CONSERVATION GENETICS OF THE YELKOUAN SHEARWATER (*Puffinus yelkouan*) IN THE MEDITERRANEAN SEA

by

Dilek Şahin

B.S. in Biology, Çanakkale Onsekiz Mart University, 2009

M.Sc. in Forest Entomology and Conservation Program, Istanbul University, 2012

Submitted to the Institute of Environmental Sciences in partial fulfilment of

the requirements for the degree of

Doctor of Philosophy

in

Environmental Sciences

Boğaziçi University

2024

CONSERVATION GENETICS OF THE YELKOUAN SHEARWATER (*Puffinus yelkouan*) IN THE MEDITERRANEAN SEA

APPROVED BY:

Prof. Dr. Raşit Bilgin . . . . . .
Dissertation Advisor

Prof. Dr. Andrzej Furman . . . . . .

Assoc. Prof. Dr. Ulaş Tezel . . . . . .

Asst. Prof. Dr. Korhan Özkan . . . . . .

Asst. Prof. Dr. İsmail Sağlam . . . . . .

DATE OF APPROVAL: 15/01/2024

ACKNOWLEDGEMENTS

My Ph.D. journey was a long one. There are so many people who directly or indirectly helped me in this journey. I cannot fit the list of all the names here but would like to mention the ones who nudged me to take the leap of faith in the most critical moments.

First, I am grateful for my thesis advisor, **Dr. Raşit Bilgin**, for being open to ideas, for being a facilitator and supportive in all aspects of this journey. Thank you for choosing to be a collaborator! Thank you, thesis committee members, especially **Dr. İsmail Sağlam** for providing valuable feedback on my research along the way. I cannot thank enough **Dr. Karen Bourgeois** and the **Hellenic Ornithological Society** team led by **Danae Portolou** for providing the blood samples, which were the basis of this work. Many thanks also to the researchers who spent many nights in the cliffs to collect these samples. The numerical calculations reported in this thesis were fully performed at **TUBITAK ULAKBIM High Performance and Grid Computing Center** (TRUBA resources). I am grateful for the TRUBA team for providing free resources to graduate students and facilitating scientific research in the country! Part of this study was supported by a grant from **The Scientific and Technological Research Council of Türkiye** TUBITAK (No: 120N122) to RB.

Thank you, **Dr. Steffen Oppel**, for dedicating your time to help me produce a research proposal and for introducing me to your network of great seabird scientists.

Big thanks to the whole **Conservation Leadership Programme** team but especially **Maureen Ryan** and **Christina Imrich** for your mentorship and all the training opportunities you provided that helped me know myself better!

Taha Akkuş, thank you for being the biggest supporter in rain and shine! Thank you for providing me with great resources that helped me to hold on to this journey. I could not do this without your support!

My friends, **Sezgi, Leyla, İso, Melih, Haydar** and **Yahya**, thank you for the mental and financial support you provided all along this journey! Seeing your respect for my work made me believe in myself more. Thank you, my dear friend **Ayça Eleman**, for introducing me to a very critical person in 2020! My lab mate **Dr. Merve Sıvacı**, I am so grateful for the endless discussions we had in the lab. Thank you for teaching me the nuts and bolts of lab work and for completing my experiments

when needed. Big thanks to my highly-skilled friend **Tora Benzeyen**, for producing the artwork for this thesis without any hesitation!

Thank you, two brilliant scientists, **Matthew DeSaix** and **Dr. Matthew Kling** for allocating your valuable time to help me improve the genetic assignment and seascape genomics analyses, respectively. I am grateful for the discussions and for the programs you developed!

Mahalo nui loa to my family on Kauaʻi, **Mike McFarlin**, for being a great companion in the most critical phase of this journey, for hours of help in installing poorly instructed programs and for facilitating the "big data crisis" in seascape genomics analyses. You still owe me so many dinners though! Thanks to the two fat cats of this family, **Bully** and **Narley**, for being major stress absorbers during the final phase (and of course for sticking to your gravy and shreds instead of killing seabirds).

The biggest thanks are to **my mother, Münevver,** and to **my brother, Ümit**. You always supported me in achieving my goals, in my decisions of leaving countless good non-seabird conservation jobs or moving to the other side of the world for seabird conservation. Thank you, mom, for encouraging me for the things you yearned to achieve but couldn't, for always providing me with a safe place to return. I could not be where I am today without such unconditional support. If I succeed a little in advancing seabird conservation in the world, I am dedicating it to my dear family.

Long live Yelkies, long live seabirds!

ABSTRACT

CONSERVATION GENOMICS OF THE YELKOUAN SHEARWATER (*Puffinus yelkouan*) IN THE MEDITERRANEAN SEA

Seabird populations globally face decline from threats on breeding grounds and at sea. Seabirds form spatially structured populations and limited understanding of population connectivity hinders their conservation. Genomic approaches, generally underutilised in seabird conservation, allows to assess population connectivity and the driving historical and contemporary factors and aids in identifying relevant conservation units. This research focuses on the Yelkouan Shearwater, a Mediterranean-endemic seabird facing conservation challenges. Employing next generation sequencing and diverse analytical methods, this study unveils, for the first time, the population genomic structure of the Yelkouan Shearwater. It also evaluates contemporary factors influencing this structure and tests its utility for the conservation of the Yelkouan Shearwater. The results of this dissertation highlight the pivotal role of comprehensive range-wide sampling and high-resolution sequencing data in unravelling the population genomic structure of seabirds. It also underscores the intricate relationship between seascape features and seabird genomic structure. Given the high pace of global change and biodiversity loss, these findings should inform future seabird research and conservation management, extending their relevance beyond the Mediterranean.

ÖZET

YELKOVAN KUŞLARININ (*Puffinus yelkouan*) AKDENİZ'DE KORUMA GENETİĞİ

Deniz kuşu popülasyonları üreme alanlarındaki ve denizdeki birçok tehditten ötürü küresel ölçekte azalmaktadır. Deniz kuşları mekânsal yapılanma gösteren popülasyonlar halinde yaşarlar ve bu popülasyonlar arasındaki bağlantıların yeteri kadar anlaşılmamış olması koruma çalışmalarına ket vurmaktadır. Deniz kuşu koruma çalışmalarında henüz yeteri kadar kullanılmayan genomik yöntemler popülasyonlar arasındaki bağlantıların, bu bağlantıları oluşturan tarihi ve güncel faktörlerin anlaşılmasına ve buna bağlı olarak ilgili koruma ünitelerinin belirlenmesine izin verir. Bu araştırma Akdeniz'e endemik ve nesli tehlike altında olan bir deniz kuşuna, Yelkovanlara odaklanmaktadır. Yeni nesil dizileme yöntemleri ve çeşitli analitik yöntemleri kullanarak bu çalışma Yelkovan kuşlarının popülasyon genomik yapısını ilk defa ortaya koymaktadır. Bu çalışma aynı zamanda Yelkovanların popülasyon genomik yapısını etkileyen güncel faktörleri incelemekte ve bu genomik yapının türün koruma çalışmalarında kullanılabilirliğini test etmektedir. Bu çalışma ile elde edilen bulgular, türün dağılım alanını kapsayan örneklemenin ve yüksek çözünürlüklü dizileme verisinin deniz kuşlarında popülasyon genomik yapısının anlaşılmasında son derece önemli olduğunu göstermektedir. Bulgular aynı zamanda deniz ekosistemi bileşenleri ile deniz kuşu popülasyon genomik yapısı arasındaki ilişkinin karmaşıklığını da vurgulamaktadır. Küresel değişimin ve biyolojik çeşitlilik kayıplarının hızı düşünüldüğünde bu çalışmada elde edilen bulguların Akdeniz ve diğer bölgelerde gelecekteki deniz kuşu araştırma ve koruma çalışmalarında kullanışlı alacağı beklenmektedir.

TABLE OF CONTENTS

ABSTRACT ……...V
TABLE OF CONTENTS ……VII
LIST OF FIGURES ……...IX
LIST OF TABLES ……X
LIST OF SYMBOLS/ABBREVIATIONS ……...XI
1. INTRODUCTION ……..1
1.1. The Yelkouan Shearwater: An Endangered Mediterranean Endemic Species.....3
1.2. Population Genetics of the Mediterranean Seabirds.....4
1.3. Seabird Population Genomics in the Light of Isolation-by-Resistance.....6
1.4. Using Genomics for the Conservation of Seabirds 9
1.5. Research Objectives and Methodological Overview.....11
2. MATERIALS AND METHODS 13
2.1. Ethical Statement.....13
2.2. Sample Collection and Markers Used 13
2.3. DNA Extraction and Sequencing 14
2.4. Bioinformatics and Data Analyses 16
2.4.1. Population Genomics Data Analysis.....17
2.4.2. Seascape Genomic Data Analysis.....19
2.4.3. Genetic Assignment Data Analysis.....22
3. RESULTS ……25
3.1. Population Genomic Structure of the Yelkouan Shearwater.....25
3.1.1. mtDNA.....25
3.1.2. RAD-Seq Data 27
3.2. Contemporary Environmental Factors Influencing the Genomic Structure of the Yelkouan Shearwater in the Mediterranean Seascape 31
3.3.1. Estimated Effective Migration Surfaces 31
3.3.2. Isolation-by-Distance and Isolation-by-Resistance 32
3.2.3. Windscape as Resistance to Gene Flow.....33
3.3. The Utility of Assignment Tests in Weakly Structured Yelkouan Shearwater Population...34
3.3.1. Leave-one-out Cross Validation 35
3.3.2. Assignment of Unknown-origin Samples.....37
4. DISCUSSION ……...39

4.1. ddRAD Data Reveals Fine-scale Population Genomic Structure 39
4.2. Seascape Habitat Configuration as Resistance to Gene Flow 43
4.3. Utility of Assignment Tests in Weakly Structured Yelkouan Shearwater Metapopulation..45
5. CONCLUSIONS AND RECOMMENDATIONS ……...50
REFERENCES ……..52
APPENDIX ……...71

LIST OF FIGURES

Figure 2.1. The distribution of Yelkouan Shearwater breeding colonies and sampling locations …….....14

Figure 3.1. Haplotype network of samples grouped according to PCA and admixture results …….26

Figure 3.2. Principal Component Analysis plots showing the projection in the first four components ……...28

Figure 3.3. Admixture plots for K2 to K8 calculated from genotype likelihoods …….29

Figure 3.4. Colony-based nucleotide diversity, Tajima's D, and heterozygosity ……...30

Figure 3.5. Genetic differentiation statistics (Fst) between each pair of colonies and correlation between genetic distance (Fst) and geographical distance calculated as least cost path over the water ……31

Figure 3.6. Barriers to gene flow and regions of sustained genetic connectivity in the Yelkouan Shearwater metapopulation …….32

Figure 3.7. Windscape of the Mediterranean estimated from 30 years-long wind data …….34

Figure 3.8. The relationship between windscape parameters and distance for colony locations …..34

Figure 3.9. Distribution and frequency of assignments in leave-one-out cross validation tests ……..36

LIST OF TABLES

Table 2.1. Sequences of adapters and primers used in ddRAD library preparations…..…….....16

Table 2.2. Barcode and index sequences used in ddRAD library preparation……..…..…….....16

Table 3.1. Distribution and number of samples according to colonies and markers used in this study ……...25

Table 3.2. Summary statistics of mtDNA sequences ……27

Table 3.3. Colony-based summary of nucleotide diversity and Tajima's D values ……29

Table 3.4. The details of datasets used and results of Mantel tests applied ……32

Table 3.5. Resistance model selection results ……33

Table 3.6. Details of different subsets used and results of leave-one-out assignment tests ……35

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation
µl	Microliter
$\chi 2$	Chi-squared Test
°C	Degree Celsius
Fst	Fixation Index
nM	nanoMolar
ng	Nanogram

Abbreviation	Explanation
AFLP	Amplified Fragment Length Polymorphism
BAM	Binary Alignment Map
bp	Base Pair
Cytb	Cytochrome b Gene
DNA	Deoxyribonucleic Acid
ddRAD	Double Digest Restriction-site Associated DNA
EDTA	Ethylenediaminetetraacetic Acid
EEMS	Estimating Effective Migration Surfaces
IUCN	International Union for Conservation of Nature
MCMC	Markov Chain Monte Carlo
mtDNA	Mitochondrial DNA
MYA	Million Years Ago
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
SNP	Single Nucleotide Polymorphism
qPCR	Quantitative PCR



Image courtesy of Ercan Üç

1. INTRODUCTION

Life-history theory attempts to explain the relationship between traits affecting species' reproduction, survival and population growth and broadly classifies species on a fast-slow turnover gradient; some species have low fecundity but high survival, others have high fecundity but lower survival. Seabird survival depends on marine resources, and the life-history of seabirds has been greatly shaped by this dependence (Hamer et al., 2001). The availability and abundance of food in the marine environment is one of the main factors influencing breeding frequency, clutch size, breeding success and growth rate in seabirds. It is suggested that dependency on patchy and dynamic marine resources causes seabirds to invest in developing the ability of finding reliable food sources that would maximise offspring survival (Weimerskirch, 2001). As a result, most seabird species exhibit delayed sexual maturation, spending their first few years at the sea discovering. They also exhibit low annual fecundity, biparental care, and an extended chick-rearing period, which leads to low population growth rate and slow response to changes due to slow turnover. These life-history traits make seabirds vulnerable to global change as populations are unable to respond to the rapid changes in habitat availability and suitability, predation, or prey abundance. Indeed, seabirds are one of the most threatened groups of birds (Croxall et al., 2012; Dias et al., 2019) due to variety of threats operating in both terrestrial and marine environments including but not limited to climate change, overfishing, fisheries bycatch, diseases, depredation of eggs and chicks by introduced species, habitat loss and light pollution.

Protecting seabird populations is important as seabirds are an essential part of coastal and island ecosystems. Seabirds act as ecological connectors, linking marine and terrestrial ecosystems by nesting, foraging, and migrating. They play a vital role by bringing essential marine nutrients to land and contributing to the complex dynamic relationships in these ecosystems (e.g., Benkwitt et al., 2022; Liu et al., 2023). A meta-analysis conducted by Grant et al. (2022) concluded that seabird guano increased the nutrient level in colony locations when compared to the soil in control locations. Moreover, increasing number of studies show that nutrient-transfer by seabirds also enhance coastal marine habitats such as coral reefs (Benkwitt et al., 2023). Besides being nutrient engineers, seabirds are also considered good indicators for the health of marine environment due to their dependency to marine resources (Cairns, 1987; Frederiksen et al., 2007; Piatt et al., 2007). Considering their role in marine ecosystem and accessibility for monitoring and research, seabirds have been integrated into several marine ecosystem management policies as policy-relevant ecosystem indicators (e.g.,

OSPAR, HELCOM, MSFD, and UKMMAS). However, in practice seabirds, particularly those belonging to the order of Procellariiformes (albatrosses, petrels, shearwaters and storm-petrels) are a challenge for population ecology and conservation studies.

Procellariiform seabirds are wide-ranging species, spending most of their life at sea and nesting on remote islands that may be located in different political boundaries. They have spatially structured populations where subpopulations are connected at various spatial scales and driven by complex dynamics (e.g., Dayton & Szczys, 2021; Jeglinski et al., 2023). Due to delayed sexual maturity in seabirds, a large non-breeding segment of the population (sometimes >50% of the breeding population (Klomp & Furness, 1992) wanders around breeding sites. Several factors might influence the recruitment of these prospecting individuals at a certain colony such as philopatry, density dependence and conspecific attraction (Fay et al., 2015; Gauthier et al., 2010; Inchausti & Weimerskirch, 2002). In addition, a wide variety of threats and quickly changing habitat quality can cause further shifts in recruitment in colonies (Horswill et al., 2022). Eventually, these interactions impact the population structure. The form and importance of population structure is one of the research priorities for seabirds (Lewison et al., 2012) and the basis for improved understanding of seabird population regulation and response to the changing environment. Furthermore, understanding the structure allows for identifying significant conservation units within the metapopulation as an important first step in the effective management of seabird populations.

Population genomics is a powerful and growing approach for this task and provides new insights into demographic history, population structure, adaptation, and the consequences of genetic diversity (Barbosa et al., 2020). Information obtained from population genomics helps delineate units for conservation, maintain genetic diversity in wild populations, and predict adaptive potential. With the advancement of sequencing technologies and analytical tools, genomics became one of the essential approaches in studying non-model organisms (Ellegren, 2014). Until recently, the literature of seabird studies using DNA sequencing was focusing on low resolution DNA markers such as mtDNA, microsatellites and AFLPs. These studies addressed questions on phylogenetics (Austin et al., 2004; Martínez-Gómez et al., 2015) taxonomic status (Heidrich et al., 1998; Penhallurick & Wink, 2004), population differentiation and hybridizations (Brown et al., 2010). Increasing number of studies using reduced genomic representation or low coverage whole genome sequencing approaches are revealing fine-scale and complex population genomic structures of seabird populations and challenging the results of previous studies carried out using single gene markers (Byerly et al., 2023; Cristofari et al., 2019; Kersten et al., 2021; Rexer-Huber et al., 2019). In this context, the Mediterranean seabird population genomics and its integration into seabird population management continue to be a gap in

the knowledge and practice, respectively. This thesis aims to contribute to filling this gap for an endemic and endangered Mediterranean seabird species; the Yelkouan Shearwater. The objectives of this thesis are to understand the population genomic structure of the Yelkouan Shearwaters, the influence of habitat configuration on this structure and the utility of this structure in the applied conservation of the species.

1.1. The Yelkouan Shearwater: An Endangered Mediterranean Endemic Species

Yelkouan Shearwater is a Mediterranean-endemic pelagic seabird, once considered the sub-species of its Atlantic counterpart Manx Shearwater (*Puffinus puffinus*) and separated from its Mediterranean sibling Balearic Shearwater (*Puffinus mauretanicus*) and Manx Shearwater recently (Sangster et al., 2002). Yelkouan Shearwater occurs in a few large and many small breeding colonies on the Mediterranean Islands (BirdLife International, 2023; Bourgeois & Vidal, 2008). Current population is estimated at 15,337-30,519 breeding pairs (Derhé, 2012).

Yelkouan Shearwater is a middle-size and coastal shearwater, depending on seasonality in the Mediterranean. Breeding starts in January and spans until June (Bourgeois & Vidal, 2008); this phenology is appeared to be a month later in the Eastern Mediterranean (Portolou, pers. com.). Breeding birds show a high degree of site fidelity, returning to the same nest in consecutive years. Females lay a single egg per season and both parents invest in rearing the chick. After fledged, young birds spend their time at sea until they reach breeding maturity in a few years (Sultana et al., 2011). Breeding birds utilize the areas close to breeding colonies for foraging during the breeding season (Gatt et al., 2019; Péron & Grémillet, 2013; Pezzo et al., 2021) but use a much wider area in the Mediterranean and Black Seas during the non-breeding season (Militão et al., 2013; Raine et al., 2013).

The IUCN status of the Yelkouan Shearwater is Vulnerable (VU), with the major threats operating at almost all known colonies are the predation of eggs and chicks by invasive mammals, mortality due to fisheries by-catch, overfishing, climate change, habitat loss and light pollution (BirdLife International, 2023). The population is declining rapidly in colonies where threats are widespread (Oppel et al., 2011) and some other colonies remain stable despite the presence of high predation by cats (Bonnaud et al., 2009), an indication of the influence of immigration on population trends. Conservation effort for the Yelkouan Shearwater is not evenly distributed in the Mediterranean due to limited access, funding, and human resources. There are significant gaps in the

knowledge on species distribution and breeding status, especially in the Southern and Eastern Mediterranean.

The current conservation actions focus particularly on increasing the breeding success by removing invasive mammals from the breeding colonies and increasing survival by monitoring and mitigating fisheries bycatch. Conservation organizations like BirdLife recently started to invest subregion-scale coordinated conservation actions to protect the metapopulation (e.g. LIFE PanPuffinus Project, https://lifepanpuffinus.org/). Yet, population genomic structure of the Yelkouan Shearwater to evaluate the geographic scale of conservation strategies remain to be a knowledge gap. Moreover, even taxonomy within the species remains unresolved as more information is obtained through next generation sequencing (e.g., Ferrer-Obiol et al., 2023). It also remains to be unknown whether the observed phenological differences between the Western and Eastern Mediterranean is reflected in genomic structure and has future evolutionary potential. A few genetic studies included the Yelkouan Shearwater populations with limited sampling. Among these, Heidrich et al. (1998) found colony-specific variance in mtDNA cytb haplotypes in Maltese and Greek populations of the Yelkouan Shearwater and Genovart et al. (2007) found a hybrid zone with the Balearic Shearwater. A comprehensive genomic study utilising the more recent sequencing technologies and covering the range of the species is needed. Filling these knowledge gaps could facilitate the designation of more effective conservation strategies for the species in the Mediterranean, as well as contributing to our understanding of dynamics between the population trends of the Yelkouan Shearwater, its ecology, and the marine environment.

1.2. Population Genetics of the Mediterranean Seabirds

The Mediterranean is a small, semi-closed ocean, exchanging waters with the Atlantic via the Strait of Gibraltar, and with the Black Sea via the Turkish Strait System. The Mediterranean was separated from the Atlantic during the Messinian Salinity Crisis (5.97 to 5.33 MYA; Hsü et al., 1973) and filled once again to today's level when the connection was restored (van Dijk et al., 2023). The Mediterranean has complex oceanographic features; this saline ocean gets cold inflow via Gibraltar Strait (Bormans et al., 1986). Its long stretch of coast, combined with many small islands, the freshwater influx from several rivers along the coast and strong localised winds causes heterogenous water temperature, salinity, and primary productivity. These characteristics of the Mediterranean are considered as important factors creating naturally uneven and fragmented spatial distribution in seabird populations (Carboneras & Requena-Moreno, 2010; Oro, 2003). In comparison to other

oceans, seabird diversity in the Mediterranean is low, but shows high endemism (Zotier et al., 1999); nearly all taxa in the Mediterranean have an endemic species or subspecies.

Genetic studies on Mediterranean seabirds focused on several different species from coastal shags and gulls to pelagic, long-distance migrant shearwaters, representing different ecological guilds. These studies mostly covered Mediterranean and Atlantic populations of the European seabirds, revealing both longitudinal (west-east; e.g., Cagnon et al., 2004; Genovart et al., 2013) and latitudinal (north-south; e.g., Thanou et al., 2017) genetic structuring. Both historical and contemporary factors have been cited in these studies as an operating mechanism for current genetic structuring, especially for strong differentiation between the Atlantic and Eastern Mediterranean populations. Of these, the most important historical events are the formation of Gibraltar and glaciations during the Pleistocene.

Higher genetic diversity in the Atlantic populations found in these studies suggests that Gibraltar was the route for colonising the Mediterranean. Cagnon et al. (2004) and Gomez-Diaz et al. (2006) supported this view for Mediterranean Storm Petrel and Scopoli's Shearwater, respectively. Cagnon et al. (2004) found that the timing of coalescence between Atlantic and Mediterranean populations of Mediterranean Storm-petrel is in line with long-term geological isolations between two basins and Gibraltar was probably the route to invade the Mediterranean and following glacial events created isolation between two basins. However, occasional long-distance dispersal of Mediterranean Storm Petrel continues to enable gene flow between the two basins. Potential for this dispersal was supported in Lago et al. (2019) by showing the movement of the species between two oceans. A similar mechanism operates for Scopoli's Shearwater; in addition to the Almerian front near Gibraltar acting as a genetic break, Gomez-Diaz et al. (2006) suggested glacial events as important factors in shaping within basins gene flow historically.

The Mediterranean has a strong oceanographic variation in the east-west gradient, mainly driven by two sills and thermohaline circulation, which has proven as a genetic break for several marine organisms (see review of Pascual et al., 2017) but the range-wide evaluation of population genetic structure of Mediterranean seabirds in this context is lacking. Two notable exceptions are Gomez-Diaz et al. (2006) and Thanou et al. (2017). Thanou et al. (2017) found within-Mediterranean differentiation in European Shag populations despite no apparent physical barriers. The observed genetic structuring within the Mediterranean corresponds to four regions: Corsica, Adriatic, Aegean and Crete. The authors suggested that this is probably related to the coastal behaviour of the species and to local adaptation to prey movements rising from the oceanography influencing prevailing

current flows. Gomez-Diaz et al. (2006) on the other hand, pointed out a within-Mediterranean structure in the east-west gradient is captured in biometry but not revealed in genetic analysis for Scopoli's Shearwater. Genovart et al. (2013) later found weak within-Mediterranean structure in the same species with significant Fst values between Eastern Mediterranean (Crete), Central (Sardinian Colonies) and Western Mediterranean (French colonies).

It should be noted that so far only mtDNA and microsatellite markers have been used to measure the genetic variation and structure in the Mediterranean seabird populations. Despite providing valuable information, these markers limit the evaluation of population structure to a single or a few loci. With the development of methods, bi-allelic SNPs in reduced representation sampling or whole genome sequencing are replacing these markers and providing a higher resolution information on genomic variation. Moreover, as SNPs cover much higher number of loci, they provide new insights into fine-scale structures in weakly structured populations. SNPs have only recently started to be used in seabird genomic studies (references herein) and not so far in the Mediterranean seabird studies except for the endangered Balearic Shearwater (Cuevas-Caballé et al., 2021; Ferrer-Obiol et al., 2023).

So far a thorough genomic analysis to measure the variation and connectivity among different Yelkouan Shearwater populations in the Mediterranean has not been carried out. Bourgeois (unpublished data, 2020) analysed the difference in cytochrome b genes of the mtDNA across its breeding range and found very low differentiation, yet with some potentially important variation between the Eastern and Central Mediterranean populations. Considering its conservation status and apparent phenological differences between the Western and Eastern Mediterranean, understanding the genomic structure of the Yelkouan Shearwater metapopulation in the Mediterranean using markers providing higher resolution data would assist the management decisions for this species. In addition, population genomic structure can potentially help defining sufficiently diverged units for management and can lead to new research questions that would integrate genomics into the conservation of the species. First objective of this thesis is to quantify the genomic variation and structure across the breeding range of this species using mitochondrial DNA (mtDNA) and high-resolution nuclear DNA (double-digest RAD sequencing) markers.

1.3. Seabird Population Genomics in the Light of Isolation-by-Resistance

Procellariiform seabirds exhibit life-history traits such as low fecundity, high natal philopatry and breeding site fidelity that may significantly influence gene flow among populations. Most species

in this group also have high wing-loading that enables low-cost flight, which may facilitate dispersal and therefore gene flow. However, as shown by the literature (references herein) mechanisms leading to the observed genetic differentiation within this group is not a simple function of isolation due to philopatry or complete genetic mixing due to mobility. As discussed by Milot et al. (2008) in detail, the paradox between high movement capacities and apparent high philopatry raises questions on the impact of non-visible barriers to gene flow in seabird populations. Identifying barriers to gene flow reliably is important to understand how sub-populations evolve through genetic structuring and how to conserve the evolutionary potential of a species. Several studies tried to shed light on the common factors shaping the genetic structure in seabirds in the absence of physical barriers (see reviews of Friesen et al., 2007; Lombal et al., 2020) and provided valuable insight. However, increasing use of NGS and whole genome sequencing in seabird genomics is providing new perspectives, particularly into fine-scale structure and disclosing new questions.

Genetic structure and barriers to gene flow in seabird populations have rarely been investigated in relation to isolation that is a product of resistance posed by landscape heterogeneity. All habitats have spatiotemporal variation across different scales, impacting different processes in landscape ecology and population genetics by modifying environmental selection and gene flow. For example, fragmentation in habitats (natural or human-made) creates discontinuities in the distribution of resources like food, and/or environmental conditions like microclimates, which in turn may impact movement of species and gene flow. For instance, Pascual et al. (2017)'s review showed that major fronts in the Mediterranean, as discontinuities in the marine habitat, increase genetic structuring in several different marine organisms with relatively longer larval life cycles. Regions with fronts that exhibit inter-annual variability exhibit a degree of genetic homogenisation suggesting that accessibility in such oceanographic discontinuities dictates this genetic structuring.

In this perspective, it should be noted that seabirds have much higher vagility in comparison to low-level marine organisms. Their dispersal capacities are theoretically maximised through lowered flight cost (Sachs et al., 2012) and the long time period young birds have to explore a large geographical area until reaching sexual maturity. However, despite high dispersal capacity, several studies showed that seabird distribution is too limited by several environmental factors and distribution of prey at varying scales (e.g., Crespin et al., 2006; Fauchald, 2010; Weimerskirch et al., 2006).

Wind is one of the most-studied environmental factors in relation to seabird movement. Wind has a high influence on the energetic cost of flight and therefore the birds' response to the wind has

significant implications for fitness (Kranstauber et al., 2015). Seabirds are central place foragers during the breeding season where they need to offset the cost of travelling to farther productive areas (i.e., the forager must bring the captured food items or semi-digested food to the colony, the central place). Several studies showed that seabird movement is highly related to the wind conditions that minimise the cost of flight during their foraging trips in the breeding season (e.g., Gómez Laich et al., 2012; Weimerskirch et al., 2000). Wind use in seabirds can, in turn, impact important population processes that have fitness consequences. For example, Czapanskiy (2018) tracked breeding Red-footed Booby movements and quantified the cost of movement using energy windscape around different breeding colonies. They found that the windscape predicts the foraging distribution of the breeding Red-footed Boobies in the Hawaiian Archipelago such that different colonies use optimal cost paths according to wind patterns. In another study, Elliott et al. (2014) showed that two seabird species adjusted their airspeed to reduce the energetic cost during high winds, however, this in turn impacted their offspring feeding strategy as high winds prevented foraging on prey that requires substantial above-water searching. At the larger geographical scale, wind is known to shape migration routes of seabirds (González-Solís et al., 2009). The strong influence of wind on flight path and cost can potentially influence demographic and evolutionary processes among seabird populations such as dispersal and gene flow. Wind patterns at the global scale are shaped by atmospheric circulation and recent studies suggest that these patterns have not changed substantially in the long-term (Kling & Ackerly, 2020). The relationship between genetic structure and such a long-term putative selection process has never been tested in seabirds.

Similarly, the topography of seabed and a wide variety of environmental factors related to that topography is known to be shaping seabird distribution at sea. Mid-sized shearwaters are particularly associated with shallow waters (Amorim et al., 2009; Carvalho & Davoren, 2019; Louzao et al., 2006; Yen et al., 2004), which can inhibit the movement over deep-sea areas and contact between colonies in return. Indeed, Hipfner et al. (2020) suggested that deep ocean habitat along the northern continental shelf (Alaska to Asia) inhibits movement between Rhinoceros Auklet populations in the eastern and western Pacific Ocean and is the main cause for genetic differentiation between these two regions. The Mediterranean is relatively a miniature ocean to pose a major barrier to shearwater flight capabilities. However, the Mediterranean is also characterised by the scarcity of primary productivity areas (Coll et al., 2010; Cossarini et al., 2021; Lazzari et al., 2012), which creates “islands” of foraging habitat for seabirds. Tracking studies have shown that Yelkouan Shearwater populations use areas nearby their breeding colonies for forging during the breeding season (Metzger et al., 2015; Péron & Grémillet, 2013; Pezzo et al., 2021). Whether and how much the potential resistance caused by the seabed topography and productivity translates into barriers to dispersal and gene flow in Yelkouan

Shearwater population is unknown. As shown by studies cited here, genetic structure in seabirds is probably a product of complex and interacting processes, with the environment and/or landscape having a significant role for some species, however so far the literature was shy of detailed evaluations, especially quantifications of the relationship between the seabird ecology and genetic structure. The second objective of this thesis is to quantify the relationship between habitat configuration and genomic structure of the Yelkouan Shearwater.

1.4. Using Genomics for the Conservation of Seabirds

Seabirds of the Procellariidae family nest on offshore islands that are considered as naturally fragmented habitats. Sub-populations located on these islands are interlinked at various but often large spatial scales (Milot et al., 2008), which drives the viability of the metapopulation through demographic connectivity, colonisations, and extinctions (Thomas & Kunin, 1999). Understanding the factors affecting these processes in each subpopulation is important in producing robust assessments of metapopulation viability. Identifying colony-specific terrestrial threats to seabird populations and measuring their impact on colony persistence is relatively achievable. However, one of the major challenges in seabird ecology and conservation is related to observing their life stages spent at sea and linking the threats during these stages to breeding colonies. Despite being central place foragers in the breeding season (i.e., the forager must bring the captured food items or semi-digested food to the colony, the central place), Procellariid seabirds travel long distances for foraging. For example, average foraging distance of Wandering Albatross during the incubation phase is 1,284 km (Weimerskirch et al., 1993) while much smaller and coastal Yelkouan Shearwater travels up to 579 km in chick rearing phase (Pezzo et al., 2021). Similarly in the non-breeding season, most species migrate long distances (e.g., Dias et al., 2012), while others stay closer to their breeding grounds but stay at sea (e.g., Raine et al., 2013). Moreover, young seabirds in many species disperse long distances in the post-fledging phase and spend years in the sea travelling vast distances (Péron & Grémillet, 2013). Major threats that seabirds are exposed to in this difficult-to-observe stage include mortality from oil spills, offshore wind farms, fisheries bycatch, and diseases (Croxall et al., 2012; Dias et al., 2019). However, colony-specific impacts of these threats are often unknown.

Assignment tests are one of the standard tools in molecular ecology with broad applications such as estimating contemporary rates of gene flow (Berry et al., 2004), identifying taxonomic (Sekino & Hara, 2007) or population boundaries (Manel et al., 2007), hybridization (Steeves et al., 2010), and cryptic species (García-Dávila et al., 2013). Assignment tests calculate the likelihood of sampled genotypes originating from the allele frequency of the sampled source population (Paetkau et al.,

1995). In the context of seabird ecology and conservation, assignment tests have been used mainly in identifying dispersal between genetically distinct populations (Bicknell et al., 2012; Dearborn et al., 2003; Hailer et al., 2010). Dantas et al. (2019) used assignment tests to confirm philopatry in Humboldt Penguin populations and Munilla et al. (2016) used them to help understand colony foundation in an establishing Cory's Shearwater breeding population. Besides answering ecological and evolutionary questions, assignment tests are also becoming a common tool in applied conservation problems, particularly in identifying genetic stocks, from which an individual in question originates. For example, Baetscher et al. (2022) genotyped Northern Fulmar bycatch samples obtained through an 11-year long fisheries observation program and showed how bycatch rates were disproportionately different than expected values based on colony size. Abbott et al. (2006) used genetic assignment tests to quantify the impact of fisheries bycatch on two albatross species that are otherwise difficult to identify morphologically to the species level.

The Mediterranean Sea is considered one of the hotspots of fisheries bycatch (Lewison et al., 2014) and several studies quantified the scale of seabird bycatch, especially in the Western Mediterranean (Barcelona et al., 2010; Cortés et al., 2017; Genovart et al., 2017). The Mediterranean Sea is also a focal region for offshore wind farm development, including in marine protected areas that have been designated using seabird distribution data (Lloret et al., 2023). Determining population-level impacts of these threats for seabird populations is becoming more important as they may lead to loss of genetic diversity for the areas with reduced gene flow or reduce the effective population size by changing the sex ratio (Gianuca et al., 2017). This is particularly important considering the high endemism rate and poor conservation status of the Mediterranean seabirds. In this perspective, assignment tests can be valuable tools in the conservation toolbox and help in closing the knowledge gaps in these areas in the Mediterranean (Manel et al., 2005).

A common observation of the Procellariid seabirds' genetic make-up, namely weak population structure, has been posing challenges in assignment tests when using low-resolution markers such as mtDNA or microsatellites. In the case of NGS data, assignment tests used to require called genotypes to achieve high accuracy assignment (Manel et al., 2005). However, advanced sequencing methods enabled the use of thousands of markers in assignment tests (Meek et al., 2016) and concomitantly advanced analysis methods can achieve high assignment accuracy in weakly structured populations using genotype likelihoods (DeSaix et al., 2023).

From stochastic mass mortalities to sustained mortality caused by fisheries bycatch, assignment tests can provide valuable insight into the impact of at-sea threats on different breeding seabird

populations. The third objective of this thesis is to test the accuracy of assignment tests by using samples with known origin to understand if the quantified level of genetic differentiation in the Yelkouan Shearwater metapopulation yields high-accuracy assignment results for future applications. In the light of results, this thesis also evaluates the assignment of unknown-origin samples collected in Türkiye.

1.5. Research Objectives and Methodological Overview

This thesis aims to address a knowledge gap on the form and potential use of genomic structure of the Mediterranean endemic Yelkouan Shearwater metapopulation. The outcomes of this thesis are expected to lead to further research and conservation planning for this species of conservation concern. To reach its goal, the thesis attempts to answer three research questions:

(A) What is the contemporary population genomic structure of the Yelkouan Shearwater across its breeding range?
(B) Given the resistance posed to seabird movement by certain environmental factors across the seascape, if and how these factors might be influencing the genomic structure of the Yelkouan Shearwater metapopulation?
(C) Can we use the contemporary observed population genomic structure of the Yelkouan Shearwater metapopulation in conservation practice of this endangered species?

To answer these research questions, high throughput sequencing and a broad range of genomic analysis tools were utilised, and the results were given in three subsequent sections. In the first section, approaches utilising genotype likelihoods were taken to identify pairwise genomic differences among sub-populations. Based on the result, broad-level underlying factors for the obtained genomic structure were discussed. In the second section, individual and colony-based genetic distance measures were used to understand the impact of habitat configuration on gene flow. To achieve this, landscape genetics framework was adopted to test isolation-by-resistance hypotheses. Landscape genetics approaches have not been used in bird population genomic studies widely (Kozakiewicz et al., 2018) and never in seabird studies. This section is therefore providing an important primary example. The final section used individual-based estimators of gene flow and genetic assignment approaches to identify the origin of colony of any given individual. Genetic assignment was tested in a weakly structured population and using genotype likelihoods, which is a recently emerging methodology.


tora'23

2. MATERIALS AND METHODS

2.1. Ethical Statement

Blood and tissue samples of Yelkouan Shearwaters used in this study were collected under the following permits.

(A) Dr. Karen Bourgeois (France). Permit by the Regional Directorate of the Environment, Planning and Housing in Provence-Alpes-Côte d'Azur dated 16/01/2025 and Marseille dated 28/01/2015.

(B) Association BIOM (Croatia). Permit 517-07-1-1-1-16-3 by the Ministry of Environment and Nature Protection in Croatia dated 10/03/2016.

(C) Dr. Karen Bourgeois (Italy). Permit 12498-255 by the Autonomous Region of Sardinia, Regional Department of Environmental Defence dated 28/06/2016.

(D) ISPRA, Italian Institute for Environmental Protection and Research (Italy). Permit 37088 by the Puglia Department of Agriculture, Rural Development and Environmental Protection dated 21/06/2016.

(E) Prof. Dr. Bruno Massa and Dr. Karen Bourgeois (Italy). Permit 11382 by the Sicilian Regional Department of Agriculture, Rural Development, and Mediterranean Fisheries dated 09/05/2016.

(F) Dr. Benjamin Metzger (Malta). Permits by the Parliamentary Secretariat for Agriculture, Fisheries and Animal Rights dated 10/03/2014 and 05/02/2016.

2.2. Sample Collection and Markers Used

Blood and tissue samples were collected between 2015 to 2021 from 15 different breeding colonies representing 71 to 74% of the global breeding population and one migration spot (Figure 2.1). A total of 218 blood and 29 tissue samples were collected from breeding birds and tissue samples were collected from ten dead specimens bycaught by recreational fishers in the Bosporus, Türkiye. Up to 0,5 µl blood was taken from metatarsal vein of the birds captured on their way in to the burrows. Blood samples were stored in EDTA, alcohol or blood cards and in -20 °C after the collection. Liver samples were collected from Bosporus birds, which were stored in -80 °C before necropsy. Distribution of sampling locations in relation to breeding colonies are given in Figure 2.1 together with sample and breeding colony sizes.

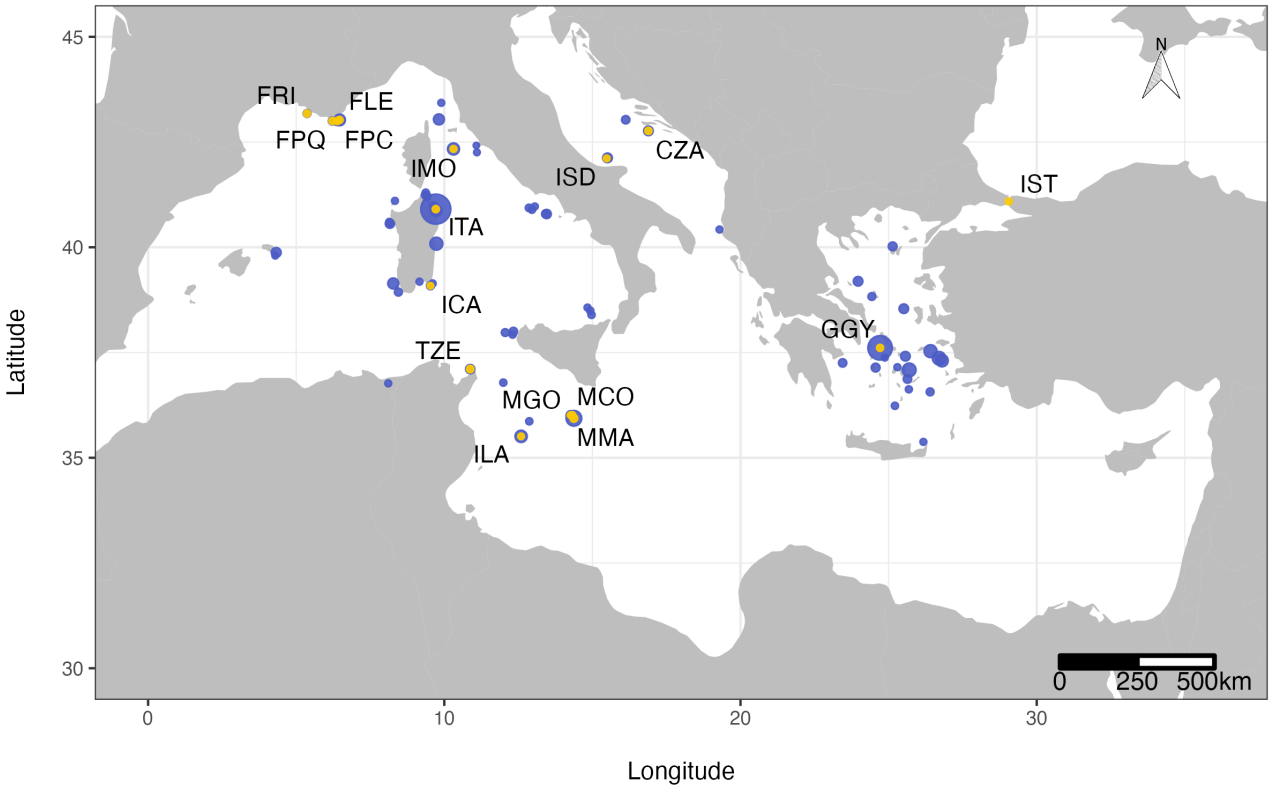


Figure 2.1. The distribution of Yelkouan Shearwater breeding colonies (blue) and sampling locations (yellow). Size of the blue circles indicate the size of the breeding population on the specific island and size of the yellow circle indicates the sampling size. CZA: Croatia Zaklopatica, FLE: France Levant, FPC: France Port-Cros, FPQ: France Porquerolles, FRI: France Riou, GGY: Greece Gyaros, ICA: Italy Cavoli, IMO: Italy Montecristo, ISD: Italy San Domino, IST: Türkiye Istanbul, ITA: Italy Tavolara, MCO: Malta Comino and Cominotto, MGO: Malta Gozo, MMA: Malta Main Island, TZE: Tunisia Zembretta.

Both nuclear and mitochondrial DNA (mtDNA) markers were used in this study to overcome the bias of sex-mediated gene flow. For the nuclear DNA, Single Nucleotide Polymorphisms estimated through restriction site reduced representation sampling was used, whereas for mtDNA Domain I was sequenced.

2.3. DNA Extraction and Sequencing

Three different DNA extraction kits were used in isolating DNA from blood, muscle, and tissue samples: Qiagen MagAttract High Molecular Weight DNA Kit and DNeasy Blood and Tissue Kit and AMBRD Genomic DNA Kit (Istanbul, Türkiye). Modifications were made to the protocols of all three kits by doubling the amount of Proteinase K and increasing the lysis time to digest the old blood samples more efficiently. The isolated DNA were quantified using Qubit v3 and integrity of the DNA

extractions was checked with a 0.7% agarose gel electrophoresis. Some samples were processed more than once using a different kit if they give low DNA yield in the first round.

For the mitochondrial control region Domain I amplification, two of the three primers designed by Gomez-Diaz et al. (2006) were used (CAL1L (5'GGTCCTGAAGCTAGTAATAC3') and CAL2H (5'CATCCCATCCAACTTAAG3')). The final PCR volume of 25 µL solution consisted of 1 µL of each 10 nM primer and 15 to 60 ng/µL of DNA template depending on the concentration of the sample and the PCR mix of AMBRD (Istanbul, Türkiye). PCR reaction is carried out in five phases: 4 minutes start at 94 °C, 40 cycles of 45 seconds of denaturation at 94 °C, 40 cycles of 45 seconds of annealing at 60 °C, 40 cycles of 90 seconds of extension at 72 °C and 5 minutes of final extension at 72 °C. PCR products were commercially sequenced using the same primers at Macrogen, Inc in Korea.

Libraries for nDNA SNP data was prepared using double digest restriction-site associated DNA method (Peterson et al., 2012) and following the protocol of Brelsford et al. (2016), which is a modification of Peterson et al (2012)'s protocol. Using Fragmatic software (Chafin et al., 2018) and the reference genome of the closely related Cory's Shearwater (GenBank Accession Number GCA_013401115.1), *in silico* analyses were carried out to identify the optimal restriction enzyme combinations and the number of loci expected to be recovered. Among the three pairs of enzymes tested, PstI-HF, which cuts rarely and MspI, which cuts frequently, yielded the highest number of loci, and were subsequently used for digestion of the DNA.

The genomic DNA was double digested with the two selected restriction enzymes, PstI (NEB, UK) as a rare cutter and MspI (NEB, UK) as a frequent cutter. After shearing DNA from restriction sites, unique combinations of short DNA sequences were ligated to the sticky ends with T4 DNA ligase (NEB, UK) together with MseI adapters for dual indexing. Adapter and primer sequences used in library preparation are given in Table 2.1 and unique index and barcode sequences are given in Table 2.2. Following adapter ligation to the fragmented DNA, short fragments were removed using Agencourt AMPure and target fragments were amplified using PCR primer of i5 index (Table 2.1) and 12 indexed i7 PCR primers (Table 2.2). PCR products were run on a 1.5% agarose gel for quality control. Samples that failed to amplify were excluded from the pool and re-processed. Size selection of fragments in the 450-550 bp range were carried out on a 2% agarose gel and selected fragments were purified using MinElute Gel Extraction Kit. Final samples were then quantified using qPCR (Applied Biosystems, USA) using KAPA Library Quantification Kit (Roche Diagnostics, Swiss). For quantification, the instructions of the manufacturer were followed. Based on qPCR results samples

were pooled equimolarly. Libraries were sent to Macrogen, Inc (Korea) and Gen-Era Ltd. (Türkiye) for sequencing in Illumina platforms in three different pools.

Table 2.1. Sequences of adapters and primers used in ddRAD library preparations. nnnnnn represents unique barcode and zzzzzz represents unique index sequences, which are provided in Table 2.2.

Name	Sequence (5' to 3')
MseI Adapter 1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
MseI Adapter 2	TAAGATCGGAAGAGCGAGAACAA
PstI Adapter	ACACTCTTTCCCTACACGACGCTCTTCCGATCTnnnnnnTGCA
PCR Primer 1	A*A*TGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG
PCR Primer 2	C*A*AGCAGAAGACGGCATACGAGAzzzzzzzzGTGACTGGAGTTCAGACGTGTGC

Table 2.2. Barcode and index sequences used in ddRAD library preparation.

P5 Barcode	**i7 Index**
ACAGGT	ATCACG
ACGATG	CGATGT
AGAGGA	TTAGGC
AGTCCT	TGACCA
AGTCGA	ACAGTG
CACAGA	GCCAAT
CTCAGT	CAGATC
CTTCAG	ACTTGA
GACACA	GATCAG
GTGTCT	TAGCTT
TCTCGA	GGCTAC
TGTCCA	CTTGTA

In genomic studies date of data collection (e.g., DNA isolation or PCR) could be an important batch effect due to technology, assay, collection conditions variations (Leek et al., 2010). To minimise batch effects all libraries were prepared in a few months between 2021 and 2022 using the same reagents and equipment as much as possible. In addition, negative and positive controls in mtDNA amplification were used to minimise the effect from processing technology and kits.

2.4. Bioinformatics and Data Analyses

2.4.1. Population Genomics Data Analysis

mtDNA chromatograms were manually edited using Geneious Prime V 2023.1.2. Regions with more than 3% chance of an error per base in individual sequences were batch trimmed. The forward and reverse reads of the trimmed sequences were assembled in Geneious to create consensus sequences (contigs) for each individual and ambiguities in these contigs were checked and edited manually. All contigs then aligned in Geneious using MUSCLE algorithm (Edgar, 2004) with default settings. Final alignments were checked for ambiguities, trimmed to the same length of 242 bp, and exported as a fasta file.

Fasta file that contains the consensus sequences was used to find the best DNA model for the data in Mega V 11.0.13 (Tamura et al., 2021). For model selection maximum likelihood method and neighbour-joining tree was used. Based on the model selected pairwise distances between individuals were calculated in Mega, which produces the distance matrix for phylogenetic tree building. Phylogenetic tree was built using the maximum likelihood method and bootstrapping 1000 replicates. Popart V 1.7.2 (Leigh & Bryant, 2015) was used to construct a haplotype network, based on groupings suggested by the ddRAD analysis results (see below). Nucleotide diversity (Pi) and tests of selection (Tajima's D) for each of these groups and total data set were undertaken with the same software.

Data quality of the raw ddRAD sequences was checked using FastQC (Andrews, 2010) software. Stacks V 2.58 (Catchen et al., 2013) was used for demultiplexing the data and obtaining sequences of individual samples in separate files. Demultiplexing was carried out with the process_radtags program using default parameters for window size, score limit and barcode mismatches. Reads with uncalled bases and very low-quality scores were removed.

The remaining reads were aligned to the genome of the closely related Balearic Shearwater (*Puffinus mauritanicus*), published by Cuevas-Caballé et al. (2021). Burrow-Wheeler Aligner V 0.7.17 (Li & Durbin, 2009) MEM algorithm and default parameters were used for alignment. The Sequence Alignment Map (SAM) files were validated using GATK Picard Tools V 3.0.0 ValidateSamFile program (Broad Institute, 2023). Errors found in 12 SAM files were related to non-matching forward and reverse read sequence orders in fastq files and fixed using Fastq_pair (Edwards & Edwards, 2019). Samtools V 1.17 (Danecek et al., 2021) were used to sort SAM files by coordinate and convert them to BAM files, as well as to calculate alignment statistics. Based on alignment statistics, samples with less than 90% of the sequences aligned and/or with lower than 100,000 aligned

reads were excluded from further analyses. This final quality filtering with arbitrary cut off was based on the relative number of aligned reads for each sample and resulted in 188 samples to use in downstream analyses.

Alignment statistics revealed very low depth of 1x to 5x, shown to lead to high error rates in genotype calls, for instance by underestimating the heterozygosity (Bresadola et al., 2020). Therefore, ANGSD V 0.940 (Korneliussen et al., 2014) were used for downstream analyses. ANGSD is a probabilistic framework that does not require calling genotypes and is suitable for low depth NGS data. Using ANGSD, genotype likelihoods and allele frequencies were calculated at each site for each individual. Genotype likelihoods were calculated using -doGeno 5, -doPost 1, -doMAF 1, -doMajorMinor 1, -GL 1, -minMaf 0.05, -SNP_pval 1e-12, -postCutoff 0.85 and -mInd 50% parameters and filters.

Principal Component Analysis (PCA) was run to identify population structure in the sampled data. Covariance matrix between individuals based on genotype likelihoods was estimated using PCangsd V 1.10 (Meisner & Albrechtsen, 2018). The number of significant eigenvalues to be used in modelling of individual allele frequencies was set to seven (-e parameter). PCangsd starts with a rough guess for this parameter and iteratively tries to improve it by using Minimum Average Partial (MAP) tests. By setting the -e parameter to seven, assumption of eight isolated populations was made. When compared to the PCA made with default values, -e 7 analysis captured more variance in the first seven components. Eigendecomposition of the matrix and plotting the principal components was carried out on R using base (R Core Team, 2023) and tideyverse (Wickham et al., 2019) packages. The first four principal components were plotted in combinations (e.g., PC1 and PC2, PC2 and PC3, PC3 and PC4) to check the clustering of the data. PC loadings were calculated with PCangsd V 1.10 (Meisner & Albrechtsen, 2018).

Site frequency spectrum (SFS) shows the proportions of sites with different derived allele frequencies and was used for calculating population genetics statistics. SFS was derived for each sampled colony separately from genotype likelihoods using ANGSD V 0.940 (Korneliussen et al., 2014) and winSFS (Rasmussen et al., 2022). The SFS was based on minor allele frequencies (folded SFS) since the genetic history of the species is not well known to estimate derived allele frequency from an ancestral outgroup.

Genetic diversity and neutrality statistics were calculated for each colony using ANGSD V 0.940 (Korneliussen et al., 2014). Individual heterozygosity was calculated by dividing the number of

polymorphic sites by the number of total sites in the folded SFS estimated using individual BAM files.

To determine the ancestry between populations admixture analysis was carried out with NGSadmix V 32 (Skotte et al., 2013) using genotype likelihoods. Admixture is a model-based clustering analysis and estimates ancestral proportions of each individual by assuming a model of discrete ancestral populations, which is represented with K. Admixture was run for K= 2 to 10 and for 10 replicates for each K value. Evanno et al. (2005)'s method was implemented using log likelihood of each K in these replicates to determine the most likely K. Admixture results were visualised using the pophelper package in R (Francis, 2017).

To identify outlier loci that exhibit highly different patterns of genetic variation among individuals, selection scans were made with two commonly used methods, FastPCA (Galinsky et al., 2016) and pcadapt (Luu et al., 2017), implemented in PCAngsd V 1.10 (Meisner & Albrechtsen, 2018). Both FastPCA and pcadapt identify variants that have significantly greater differentiation along top PCs than the null distribution of genetic drift but differ in derived statistics. The pcadapt method uses robust Mahalanobis distance computed for each site to evaluate all top PCs for estimating z-scores and FastPCA tests normalised variant weights for each PC separately. Both test statistics follow $\chi 2$ distributions from which a p-value for each polymorphic site is obtained. PCAngsd extends the selection statistics of both programs to genotype likelihood data.

Fst values are calculated from site frequency spectrum of each pair of population using realSFS (Korneliussen et al., 2014) and visualised in R. Linear relationship between Fst values and geographical distance was checked with a simple linear regression. Geographic distance between different colonies was calculated as the least-cost path over the water using the marmap package in R (Pante & Simon-Bouhet, 2013).

2.4.2. Seascape Genomic Data Analysis

All analyses in this study were run on the ddRAD data set described above. In addition to pairwise colony-based Fst values, individual-based measures of genetic structure were included in the analyses. Accordingly, pairwise estimates of kinship for 188 individuals were calculated in ANGSD V 0.940 (Korneliussen et al., 2014) using IBSrelate method (*-doIBS 2 -makeMatrix 1*) (Waples et al., 2019), and by sampling the consensus base. Pairwise Fst values calculated for each pair of colonies is used in Mantel tests.

An initial analysis of gene flow and seascape association was conducted using program Estimating Effective Migration Surfaces (EEMS; Petkova et al., 2016). EEMS uses the concept of 'effective migration' to model the relationship between genetics and geography with the premise of effective migration is low in regions where genetic similarity decays quickly. EEMS calculates the effective migration by incorporating resistance distances, therefore, the output of EEMS analysis was used to visually evaluate reduced regions of gene flow and in selecting environmental variables for further isolation by resistance analyses.

The pairwise individual differences matrix was used in 10 independent EEMS runs (runeems_snps), each consisting of one MCMC chain of two million iterations with one million burn-in iterations to discard at the start and 9999 thinning iterations between two writing steps. As the approximate number of demes in the population graph, a thousand demes were specified. A polygon of the Mediterranean and the Marmara Sea area was provided as habitat, in which spatial structure will be evaluated. Default parameters were used for the rest of the arguments. Results were visualised using rEEMSplots package (Petkova et al., 2016) in R.

Environmental variables for resistance models were selected based on the visual evaluation of the EEMS analysis outcome, as well as available information on environmental factors influencing the distribution and movement of closely related pelagic seabirds.

Multidecadal wind data covering the period from January 1980 to January 2010 was used in windscape modelling. Wind data in hourly resolution was downloaded from the Climate System Forecast Reanalysis (Saha et al., 2014), which has a spatial resolution of 35 km. The data was downloaded in GRIB format and converted to NetCDF using Climate Data Operators (CDO) V 2.3.0 (Schulzweida, 2023). The NetCDF files were then processed in R using processNC (RS-eco, 2023) to merge the files and terra package (Hijmans, 2023) to subset the Mediterranean data, which ended up with 16,372 timesteps in every location.

The ETOPO Global Relief Model of NOAA was used for bathymetry (NOAA National Centers for Environmental Information, 2022). The data and raster were downloaded using the marmap package (Pante & Simon-Bouhet, 2013) in R in 1 arc-minute resolution (ca. 1 km).

For isolation-by-distance analyses, distance is calculated as the least cost path over the water as implemented by the marmap package (Pante & Simon-Bouhet, 2013). Accordingly, two distance

matrices were calculated, one is least cost over the water without any limitation to movement and the second was least cost path when the movement is restricted to continental shelf (-200 m depth). Isolation-by-distance was investigated with Mantel and partial Mantel tests in vegan package (Oksanen et al., 2022) in R, using 9999 permutations to determine the significance of the Mantel correlation coefficient. Based on the monotonic shape of Fst (Fst / 1-Fst) and distance (log) values, and skewed shape of log distance values, Kendall's rank correlation method was used in computation of the correlation coefficient. Low sample size colonies were excluded as they were inflating Fst values.

For isolation-by-resistance analyses, high resolution (1 arcminute) bathymetry raster was created using ETOPO22 data and the marmap package. Instead of assigning seascape resistance values manually, which can create biases (Peterman et al., 2014), parameterisation of the resistance surface was carried out using resistanceGA package (Peterman, 2018) in R. resistanceGA uses genetic algorithms in optimising the resistance values and identify the best statistical relationship between pairwise genetic distances and effective distances without priori assumptions. reistanceGA conducts model selection based on Akaike Information Criterion (AIC). Accordingly, bathymetry raster was used in single surface resistance optimization using colony locations and the genetic distance matrix (pairwise Fst values). Genetic algorithms maximizing the log-likelihood of the fitted model were applied in optimizations (method = 'LL'). The resistance matrix calculated by resistanceGA was used in standard and partial Mantel tests following the examination of diagnostic plots of the optimization.

For the windscape modelling, methodology of Kling and Ackerly (2020) was followed using the windscape package (Kling, 2023) in R. Accordingly, wind data was split into monthly data for the whole time period (30 years) to make the computations more manageable, then combined to obtain a single file before least cost path calculation (see below). Using each monthly data in 1 hour resolution, a wind connectivity graph was constructed by connecting each grid cell with its eight neighbours. The level of connectivity for each direction was calculated in proportion to the wind frequency and speed in that direction. Wind conductance over terrestrial cells were then downweighed to zero to exclude connectivity paths over the land as pelagic seabird do not disperse over land. This was achieved by providing a masking raster layer specifying land and sea areas and filtering out the land areas from connectivity graph.

The remaining connectivity graph over the Mediterranean Sea was used to identify least cost path between colony locations. The cost was defined as wind hours: "average estimated number of hours wind would take to diffuse to that location based on the full spatiotemporal distribution of wind

conditions across the landscape" (Kling & Ackerly, 2021, p. 9). As the spatial resolution of wind data was coarse (35 km), the calculation of wind hours was scaled down 10x to interpolate values in finer spatial resolution. The wind flow between any two points (e.g., pair of colony locations) was quantified as the inverse of wind hours and calculated in each direction (upwind and downwind). Wind flow was calculated in both directions for each pair of colony locations. The wind connectivity matrix was then calculated by taking the average of wind hours in both directions for each pair of sites to obtain a symmetric matrix.

Correlation between wind connectivity and genetic similarity was tested using partial Mantel test. Tests was applied on normalized pairwise Fst, log mean wind flow matrices and accounting for isolation by distance by including geographical distance matrix used in isolation-by-distance tests. The windscape package functions were used in the Mantel tests (Kling, 2023).

2.4.3. Genetic Assignment Data Analysis

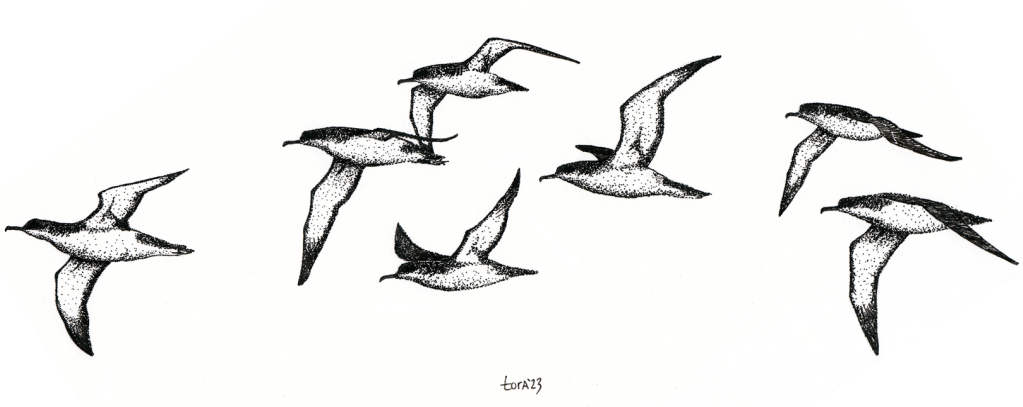
All analyses of genetic assignment were run on the ddRAD data set described above. Different subsets were generated (explained below), and genotype likelihoods from these subsets were calculated separately using ANGSD V 0.940 (Korneliussen et al., 2014). WGSassign V.1.0 (DeSaix et al., 2023), which implements probabilistic methods accounting for the uncertainties in genotype likelihood data, were used in all assignment tests without the need of called-genotypes. The training-set/hold-out protocol was followed to overcome "high-grading bias" i.e., overestimation of assignment accuracy due to the identical training and classifying datasets (Anderson, 2010). Accordingly, two datasets were generated for each subset: "reference population(s)" that the samples were assigned to and "assignment population(s)" that consist of samples of assumed unknown or unknown origin, which consisted of only Türkiye samples. Similar to training-set/hold-out protocol, "leave one out" (LOO) cross validation approach was used in reference population datasets. In LOO, WGSassign removes one sample from the dataset at a time, calculates allele frequencies of the remaining population and then calculates the likelihood of origin of the removed individual based on these allele frequencies. The LOO approach was used to test the assignment accuracy of known-origin samples in the reference datasets and used to inform further subsetting criteria.

As the population genetic structure of the Yelkouan Shearwater showed a continuous structure, performance of assignment tests was examined at varying spatial scales. In each subset, samples were grouped under colonies, countries, and the Mediterranean Sea subregions. These spatial scales were selected based on their practical use in conservation management. For the subregional dataset,

samples were grouped under four marine regions that are included in the marine ecosystem management policies such as EU's Marine Strategy Framework Directive (European Parliament, 2008) or UNEP's Barcelona Convention/Mediterranean Action Plan (United Nations Environment Programme, 1976), as these plans include assessment and monitoring criteria at the subregional level. Performance of assignment tests were then tested at all of these scales for each subset.

WGSassign uses allele frequencies in reference populations for assignment and calculates these allele frequencies from read data (i.e., BAM files) using maximum likelihood. Accuracy of the assignment is highly linked to the accuracy of allele frequency estimations. One recommendation to increase the accuracy of allele frequency estimation is to increase the sample size to compensate for low coverage. Specifically, Lou and Therkildsen (2022) suggest using 1x depth or otherwise having at least 10 individuals with varying depths that would make 10x population coverage. Although this is recommended for low-coverage Whole Genome Sequencing (lcWGS), which typically has higher number of SNPs than ddRAD data, it is still applicable for ddRAD data as in essence it aims to maximise the information for improved allele frequency estimation. Accordingly, several approaches were adopted in generating reference datasets to test the impact of sample size and sample quality on assignment accuracy: (a) highest quality five samples from each colony (based on sequencing depth and number of aligned reads, n=60)); (b) randomly selected five samples from each colony (n=60); (c) effective sample size from each colony (n=106), as determined by WGSassign –ne_obs parameter; and (d) unequal sample sizes, which includes all 188 samples in the global dataset. Remaining samples in each option but (d) were treated as "unknown origin" and used as the assignment set. Only Türkiye samples were included in the assignment test for the (d) subset.

Using genotype likelihoods estimated by ANGSD for each subset, allele frequencies and effective sample sizes for reference populations were calculated using WGSassign as well as assignment likelihood values for each sample in the leave-one-out cross validation test. Assignment of unknown origin samples were then tested using the reference population allele frequencies and genotype likelihoods of assignment datasets. Results were visualised in R using the tidyverse packages (Wickham et al., 2019).


tora'23

3. RESULTS

3.1. Population Genomic Structure of the Yelkouan Shearwater

3.1.1. mtDNA

Manual editing and filtering out poor quality samples resulted in retaining of 196 individuals out of 249 samples for the subsequent analyses. Distribution of these samples according to countries and islands are given in Table 3.1.

Table 3.1. Distribution and number of samples according to colonies and markers used in this study.

Country	Island	Code	mtDNA Samples	ddRAD Samples
Croatia	Zaklopatica	CZA	14	16
France	Levant	FLE	9	8
	Port-Cros	FPC	24	25
	Porquerolles	FPQ	14	10
	Riou	FRI	13	12
Greece	Gyaros	GGY	16	15
Italy	Cavoli	ICA	3	2
	Lampedusa	ILA	2	3
	Montecristo	IMO	5	8
	San-Domino	ISD	3	1
	Tavolara	ITA	16	13
Türkiye	Bosporus	IST	9	10
Malta	Comino & Cominotto	MCO	16	15
	Gozo	MGO	5	4
	Malta	MMA	22	23
Tunisia	Zembretta	TZE	25	23
TOTAL			**196**	**188**

Among these 194 sequences of 242 bp long mtDNA control region, 13 out of 31 (42%) variable sites were parsimony informative (Table 2). Phylogenetic tree of sampled colonies did not show any structure or distinct clades based on colonies, countries, or sub-regions. The haplotype network defined 16 haplotypes when samples were grouped based on the grouping obtained in PCA and admixture analyses (Figure 3.1). The haplotype network revealed a pattern indicating gene flow between populations and an old population with a few ancestral haplotypes and a few rare. Most individuals had the same haplotypes and there were four unique and distant haplotypes in the Eastern Mediterranean group (Greece and Türkiye) and Central Mediterranean group (Tunisia and Malta). Maximum likelihood (ML, Figure A.1) and neighbor joining (NJ, Figure A.2) trees produced similar

results except that ML tree revealed two groups: one including haplotype five and 13 and the other including the rest of the haplotypes. Haplotype five is found in Malta, Croatia, France and Greece samples and 13 is only in Greece samples. Overall, the structure obtained with mtDNA sequences does not show the fine scale structure captured with ddRAD sequences (see below).

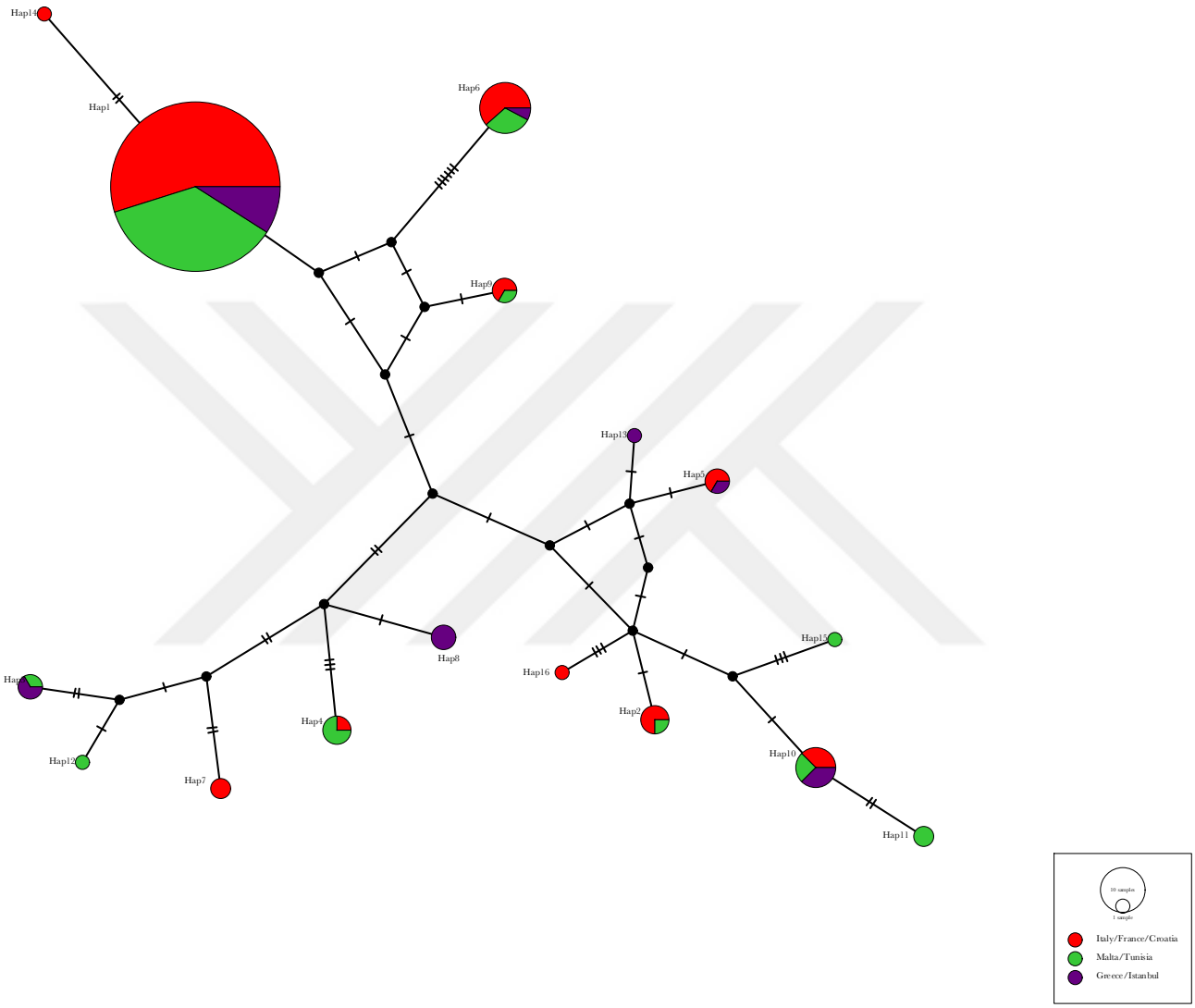


Figure 3.1. Haplotype network of samples grouped according to PCA and admixture results. Accordingly, samples were grouped in two subregions of the Mediterranean, Eastern, Central and Western except for the Croatian samples, which are grouped under Western Mediterranean.

Tajima's D values for all samples were -1.33 and ranged between -0.51 to -1.42, suggesting that the mtDNA sequences of the Yelkouan Shearwater agree with neutrality expectations (Table 3.2). Nucleotide diversity values of each group were very similar (Table 3.2).

Table 3. 2 Summary statistics of mtDNA sequences.

Group	Tajima's D	Nucleotide Diversity	#Samples	#Haplotypes	#Segregating Sites	#Parsimony Informative Sites
All	-1.331	0.013	194	16	31	13
IFC	-1.416	0.011	102	11	24	10
MT	-1.365	0.013	68	10	24	12
GI	-0.513	0.019	24	7	20	8

3.1.2. RAD-Seq Data

Out of 256 sequenced samples 229 samples passed the quality filters and the number of reads for these samples varied between one and 48 million bp. Following alignment to the reference genome, further filtering of samples was carried out based on BAM file validation and calculation of alignment statistics, which resulted in 188 samples to use in downstream analyses (Table 3.1). These samples had alignment percentages of 96% or higher, aligned number of reads ranging from one to six million and an average depth of 3x (range 1.67x to 5.35x).

Genotype likelihood calculation based on the sites that are present in at least 50% of the sampled population resulted in 752,480 polymorphic sites in 2,172 scaffolds.

Variation across 196 Yelkouan Shearwater mtDNA D-Loop sequences (242 bp) defined a few unique haplotypes but no significant population structure. For the nuclear DNA, the PCA applied on genotype likelihoods resulted in two main clusters, each including all populations (Figure 2 A). The second and third component, on the other hand, clustered samples based on regions in a gradient where France, Italy, Croatia projected more closely, Tunisia in between two groups and followed by Malta, Greece and Türkiye samples (Figure 3.2 B and 3.2 C). Greece and Türkiye formed their own cluster in PC3. Model-based clustering (admixture) results showed a clustering similar to PCA results (Figure 3.3). Based on the method of Evanno et al. (2005) the most optimal K was identified as 3. At K=3, admixture plots show a central group with shared ancestry at each location and presence of a third ancestral cluster in the majority of the groups but lowest in the Greece and Türkiye samples (Figure 3.3). Plots of log probability of the data given K (Ln P(X|K)) and the second order rate of change in Ln P(X|K) are given in Figure A.3.

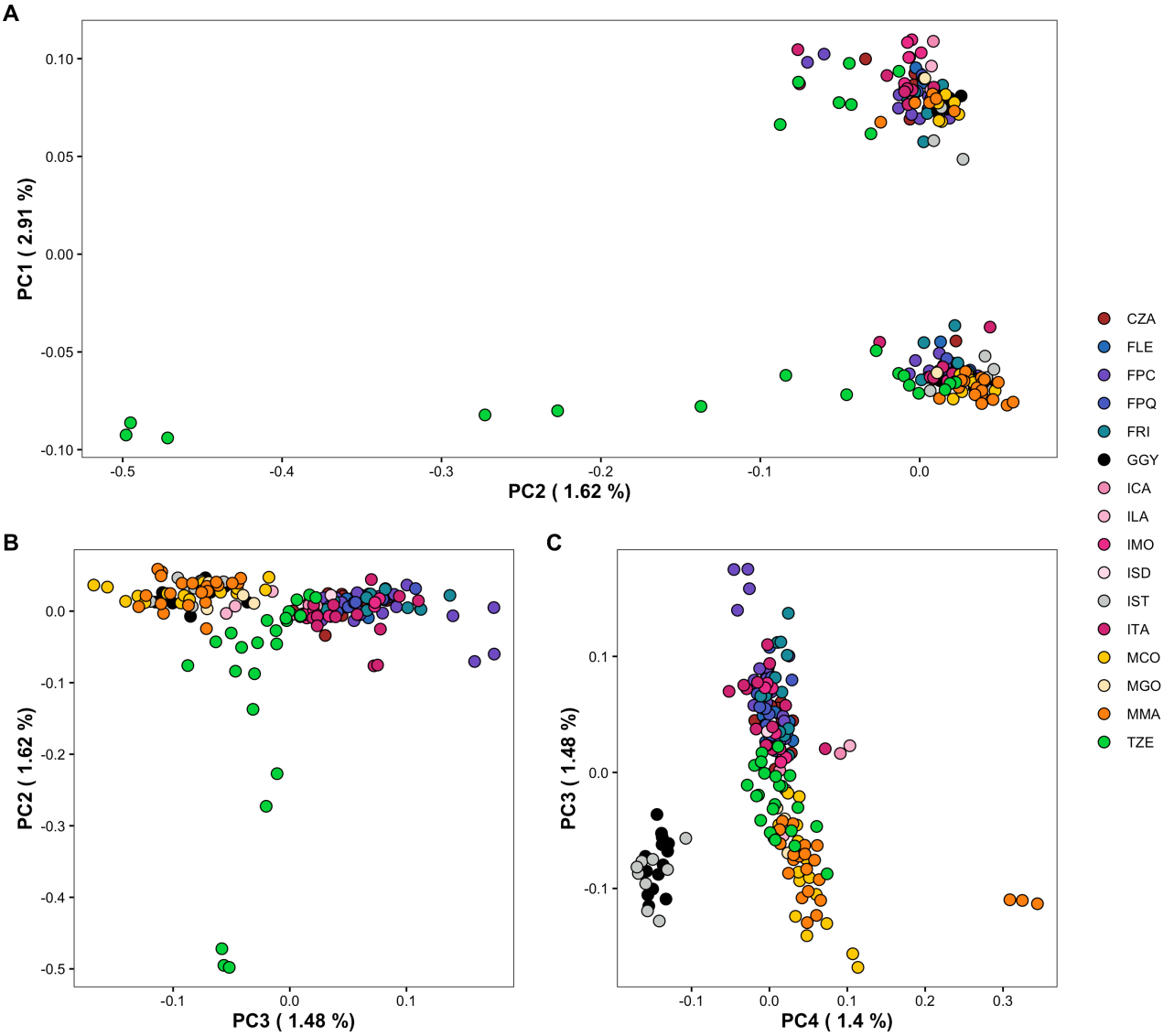


Figure 3.2. Principal Component Analysis plots showing the projection in the first four components. Colonies are color-coded. CZA: Croatia Zaklopatica, FLE: France Levant, FPC: France Port-Cros, FPQ: France Porquerolles, FRI: France Riou, GGY: Greece Gyaros, ICA: Italy Cavoli, IMO: Italy Montecristo, ISD: Italy San Domino, IST: Türkiye Istanbul, ITA: Italy Tavolara, MCO: Malta Comino and Cominotto, MGO: Malta Gozo, MMA: Malta Main Island, TZE: Tunisia Zembretta.

The low variance explained by the significant principal components highlights the amount of variability in the dataset and might be related to low fixation rate across populations in comparison to total variation in the metapopulation. Colony-based SFS agrees with this result by showing a high number of unique sites (i.e., rare allele abundance) in several populations (Figure A.4). The inspection of PC loadings to understand which SNPs are strongly contributing to the split in two clusters (SNP weights) did not reveal a specific region causing the split (Figure A.5).

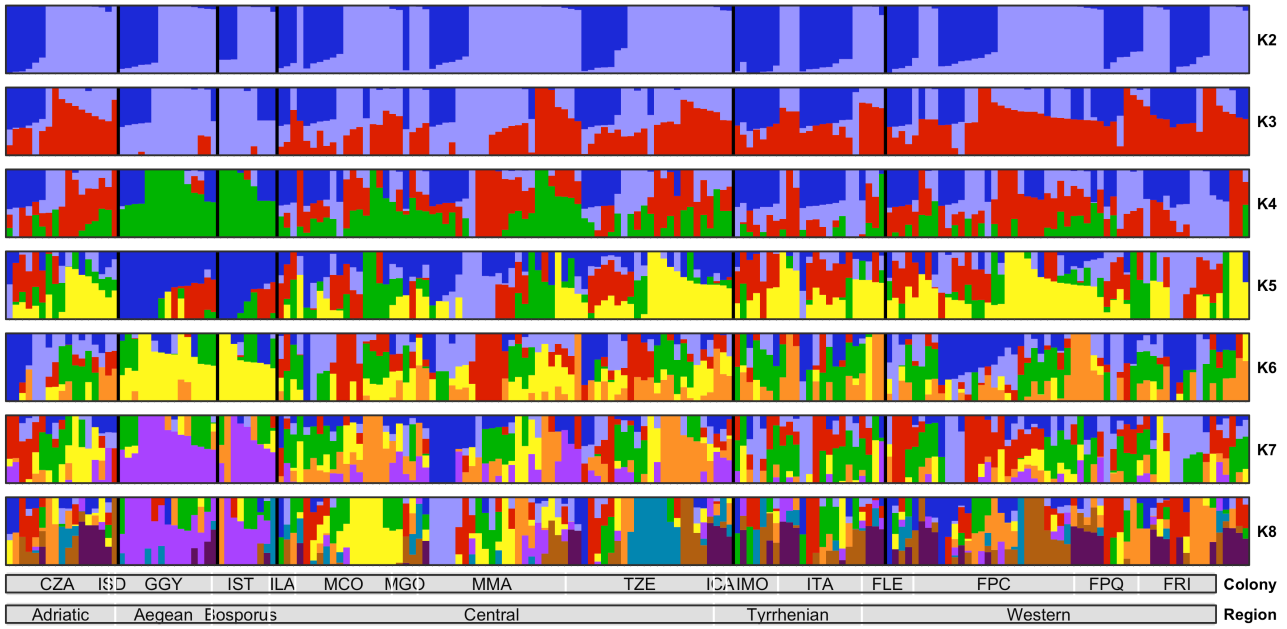


Figure 3.3. Admixture plots for K2 to K8 calculated from genotype likelihoods. Colony and region-based grouping is provided in the bottom panel, region-based grouping is also shown with black lines between bars. Maximum K (8) was based on the total number of countries present in the dataset.

Colony-based mean nucleotide diversity varied between 0.005 and 0.010 (Table 3.3, Figure 3.4). Site-based nucleotide diversity varied slightly among sampled colonies (Figure A.6). Colony-based observed heterozygosity did not differ significantly between locations (Kruskal-Wallis chi-squared = 12.954, df = 14, p-value = 0.530). Mean value of Tajima's D was not lower than neutral expectations for sampled colonies except FPC, MMA and TZE.

Table 3.3. Colony-based summary of nucleotide diversity and neutrality test (Tajima's D) values. CZA: Croatia Zaklopatica, FLE: France Levant, FPC: France Port-Cros, FPQ: France Porquerolles, FRI: France Riou, GGY: Greece Gyaros, ICA: Italy Cavoli, IMO: Italy Montecristo, ISD: Italy San Domino, IST: Türkiye Istanbul, ITA: Italy Tavolara, MCO: Malta Comino and Cominotto, MGO: Malta Gozo, MMA: Malta Main Island, TZE: Tunisia Zembretta.

Colony	**Nucleotide Diversity ± Standard Deviation ($x10^{-2}$)**	**Tajima's D ± Standard Deviation**
CZA	0.776 ± 0.234	-1.685 ± 0.493
FLE	0.718 ± 0.272	-1.393 ± 0.425
FPC	0.918 ± 0.211	-1.969 ± 0.464
FPQ	0.707 ± 0.256	-1.475 ± 0.470
FRI	0.639 ± 0.232	-1.416 ± 0.502
GGY	0.761 ± 0.251	-1.657 ± 0.492
ICA	0.549 ± 0.267	-0.920 ± 0.484
IMO	0.579 ± 0.256	-1.089 ± 0.524
IST	0.696 ± 0.239	-1.479 ± 0.405

ITA	0.741 ± 0.221	-1.560 ± 0.478
MCO	0.839 ± 0.246	-1.696 ± 0.400
MGO	0.589 ± 0.309	-0.941 ± 0.290
MMA	1.015 ± 0.233	-2.002 ± 0.383
TZE	0.867 ± 0.222	-1.883 ± 0.473

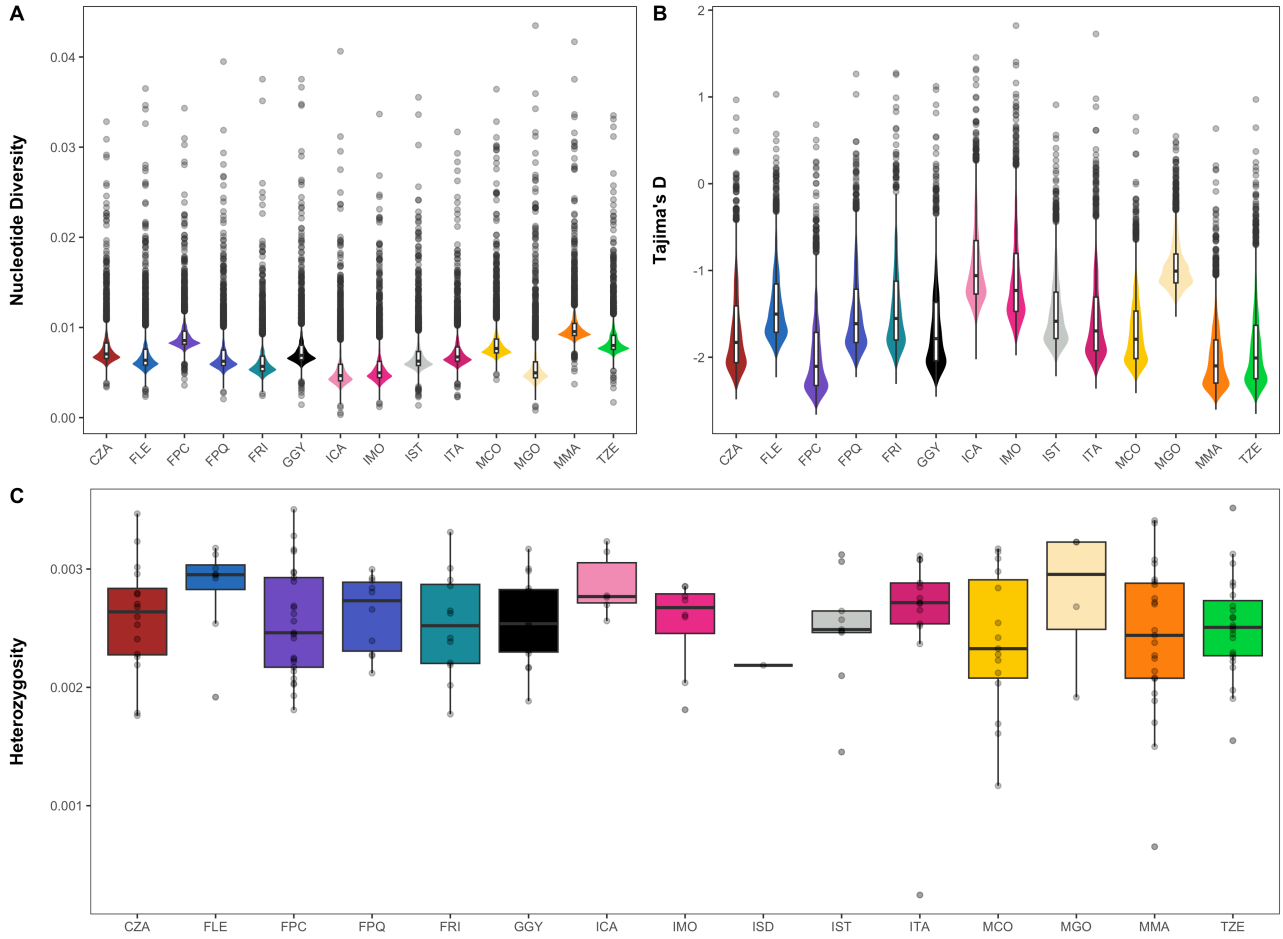


Figure 3.4. Colony-based nucleotide diversity (A), Tajima's D (B) and heterozygosity (C). Each colony is colour-coded. CZA: Croatia Zaklopatica, FLE: France Levant, FPC: France Port-Cros, FPQ: France Porquerolles, FRI: France Riou, GGY: Greece Gyaros, ICA: Italy Cavoli, IMO: Italy Montecristo, ISD: Italy San Domino, IST: Türkiye Istanbul, ITA: Italy Tavolara, MCO: Malta Comino and Cominotto, MGO: Malta Gozo, MMA: Malta Main Island, TZE: Tunisia Zembretta.

Fst values calculated for each pair of populations show low but some degree of differentiation between colonies (Figure 3.5 A). A simple linear regression model revealed a significant linear relationship between Fst values and geographic distance between the colonies (Figure 3.5 B, R-square: 0.2032, Residual standard error: 0.006 on 64 df, p-value: 0.000).

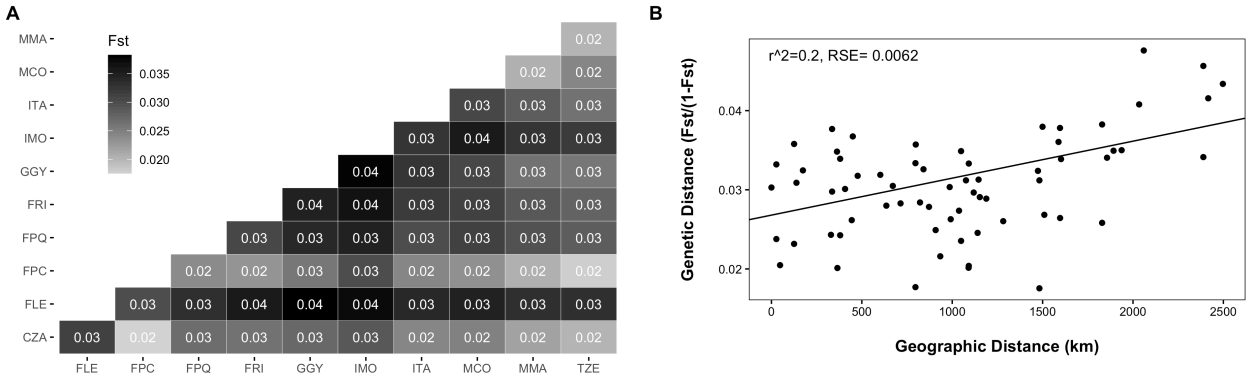


Figure 3.5. Genetic differentiation statistics (Fst) between each pair of colonies (A). CZA: Croatia Zaklopatica, FLE: France Levant, FPC: France Port-Cros, FPQ: France Porquerolles, FRI: France Riou, GGY: Greece Gyaros, ICA: Italy Cavoli, IMO: Italy Montecristo, ISD: Italy San Domino, IST: Türkiye Istanbul, ITA: Italy Tavolara, MCO: Malta Comino and Cominotto, MGO: Malta Gozo, MMA: Malta Main Island, TZE: Tunisia Zembretta. (B) Correlation between genetic distance (Fst) and geographical distance calculated as least cost path over the water.

3.2. Contemporary Environmental Factors Influencing the Genomic Structure of the Yelkouan Shearwater in the Mediterranean Seascape

3.3.1. Estimated Effective Migration Surfaces

EEMS analysis revealed higher-than-average historical gene flow between Central Mediterranean and the Adriatic and within the Aegean Sea (Figure 3.6). Genetic dissimilarities within the Yelkouan shearwater population were related to short distances in transition areas between different subregions in the Mediterranean and partially related to deep sea habitats as barriers to effective migration. The estimated migration surfaces indicated the role of other factors than geographic distance driving the population structure, also supported by further seascape genetics analyses (see below). In pure isolation-by-distance (IBD) a nearly uniform estimated surface is expected as the relationship between genetic similarity and geographic distance will remain constant across the habitat.

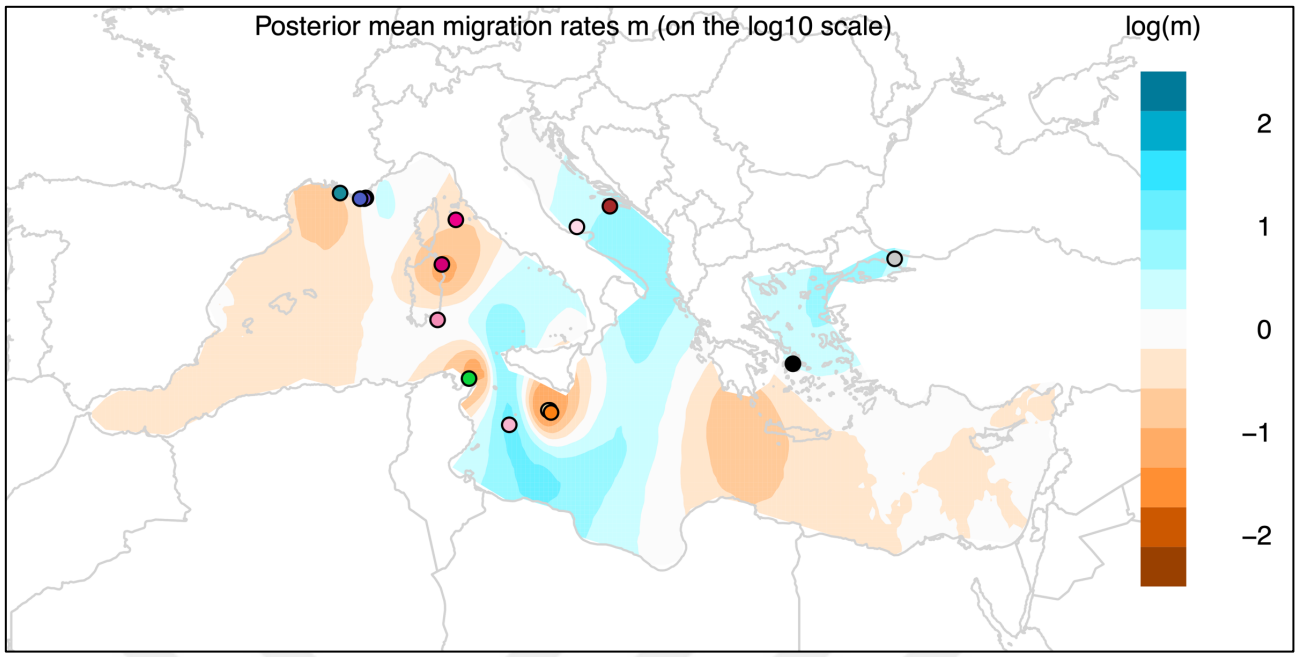


Figure 3.6. Barriers to gene flow and regions of sustained genetic connectivity in the Yelkouan Shearwater meta population. The colour scale of blue to orange represents the log10-transformed effective migration rates (m); the blue indicates areas with relatively higher migration, and orange indicates areas with relatively lower migration. Coloured circles represent the sampling locations. Same colouring used for the locations throughout this study.

3.3.2. Isolation-by-Distance and Isolation-by-Resistance

Standard Mantel test applied on least-cost path over the water (based on geographic distance) and pairwise Fst values from 12 colonies revealed a significant but weak correlation when small sample size colonies were excluded from the analysis (p-value = 0.058, r value = 0.230). The weak correlation in IBD was majorly driven by Türkiye samples; exclusion of these samples resulted in a non-significant test statistic. Evaluation of IBD within the Central and Western Mediterranean alone did not increase the strength of correlation. Least-cost path calculated by limiting the seabird movement to continental shelf did not show a significant correlation with the genetic distance (Table 3.4).

Table 3.4. The details of datasets used and results of Mantel tests applied.

Hypothesis	Dataset	Test	p-value	Mantel statistic
IBD	LC path over the water distance and Fst	Standard Mantel	0.058	0.230
IBD	LC path over the water restricted to continental shelf and Fst	Standard Mantel	0.393	0.040

IBR	Bathymetry as resistance and Fst, geographic distance as control	Partial Mantel	0.353	0.085
IBR	Wind connectivity and Fst, least cost path over the water distance as control	Partial Mantel	0.222	0.065

Model selection based on single surface resistance optimization, using only bathymetry as the resistance surface and distance as control, ranked geographical distance as the stronger statistical relationship to the genetic distance (Table 3.5). Bathymetry-related resistance matrix did not show a significant correlation to genetic distance in a partial Mantel test when distance was accounted for, but the relationship was significant in a standard Mantel test.

Table 3.5. Resistance model selection results. k indicates the number of parameters fit in each model. AIC is the original AIC value generated from the MLPE mixed effects model, while AICc is the AIC value adjusted for the number of populations sampled and the number of parameters optimized. R2m and R2c are the marginal and conditional R2 values of the fitted MLPE model, respectively. LL is the log-likelihood of the MLPE model.

Surface	Obj func LL	k	AIC	AICc	R2m	R2c	LL
Distance	398.258	2	-792.527	-791.517	0.003	0.985	398.258
Bathymetry	399.259	4	-790.518	-786.518	0.007	0.985	399.259
Null	392.210	1	-782.421	-782.113	0	0.983	392.210

3.2.3. Windscape as Resistance to Gene Flow

Windscape modelling revealed a strong directional asymmetry and geographical variation in wind speed (Figure 3.7). Wind conductance between pairs of colony locations showed variable wind travel time, correlating with distance as expected. Strength of wind connectivity (pairwise mean wind diffusion speed), which was the only evaluated parameter from windscape, varied by a factor of more than 10 across the populations, and also showed strong correlation with distance (Figure 3.8 A). Wind flow asymmetry had a U-shaped relationship with distance, where high asymmetry was present in short and long distances (Figure 3.8 B).

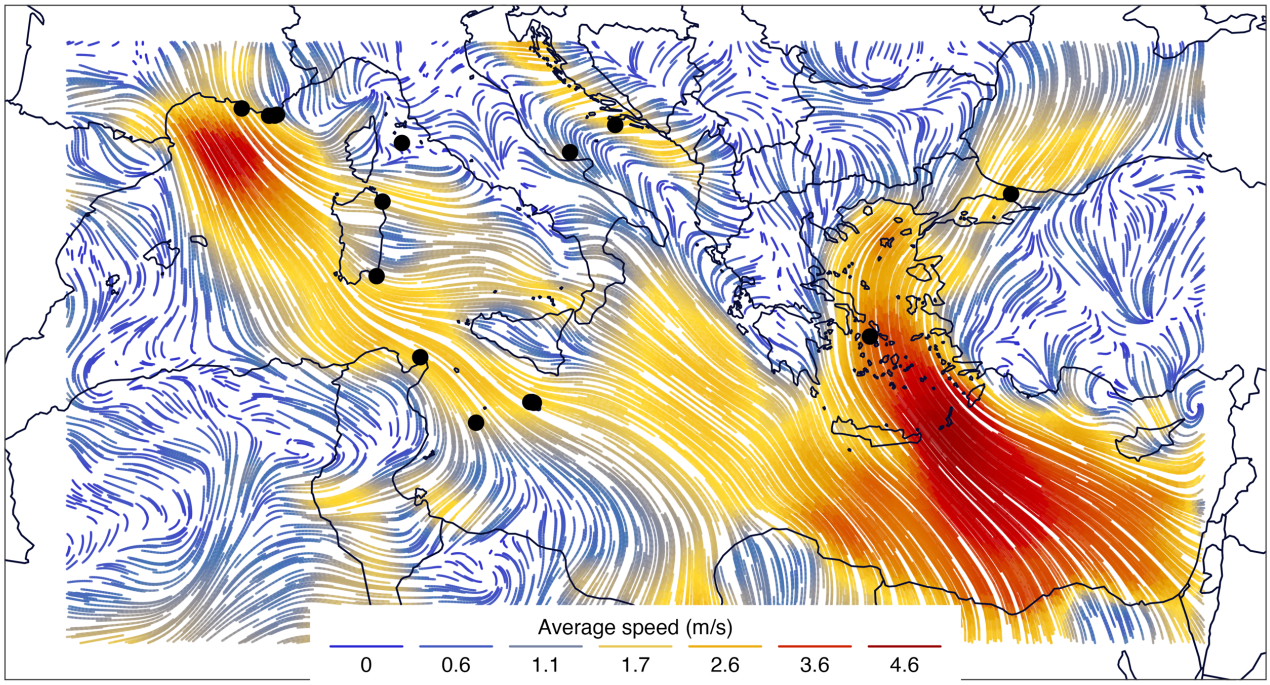


Figure 3.7. Windscape of the Mediterranean estimated from 30 years-long wind data.

Standard and partial Mantel tests were applied on pairwise Fst values, wind connectivity and least-cost path over the water distances between colony locations. Both standard and partial Mantel tests were resulted in non-significant correlation statistic based on Kendall's rank correlation (Table 3.4).

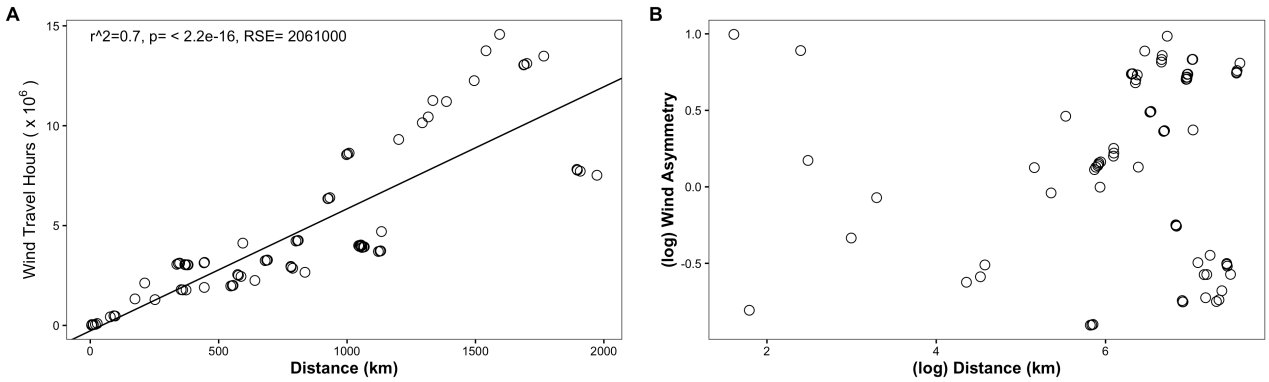


Figure 3.8. The relationship between windscape parameters and distance for colony locations. (A) Wind flow and distance and (B) wind asymmetry and distance.

3.3. The Utility of Assignment Tests in Weakly Structured Yelkouan Shearwater Population

Subsets used in assignment tests varied in number of samples and polymorphic sites (Table 3.6). First subset (all samples, uneven sample size) had 241,666 polymorphic sites and consisted of all

samples collected in breeding colonies as the reference population (n=179), Türkiye samples (n=9) as the assignment population. The second subset (high quality, equal sample size) had 759,979 sites and consisted of five highest quality samples from 12 populations (n=60) as the reference population, selected based on alignment metrics and depth and remaining samples (n=143) as the assignment population. The third subset (effective sample size) had 752,482 polymorphic sites and consisted of 106 reference samples selected based on effective sample size for each colony calculated by WGSassign, and the remaining samples (n=82) as the assignment population.

Table 3.6. Details of different subsets used and results of the leave-one-out assignment test. Leave-one-out cross validation was only applied to reference datasets and on three different spatial levels: colony, country and subregion. Groups that had the highest assignment accuracy at each level are given in the last column; FPC: France, Port-Cros colony, GGY: Greece, Gyaros colony, FR: France, WES: Western Mediterranean and CEN: Central Mediterranean.

Subset	**Sample Size**	**% LOO mean assignment accuracy**			**Site with highest assignment accuracy**
		Colony	**Country**	**Subregion**	
All samples, uneven sample size across colonies	179	32 (0 – 100)	48 (0 – 1)	42 (0 – 96)	FPC and GGY, FR and WES
Highest quality five samples based on number of aligned reads, even sample size across colonies	60	25 (0 – 80)	21 (0 – 100)	20 (0 – 100)	GGY, FR and CEN
Effective sample size from each colony as calculated by WGSassign	106	24 (0 – 100)	35 (0 – 100)	39 (0 – 92)	GGY, FR and WES

3.3.1. Leave-one-out Cross Validation

The mean accuracy of leave-one-out cross validation assignment varied between different subsets (Table 3.6). The highest accuracy was achieved in the global dataset that included all known-origin samples. Among the colonies, GGY and FPC had the highest assignment accuracy in all subsets. All individuals that were incorrectly assigned in all subsets were either assigned to a population with higher effective sample size or nearby colonies/countries (Figure 3.8). Total number of assignments to France, particularly Port-Cros (FPC) was very high in comparison to other colonies and countries in the dataset. Similarly, Western Mediterranean, which includes French colonies, was the highest assignment subregion, followed by the Central Mediterranean (Figure 3.8 A, 3.8 D and 3.8 G).

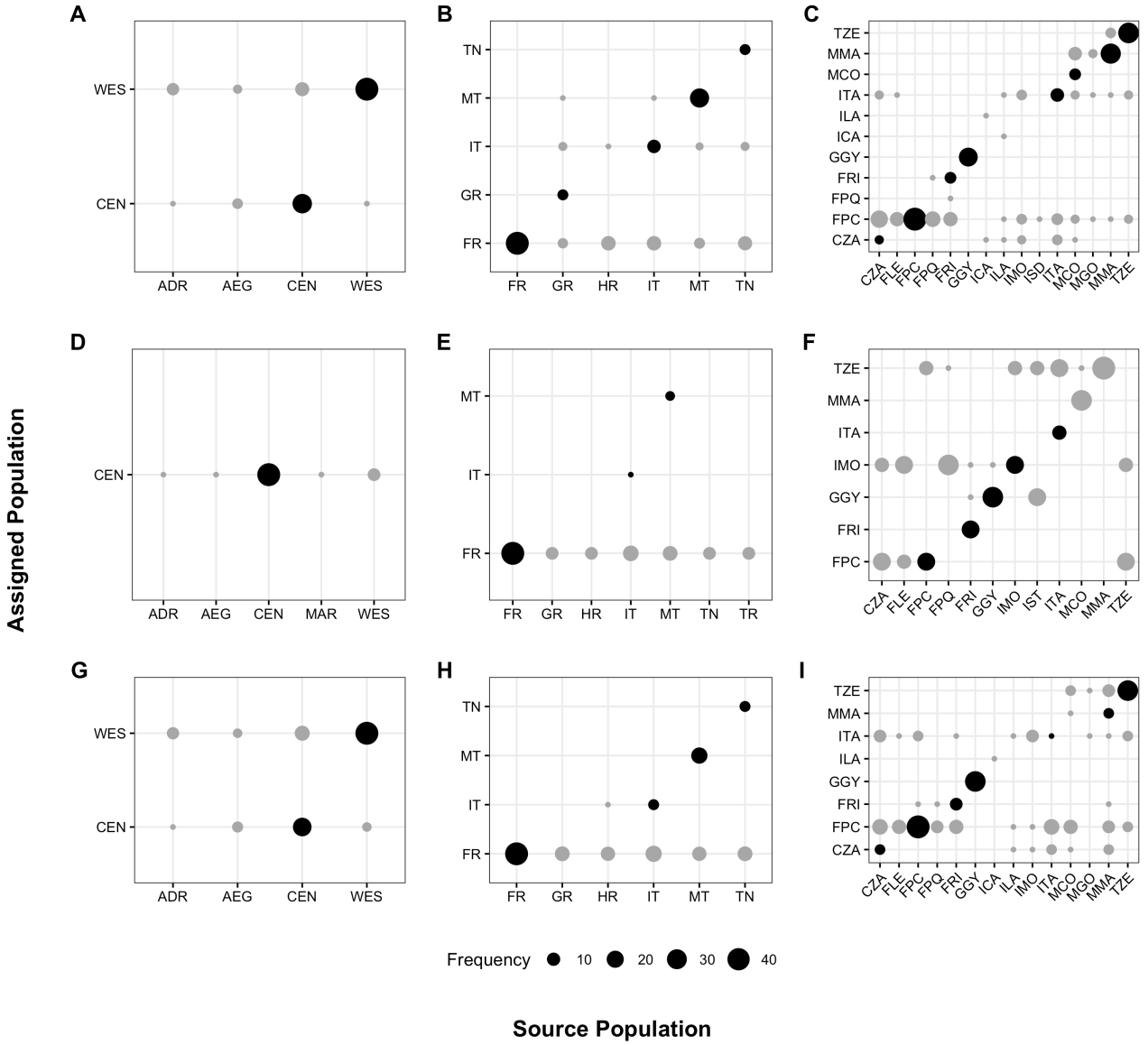
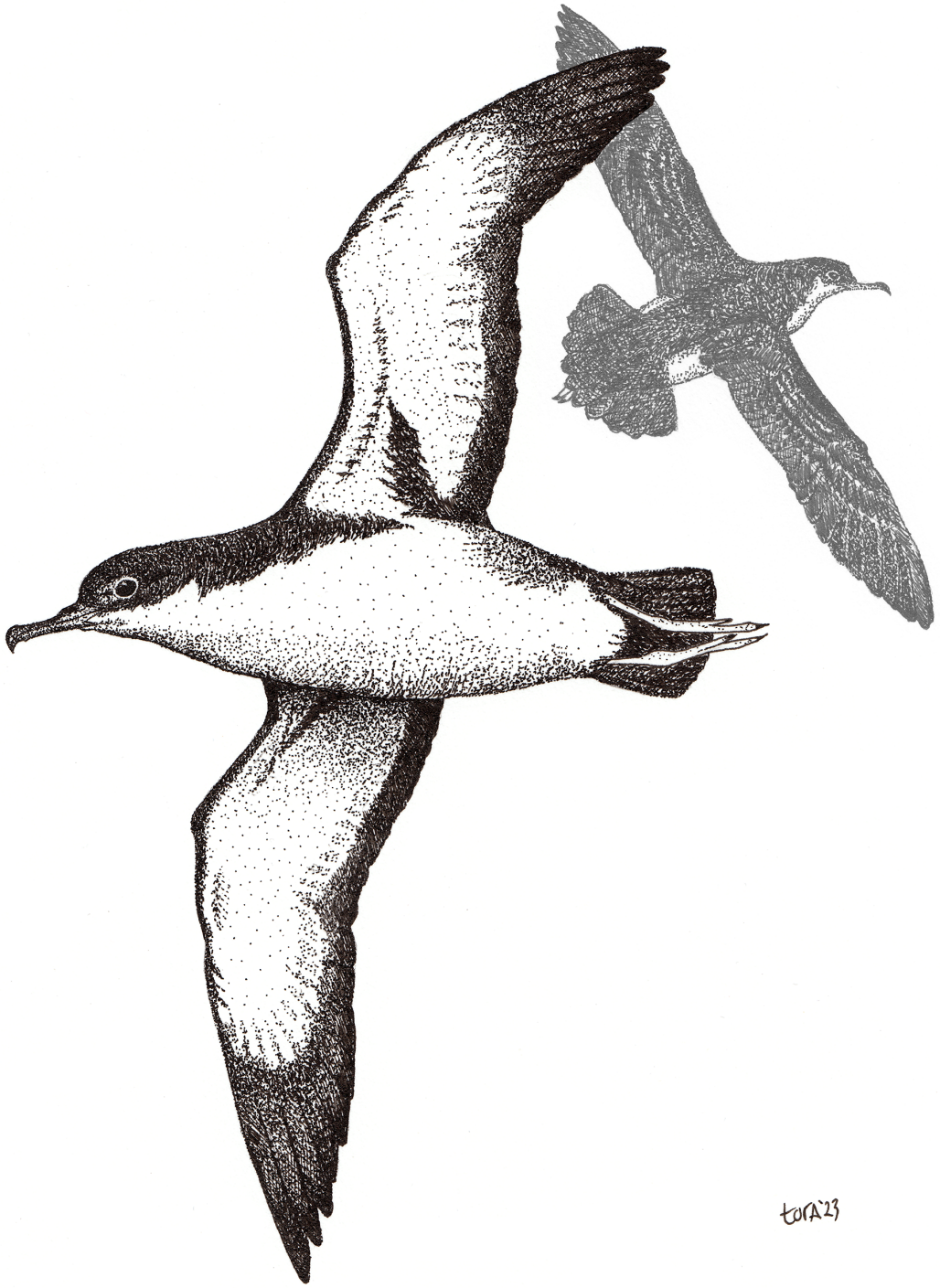


Figure 3. 9 Distribution and frequency of assignments in leave-one-out cross validation tests. Each row in the figure represents different subsets: A to C: all samples, D to F: high quality samples subset with even sample sizes and G to I effective sample size subset. Each column of figures from left to right represents subregion-based, country-based and colony-based assignment results. Subregions are based on MSFD and SPA/RAC regional marine conservation directives. Size of the points represents frequency (count) of assignment to each group (or population). Black points represent correct assignments whereas grey points represent incorrect assignments where the source population is different from the population that the sample was assigned. ADR: Adriatic, AEG: Aegean, CEN: Central Mediterranean, WES: Western Mediterranean; FR: France, GR: Greece, HR: Croatia, IT: Italy, MT: Malta, TN: Tunisia; CZA: Croatia Zaklopatica, FLE: France Levant, FPC: France Port-Cros, FPQ: France Porquerolles, FRI: France Riou, GGY: Greece Gyaros, ICA: Italy Cavoli, IMO: Italy Montecristo, ISD: Italy San Domino, ITA: Italy Tavolara, MCO: Malta Comino and Cominotto, MGO: Malta Gozo, MMA: Malta Main Island, TZE: Tunisia Zembretta.

3.3.2. Assignment of Unknown-origin Samples

The assignment of the samples with known origin resulted in similar, but slightly lower assignment accuracy than the LOO values in all subsets. In the global dataset where all samples except those from Türkiye was included in the reference population, all Türkiye samples were assigned to GGY in colony-based analysis, to Greece (n=6/9), France (n=2/9) and Malta (n=1/9) in the country-based analysis and to Central (n=6/9) and Western (n=3/9) Mediterranean in the subregion-based analysis. Similarly, in the effective sample size dataset, all Türkiye samples were assigned to GGY, to Greece (n=4/9) and France (n=5/9), and to Central (n=6/9) and Western (n=3/9) Mediterranean.


tora'23

4. DISCUSSION

4.1. ddRAD Data Reveals Fine-scale Population Genomic Structure

Despite being an endemic and endangered species, the Yelkouan Shearwater metapopulation is poorly known from various perspectives including genetic diversity and structure. This study presents the genetic diversity and structure of the Yelkouan Shearwater metapopulation for the first time, with samples from the majority of known and accessible colonies in the Mediterranean and investigated through two different markers and high throughput sequencing. mtDNA data from 196 individuals did not show significant population structure. RAD-seq data from 189 individuals and 752,480 variable sites show a gradient of clusters that could be assigned to broad subregions in the Mediterranean: (a) on the west, France, Italy and Croatia; (b) Tunisia as a transition between two groups; and (c) on the east Malta, Greece and Türkiye. Furthermore, when colonies with small sample sizes are excluded from the analysis, genetic differentiation within the metapopulation indicates signs of isolation by distance. The weak population structure is in line with the genetically and geographically close Balearic Shearwater (Genovart et al., 2007). mtDNA cytochrome b and Control Region sequences (n=113) from most breeding distribution of Balearic Shearwater and from 29 segregating sites showed weak population structure.

Similarly, Gómez-Díaz et al. (2009) found low genetic structuring in the Mediterranean population of the Scopoli's Shearwater based on mtDNA Domain I, but statistically significant differentiation around the Siculo-Tunisian Strait between eastern and western Mediterranean populations. This region is identified as a transition area for genetic differentiation also in marine organisms (Fadhlaoui-Zid et al., 2021). The results of this study show signals of this transition in clustering (see Figure 3.2 and 3.3 above) and genetic distance (Figure 3.5 above) analyses where Tunisian and to some degree Maltese populations are located intermediately between West and East Mediterranean. An exception here was for the Croatian population, which is genetically closer to the Western Mediterranean populations despite being located to the east of this channel. However, this is not counterintuitive as the Siculo-Tunisian Strait has complex oceanographic features such as seasonally different strong marine currents in the north and south of the channel (Ciappa, 2009; Reyes Suarez et al., 2019) that might influence the dispersal of organisms in different ways. Yelkouan Shearwater is also known with its varying phenology between the western and eastern Mediterranean with birds nesting on the Greek Islands showing delayed phenology in comparison to those breeding on Maltese and French Islands (Portolou pers. com.), which may be reflected in genomic data. As a

coastal species depending on seasonality (Zotier et al., 1999) it is possible that contemporary factors such as oceanography in different subregions in the Mediterranean might be contributing to the observed genetic structure. Interestingly, the EEMS results (Figure 3.6 above) indicated higher than average gene flow around the Siculo-Tunisian Strait, probably driven by the level of genetic differentiation within the Maltese Archipelago and low sample size from the Lampedusa colony (ILA). Increased sample size from these colonies would further improve the resolution of gene flow around this region.

There are no published philopatry measures of the Yelkouan Shearwater populations, however Sultana and Borg (2006) mentions higher apparent philopatry in the Maltese male Yelkouan Shearwaters in comparison to females based on unknown data sources. Genovart et al. (2012) found a degree of structuring despite observed philopatry in Balearic Shearwater and suggested that high level of inbreeding observed in the species reflects the effect of philopatry. This study also demonstrated a degree of differentiation in the Yelkouan Shearwater metapopulation that indicates historical or current gene flow between colonies. Although the level of gene flow could not be quantified, Wright (1931) predicted that as low as one migrant per generation between populations may be sufficient to prevent genetic differentiation by genetic drift. Further demographic and genetic research are needed to understand the level of philopatry and gene flow in the Yelkouan Shearwater metapopulation and its relationship to the currently observed level of genetic differentiation.

The PCA results showed two clusters in the first principal component, each including all sampled populations. The inspection of PC loadings to understand which SNPs are strongly contributing to the split in two clusters (SNP weights) did not reveal a specific region causing the split. The analysis showed a total of 260 sites that were distributed throughout the genome pulling in one direction and only one site (a gene that is involved in a variety of cellular processes) pulling in the other direction. The samples in this study were not sexed at the beginning and the coverage and depth of the sequences were not adequate for determining the sex from sequencing. These two clusters in the PCA could potentially be due to genetic split between males and females, as found in another study on White-tailed Eagles (Hansen et al., 2022) or recent colonization from two source colonies in the Mediterranean. However, further analyses are required to test these hypotheses.

Nucleotide diversity values and variation in Yelkouan Shearwater colonies were similar to those found in Balearic Shearwater mtDNA cytochrome b and control region sequences (Genovart et al., 2007). Mean value of the heterozygosity in Yelkouan Shearwater colonies ranged between 0.002 to 0.003, which is including the genome-wide heterozygosity value of closely related, endangered

Balearic Shearwater (0.002; Cuevas-Caballé et al., 2021) and within the range of genome-wide heterozygosity estimated by Cuevas-Caballé et al. (2021) for other Procellariiformes that have reference genomes available (0.001 to 0.004). Although not directly comparable due to different methods and markers used, these results suggest that genetic diversity values presented by the ddRAD sequences of the Yelkouan Shearwater metapopulation are not outside of the range of reported values for similar species that have varying population sizes.

Considering deviations from neutrality, all colonies showed negative Tajima's D values, indicating the lack of mutation-drift equilibrium. However, Tajima's D values were not outside of neutral expectations (between 2 to -2) for sampled colonies except Port-Cros (FPC), Malta (MMA) and Zembretta (TZE), suggesting an excess of low frequency polymorphism (singletons) in these three populations. Departure from neutrality in these populations suggests a recent bottleneck/founder event or genetic sweep further back in population history (Tajima, 1989). However, PCA-based selection scans on all populations for outlier loci did not reveal p-values significantly different than average, suggesting that the strong negative Tajima's D values might be related to bottleneck or genetic sweeps in the population history. All three colonies had relatively large sample sizes but the effect of sample sizes on genetic diversity and Tajima's D values cannot be ignored. Subramanian (2016) showed the negative relationship between Tajima's D values with sample size and positive change in theta values with sample size. The values obtained in this study are consistent with this pattern; larger sample colonies showed higher values of theta and lower Tajima's D.

Genetic differentiation in Procellariiform seabird populations has a wide range of Fst values (see the review by Friesen et al., 2007). It is not uncommon to see low values of Fst despite distinct clustering in PCA and shared ancestry analyses, especially in nuclear DNA markers (e.g., Rexer-Huber et al., 2019; Kersten et al., 2021; Herman et al., 2022; Ferrer Obiol et al., 2023). A review undertaken by Lombal et al. (2020) suggested lack of mutation-drift equilibrium as one of the dominating factors for the observed weak population genetic differentiation in seabirds. The Fst values obtained from nuclear DNA markers in this study indicate low differentiation between colonies but this structure corroborates the clustering obtained in PCA. However, one potential limitation is uneven sample size among colonies, which can affect Fst estimators. Bhatia et al. (2013)'s analysis using simulations and 1000 genome data cautions that commonly used Fst estimators can give inflated single SNP estimates of Fst when the sample size is small, resulting in false-positive signals of selection. To minimise these biases, we implemented methods taking small sample sizes into consideration (realSFS -whichFst 1 option, Korneliussen et al., 2014), however larger and more

evenly distributed sample sizes especially from small and unmanaged colonies could improve the inferences on genetic diversity values and Fst estimators.

The results of this study have conservation implications for the endangered Yelkouan Shearwater. The level of genetic differentiation found within a colony, such as within the Maltese archipelago, is comparable to the differentiation between distant colonies. However, current conservation management effort for this species is patchy throughout its range and greatly limited by the access to the colonies and available conservation funding. Moreover, there are still significant knowledge gaps in its breeding distribution in the Southern and Eastern Mediterranean. The weak but continuous genetic differentiation obtained in this study indicates connectivity between various populations. Therefore, Yelkouan Shearwater concerted research and conservation management efforts throughout the Mediterranean rather than major colonies or certain subregions, will be critical to preserve the genetic diversity of the species. Due to limitations in the data (i.e. genotype likelihoods instead of called genotypes), reliable estimate of effective population size and evolutionary significant units was not possible (see the review by Waples, 2024). Therefore, until further research determines evolutionary significant units, the conservation of the species should cover the extent of metapopulation.

Besides the conservation status and biology of the species, taxonomic status of the Yelkouan and Balearic Shearwaters is under discussion. The two species were delineated in 2002 (Sangster et al., 2002), however, a recent study suggested that two species are in the initial process of speciation based on the clustering, species delineation, phylogenetic and coalescent analyses with RAD-seq data and they should be merged, and be considered as subspecies (Ferrer-Obiol et al., 2023). The detailed structure representing more than 70% of the species range in this study challenges subspecies recommendation by Ferrer Obiol et al. (2023), which is based on samples collected from France, Tunisia and Croatia colonies only. This study shows that these groups cluster tightly together, except for some Tunisian samples that are distinct from France and Croatia samples, despite representing Western and Central Mediterranean and that metapopulation structure is richer and more complex when the Eastern Mediterranean samples are added. Combining the samples from two studies and re-analysing the genomic distance and exchange between the two species would provide more detailed insight into their taxonomic status.

This study is valuable in using RAD-seq data since most of the seabird population genetics studies, especially in the Mediterranean used microsatellites and mitochondrial DNA genes as markers. Only recently seabird studies started to utilize high throughput sequencing and this study is

the first one for the Yelkouan Shearwater in the Mediterranean. However, even though it is providing a much higher resolution information on the species population genomics, higher and more evenly distributed sample size and more advanced sequencing techniques could improve the results obtained in this study. The true range of this species is not well known and therefore the sampling could be improved as the knowledge on its breeding distribution advances, especially in the Eastern Mediterranean.

4.2. Seascape Habitat Configuration as Resistance to Gene Flow

Given the prevalence of weak structure found in pelagic seabird populations (Friesen et al., 2007; Lombal et al., 2020), an important knowledge gap remains is whether and to what degree the resistance posed by habitat configuration is driving the fine-scale genetic structure. This study evaluated the relationship between seascape features and genetic differentiation in the Yelkouan Shearwater metapopulation in the Mediterranean. Isolation-by-distance and isolation-by-resistance hypotheses tested colony-based pairwise Fst values against geographic distance and seascape resistance surfaces. The results revealed areas of low historical gene flow in different subregions of the Mediterranean, but isolation-by-distance as the strongest statistical relationship among the tested resistance surfaces in this study. Based on estimated effective migration and resistance surfaces evaluated against Fst, the results highlight the need for building more complex resistance models and using different genetic distance measures such as gene flow direction and strength for an improved understanding of the relationship between seascape and genetic structure in the Yelkouan Shearwater metapopulation.

Least-cost path over the water as geographic distance weakly correlated with genetic differentiation when the low-sample size colonies, which were characterized by high Fst values in short geographic distances, were removed from the analysis. Weak isolation-by-distance was found in other similar species such as Cory's Shearwater in the Mediterranean and Atlantic (Genovart et al., 2013), Atlantic Puffin in the North Atlantic (Kersten et al., 2021) and Band-rumped Storm-petrels in North-east Atlantic (Smith et al., 2007), and was suggested as a weak explanation for the population genetic structure in seabirds (Friesen et al., 2007).

Isolation-by-distance is considered as a baseline pattern that is observed when the landscape elements do not have additional effects on gene flow (Balkenhol et al., 2009). However, through simulations, van Strien et al. (2015) showed that habitat configuration can have a strong influence on isolation-by-distance patterns and therefore these patterns should not be regarded as resulting from

only distance-constrained migration. Given their high mobility, it is possible that the observed weak isolation-by-distance in Yelkouan Shearwater metapopulation is a product of habitat configuration rather than limited migration. Indeed, the EEMS results showed areas of lowered historical gene flow and not a uniform migration surface that would be expected under pure isolation-by-distance scenario (Petkova et al., 2016). Other than land masses acting as barriers to gene flow, mechanisms related to habitat configuration limiting dispersal in large distances in seabird populations could be segregating distribution at sea due to proximity to productive areas. This pattern could also be due to environmental factors increasing the cost of movement such as vast areas of deep-sea habitats or unfavourable winds.

Based on lowered gene-flow around deep-sea habitats in EEMS results, evaluation of bathymetry as a resistance surface did not reveal a significant relationship with pairwise Fst values. Shearwater distributions at sea are known to be associated with different aspects of bathymetry. For example, Cory's Shearwater distribution was occurring in greater intensity around some seamounts (Amorim et al., 2009), whereas Balearic Shearwater distribution in the breeding season was related to frontal systems along the continental shelf waters (depth 200 m) with elevated chlorophyll-a concentrations (Louzao et al., 2006). For the Yelkouan Shearwater, a detailed analysis on the relationship of its distribution to bathymetry is lacking but the movement and at-sea count data show that at-sea distribution during the breeding season is associated with shallow marine habitats close to breeding grounds (Metzger et al., 2015; Pezzo et al., 2021). However, movement data also show that Yelkouan Shearwaters cross these areas in post-breeding migration (Raine et al., 2013). It is possible that despite utilizing shallow habitats during the breeding season, high mobility and being able to rest on the water during migration might be facilitating the deep-sea crossing for Yelkouan Shearwater and preventing the association with shallow habitats to translate into gene flow.

Another environmental variable hypothesised in this study as a potential contributor to resistance to gene flow was wind. The evaluation of wind flow between pairs of colony locations did not reveal a significant relationship with genetic differentiation between these points. The hypothesis tested here was that the level of wind flow between colonies influences the genetic differentiation, positively or negatively. The lack of correlation is possibly due to strong asymmetry in wind direction and strength, as revealed by the windscape model. Due to limitations in quality of genetic data, it was not possible to quantify gene flow between different population pairs, which is an asymmetric measure, to test for the relationship between wind asymmetry and asymmetric gene flow in this study. The lack of correlation could also stem from limitations in wind data. Specifically, the available data for wind altitude used in this study was 10 meters. Seabirds use wind in lower altitudes, right above the surface

of ocean. As an alternative to wind, surface waves, which are driven by wind (Pizzo et al., 2021) or wind waves could provide more relevant resistance surfaces and requires further analyses. So far, there are no studies evaluating the seabird genetic differentiation or gene flow in the light of wind patterns. Kling and Ackerly (2021) demonstrated the strong influence of wind on genetic structure of wind-dispersing plant populations and found that populations linked by stronger wind patterns show higher genetic similarity, whereas wind asymmetry was related to directional gene flow. Wind dispersal models have been developed mainly for plants, assuming isotropic and spatially uniform potential (Kuparinen, 2006; Nathan et al., 2011). The way seabirds use the wind is far more complex than passively dispersing organisms (Thorne et al., 2023). In this perspective, this study provides the first attempt in quantifying the impact of wind on genetic structure explicitly, however, further research should focus on wind and gene flow asymmetry and building more complex wind models e.g., by including crosswind component. Finally, this study used wind resistance as a contemporary factor shaping the genetic structure in Yelkouan Shearwater metapopulation. It is also possible that the quantified structure is dominated by historical factors (Lombal et al., 2020).

Despite limitations in the genetic data, this study provides a valuable contribution to literature, as it takes the first step in evaluating the relationship between seascape features and genetic structure in the Yelkouan Shearwater population and is the first study attempting to quantify the wind-genetic structure relationship in pelagic seabirds. Although covering the extent of Yelkouan Shearwater range, the sampling does not include all colonies that Yelkouan Shearwater is known to be nesting. Adding samples from these colonies may improve the isolation-by-distance or -resistance analyses, through improving the discrete sampling of the IBD and IBR gradients, respectively. In addition, obtaining different genetic differentiation measures such as direction and strength of gene flow and genetic diversity measures could allow for testing more hypothesis related to isolation-by-resistance. However, despite being an important determinant in seabird distribution, it is possible that selected resistance surfaces do not relate to the gene flow meaningfully across the seascape regardless of the quality of genetic data. Dispersal is still one of the least known aspects of seabird biology (Lewison et al., 2012), however, advancing landscape genomics models and methods hold a great potential in investigating the complex relationship between seabird movement, seascape, and gene flow. This study provides an essential first analysis in explicit quantification of gene flow-seascape relationship and highlights the need for developing the resistance models geared towards seabird movement and obtaining different genetic distance measures to achieve more robust analyses.

4.3. Utility of Assignment Tests in Weakly Structured Yelkouan Shearwater Metapopulation

This study used genotype likelihoods estimated from low-quality ddRAD data and weak population structure to test the performance of assignment tests. The results highlighted the importance of sample size and evenness in increasing the performance of population assignment models. Based on different subsets of data with varying sample size and sample quality, this study demonstrated weak but consistent and promising performance of assignment tests in assigning individuals to the colony of origin based on the ddRAD sequencing data and genotype likelihoods obtained across the breeding range of the Yelkouan Shearwater.

Reliable genetic assignment tests in weakly structured populations recently became possible as the NGS technology increased the number of available markers significantly and enabled selecting informative loci (e.g. Benestan et al., 2015; Ruegg et al., 2014). However, the number of comparable studies using genotype likelihoods rather than called genotypes in assignment tests is still limited. DeSaix et al. (2023) used genotype likelihoods in weakly structured American Redstart populations (Fst 0.004 to 0.018) and obtained 89% LOO and 90% assignment accuracy by using effective sample sizes. Similarly, Baetscher et al. (2022) obtained up to 91% LOO assignment success in Northern Fulmars (Fst 0.003 to 0.01) using downsampled genotype data to overcome the bias stemming from uneven sample sizes. The assignment accuracy obtained in the Yelkouan Shearwater data is lower, however, given lower Fst values in these studies, it can be argued that lower assignment accuracy obtained in this study is due to sampling effects rather than weak population structure. Indeed, even LOO assignment accuracy did not approach to 1 for many groups and the incorrect assignments were majorly to the large sample size groups in the dataset. However, when groups were combined under countries or regions (i.e., increasing the sample size in each group), higher LOO assignment accuracies were obtained, which supports the hypothesis of influence of sample size on accuracy. These results point out a potential bias large sample size groups are creating in estimated allele frequencies. Both Baetscher et al. (2022) and DeSaix et al. (2023) also obtained higher assignment accuracy for the larger sample size groups in their studies. Benestan et al. (2015) demonstrated the impact of sample size on assignment accuracy in American Lobster ddRAD data with very weak population structure (mean Fst= 0.002). They reached assignment accuracy of 50% or higher with more than 20 individuals per sampling location and after downsampling based on highest Fst SNPs (i.e., informative sites). In this study, the highest sample size in any colony was 25 (FPC) and therefore it was not possible to test the assignment accuracy in similar sample sizes without excluding most of the colonies, which would introduce further bias due to unsampled locations in the dataset. However, obtaining the highest accuracy values in the subset with all samples is an indication that similar results would be reached if there were more samples from each colony.

DeSaix et al. (2023) reached the best assignment results by using the effective sample size dataset. The subset with effective sample sizes resulted in lower assignment accuracy in this study when compared to results obtained by using all samples. This might be due to the lack of a robust selection criteria for the samples. WGSassign determines effective sample size for each group based on calculated Fisher information and the effective sample size is defined as "number of samples with completely observed genotypes that would yield the same amount of statistical information for estimating allele frequency as the observed genotype likelihood in a dataset" (DeSaix et al., 2023, p. 5). However, WGSassign does not produce the list samples that are contributing to the effective sample size, instead it also outputs the effective sample size for each individual in the same run. In this study, these individual effective sample sizes were used to select the samples for the effective sample size dataset. Accordingly, individual effective sample sizes were first downsampled to the range of 0.5 to 0.85 and the highest effective sample size individuals were selected in each group to reach the required group effective sample size. However, selection of samples that will form the effective sample size subset requires the identification of samples that are maximising the allele frequency estimation in the group and needs further research and development of methods.

Assignment accuracy of assumed-unknown-origin samples were close but slightly lower than reference samples tested in LOO cross validation, likely due to same reasons discussed above. Given the highest average assignment accuracy obtained in the global dataset and colony-based assignment, the probability that Türkiye samples originated from Greece colonies is the highest. Indeed, from the recent tracking studies (Hellenic Ornithological Society, pers. com.), Yelkouan Shearwaters are found to be crossing the Dardanelles and Bosporus frequently in their foraging trips during the breeding season. These specimens were collected in May, which is the incubation period for Yelkouan Shearwaters breeding in Greece. However, there is no way to tell whether the sampled birds were breeding adults or prospecting non-breeding birds. Considering the weak genomic structure obtained with the ddRAD data and moderate assignment accuracy, the origin of Türkiye samples is most likely Greece, but this inference is not certain.

The available literature shows that risk of bycatch for Mediterranean seabirds is present in all subregions in the Mediterranean but is particularly high in the Western Mediterranean (Barcelona et al., 2010; Belda & Sanchez, 2001; Cortés et al., 2017; Laneri et al., 2010). With the assumption of birds breeding in the Western Mediterranean (French and Italian colonies in this study) are exposed to the highest bycatch risk in the region, the results of this study show that using the current dataset, it is possible to assign Western Mediterranean samples at least in the country-level with relatively high accuracy.

The Yelkouan Shearwater ddRAD data used in this study has several limitations, primarily low coverage and uneven sample sizes among groups but also the potential effect of unsampled colonies. Although covering the extent of breeding range in the Mediterranean, the dataset used in this study does not include data from several other islands that Yelkouan Shearwater is known to be nesting, particularly in Italy and Greece. Assignment accuracy for the Greek colony was high in all colony-level analyses, possibly due to higher differentiation of this group than Central and Western Mediterranean colonies as supported by PCA, Admixture and Fst values. However, assignment accuracy of Italian colonies could potentially be improved with the additional samples and sampling locations. At the subregion level, the importance of having even sample sizes in each group was more pronounced; despite giving good results at the colony level, Greek samples did not produce high accuracy values at the subregion and country level analyses due to much smaller sample size in comparison to the other groups.

Considering the use of old blood samples (5+ years old) that went through several freeze-thaw cycles (i.e. level of DNA degradation in samples), this study was able to obtain promising results even at the colony-level assignment. The results of this study suggest that bycatch specimens with some level of DNA degradation would yield sufficient and useful results in assignment tests, especially when there is a more robust reference dataset. This study is also valuable as it is making a dataset available for future assignment studies using NGS. This is especially challenging for non-model organisms and happens in incremental steps as more sequencing is done. Mediterranean seabirds particularly suffer from the lack of such datasets and therefore the assignment tests with limited number of samples and low-coverage data is expected to provide a good reference and starting point for the future studies. To maximise the assignment accuracy by using similar data for the Yelkouan Shearwater population, sample size should be evenly distributed across the sampling locations and arguably more sampling locations should be covered in the Central and Eastern Mediterranean.


torA`23

5. CONCLUSIONS AND RECOMMENDATIONS

This study presents weak but continuous population genomic structure in the Yelkouan Shearwater metapopulation with the indication of Siculo-Tunisian strait acting as a transition zone. Subset of the dataset that is more balanced in sample size distribution supports the signs of isolation-by-distance as the weak, but main driving force of this structure. However, the apparent impact of geographic distance on the genetic distance might also be related to the resistance posed by complex habitat configuration to gene flow. Indeed, the results of this study offer insight into the complexity of factors underlying the genomic structure in the Yelkouan Shearwater metapopulation. While the selected resistance surfaces did not show a significant relationship with genetic structure, the study highlights the need for more complex resistance models and different genetic distance parameters in the implementation of seascape genomics methods in seabird populations.

Investigating all possible factors leading to the observed structure is beyond the scope of this study, however, higher resolution data such as whole genome sequences and more and equal sample size from each colony would improve the inference of this structure and its relationship to the historical and temporary factors. This study reveals the continuous genetic structure of the species and supports the hypothesis of Central Mediterranean being a genetic transition zone between the Western and Eastern Mediterranean, and therefore, management of the species should focus evenly on these three subregions. Investigation on the presence of population-specific variable sites between the Eastern and Western Mediterranean populations could improve the management decisions for the endemic and threatened Yelkouan Shearwater.

This study demonstrated that low quality ddRAD data obtained across the breeding range of the Yelkouan Shearwater is suited for moderately accurate assignment performance. This is an important finding as the bycatch specimens are not always processed immediately after mortality and therefore can provide low quality results in DNA extraction and sequencing. In this sense, this study demonstrates promising outcomes in such scenarios. The results of the assignment tests with different subsets highlighted the need for improved sampling, particularly evenly distributed sample sizes in increasing the assignment accuracy for this metapopulation. Further sequencing should add more ddRAD samples or upgrade the sequencing to be able to make robust reference data available for applied studies.

Due to limitations in the genetic data, this study could not make further conservation recommendations, however, demonstrates that genomic data of the Yelkouan Shearwater metapopulation could play a significant role in conservation management of the species.

REFERENCES

Abbott, C. L., Double, M. C., Gales, R., Baker, G. B., Lashko, A., Robertson, C. J. R., & Ryan, P. G. (2006). Molecular provenance analysis for shy and white-capped albatrosses killed by fisheries interactions in Australia, New Zealand, and South Africa. *Conservation Genetics*, *7*(4), 531–542.

Amorim, P., Figueiredo, M., Machete, M., Morato, T., Martins, A., & Serrão Santos, R. (2009). Spatial variability of seabird distribution associated with environmental factors: a case study of marine Important Bird Areas in the Azores. *ICES Journal of Marine Science*, *66*(1), 29–40.

Anderson, E. C. (2010). Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, *10*(4), 701–710.

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

Austin, J. J., Bretagnolle, V., & Pasquet, E. (2004). A Global Molecular Phylogeny of the Small Puffinus Shearwaters and Implications for Systematics of the Little-Audubon's Shearwater Complex. *The Auk*, *121*(3), 847–864.

Baetscher, D. S., Beck, J., Anderson, E. C., Ruegg, K., Ramey, A. M., Hatch, S., Nevins, H., Fitzgerald, S. M., & Carlos Garza, J. (2022). Genetic assignment of fisheries bycatch reveals disproportionate mortality among Alaska Northern Fulmar breeding colonies. *Evolutionary Applications*, *15*(3), 447–458.

Balkenhol, N., Waits, L. P., & Dezzani, R. J. (2009). Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography*, *32*(5), 818–830.

Barbosa, S., Hendricks, S. A., Funk, W. C., Rajora, O. P., & Hohenlohe, P. A. (2020). Wildlife Population Genomics: Applications and Approaches. In P. A. Hohenlohe & O. P. Rajora (Eds.), *Population Genomics: Wildlife*. Springer.

Barcelona, S. G., Ortiz de Urbina, J. M., de la Serna, J. M., Alot, E., & Macías, D. (2010). Seabird bycatch in Spanish Mediterranean large pelagic longline fisheries, 2000-2008. *Aquat. Living Resour.*, *23*(4), 363–371.

Belda, E. J., & Sanchez, A. (2001). Seabids mortality on longline fisheries in the western Mediterranean: factors affecting bycatch and proposed mitigation measures. *Biological Conservation*, *98*, 357–363.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (Homarus americanus). *Molecular Ecology*, *24*(13), 3299–3315.

Benkwitt, C. E., Carr, P., Wilson, S. K., & Graham, N. A. J. (2022). Seabird diversity and biomass enhance cross-ecosystem nutrient subsidies. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1974), 20220195.

Benkwitt, C. E., D'Angelo, C., Dunn, R. E., Gunn, R. L., Healing, S., Mardones, M. L., Wiedenmann, J., Wilson, S. K., & Graham, N. A. J. (2023). Seabirds boost coral reef resilience. *Science Advances*, *9*(49), eadj0390.

Berry, O., Tocher, M. D., & Sarre, S. D. (2004). Can assignment tests measure dispersal? *Molecular Ecology*, *13*(3), 551–561.

Bhatia G, Patterson N, Sankararaman S, & Price A L. (2013). Estimating and interpreting FST: the impact of rare variants. *Genome Research*, *23*(9), 1514–1521.

Bicknell, A. W. J., Knight, M. E., Bilton, D., Reid, J. B., Burke, T., & Votier, S. C. (2012). Population genetic structure and long-distance dispersal among seabird populations: Implications for colony persistence. *Molecular Ecology*, *21*(12), 2863–2876.

BirdLife International. (2023). *Species factsheet: Puffinus yelkouan*. http://datazone.birdlife.org/species/factsheet/yelkouan-shearwater-puffinus-yelkouan

Bonnaud, E., Bourgeois, K., Vidal, E., Legrand, J., & Le Corre, M. (2009). How can the Yelkouan shearwater survive feral cat predation? A meta-population structure as a solution? *Population Ecology*, *51*(2), 261–270.

Bormans, M., Garrett, C., & Thompson, K. (1986). Seasonal variability of the surface inflow through the strait-of-gibraltar. *Oceanologica Acta*, *9*(4), 403–414.

Bourgeois, K., & Vidal, E. (2008). The endemic Mediterranean yelkouan shearwater *Puffinus yelkouan*: distribution, threats and a plea for more data. *Oryx*, *42*(02), 187–194.

Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (Hyla arborea) identify the sex chromosome without information on offspring sex. *Heredity*, *116*(2), 177–181.

Bresadola, L., Link, V., Buerkle, C. A., Lexer, C., & Wegmann, D. (2020). Estimating and accounting for genotyping errors in RAD-seq experiments. *Molecular Ecology Resources*, *20*(4), 856–870.

Broad Institute. (2023). *Picard Tools*. http://broadinstitute.github.io/picard. Accessed January 15, 2024; Version 3.1.1.

Brown, R. M., Nichols, R. A., Faulkes, C. G., Jones, C. G., Bugoni, L., Tatayah, V., Gottelli, D., & Jordan, W. C. (2010). Range expansion and hybridization in Round Island petrels (Pterodroma spp.): evidence from microsatellite genotypes. *Molecular Ecology*, *19*(15), 3157–3170.

Byerly, P. A., Chesser, R. T., Fleischer, R. C., McInerney, N., Przelomska, N. A. S., & Leberg, P. L. (2023). Conservation genomics reveals low connectivity among populations of threatened roseate terns (Sterna dougallii) in the Atlantic Basin. *Conservation Genetics*, *24*(3), 331–345.

Cagnon, C., Lauga, B., Hemery, G., & Mouches, C. (2004). Phylogeographic differentiation of storm petrels (Hydrobates pelagicus) based on cytochrome b mitochondrial DNA variation. *Marine Biology*, *145*(6), 1257–1264.

Cairns, D. K. (1987). Seabirds as indicators of marine food supplies. *Biological Oceanography*, *5*(August), 261–271.

Carboneras, C., & Requena-Moreno, S. (2010, May). GIS mapping of seabird distribution--a pan-Mediterranean perspective. *Conference: Rapport Commission International Mer Méditerranée*.

Carvalho, P. C., & Davoren, G. K. (2019). Associations of non-breeding shearwater species on the northeastern Newfoundland coast. *Marine Ecology Progress Series*, *627*, 1–12.

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.

Chafin, T. K., Martin, B. T., Mussmann, S. M., Douglas, M. R., & Douglas, M. E. (2018). FRAGMATIC: in silico locus prediction and its utility in optimizing ddRADseq projects. *Conservation Genetics Resources*, *10*(3), 325–328.

Ciappa, A. C. (2009). Surface circulation patterns in the Sicily Channel and Ionian Sea as revealed by MODIS chlorophyll images from 2003 to 2007. *Continental Shelf Research*, *29*(17), 2099–2109.

Coll, M., Piroddi, C., Steenbeek, J., Kaschner, K., Ben Rais Lasram, F., Aguzzi, J., Ballesteros, E., Bianchi, C. N., Corbera, J., Dailianis, T., Danovaro, R., Estrada, M., Froglia, C., Galil, B. S., Gasol, J. M., Gertwagen, R., Gil, J., Guilhaumon, F., Kesner-Reyes, K., … Voultsiadou, E. (2010). The Biodiversity of the Mediterranean Sea: Estimates, Patterns, and Threats. *PLOS ONE*, *5*(8), e11842.

Cortés, V., Arcos, J. M., & González-Solís, J. (2017). Seabirds and demersal longliners in the northwestern Mediterranean: factors driving their interactions and bycatch rates. *Marine Ecology Progress Series*, *565*, 1–16.

Cossarini, G., Feudale, L., Teruzzi, A., Bolzon, G., Coidessa, G., Solidoro, C., Di Biagio, V., Amadio, C., Lazzari, P., Brosich, A., & Salon, S. (2021). High-Resolution Reanalysis of the Mediterranean Sea Biogeochemistry (1999–2019). *Frontiers in Marine Science*, *8*.

Crespin, L., Harris, M. P., Lebreton, J. D., Frederiksen, M., & Wanless, S. (2006). Recruitment to a seabird population depends on environmental factors and on population size. *Journal of Animal Ecology*, *75*(1), 228–238.

Cristofari, R., Plaza, P., Fernández, C. E., Trucchi, E., Gouin, N., Le Bohec, C., Zavalaga, C., Alfaro-Shigueto, J., & Luna-Jorquera, G. (2019). Unexpected population fragmentation in an endangered seabird: the case of the Peruvian diving-petrel. *Scientific Reports*, *9*(1), 2021.

Croxall, J. P., Butchart, S. H. M., Lascelles, B., Stattersfield, A. J., Sullivan, B., Symes, A., & Taylor, P. (2012). Seabird conservation status, threats and priority actions: a global assessment. *Bird Conservation International*, *22*(01), 1–34.

Cuevas-Caballé, C., Obiol, J. F., Vizueta, J., Genovart, M., Gonzalez-Solís, J., Riutort, M., & Rozas, J. (2021). The genome of the Balearic shearwater (*Puffinus mauretanicus*), a Critically Endangered seabird: a valuable resource for evolutionary and conservation genomics. *BioRxiv*, 2021.12.17.473171.

Czapanskiy, M. F. (2018). *Using energy landscapes to understand seabird movement and spatial ecology* [Masters Thesis]. San Francisco State University.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008.

Dantas, G. P. M., Oliveira, L. R., Santos, A. M., Flores, M. D., Melo, D. R. de, Simeone, A., González-Acuña, D., Luna-Jorquera, G., Le Bohec, C., Valdés-Velásquez, A., Cardeña, M., Morgante, J. S., & Vianna, J. A. (2019). Uncovering population structure in the Humboldt penguin (*Spheniscus humboldti*) along the Pacific coast at South America. *PLOS ONE*, *14*(5), e0215293.

Dayton, J., & Szczys, P. (2021). Metapopulation connectivity retains genetic diversity following a historical bottleneck in a federally endangered seabird. *Ornithological Applications*, *123*(4), duab037.

Dearborn, D. C., Anders, A. D., Schreiber, E. A., Adams, R. M. M., & Mueller, U. G. (2003). Inter-island movements and population differentiation in a pelagic seabird. *Molecular Ecology*, *12*(10), 2835–2843.

Derhé, M. A. (2012). Population assessment for the Yelkouan Shearwater Puffinus yelkouan. In *Methodology for Bird Species Recovery Planning in the European Union. Final Report to the European Commission.*

DeSaix, M. G., Anderson, E. C., Bossu, C. M., Rayne, C. E., Schweizer, T. M., Bayly, N. J., Narang, D. S., Hagelin, J. C., Gibbs, H. L., Saracco, J. F., Sherry, T. W., Webster, M. S., Smith, T. B.,

Marra, P. P., & Ruegg, K. C. (2023). Low-coverage whole genome sequencing for highly accurate population assignment: Mapping migratory connectivity in the American Redstart (Setophaga ruticilla). *Molecular Ecology*, *32*(20), 5528–5540.

DeSaix, M., Rodriguez, M., Ruegg, K., & Anderson, E. (2023). Population assignment from genotype likelihoods for low-coverage whole-genome sequencing data. *Authorea*.

Dias, M. P., Granadeiro, J. P., & Catry, P. (2012). Do Seabirds Differ from Other Migrants in Their Travel Arrangements? On Route Strategies of Cory's Shearwater during Its Trans-Equatorial Journey. *PLOS ONE*, *7*(11), e49376.

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*, 525–537.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797.

Edwards, J. A., & Edwards, R. A. (2019). Fastq-pair: efficient synchronization of paired-end fastq files. *BioRxiv*, 552885.

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63.

Elliott, K. H., Chivers, L. S., Bessey, L., Gaston, A. J., Hatch, S. A., Kato, A., Osborne, O., Ropert-Coudert, Y., Speakman, J. R., & Hare, J. F. (2014). Windscapes shape seabird instantaneous energy costs but adult behavior buffers impact on offspring. *Movement Ecology*, *2*(1), 17.

European Parliament. (2008). Establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive), Pub. L. No. Directive 2008/56/EC (2008).

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

Fadhlaoui-Zid, K., Cossu, P., Sanna, D., Scarpa, F., Lai, T., Castelli, A., Casu, M., & Maltagliati, F. (2021). Spatial genetic patterns of Octopus vulgaris Mediterranean populations support the hypothesis of a transitional zone across the Siculo-Tunisian Strait. *Hydrobiologia*, *848*(18), 4225–4240.

Fauchald, P. (2010). Predator–prey reversal: A possible mechanism for ecosystem hysteresis in the North Sea? *Ecology*, *91*(8), 2191–2197.

Fay, R., Weimerskirch, H., Delord, K., & Barbraud, C. (2015). Population density and climate shape early-life survival and recruitment in a long-lived pelagic seabird. *Journal of Animal Ecology*, *84*(5), 1423–1433.

Ferrer-Obiol, J., Herranz, J. M., Paris, J. R., Whiting, J. R., Rozas, J., Riutort, M., & González-Solís, J. (2023). Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. *Molecular Phylogenetics and Evolution*, *179*, 107671.

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, *17*(1), 27–32.

Frederiksen, M., Mavor, R. a., & Wanless, S. (2007). Seabirds as environmental indicators: The advantages of combining data sets. *Marine Ecology Progress Series*, *352*, 205–211.

Friesen, V. L., Burg, T. M., & McCoy, K. D. (2007). Mechanisms of population differentiation in seabirds: Invited review. *Molecular Ecology*, *16*(9), 1765–1785.

Galinsky, K. J., Bhatia, G., Loh, P.-R., Georgiev, S., Mukherjee, S., Patterson, N. J., & Price, A. L. (2016). Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *The American Journal of Human Genetics*, *98*(3), 456–472.

García-Dávila, C., Duponchelle, F., Castro-Ruiz, D., Villacorta, J., Quérouil, S., Chota-Macuyama, W., Núñez, J., Römer, U., Carvajal-Vallejos, F., & Renno, J.-F. (2013). Molecular identification of a cryptic species in the Amazonian predatory catfish genus Pseudoplatystoma (Bleeker, 1962) from Peru. *Genetica*, *141*(7), 347–358.

Gatt, M. C., Lago, P., Austad, M., Bonnet-Lebrun, A.-S., & Metzger, B. J. (2019). Pre-laying movements of Yelkouan Shearwaters (Puffinus yelkouan) in the Central Mediterranean. *Journal of Ornithology*, *160*(3), 625–632.

Gauthier, G., Milot, E., & Weimerskirch, H. (2010). Small-scale dispersal and survival in a long-lived seabird, the wandering albatross. *Journal of Animal Ecology*, *79*(4), 879–887.

Genovart, M., Doak, D. F., Igual, J. M., Sponza, S., Kralj, J., & Oro, D. (2017). Varying demographic impacts of different fisheries on three Mediterranean seabird species. *Global Change Biology*, *October 2016*, 3012–3029.

Genovart, M., Juste, J., Contreras-Diaz, H., & Oro, D. (2012). Genetic and phenotypic differentiation between the critically endangered Balearic shearwater and neighboring colonies of its sibling species. *The Journal of Heredity*, *103*(3), 330–341.

Genovart, M., Oro, D., Juste, J., & Bertorelle, G. (2007). What genetics tell us about the conservation of the critically endangered Balearic Shearwater? *Biological Conservation*, *137*(2), 283–293.

Genovart, M., Thibault, J.-C., Igual, J. M., Bauza-Ribot, M. del M., Rabouam, C., & Bretagnolle, V. (2013). Population structure and dispersal patterns within and between Atlantic and Mediterranean populations of a large-range pelagic seabird. *PloS One*, *8*(8), e70711.

Gianuca, D., Phillips, R. A., Townley, S., & Votier, S. C. (2017). Global patterns of sex- and age-specific variation in seabird bycatch. *Biological Conservation*, *205*, 60–76.

Gómez Laich, A., Quintana, F., Shepard, E. L. C., & Wilson, R. P. (2012). Intersexual differences in the diving behaviour of Imperial Cormorants. *Journal of Ornithology*, *153*(1), 139–147.

Gómez-Díaz, E., González-Solís, J., & Peinado, M. A. (2009). Population structure in a highly pelagic seabird, the Cory's shearwater *Calonectris diomedea*: an examination of genetics, morphology and ecology. *Marine Ecology Progress Series*, *382*, 197–209.

Gomez-Diaz, E., Gonzalez-Solis, J., Peinado, M. A., & Page, R. D. M. (2006). Phylogeography of the Calonectris shearwaters using molecular and morphometric data. *Molecular Phylogenetics and Evolution*, *41*(2), 322–332.

González-Solís, J., Felicísimo, A., Fox, J. W., Afanasyev, V., Kolbeinsson, Y., & Muñoz, J. (2009). Influence of sea surface winds on shearwater migration detours. *Marine Ecology Progress Series*, *391*, 221–230.

Grant, M. L., Bond, A. L., & Lavers, J. L. (2022). The influence of seabirds on their breeding, roosting and nesting grounds: A systematic review and meta-analysis. *Journal of Animal Ecology*, *91*(6), 1266–1289.

Hailer, F., Schreiber, E. A., Miller, J. M., Levin, I. I., Parker, P. G., Chesser, R. T., & Fleischer, R. C. (2010). Long-term isolation of a highly mobile seabird on the Galapagos. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1707), 817–825.

Hamer, K. C., Schreiber, E. A., & Burger, J. (2001). Breeding biology, life histories, and life history-environment interactions in seabirds. In E. A. Schreiber & J. Burger (Eds.), *Biology of marine birds* (Vol. 45, pp. 217–261). CRC Press.

Hansen, C. C. R., Westfall, K. M., & Pálsson, S. (2022). Evaluation of four methods to identify the homozygotic sex chromosome in small populations. *BMC Genomics*, *23*(1), 160–174.

Heidrich, P., Amengual, J., & Wink, M. (1998). Phylogenetic relationships in mediterranean and North Atlantic shearwaters (Aves: Procellariidae) based on nucleotide sequences of mtDNA. *Biochemical Systematics and Ecology*, *26*(2), 145–170.

Herman, R. W., Winger, B. M., Dittmann, D. L., & Harvey, M. G. (2022). Fine-scale population genetic structure and barriers to gene flow in a widespread seabird (Ardenna pacifica). *Biological Journal of the Linnean Society*, *137*(1), 125–136.

Hijmans, R. (2023). *terra: Spatial Data Analysis* (1-7.55). R package version 1.7-65. https://CRAN.R-project.org/package=terra.

Hipfner, J. M., Prill, M. M., Studholme, K. R., Domalik, A. D., Tucker, S., Jardine, C., Maftei, M., Wright, K. G., Beck, J. N., Bradley, R. W., Carle, R. D., Good, T. P., Hatch, S. A., Hodum, P. J., Ito, M., Pearson, S. F., Rojek, N. A., Slater, L., Watanuki, Y., … Burg, T. M. (2020). Geolocator tagging links distributions in the non-breeding season to population genetic structure in a sentinel North Pacific seabird. *PLOS ONE*, *15*(11), e0240056.

Horswill, C., Wood, M. J., & Manica, A. (2022). Temporal change in the contribution of immigration to population growth in a wild seabird experiencing rapid population decline. *Ecography*, *2022*(11), e05846.

Hsü, K. J., Ryan, W. B. F., & Cita, M. B. (1973). Late Miocene Desiccation of the Mediterranean. *Nature*, *242*(5395), 240–244.

Inchausti, P., & Weimerskirch, H. (2002). Dispersal and metapopulation dynamics of an oceanic seabird, the wandering albatross, and its consequences for its response to long-line fisheries. *Journal of Animal Ecology*, *71*(5), 765–770.

Jeglinski, J. W. E., Wanless, S., Murray, S., Barrett, R. T., Gardarsson, A., Harris, M. P., Dierschke, J., Strøm, H., Lorentsen, S.-H., & Matthiopoulos, J. (2023). Metapopulation regulation acts at multiple spatial scales: Insights from a century of seabird colony census data. *Ecological Monographs*, *93*(2), e1569.

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

Kling, M. (2023). *windscape: Landscape connectivity by wind* (0.1.0). R package version 0.1.0.

Kling, M. M., & Ackerly, D. D. (2020). Global wind patterns and the vulnerability of wind-dispersed species to climate change. *Nature Climate Change*, *10*(9), 868–875.

Kling, M. M., & Ackerly, D. D. (2021). Global wind patterns shape genetic differentiation, asymmetric gene flow, and genetic diversity in trees. *Proceedings of the National Academy of Sciences*, *118*(17), e2017317118.

Klomp, N. I., & Furness, R. W. (1992). The dispersal and philopatry of great skuas from foula, Shetland. *Ringing and Migration*, *13*(2), 37–41. https://doi.org/10.1080/03078698.1992.9674022

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*(1), 356.

Kozakiewicz, C. P., Carver, S., & Burridge, C. P. (2018). Under-representation of avian studies in landscape genetics. *Ibis*, *160*(1), 1–12.

Kranstauber, B., Weinzierl, R., Wikelski, M., & Safi, K. (2015). Global aerial flyways allow efficient travelling. *Ecology Letters*, *18*(12), 1338–1345.

Kuparinen, A. (2006). Mechanistic models for wind dispersal. *Trends in Plant Science*, *11*(6), 296–301.

Lago, P., Austad, M., & Metzger, B. (2019). Partial migration in the Mediterranean Storm Petrel Hydrobates pelagicus melitensis. *Marine Ornithology*, *47*, 105–113.

Laneri, K., Louzao, M., Martínez-Abraín, A., Arcos, J. M., Belda, E. J., Guallart, J., Sánchez, A., Giménez, M., Maestre, R., & Oro, D. (2010). Trawling regime influences longline seabird bycatch in the Mediterranean: New insights from a small-scale fishery. *Marine Ecology Progress Series*, *420*, 241–252.

Lazzari, P., Solidoro, C., Ibello, V., Salon, S., Teruzzi, A., Béranger, K., Colella, S., & Crise, A. (2012). Seasonal and inter-annual variability of plankton chlorophyll and primary production in the Mediterranean Sea: a modelling approach. *Biogeosciences*, *9*(1), 217–233.

Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews Genetics*, *11*(10), 733–739.

Leigh, J. W., & Bryant, D. (2015). popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116.

Lewison, R. L., Crowder, L. B., Wallace, B. P., Moore, J. E., Cox, T., Zydelis, R., McDonald, S., DiMatteo, A., Dunn, D. C., Kot, C. Y., Bjorkland, R., Kelez, S., Soykan, C., Stewart, K. R., Sims, M., Boustany, A., Read, A. J., Halpin, P., Nichols, W. J., & Safina, C. (2014). Global patterns of marine mammal, seabird, and sea turtle bycatch reveal taxa-specific and cumulative megafauna hotspots. *Proceedings of the National Academy of Sciences*, *111*(14), 5271–5276.

Lewison, R., Oro, D., Godley, B. J., Underhill, L., Bearhop, S., Wilson, R. P., Ainley, D., Arcos, J. M., Boersma, P. D., Borboroglu, P. G., Boulinier, T., Frederiksen, M., Genovart, M., González-Solís, J., Green, J. A., Grémillet, D., Hamer, K. C., Hilton, G. M., Hyrenbach, K. D., … Yorio, P. (2012). Research priorities for seabirds: Improving conservation and management in the 21st century. *Endangered Species Research*, *17*(2), 93–121.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Liu, Y., Tao, S., Sun, R., You, C.-F., Wang, T.-H., Felis, T., Zheng, W., Sun, S., Liu, X., Shi, Q., Zhao, J., & Yu, K. (2023). Coral growth over the past 550 years in the central South China Sea linked to monsoon- and seabird-induced nutrient stress. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *617*, 111488.

Lloret, J., Wawrzynkowski, P., Dominguez-Carrió, C., Sardá, R., Molins, C., Gili, J. M., Sabatés, A., Vila-Subirós, J., Garcia, L., Solé, J., Berdalet, E., Turiel, A., & Olivares, A. (2023). Floating offshore wind farms in Mediterranean marine protected areas: a cautionary tale. *ICES Journal of Marine Science*, fsad131.

Lombal, A. J., O'dwyer, J. E., Friesen, V., Woehler, E. J., & Burridge, C. P. (2020). Identifying mechanisms of genetic differentiation among populations in vagile species: historical factors dominate genetic differentiation in seabirds. *Biological Reviews*, *95*(3), 625–651.

Lou, R. N., & Therkildsen, N. O. (2022). Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources*, *22*(5), 1678–1692.

Louzao, M., Hyrenbach, K. D., Arcos, J. M., Abelló, P., de Sola, L. G., & Oro, D. (2006). Oceanographic habitat of an endangered Mediterranean Procellariiform: implications for marine protected areas. *Ecological Applications*, *16*(5), 1683–1695.

Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77.

Manel, S., Berthoud, F., Bellemain, E., Gaudeul, M., Luikart, G., Swenson, J. E., Waits, L. P., Taberlet, P., & Consortium, I. (2007). A new individual-based spatial approach for identifying genetic discontinuities in natural populations. *Molecular Ecology*, *16*(10), 2031–2043.

Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, *20*(3), 136–142.

Martínez-Gómez, J. E., Matías-Ferrer, N., Sehgal, R. N. M., & Escalante, P. (2015). Phylogenetic placement of the critically endangered Townsend's Shearwater (Puffinus auricularis auricularis): evidence for its conspecific status with Newell's Shearwater (Puffinus a. newelli) and a mismatch between genetic and phenotypic differentiation. *Journal of Ornithology*, *156*(4), 1025–1034.

Meek, M. H., Baerwald, M. R., Stephens, M. R., Goodbla, A., Miller, M. R., Tomalty, K. M. H., & May, B. (2016). Sequencing improves our ability to study threatened migratory species: Genetic population assignment in California's Central Valley Chinook salmon. *Ecology and Evolution*, *6*(21), 7706–7716.

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731.

Metzger, B., Oppel, S., Carroll, M., Meirinho, A., Dias, M., Barbara, N., & Lago, P. (2015). *Malta Marine IBA Inventory Report*. BirdLife Malta.

Militão, T., Bourgeois, K., Roscales, J. L., & González-Solís, J. (2013). Individual migratory patterns of two threatened seabirds revealed using stable isotope and geolocation analyses. *Diversity and Distributions*, *19*(3), 317–329.

Milot, E., Weimerskirch, H., & Bernatchez, L. (2008). The seabird paradox: dispersal, genetic structure and population dynamics in a highly mobile, but philopatric albatross species. *Molecular Ecology*, *17*(7), 1658–1673.

Munilla, I., Genovart, M., Paiva, V. H., & Velando, A. (2016). Colony Foundation in an Oceanic Seabird. *PLOS ONE*, *11*(2), e0147222.

Nathan, R., Katul, G. G., Bohrer, G., Kuparinen, A., Soons, M. B., Thompson, S. E., Trakhtenbrot, A., & Horn, H. S. (2011). Mechanistic models of seed dispersal by wind. *Theoretical Ecology*, *4*(2), 113–132.

NOAA National Centers for Environmental Information. (2022). *2022: ETOPO 2022 15 Arc-Second Global Relief Model*.

Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Weedon, J. (2022). *vegan: Community Ecology Package* (2.6-4).

Oppel, S., Raine, A. F., Borg, J. J., Raine, H., Bonnaud, E., Bourgeois, K., & Breton, A. R. (2011). Is the Yelkouan shearwater Puffinus yelkouan threatened by low adult survival probabilities? *Biological Conservation*, *144*(9), 2255–2263.

Oro, D. (2003). Managing seabird metapopulations in the Mediterranean: constraints and challenges. *Scientia Marina*, *67*(2), 13–22.

Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, *4*(3), 347–354.

Pante, E., & Simon-Bouhet, B. (2013). marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *PLOS ONE*, *8*(9), e73051.

Pascual, M., Rives, B., Schunter, C., & Macpherson, E. (2017). Impact of life history traits on gene flow: A multispecies systematic review across oceanographic barriers in the Mediterranean Sea. *PLOS ONE*, *12*(5), 1–20.

Penhallurick, J., & Wink, M. (2004). Analysis of the taxonomy and nomenclature of the Procellariiformes based on complete nucleotide sequences of the mitochondrial cytochrome Cytb gene. *Emu*, *104*(2), 125–147.

Péron, C., & Grémillet, D. (2013). Tracking through Life Stages: Adult, Immature and Juvenile Autumn Migration in a Long-Lived Seabird. *PLOS ONE*, *8*(8), e72713.

Peterman, W. E. (2018). ResistanceGA: An R package for the optimization of resistance surfaces using genetic algorithms. *Methods in Ecology and Evolution*, *9*(6), 1638–1647.

Peterman, W. E., Connette, G. M., Semlitsch, R. D., & Eggert, L. S. (2014). Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*, *23*(10), 2402–2413.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLOS ONE*, *7*(5), 1–11.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Nature Genetics*, *48*(1), 94–100.

Pezzo, F., Zenatello, M., Cerritelli, G., Navone, A., Giunchi, D., Spano, G., Pollonara, E., Massolo, A., Gagliardo, A., & Baccetti, N. (2021). Productivity changes in the Mediterranean Sea drive foraging movements of yelkouan shearwater *Puffinus yelkouan* from the core of its global breeding range. *Marine Ecology*, *42*(4), e12668.

Piatt, J. F., Sydeman, W. J., & Wiese, F. (2007). Introduction: A modern role for seabirds as indicators. *Marine Ecology Progress Series*, *352*, 199–204.

Pizzo, N., Deike, L., & Ayet, A. (2021). How does the wind generate waves? *Physics Today*, *74*(11), 38–43.

R Core Team. (2023). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Raine, Andre. F., Borg, J. J., Raine, H., & Phillips, R. A. (2013). Migration strategies of the Yelkouan Shearwater *Puffinus yelkouan. Journal of Ornithology*, *154*(2), 411–422.

Rasmussen, M. S., Garcia-Erill, G., Korneliussen, T. S., Wiuf, C., & Albrechtsen, A. (2022). Estimation of site frequency spectra from low-coverage sequencing data using stochastic EM reduces overfitting, runtime, and memory usage. *Genetics*, *222*(4), iyac148.

Rexer-Huber, K., Veale, A. J., Catry, P., Cherel, Y., Dutoit, L., Foster, Y., McEwan, J. C., Parker, G. C., Phillips, R. A., Ryan, P. G., Stanworth, A. J., van Stijn, T., Thompson, D. R., Waters, J., & Robertson, B. C. (2019). Genomics detects population structure within and between ocean basins in a circumpolar seabird: The white-chinned petrel. *Molecular Ecology*, *28*(20), 4552–4572.

Reyes Suarez, N. C., Cook, M. S., Gačić, M., Paduan, J. D., Drago, A., & Cardin, V. (2019). Sea Surface Circulation Structures in the Malta-Sicily Channel from Remote Sensing Data. *Water*, *11*(8).

RS-eco. (2023). *processNC: R Package for processing large NetCDF files* (0.0.1). R package version 0.0.1. https://github.com/RS-eco/processNC.

Ruegg, K. C., Anderson, E. C., Paxton, K. L., Apkenas, V., Lao, S., Siegel, R. B., DeSante, D. F., Moore, F., & Smith, T. B. (2014). Mapping migration in a songbird using high-resolution genetic markers. *Molecular Ecology*, *23*(23), 5726–5739.

Sachs, G., Traugott, J., Nesterova, A. P., Dell'Omo, G., Kümmeth, F., Heidrich, W., Vyssotski, A. L., & Bonadonna, F. (2012). Flying at No Mechanical Energy Cost: Disclosing the Secret of Wandering Albatrosses. *PLOS ONE*, *7*(9), e41449.

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M. P., van den Dool, H., Zhang, Q., Wang, W., Chen, M., & Becker, E. (2014). The NCEP Climate Forecast System Version 2. *Journal of Climate*, *27*(6), 2185–2208.

Sangster, G., Collinson, J. M., Helbig, A. J., Knox, A. G., & Parkin, D. T. (2002). The specific status of Balearic and yelkouan shearwaters. *British Birds*, *95*(12), 636–639.

Sangster, G., Knox, A. G., Helbig, A. J., & Parkin, D. T. (2002). Taxonomic recommendations for European birds. *Ibis*, *144*(1), 153–159.

Schulzweida, U. (2023). *CDO User Guide* (2.3.0). Zenodo. https://doi.org/10.5281/zenodo.10020800.

Sekino, M., & Hara, M. (2007). Individual assignment tests proved genetic boundaries in a species complex of Pacific abalone (genus Haliotis). *Conservation Genetics*, *8*(4), 823–841.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702.

Smith, A. L., Monteiro, L., Hasegawa, O., & Friesen, V. L. (2007). Global phylogeography of the band-rumped storm-petrel (*Oceanodroma castro*; Procellariiformes: Hydrobatidae). *Molecular Phylogenetics and Evolution*, *43*(3), 755–773.

Steeves, T. E., Maloney, R. F., Hale, M. L., Tylianakis, J. M., & Gemmel, N. J. (2010). Genetic analyses reveal hybridization but no hybrid swarm in one of the world's rarest birds. *Molecular Ecology*, *19*(23), 5090–5100.

Subramanian S. (2016). The effects of sample size on population genomic analyses--implications for the tests of neutrality. *BMC Genomics*, *20*(17), 123.

Sultana, J., & Borg, J. J. (2006). Population ecology and conservation of the Yelkouan Shearwater. In N. Aransay (Ed.), *Proceedings of the fi rst symposium on the mediterranean action plan for the conservation of marine and coastal birds.* (pp. 40–42). RAC/SPA.

Sultana, J., Borg, J. J., Gauci, C., & Falzon, V. (2011). *The Breeding Birds of Malta*. BirdLife Malta.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595. https://doi.org/10.1093/genetics/123.3.585

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027.

Thanou, E., Sponza, S., Nelson, E. J., Perry, A., Wanless, S., Daunt, F., & Cavers, S. (2017). Genetic structure in the European endemic seabird, *Phalacrocorax aristotelis*, shaped by a complex interaction of historical and contemporary, physical and nonphysical drivers. *Molecular Ecology*, *26*(10), 2796–2811.

Thomas, C. D., & Kunin, W. E. (1999). The spatial structure of populations. *Journal of Animal Ecology*, *68*(4), 647–657. https://doi.org/https://doi.org/10.1046/j.1365-2656.1999.00330.x

Thorne, L. H., Clay, T. A., Phillips, R. A., Silvers, L. G., & Wakefield, E. D. (2023). Effects of wind on the movement, behavior, energetics, and life history of seabirds. *Marine Ecology Progress Series*, *723*, 73–117.

1976 Convention for the Protection of the Mediterranean Sea Against Pollution, Pub. L. No. 1102 UNTS 27 (1976).

van Dijk, G., Maars, J., Andreetto, F., Hernández-Molina, F. J., Rodríguez-Tovar, F. J., & Krijgsman, W. (2023). A terminal Messinian flooding of the Mediterranean evidenced by contouritic deposits on Sicily. *Sedimentology*, *70*(4), 1195–1223.

van Strien, M. J., Holderegger, R., & Van Heck, H. J. (2015). Isolation-by-distance in landscapes: considerations for landscape genetics. *Heredity*, *114*(1), 27–37.

Waples, R. K., Albrechtsen, A., & Moltke, I. (2019). Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Molecular Ecology*, *28*(1), 35–48.

Waples, R. S. (2024). Practical application of the linkage disequilibrium method for estimating contemporary effective population size: A review. *Molecular Ecology Resources*, *24*(1), e13879.

Weimerskirch, H. (2001). Seabird Demography and Its Relationship with the Marine Environment. In E. A. Schreiber & J. Burger (Eds.), *Biology of Marine Birds* (1st ed., pp. 115–135). CRC Press.

Weimerskirch, H., Åkesson, S., & Pinaud, D. (2006). Postnatal dispersal of wandering albatrosses Diomedea exulans: implications for the conservation of the species. *Journal of Avian Biology*, *37*(1), 23–28.

Weimerskirch, H., Guionnet, T., Martin, J., Shaffer, S. a, & Costa, D. P. (2000). Fast and fuel efficient? Optimal use of wind by flying albatrosses. *Proceedings. Biological Sciences / The Royal Society*, *267*(1455), 1869–1874.

Weimerskirch, H., Salamolard, M., Sarrazin, F., & Jouventin, P. (1993). Foraging Strategy of Wandering Albatrosses Through The Breeding Season: A Study Using Satellite Telemetry. *The Auk*, *110*(2), 325–342.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, *16*(2), 97–159.

Yen, P. P. W., Sydeman, W. J., & Hyrenbach, K. D. (2004). Marine bird and cetacean associations with bathymetric habitats and shallow-water topographies: implications for trophic transfer and conservation. *Journal of Marine Systems*, *50*(1), 79–99.

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

APPENDIX

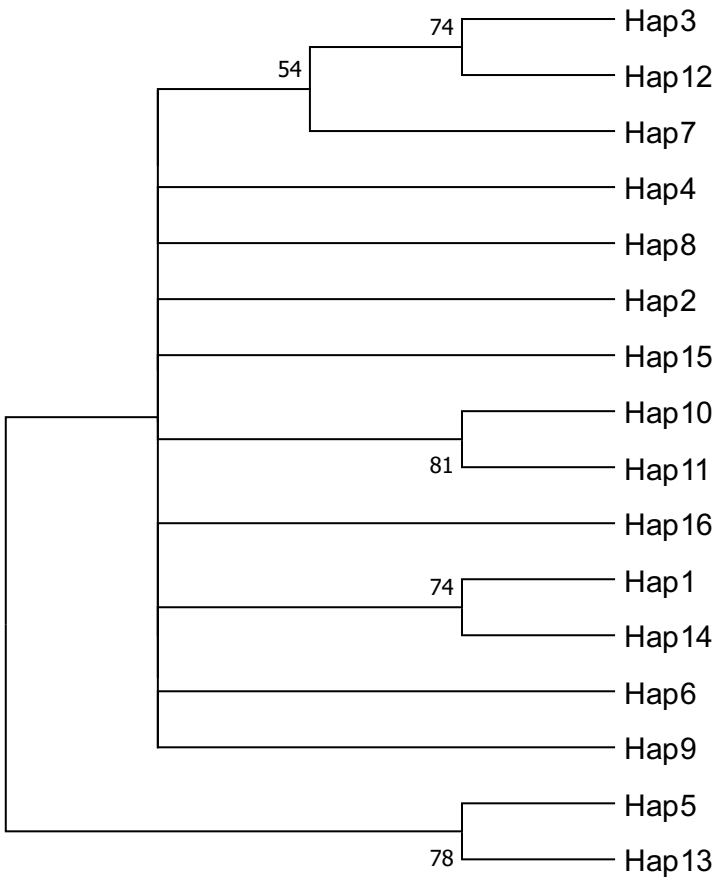


Figure A.1. Maximum-likelihood tree of 16 mtDNA control region haplotypes

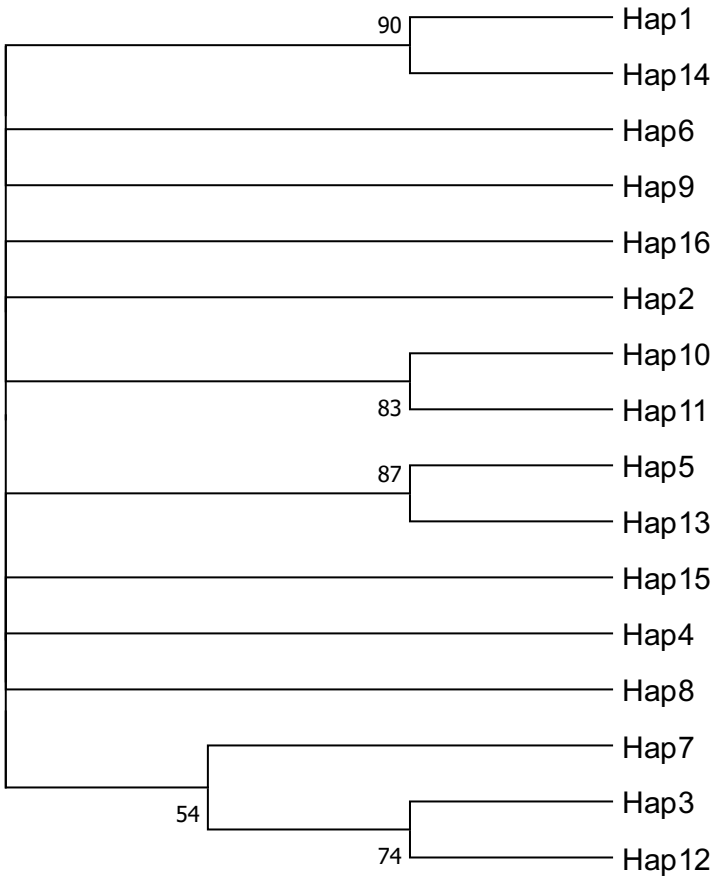


Figure A.2. Neighbor-joining tree of 16 mtDNA control region haplotypes

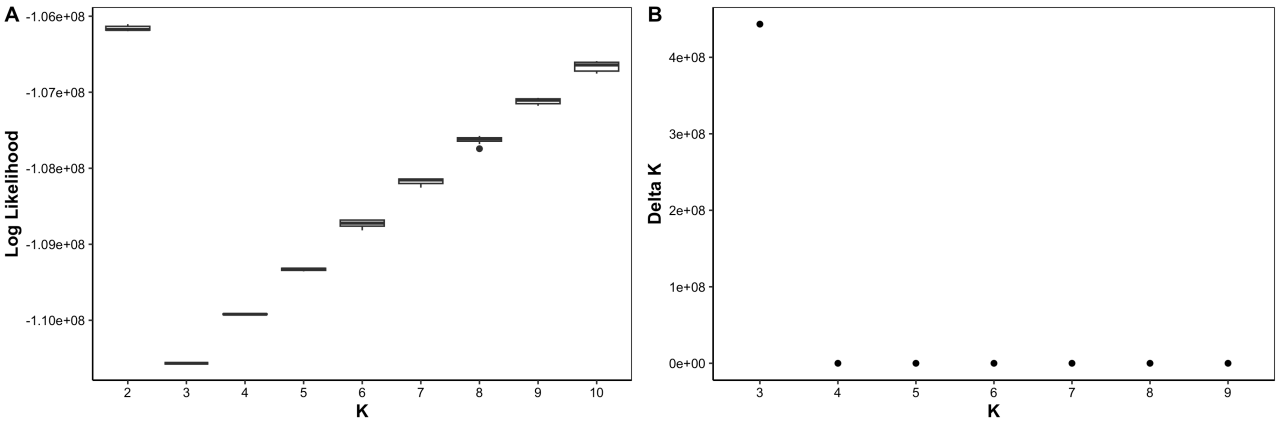


Figure A.3. Log likelihoods of admixture results. (A) Probability of data given K (Ln P(X|K)) and (B) second order rate of change of Ln P(X|K)).

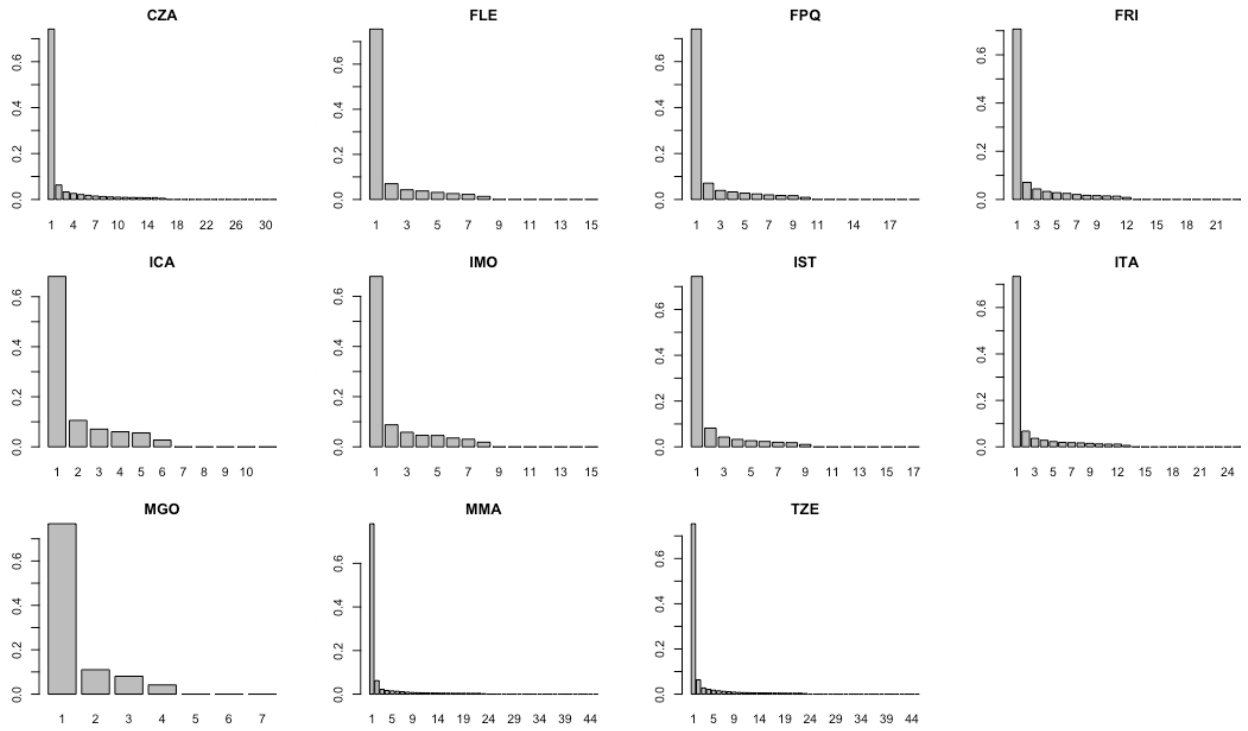


Figure A.4. Colony-based site frequency spectrum histograms.

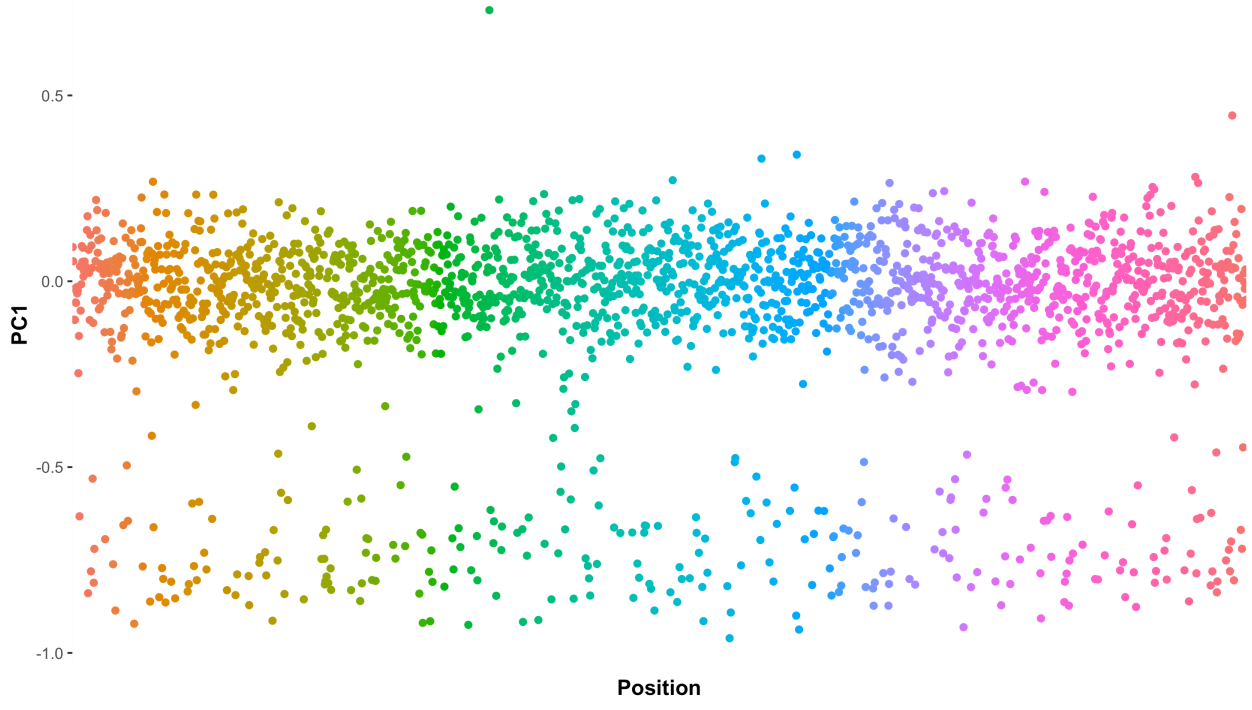


Figure A.5. PC loading showing the SNP weights. Colours represent different polymorphic sites.

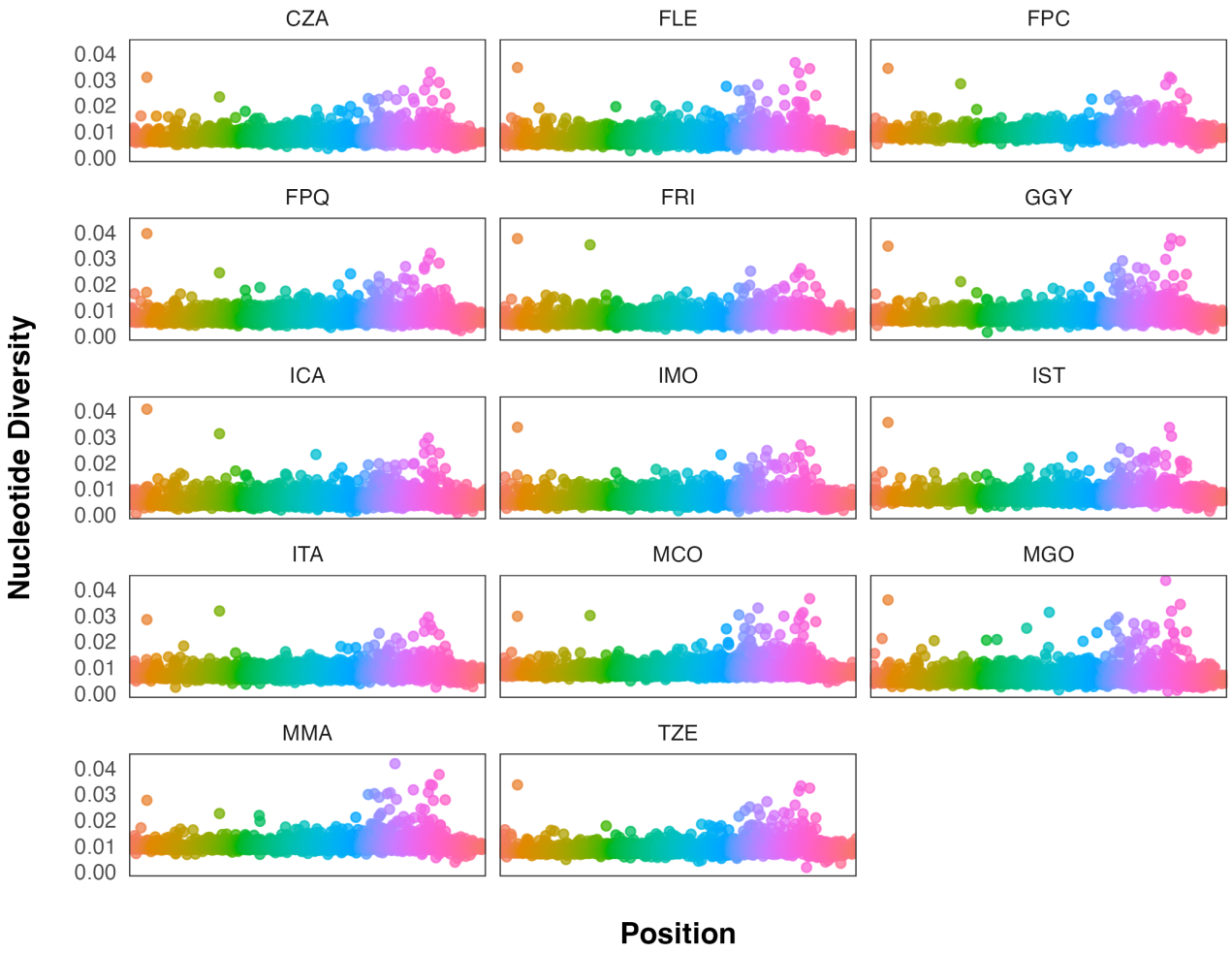


Figure A.6. Genome-wide nucleotide diversity of each Yelkouan Shearwater colony. Colours indicate different polymorphic sites.