Pujol B, Viard F, Castella V, G. J. (2007). Reliable selfing rate estimates from imperfect population genetic data. *Molecular Ecology*, *16*, 2474–2487.

Del Rey, L. (1964). *To Avenge Man*.

Dias, M. P., Martin, R., Pearmain, E. J., Burfield, I. J., Small, C., Phillips, R. A., Yates, O., Lascelles, B., Borboroglu, P. G., & Croxall, J. P. (2019). Threats to seabirds: A global assessment. *Biological Conservation*, *237*(April), 525–537. https://doi.org/10.1016/j.biocon.2019.06.033

Evans, S. R., & Sheldon, B. C. (2008). Interspecific patterns of genetic diversity in birds: Correlations with extinction risk. *Conservation Biology*, *22*(4), 1016–1025. https://doi.org/10.1111/j.1523-1739.2008.00972.x

Excoffier, L., Marchi, N., Marques, D. A., Matthey-Doret, R., Gouy, A., & Sousa, V. C. (2021). Fastsimcoal2 : Demographic Inference Under Complex Evolutionary Scenarios . *Bioinformatics*, *37*(24), 4882–4885. https://doi.org/10.1093/bioinformatics/btab468

Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*, *28*(3), 544–567. https://doi.org/10.1111/mec.14988

Foster, Y., Dutoit, L., Grosser, S., Dussex, N., Foster, B. J., Dodds, K. G., Brauning, R., Van Stijn, T., Robertson, F., McEwan, J. C., Jacobs, J. M. E., & Robertson, B. C. (2021). Genomic signatures of inbreeding in a critically endangered parrot, the kākāpō. *G3 Genes|Genomes|Genetics*, *11*(11). https://doi.org/10.1093/g3journal/jkab307

Gangloff, B., Zino, F., Shirihai, H., González-Solís, J., Couloux, A., Pasquet, E., & Bretagnolle, V. (2013). The evolution of north-east Atlantic gadfly petrels using statistical phylogeography. *Molecular Ecology*, *22*(2), 495–507. https://doi.org/10.1111/mec.12119

García-Dorado, A., & Caballero, A. (2021). Neutral genetic diversity as a useful tool for conservation biology. *Conservation Genetics*, *22*(4), 541–545. https://doi.org/10.1007/s10592-021-01384-9

Gehrman, E. (2012). *Rare Birds: The Extraordinary Tale of the Bermuda Petrel and the Man Who Brought It Back from Extinction.*

Gilroy, D. L., Phillips, K. P., Richardson, D. S., & van Oosterhout, C. (2017). Toll-like receptor variation in the bottlenecked population of the Seychelles warbler: computer simulations see the 'ghost of selection past' and quantify the 'drift debt.' *Journal of Evolutionary Biology*, *30*(7), 1276–1287. https://doi.org/10.1111/jeb.13077

Gilroy, D. L., van Oosterhout, C., Komdeur, J., & Richardson, D. S. (2017). Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. *Animal Conservation*, *20*(3), 235–250. https://doi.org/10.1111/acv.12307

Grossen, C., Guillaume, F., Keller, L. F., & Croll, D. (2020). Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14803-1

Grueber, C. E., Knafler, G. J., King, T. M., Senior, A. M., Grosser, S., Robertson, B., Weston, K. A., Brekke, P., Harris, C. L. W., & Jamieson, I. G. (2015). Toll-like receptor diversity in 10 threatened bird species: relationship with microsatellite heterozygosity. *Conservation Genetics*, *16*(3), 595–611. https://doi.org/10.1007/s10592-014-0685-x

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2013). Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. *Molecular Ecology*, *22*(17), 4470–4482. https://doi.org/10.1111/mec.12404

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2014). Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0089632

Grueber, C. E., Wallis, G. P., King, T. M., & Jamieson, I. G. (2012). Variation at Innate Immunity Toll-Like Receptor Genes in a Bottlenecked Population of a New Zealand Robin. *PLoS ONE*, *7*(9). https://doi.org/10.1371/journal.pone.0045011

Gummer, H., Taylor, G., Wilson, K. J., & Rayner, M. J. (2015). Recovery of the endangered Chatham petrel (*Pterodroma axillaris*): A review of conservation management techniques from 1990 to 2010. *Global Ecology and Conservation*, *3*, 310–323. https://doi.org/10.1016/j.gecco.2014.12.006

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10). https://doi.org/10.1371/journal.pgen.1000695

Hall, T. A. (1999). BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Hamer, K. C. (2001). Procellariiformes. In *Encyclopedia of Ocean Sciences* (Vol. 4). https://doi.org/10.1016/B978-0-12-374144-8.00218-6

Hamilton, M. B. (2009). Population Genetics. WILEY-BLACKWELL, 111 RIVER ST, HOBOKEN

07030-5774, NJ USA.

Harmon, R. S., Schwarcz, H. P., & Ford, D. C. (1978). Late Pleistocene sea level history of Bermuda. *Quaternary Research*, *9*(2), 205–218. https://doi.org/10.1016/0033-5894(78)90068-6

Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding Inbreeding Depression, Purging, and Genetic Rescue. *Trends in Ecology and Evolution*, *31*(12), 940–952. https://doi.org/10.1016/j.tree.2016.09.005

Helmus, M. R., & Behm, J. E. (2020). Island Biogeography Revisited. In *Encyclopedia of the World's Biomes* (Vol. 1). Elsevier. https://doi.org/10.1016/b978-0-12-409548-9.11945-1

Hemmings, N., West, M., & Birkhead, T. R. (2012). Causes of hatching failure in endangered birds. *Biology Letters*, *8*(6), 964–967. https://doi.org/10.1098/rsbl.2012.0655

Hoffman, J. I., Simpson, F., David, P., Rijks, J. M., Kuiken, T., Thorne, M. A. S., Lacy, R. C., & Dasmahapatra, K. K. (2014). High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(10), 3775–3780. https://doi.org/10.1073/pnas.1318945111

Huisman, J., Kruuk, L. E. B., Ellisa, P. A., Clutton-Brock, T., & Pemberton, J. M. (2016). Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(13), 3585–3590. https://doi.org/10.1073/pnas.1518046113

Imber, M. J., West, J. A., & Cooper, W. J. (2003). Cook's petrel (*Pterodroma cookii*): Historic distribution, breeding biology and effects of predators. *Notornis*, *50*(4), 221–230.

Jones, H. P., & Kress, S. W. (2012). A review of the world's active seabird restoration projects. *Journal of Wildlife Management*, *76*(1), 2–9. https://doi.org/10.1002/jwmg.240

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, *9*(10), 1205–1218. https://doi.org/10.1111/eva.12414

Keith, D. A., & Burgman, M. A. (2004). The Lazarus effect: Can the dynamics of extinct species lists tell us anything about the status of biodiversity? *Biological Conservation*, *117*(1), 41–48. https://doi.org/10.1016/S0006-3207(03)00261-1

Keller, L. F., Jeffery, K. J., Arcese, P., Beaumont, M. A., Hochachka, W. M., Smith, J. N. M., & Bruford, M. W. (2001). Immigration and the ephemerality of a natural population bottleneck: Evidence from molecular markers. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1474), 1387–1394. https://doi.org/10.1098/rspb.2001.1607

Khan, I., Maldonado, E., Silva, L., Almeida, D., Johnson, W. E., O'Brien, S. J., Zhang, G., Jarvis, E. D., Thomas, M., & Antunes, A. (2019). The vertebrate TLR supergene family evolved dynamically by gene gain/loss and positive selection revealing a host-pathogen arms race in birds. *Diversity*, *11*(8). https://doi.org/10.3390/d11080131

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press. https://doi.org/10.1017/CBO9780511623486

Kokko, H., & Ots, I. (2006). When Not To Avoid Inbreeding. *Evolution*, *60*(3), 467. https://doi.org/10.1554/05-613.1

Kumar S, Stecher G, T. K. (2016). MEGA7: molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol Biol Evol*, *33*, 1870–1874.

Lawrence, H.A., Scofield, R.P., Taylor, G.A., Lyver, P.O.B. and Gleeson, D. M. (2016). Genetic, morphological and behavioural evidence supports the treatment of Grey-faced Petrel as a separate species, *Pterodroma gouldi*. *Zoological Journal of the Linnean Society*, *179*(1), 201–216.

Laws, R. J., & Jamieson, I. G. (2011). Is lack of evidence of inbreeding depression in a threatened New Zealand robin indicative of reduced genetic load? *Animal Conservation*, *14*(1), 47–55. https://doi.org/10.1111/j.1469-1795.2010.00388.x

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M. L., Koskiniemi, J., & Vainikka, A. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. *Ecology and Evolution*, *9*(4), 2106–2120. https://doi.org/10.1002/ece3.4905

Levy, H., Fiddaman, S. R., Vianna, J. A., Noll, D., Clucas, G. V., Sidhu, J. K. H., Polito, M. J., Bost, C. A., Phillips, R. A., Crofts, S., Miller, G. D., Pistorius, P., Bonnadonna, F., Le Bohec, C., Barbosa, A., Trathan, P., Raya Rey, A., Frantz, L. A. F., Hart, T., & Smith, A. L. (2020). Evidence of Pathogen-Induced Immunogenetic Selection across the Large Geographic Range of a Wild Seabird. *Molecular Biology and Evolution*, *37*(6), 1708–1726. https://doi.org/10.1093/molbev/msaa040

Levy, H., Fiddaman, S. R., Vianna, J. A., Noll, D., Clucas, G. V., Sidhu, J. K. H., Polito, M. J., Bost, C. A., Phillips, R. A., Crofts, S., Miller, G. D., Pistorius, P., Bonnadonna, F., Le Bohec, C., Barbosa, A., Trathan, P., Rey, A. R., Frantz, L. A. F., Hart, T., & Smith, A. L. (2020). Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird. *Molecular Biology and Evolution*, *37*(6), 1708–1726. https://doi.org/10.1093/molbev/msaa040

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647. https://doi.org/10.1111/1755-0998.12995

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Liu, X., & Fu, Y. X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology*, *21*(1), 1–9. https://doi.org/10.1186/s13059-020-02196-9

Madeiros, J. (2018). *CAHOW RECOVERY PROGRAM For Bermuda's Endangered National Bird 2017 – 2018 Breeding Season Report*. 1–30.

Madeiros, J. L. (2005). *Recovery plan for the Bermuda Petrel (Cahow) Pterodroma cahow*. 90.

Maisonneuve, L., Beneteau, T., Joron, M., Smadi, C., & Llaurens, V. (2020). When do opposites attract? A model uncovering the evolution of disassortative mating. *BioRxiv*, 1–55. https://doi.org/10.1101/2020.05.19.104190

Mays, H. L., Albrecht, T., Liu, M., & Hill, G. E. (2008). Female choice for genetic complementarity in birds: A review. *Genetica*, *134*(1), 147–158. https://doi.org/10.1007/s10709-007-9219-5

Menezes, D, Oliveira, P, Ramirez, I. (2010). *Pterodromas do arquipélago da Madeira. Duas espécies em recuperação*.

Meyerhans, A., Vartanian, J. P., & Wain-Hobson, S. (1990). DNA recombination during PCR. *Nucleic Acids Research*, *18*(7), 1687–1691. https://doi.org/10.1093/nar/18.7.1687

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7). https://doi.org/10.1371/journal.pgen.1002764

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G., & Ellegren, H. (2015). Temporal dynamics of

avian populations during pleistocene revealed by whole-genome sequences. *Current Biology*, *25*(10), 1375–1380. https://doi.org/10.1016/j.cub.2015.03.047

Nandakumar, M., & Ishtiaq, F. (2020). Genetic drift and bottleneck do not influence diversity in Toll-like receptor genes at a small spatial scale in a Himalayan passerine. *Ecology and Evolution*, *10*(21), 12246–12263. https://doi.org/10.1002/ece3.6855

Nava, C., Neves, V. C., Andris, M., Dubois, M. P., Jarne, P., Bolton, M., & Bried, J. (2017). Reduced population size does not affect the mating strategy of a vulnerable and endemic seabird. *Die Naturwissenschaften*, *104*(11–12), 103. https://doi.org/10.1007/s00114-017-1523-z

O'Dwyer, T. W., Buttemer, W. A., Priddel, D. M., & Downing, J. A. (2006). Prolactin, body condition and the cost of good parenting: An interyear study in a long-lived seabird, Gould's Petrel (*Pterodroma leucoptera*). *Functional Ecology*, *20*(5), 806–811. https://doi.org/10.1111/j.1365-2435.2006.01168.x

Paleczny, M., Hammill, E., Karpouzi, V., & Pauly, D. (2015). Population trend of the world's monitored seabirds, 1950-2010. *PLoS ONE*, *10*(6), 1–11. https://doi.org/10.1371/journal.pone.0129342

Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: a road map for stacks. *Methods in Ecology and Evolution*, *8*(10), 1360–1373. https://doi.org/10.1111/2041-210X.12775

Penn, D., & Potts, W. K. (1998). Untrained mice discriminate MHC-determined odors. *Physiology and Behavior*, *64*(3), 235–243. https://doi.org/10.1016/S0031-9384(98)00052-3

Pertoldi, C., Bijlsma, R., & Loeschcke, V. (2007). Conservation genetics in a globally changing environment: Present problems, paradoxes and future challenges. *Biodiversity and Conservation*, *16*(14), 4147–4163. https://doi.org/10.1007/s10531-007-9212-4

Pew, J., Muir, P. H., Wang, J., & Frasier, T. R. (2015). related: An R package for analysing pairwise relatedness from codominant molecular markers. *Molecular Ecology Resources*, *15*(3), 557–561. https://doi.org/10.1111/1755-0998.12323

Podlaszczuk, P., Indykiewicz, P., Markowski, J., & Minias, P. (2020). Relaxation of selective constraints shapes variation of toll-like receptors in a colonial waterbird, the black-headed gull. *Immunogenetics*, *72*(4), 251–262. https://doi.org/10.1007/s00251-020-01156-8

Rayner, M. J., Carraher, C. J. F., Clout, M. N., & Hauber, M. E. (2010). Mitochondrial DNA analysis reveals genetic structure in two New Zealand Cook's petrel (*Pterodroma cookii*) populations. *Conservation Genetics*, *11*(5), 2073–2077. https://doi.org/10.1007/s10592-010-0072-1

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology*, *28*(21), 3487-3494.e4. https://doi.org/10.1016/j.cub.2018.08.066

Rozas J, Sanchez-DelBarrio JC, Messeguer X, R. R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*, 2496–2497.

Sanderson, E. W., Walston, J., & Robinson, J. G. (2018). From Bottleneck to Breakthrough: Urbanization and the Future of Biodiversity Conservation. *BioScience*, *68*(6), 412–426. https://doi.org/10.1093/biosci/biy039

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., & Caballero, A. (2020). Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution*, *37*(12), 3642–3653. https://doi.org/10.1093/molbev/msaa169

Silva, M. C., Matias, R., Wanless, R. M., & Ryan, P. G. (2015). *Understanding the mechanisms of antitropical divergence in the seabird White-faced Storm-petrel (Procellariiformes : Pelagodroma marina) using a multilocus approach*. 3122–3137. https://doi.org/10.1111/mec.13212

Silva, M. F., Smith, A. L., Friesen, V. L., Bried, J., Hasegawa, O., Coelho, M. M., & Silva, M. C. (2016). Mechanisms of global diversification in the marine species Madeiran Storm-petrel Oceanodroma castro and Monteiro's Storm-petrel O. monteiroi: Insights from a multi-locus approach. *Molecular Phylogenetics and Evolution*, *98*, 314–323. https://doi.org/10.1016/J.YMPEV.2016.02.014

Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo, J., Ekblom, R., Bennison, C., Ball, A. D., Beckerman, A. P., & Slate, J. (2010). Adaptation genomics: The next generation. *Trends in Ecology and Evolution*, *25*(12), 705–712. https://doi.org/10.1016/j.tree.2010.09.002

Stephens M, Smith NJ, D. P. (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, *68*, 978–989.

Stoffel, M. A., Esser, M., Kardos, M., Humble, E., Nichols, H., David, P., & Hoffman, J. I. (2016). inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution*, *7*(11), 1331–1339. https://doi.org/10.1111/2041-210X.12588

Szulkin, M., Bierne, N., & David, P. (2010). Heterozygosity-fitness correlations: A time for reappraisal. *Evolution*, *64*(5), 1202–1217. https://doi.org/10.1111/j.1558-5646.2010.00966.x

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(10), 1–10. https://doi.org/10.1073/pnas.2015096118

Thrasher, D. J., Butcher, B. G., Campagna, L., Webster, M. S., & Lovette, I. J. (2018). Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources*, *18*(5), 953–965. https://doi.org/10.1111/1755-0998.12771

Tizard, J., Patel, S., Waugh, J., Tavares, E., Bergmann, T., Gill, B., Norman, J., Christidis, L., Scofield, P., Haddrath, O., Baker, A., Lambert, D., & Millar, C. (2019). DNA barcoding a unique avifauna: An important tool for evolution, systematics and conservation. *BMC Evolutionary Biology*, *19*(1), 1–13. https://doi.org/10.1186/s12862-019-1346-y

Townsend, S. M., & Jamieson, I. G. (2013). Molecular and pedigree measures of relatedness provide similar estimates of inbreeding depression in a bottlenecked population. *Journal of Evolutionary Biology*, *26*(4), 889–899. https://doi.org/10.1111/jeb.12109

Trask, A. E., Ferrie, G. M., Wang, J., Newland, S., Canessa, S., Moehrenschlager, A., Laut, M., Duenas, L. B., & Ewen, J. G. (2021). Multiple life-stage inbreeding depression impacts demography and extinction risk in an extinct-in-the-wild species. *Scientific Reports*, *11*(1), 1–10. https://doi.org/10.1038/s41598-020-79979-4

Újvári, B., Madsen, T., Kotenko, T., Olsson, M., Shine, R., & Wittzell, H. (2002). Low genetic diversity threatens imminent extinction for the Hungarian meadow viper (Vipera ursinii rakosiensis). *Biological Conservation*, *105*(1), 127–130. https://doi.org/10.1016/S0006-3207(01)00176-8

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 1–12. https://doi.org/10.1093/nar/gks596

Velando, A., Barros, Á., & Moran, P. (2015). Heterozygosity-fitness correlations in a declining seabird population. *Molecular Ecology*, *24*(5), 1007–1018. https://doi.org/10.1111/mec.13092

Velová, H., Gutowska-Ding, M. W., Burt, D. W., & Vinkler, M. (2018). Toll-like receptor evolution in birds: Gene duplication, pseudogenization, and diversifying selection. *Molecular Biology and Evolution*, *35*(9), 2170–2184. https://doi.org/10.1093/molbev/msy119

Ventura, F., Granadeiro, J. P., Padget, O., & Catry, P. (2020). Gadfly petrels use knowledge of the windscape, not memorized foraging patches, to optimize foraging trips on ocean-wide scales. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1918).

https://doi.org/10.1098/rspb.2019.1775

Wang, J. (2011). COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. *Molecular Ecology Resources*, *11*, 141–145.

Wang, J. (2002). An estimator for pairwise relatedness using molecular markers. *Genetics*, *160*(3), 1203–1215. https://doi.org/10.1093/genetics/160.3.1203

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*(3), 773–777. https://doi.org/10.1093/molbev/msx335

Wedekind, C., & Furi, S. (1997). Body odour preferences in men and women: Do they aim for specific MHC combinations or simply heterozygosity? *Proceedings of the Royal Society B: Biological Sciences*, *264*(1387), 1471–1479. https://doi.org/10.1098/rspb.1997.0204

Wickham H. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag New York*. https://ggplot2.tidyverse.org

Wood, J. R., Alcover, J. A., Blackburn, T. M., Bover, P., Duncan, R. P., Hume, J. P., Louys, J., Meijer, H. J. M., Rando, J. C., & Wilmshurst, J. M. (2017). Island extinctions: Processes, patterns, and potential for ecosystem restoration. *Environmental Conservation*, *44*(4), 348–358. https://doi.org/10.1017/S037689291700039X

Zapata, D., Rivera-Gutierrez, H. F., Parra, J. L., & Gonzalez-Quevedo, C. (2020). Low adaptive and neutral genetic diversity in the endangered Antioquia wren (Thryophilus sernai). *Conservation Genetics*, *21*(6), 1051–1065. https://doi.org/10.1007/s10592-020-01313-2

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

Supplementary material

Table S 1 - Description of primers used for TLR4 and TLR5 amplification

Gene/Exon	Primer name	Primer label	Sequence (5'-3')	Reference
TLR4 / Exons 1 and 2	PengTLR4_Ex1_F2	T41F	CTCACTCCAGGTCCTTGAGTG	Levy et al., 2020
	PengTLR4_Ex2_R1	T42R	AGCCAAGACCAACAGACACC	Levy et al., 2020
TLR4 / Exon 3 - Start	PengTLR4_Ex3_Start_F	T43SF	AAGGGACAGTGTTGCATGCA	Levy et al., 2020
	cahow_tlr4_ex3_start_R	cT4-3SR	TCGAATTTCTGCCGGAAGCT	Present thesis
TLR4 / Exon 3 - Middle	avTLR4F	avT4F	GAGACCTTGATGCCCTGAG	Alcaide & Edwards, 2011
	avTLR4R	avT4R	CCATCTTRAGCACTTGCAAAG	Alcaide & Edwards, 2011
TLR4 / Exon 3 - End	cahow_tlr4_ex3_end_F	cT4-3EF	TTCACTAGCTGCTGTTCCCC	Present thesis
	PengTLR4_Ex3_EndR2	T43ER	TGCCTTCTAGCAGGACTCCT	Levy et al., 2020
TLR5 - A	TLR5_full_new_F	T5AF	AGTGTGACATTAGTCTTTTGTTCTACA	Levy et al., 2020
	TLR5_1000_R	T5AR	AGCCAAATCCCGAACCCATT	Levy et al., 2020
TLR5 - B	cahow_tlr5_B_F	cT5BF	AGTGACAATGGCTGGAGCTC	Present thesis
	cahow_tlr5_B_R	cT5BR	TGAATCCAGGCGCTTTAGCA	Present thesis
TLR5 - C	Gentoo_TLR5_sequencing_3_F	T5CF	TCCCCTGAGCCTGAAGTCTT	Levy et al., 2020
	Gentoo_TLR5_full_R	T5CR	GTGTTCATTCCTGCCATGGC	Levy et al., 2020

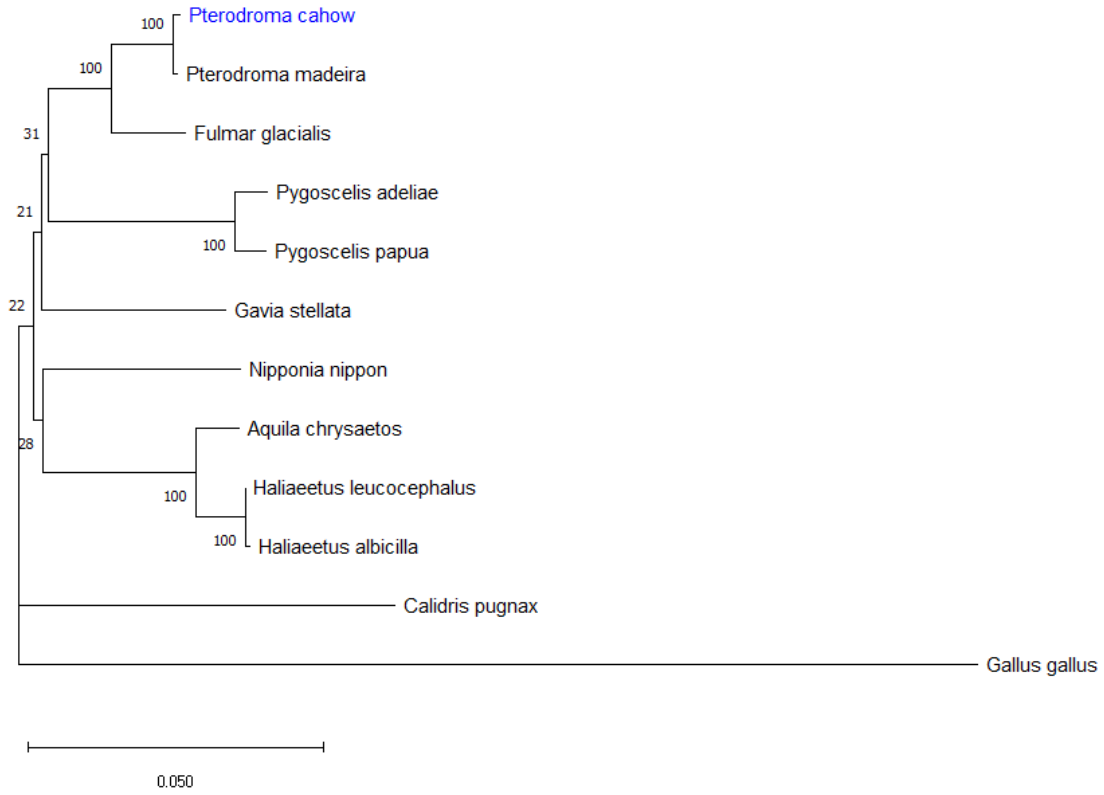
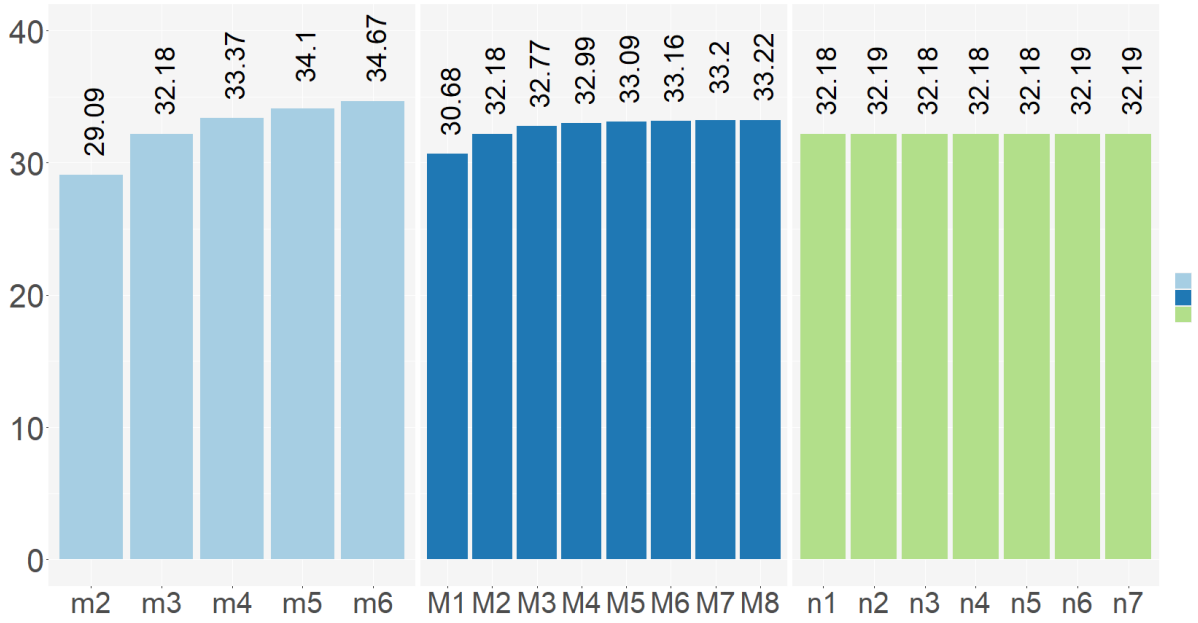


Figure S 1 - Maximum likelihood tree of ten avian species, used for detection of sites under selection. *Calidris pugnax* and *Gallus gallus* are the outgroups

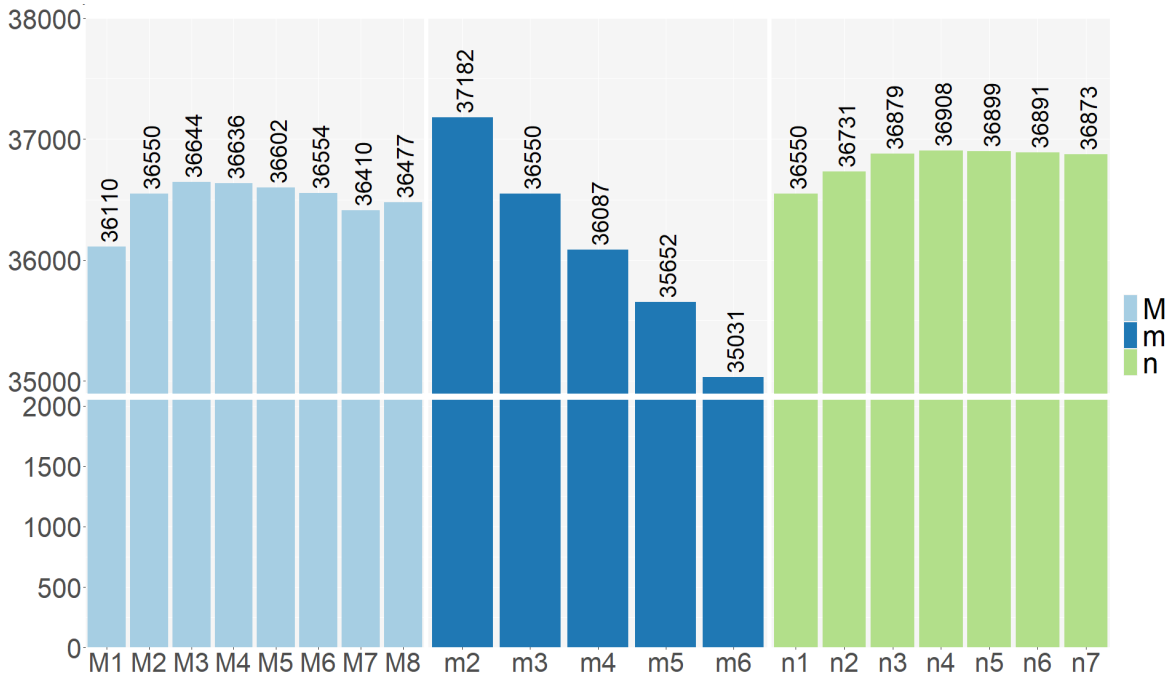
Supplementary Method S1 – Detailed description of parameter choice for Stacks assembly

The m parameter was set to 3. As expected, the number of loci and polymorphism decreased, and coverage increased with an increasing level of minimum stack depth (*m*) was. The highest number of loci and polymorphism was found when m=2 (Figure S 2), but the difference between *m*=2 and *m*=3 was very low. Therefore, the difference in coverage between these two values (Figure S 2 A) dictated the final choice for m. Parameter *M* is particularly dataset-specific, depending on the suspected degree of polymorphism of the studied population (Paris et al., 2017). Polymorphic loci peaked when *M*=4 (Figure S 2 C), although without a significant gap from *M*=3. On the other hand, the number of SNPs between *M*3 and *M*4 have a rather significant divergence (Figure S 3), which can bring us to the assumption of an increment of few loci with many SNPs. Since these new loci don't reveal high coverage (Figure S 2 A), are likely erroneously assembled loci instead of true loci. Besides, the known history of this species makes it likely to show lower polymorphism. *M*3 was thus chosen, which also comprehended a conservative option over *M*2, preventing the merging of distinct true loci. Finally, selecting the optimal value for n evolves a compromise between setting it too low and fail to merge homologous loci with fixed SNPs or setting it to high and merge false homologous loci, assuming false polymorphism (Paris et al., 2017). Once Again, deducing that this species should rather have lower genetic differences among individuals, combined with the small variations in the analysed metrics, *n* was conservatively set to 3.

A - Coverage



B – Assembled loci



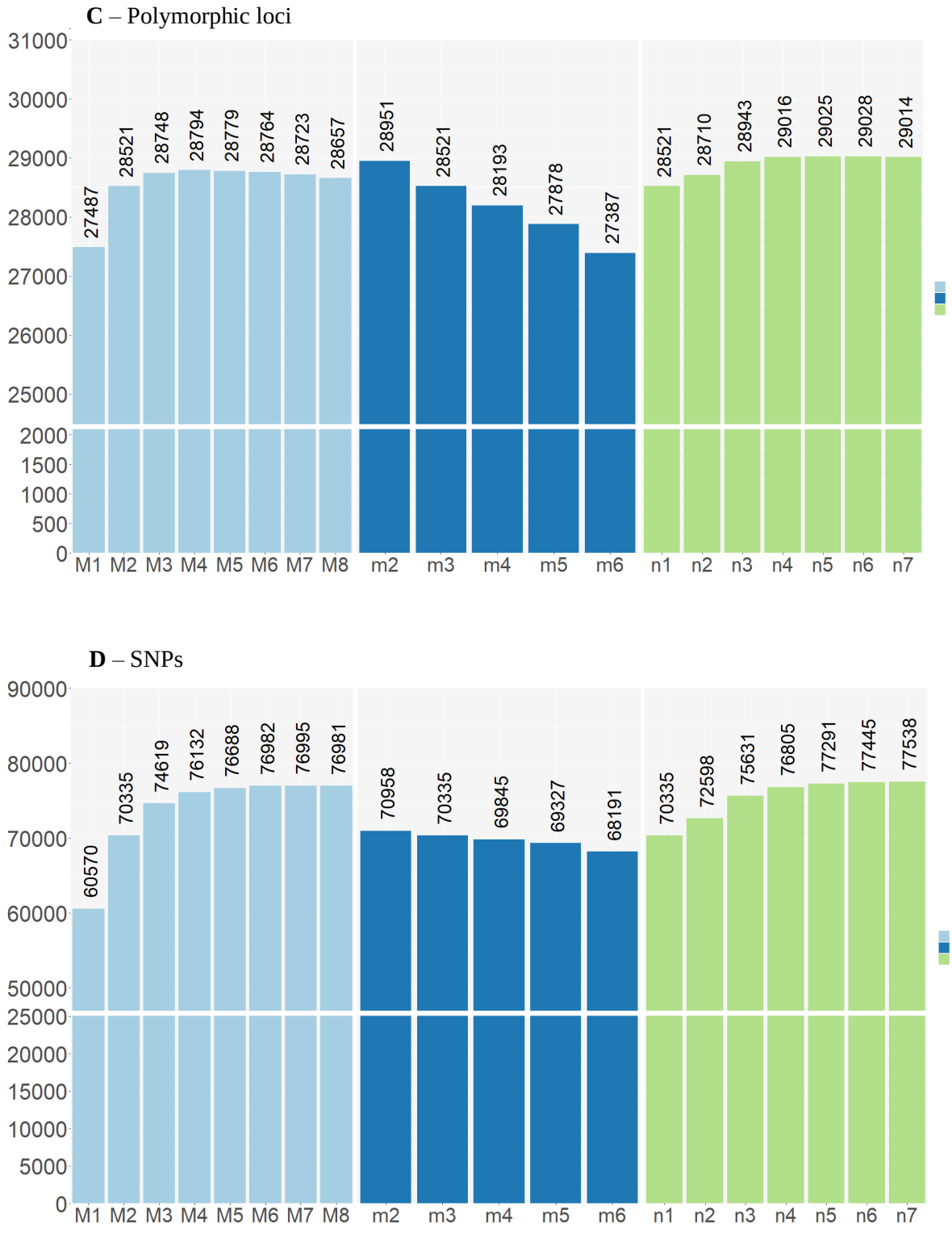


Figure S 2 - Results from optimization of Stacks's pipeline parameters. Colours represent the three parameters tested, having minimum number of raw reads required to form a putative allele (*m*) as dark blue, maximum number of mismatches accepted between alleles to consider them the same locus (*M*) as light blue and maximum number of mismatches allowed between loci of different individuals to be considered the same loci (*n*) as green. Each plot regards to the three metrics tested: A – coverage, B - total number of assembled loci, C - total number of polymorphic loci and D - total number of SNPs.

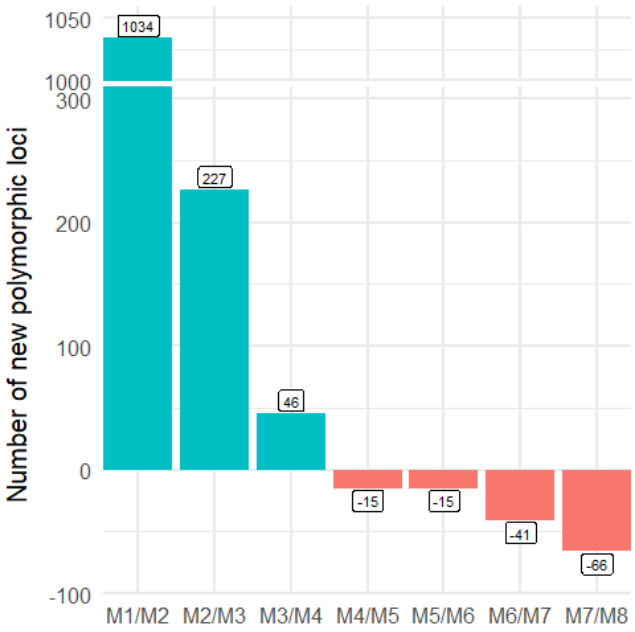


Figure S 3 - The number of new polymorphic loci (*r80 loci*) was the parameter M increased. Blue represents an increase of number of polymorphic loci when transitioning between the referred M values, while pink represents a decline.

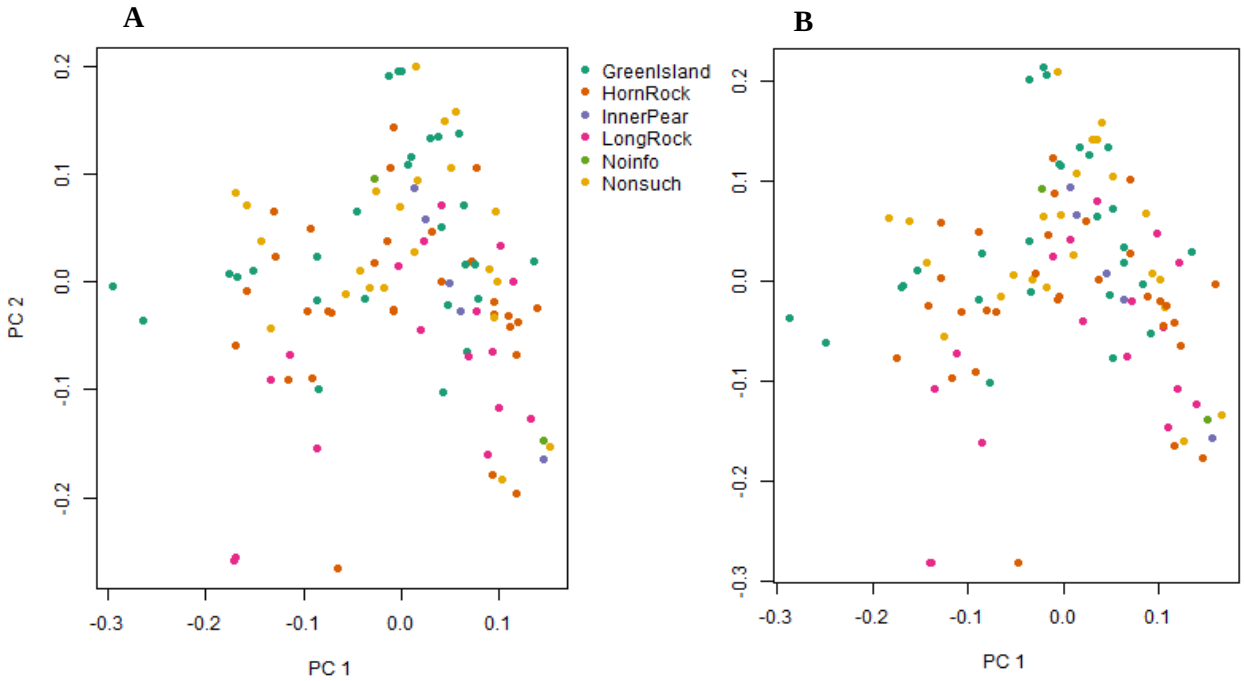


Figure S 4 - Principal Component Analysis (PCA) regarding two different MAF filter values. A –MAF < 0.05. B – MAF < 0.1. Different colours represent the island on which the individual is currently breeding.

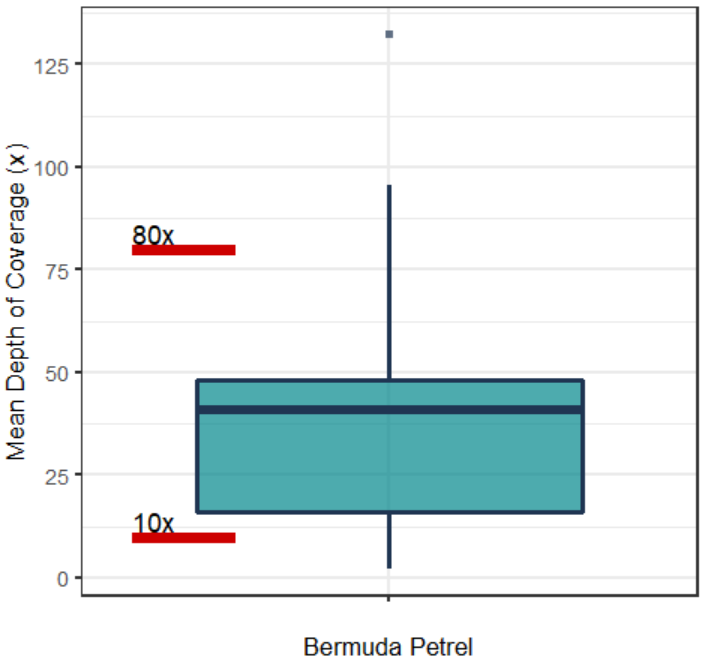


Figure S 5 - Boxplot with the mean depth of coverage per raw loci resulted from Stacks assembly, measured in times (x). Red lines represent the minimum and maximum depth of coverage filtered, as described in 3.2.1.2.2.

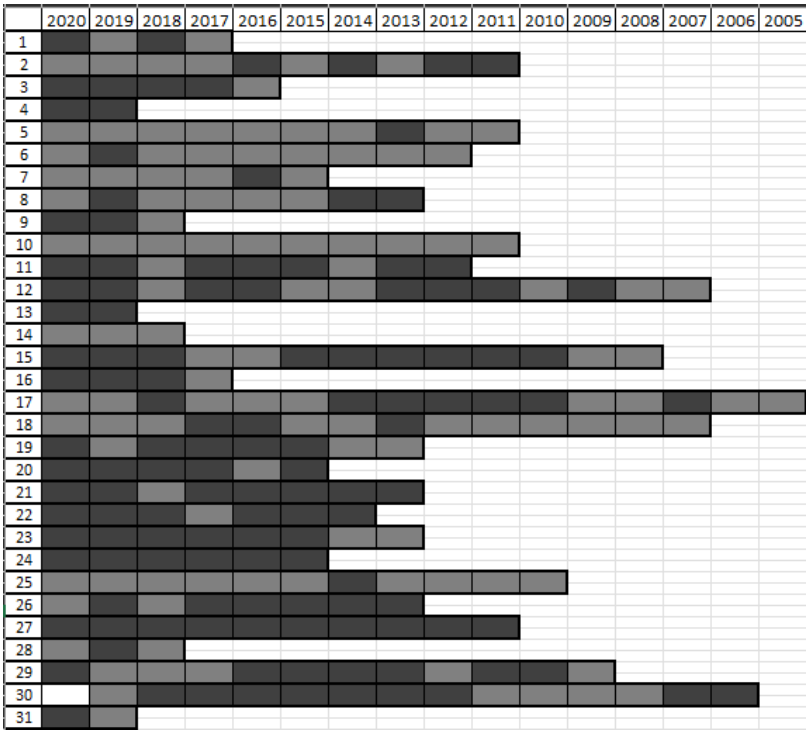


Figure S 6 - Hatching success information for the 31 couples. Dark grey squares represent successful hatches, while light grey represent failed hatches.

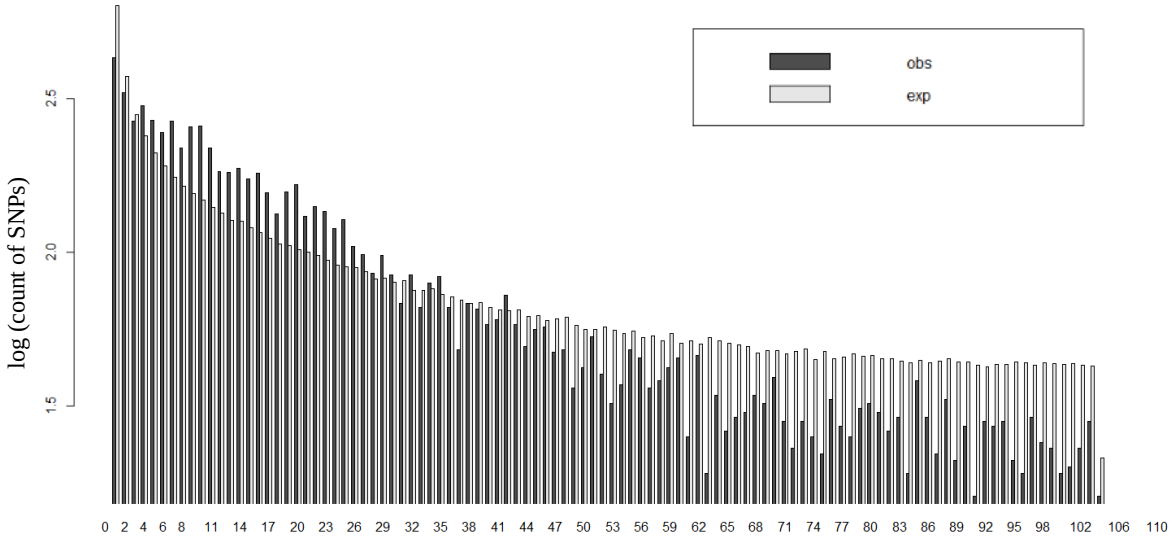


Figure S 7 - Barplot exposing the observed (dark bars) and expected (light bars) SFS resulted from fastsimcoal2 analysis.