Year : 2021

Impact of landscape barriers and isolation on the genetic and phenotypic evolution of the barn owl

Machado Ana Paula

Machado Ana Paula, 2021, Impact of landscape barriers and isolation on the genetic and phenotypic evolution of the barn owl

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive http://serval.unil.ch
Document URN : urn:nbn:ch:serval-BIB_FD58D0995E236

UNIL | Université de Lausanne
Faculté de biologie et de médecine

Département d'Ecologie et Evolution

Impact of landscape barriers and isolation on the genetic and phenotypic evolution of the barn owl

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Ana Paula MACHADO

Master of Science in Behaviour, Ecology and Evolution,
University of Lausanne

Jury

Prof. Alexandre Reymond, Président
Prof. Alexandre Roulin, Co-Directeur de thèse
Prof. Jérôme Goudet, Co-directeur de thèse
Prof. Nicolas Salamin, Expert
Dr. Reto Burri, Expert

Lausanne
(2021)

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Ecole Doctorale
Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Alexandre	**Reymond**
Directeur·trice de thèse	Monsieur	Prof.	Alexandre	**Roulin**
Co-directeur·trice	Monsieur	Prof.	Jérôme	**Goudet**
Expert·e·s	Monsieur	Prof.	Nicolas	**Salamin**
	Monsieur	Dr	Reto	**Burri**

le Conseil de Faculté autorise l'impression de la thèse de

Madame Ana Paula Oliveira Salvador Machado

Maîtrise universitaire ès Sciences en comportement, évolution et conservation,
Université de Lausanne

intitulée

Impact of landscape barriers and isolation on the genetic and phenotypic evolution of the barn owl

Date de l'examen : 6 juillet 2021

Date d'émission de l'imprimatur : Lausanne, le 15 juillet 2021

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale

Table of Contents

Acknowledgements

First and foremost, I would like to thank Alex and Jérôme. You believed in me and gave me the opportunity and independence to create my own niche between the two groups, from the field, to the lab and analyses (not to mention the travelling!). You didn't put me in a box, and I really appreciate you for it. Your complementary inputs allowed me to produce a thesis that I'm proud of, and I hope you are too. I also wish to thank the members of my committee for their insight and fruitful discussions, as well as Thomas Flatt and John Pannell who gave me a great foundation to tackle a PhD.

None of this work would have been possible without the collaborators that provided us with precious samples. Thank you all very very much! Getting to know you, hearing your stories, and seeing your passion for preserving biodiversity is deeply inspiring. A special thanks to John Lusby and David & Jaine Ramsden for welcoming me into your homes, showing me the important work that you do and for the great times we shared.

Being part of two groups means I had double the support team to whom I owe a big thank you. To Valérie, Luis and Guillaume, who helped me take my first PhD-steps. To Kim, Pauline, Arnaud, Vera, Paul, Paolo and Andrea who were there for most of the ups and downs of these five years, and some awesome (sunny) conferences. To Tristan, Elu and Alex for Goudet group v.2.0, I don't know how I'd have finished this thesis without your help and I only wish you'd arrived earlier. To Céline S. for her invaluable aid, you're indeed an angel (everyone says it, so it must be true). To all master students that made the field family bigger and better, especially those that remained, Estelle, Céline P. and Steve. To all the brilliant people that made the long days, and longer nights, in the field fun, even if they did leave you stranded at 6am (**cough cough* Kim*).

A huge thanks to the whole DEE for proving such a great work environment and great friendships since I arrived, nearly 8 years ago, with some special mentions: to Sachito, who has been there from my very first day in Lausanne (actually, Madrid); to Sarah for our little tête-à-tête that made the end of this thesis more bearable; to the Russian for sharing the pains of learning all-things-genomics and rebelling against awful courses; and to Laura for the yearly bat and rodent outings that never failed to make me happy.

Obrigada to my family that always supported me from afar, even if what I do is a bit of a mystery to you. Merci to the Renaud/Séchaud for being my Swiss family and welcoming me from the start. And lastly, but definitely not least, to Robin. Since we both did about half of each other's theses (me installing dozens of GPS for you somewhere in the Broye, you scoring carcasses for me in foreign museums on what was supposed to be our vacations), it's only fitting that we finish them together. You were and are my rock through it all, and even though words are far from enough, merci Pep. I can't wait for our next adventure.

Abstract

Patterns of genetic structure are directly dependant on the characteristics of landscape in which they occur. The presence of landscape features that act as barriers, distort otherwise smooth structure patterns by directly hindering the movement of individuals and their alleles. By influencing or outright preventing gene flow, barriers isolate populations to varying degrees. The recent advent of sequencing technology and population genomics tools allow the characterization of these mechanisms with higher resolution than ever before. In this thesis, I used whole genome sequences, capture-recapture and phenotypic data to explore the genomic and phenotypical consequences of barrier-mediated isolation in barn owl (*Tyto alba*) populations over five chapters.

In the first three chapters, I focused on the populations of three island systems with different age, size and distance from the mainland, and on the neutral and selective mechanisms that shape their *in-situ* divergence. First, we investigated the colonisation history of barn owls from the British Isles, a set of continental islands, that separated from the mainland after the last glaciation. Our aim was to understand the basis of their white plumage, a contrast with the nearest mainland populations that are rufous. We found that neutral mechanisms alone suffice to describe this phenotypical discrepancy as they were actually founded by a white population. Second, we compared the evolutionary histories of the species on two outwardly similar islands in the eastern Mediterranean, Crete and Cyprus, which have been separated from the mainland for 5 million years. Despite the similarities between the islands, we found that the populations of Crete and Cyprus had distinct origins and very different neutral outcomes, with varying population sizes and levels of inbreeding. Third, we studied the barn owls of the Canary archipelago, islands isolated for over 10 million years and far from the mainland. In addition to the substantial neutral divergence, we found striking signals of local adaptation to both the insular conditions as a whole, and to each of the lineages' local environmental conditions.

Then, in the two last chapters I investigated the impact of mountain ranges in creating and moulding barn owl genetic structure and demographic history. In the Western Palearctic, we found that the Alps and the contiguous Tauros and Zargos mountains are key barriers to gene flow, by isolating specific populations and limiting secondary contact between genetic lineages on either side of them. Finally, in North America, we showed that the Rocky Mountains force owls to contour it via the south.

To summarise, this thesis highlights the impact of large water bodies and mountain ranges on a widespread bird of prey. It provides empirical evidence of both neutral and adaptive evolutionary forces acting on isolated populations, and illustrates the power of genomic data in non-model species.

Résumé

La structure génétique des populations dépend directement des caractéristiques de l'environnement dans lequel elles évoluent. La présence d'éléments dans l'environnement qui agissent comme des barrières influence également cette structure génétique en limitant le mouvement des individus et de leurs allèles. En réduisant ou en empêchant entièrement le flux génétique, ces barrières isolent les populations à différents degrés. Le récent développement des technologies de séquençage et des outils d'analyses génomiques permettent la caractérisation des mécanismes structurant les populations comme jamais auparavant. Dans les 5 chapitres de cette thèse, j'ai utilisé des données de séquençage de génomes complets, de capture-marquage-recapture et de mesures phénotypiques pour explorer les conséquences génotypiques et morphologiques qu'ont les barrières environnementales sur les populations d'Effraie des clochers (*Tyto alba*).

Dans les trois premiers chapitres, je me suis intéressée aux mécanismes neutres et sélectifs qui ont fait évoluer des populations ayant colonisé trois systèmes d'îles qui diffèrent en âge, taille et distance du continent. Premièrement, nous avons étudié les processus de colonisation par l'Effraie des clochers des îles britanniques, un groupe d'îles qui se sont séparées du continent après les dernières glaciations. Notre objectif était de comprendre pourquoi leur plumage est blanc immaculé sur les îles, en comparaison de celui plutôt roux des populations continentales. Nous avons montré que la colonisation de ces îles s'est faite par une population blanche, et que des mécanismes neutres suffisent ensuite à expliquer le maintien de cette coloration. Deuxièmement, nous avons comparé l'histoire évolutive de populations vivant sur deux îles de la Méditerranée orientale apparemment similaires, la Crète et Chypre, toutes deux séparées du continent depuis 5 millions d'années. Malgré les similarités entre ces deux îles, nous avons constaté que leurs populations avaient des origines distinctes et des histories évolutives différentes, avec également des tailles de population et des niveaux de consanguinité variables. Troisièmement, nous avons étudié les Effraies des clochers vivant sur l'archipel des Canaries, ce dernier étant situé loin du continent et ayant été isolé depuis plus de 10 millions d'années. En plus d'une divergence neutre importante, nous avons trouvé des signaux forts d'adaptation locale aux conditions insulaires dans leur ensemble, ainsi qu'aux conditions environnementales locales de chacune des lignées.

Dans les deux derniers chapitres, j'ai étudié l'impact des chaînes de montagnes dans la création et l'évolution de la structure génétique et de l'histoire évolutive de l'Effraie des clochers. Dans le Paléarctique occidental, nous avons montré que les Alpes, ainsi que les monts Tauros et Zargos, isolent les populations de part et d'autre, et sont donc des barrières importantes aux flux de gènes. Finalement, en Amérique du nord, nous avons pu montrer que les montagnes Rocheuses étaient également une barrière pour les Chouettes effraie, qui doivent la contourner par le sud.

En résumé, cette thèse met en évidence l'influence des barrières, telles que les grandes étendues d'eau et les chaines de montagne, sur une espèce répandue de rapace. Elle fournit des données empiriques de l'action des forces évolutives neutres et adaptatives sur des populations isolées, et illustre l'utilité de données génomiques chez des espèces non modèles.

General Introduction

The variation of visible traits was the focus of the first studies about the diversity of natural populations. It was from this type of observations that, with the works of Mendel and Darwin, classical population genetics was born in the second half of the 19th century. Nearly 100 years after Mendel proclaimed his laws, molecular tools were introduced in the form of gel electrophoresis. With the revolutionary studies of Harris[1] and Hubby & Lewontin[2], came the realization that variation among individuals is much more pronounced than originally thought[3]. This dramatically affected not only the contemporaneous debates on evolutionary theory but also the fundamental nature of empirical studies as molecular population genetics emerged[4]. With the advent of molecular tools, it became possible to characterize, in higher detail than ever before, the genetic makeup of species as well as to elucidate the impact of evolutionary forces in the past history of populations.

Population structure in the landscape

The study of genetic variation within and among populations with molecular data, effectively means analysing the distribution of their allelic frequencies. Differences in distribution between groups of individuals translate into population structure, a central notion in population genetics whose description is frequently a key step in empirical studies. In sexual species, population structure essentially stems from non-random mating between individuals[5]. As most species do not have unlimited dispersal capacities within their range, mating is more likely to occur between close rather than faraway individuals. Geographical distance between individuals is thus directly linked to mate choice and, consequently, to population structure, highlighting how space and the landscape in which biological and evolutionary processes occur are crucial to their interpretation. Indeed, since most individuals mate with close neighbours, a spatial gradient in allelic frequencies will emerge over time, a ubiquitous pattern in natural populations termed isolation by distance[6,7] (seen in plants[8,9], invertebrates[10,11] and vertebrates[12,13], for example). This and other spatial genetic patterns are the specific focus of the discipline of landscape and spatial genetics[14,15].Yet, hardly any empirical population genetics studies are truly independent from these notions.

On a large scale, landscapes are rarely homogenous, and the presence of barriers – geological or ecological features that create gaps between suitable patches of habitat – restricts the movement of organisms and thus impacts their mate choice. Barriers vary drastically in dimensions and impact, from large-sized structures that impact multiple taxa, to small, species-specific ones, and some a barrier feature for on species may even be a corridor for another. At a planetary scale, continental separation (i.e. the result of plate tectonics) is the most important

barrier to terrestrial organisms, driving widespread speciation and bioregionalization[16,17], although this . More moderately-sized barriers include terrestrial features such as mountains[18,19], deserts[20], forests[21] and urban developments[22–24]; as well as aquatic ones like the open-water encircling islands[25,26] and rivers[24,27]. Barriers are an important source of population structure, as they directly hinder or impede the movement of organisms and their alleles (i.e. gene flow), and can therefore quickly promote genetic and phenotypic divergence and ultimately facilitate speciation[28]. Even though the concept of speciation with gene flow is now widely accepted[29], the study of diverging isolated (allopatric) populations was crucial to the development of early modern evolutionary theory[30], and none more so than insular populations[31–33].

Insular populations

Islands, dubbed "nature's test tubes", offer ideal settings to study isolated populations. This is due to a number of intrinsic characteristics such as their relatively small size, discrete borders, geographical remoteness, relatively young (and known) age and, finally, the sheer quantity that provides natural replication of events[34]. The evolution of the genetic composition of any given population is shaped by the interplay of mutation, gene flow, genetic drift and selection[31,35,36]. Although these processes clearly occur on the mainland as well, the unique features of islands, especially the isolation and limited carrying capacity, fundamentally impact how they operate and interact.

Despite being the fundamental source of diversity, mutation occurs slowly, even in large populations, and thus has the smallest impact of all microevolutionary mechanisms both on islands and mainland. In addition, the majority of new mutations are swiftly lost due to drift or even selection, particularly in small populations. In contrast, the other three mechanisms are interlaced in an unsteady balance that is dependent upon its effective size and has major impacts on the population's fate.

On the one hand, genetic drift is responsible for most neutral divergence and acts by sampling individuals and alleles at random at the founding of an island (i.e. founder effect or bottleneck) as well as throughout its settling and long-term establishment. The effects of drift are the most pronounced in populations of far isolated islands, colonized by a small number of individuals that then retain a reduced effective population size[36,37]. Indeed, drift will remove or fixate new alleles more quickly, be them *de novo* mutations or imported alleles by occasional gene flow, shaping allelic frequencies and accentuating divergence from neighbouring populations. A consequence of intense drift is the reduction of polymorphism and thus genetic diversity which, if coupled with low gene flow, can increase background relatedness between individuals and lead to inbreeding.

This is a common phenomenon in insular populations and is likely linked to their higher extinction rates, particularly in small islands[38,39].

On the other hand, selection acts non-randomly to direct adaptive divergence of specific alleles or morphs and can work to maintain polymorphism. Like drift, it can act both at the founding of an island, with some individuals, populations or even species being more prone to colonization, as well as after its establishment. While selection is more effective in larger populations, its contribution to divergence of insular populations is undeniable[36]. It is most often invoked to explain conspicuous morphological changes, particularly in cases of independent trait convergence (i.e. repeated evolution of similar phenotypes) and radiation (i.e. fast evolution of multiple phenotypes or species from a single one). The former is perhaps most commonly illustrated by the island rule, a trend towards intermediate body size[40] first described from studies of mammals[41,42], and the latter also boasts many examples, from spiders in Hawaii [43] to lizards in the Caribbean[44].

Finally, gene flow also plays an important role and directly influences the effects of drift and selection. Insufficient immigration may result in a too-small resident population that might easily go extinct, whereas high rates of immigration can homogenize allelic frequencies and prevent or delay both neutral and adaptive divergence. Therefore, moderate levels of gene flow typically facilitate divergence by adding enough diversity and increasing the effective population size.

Ultimately, an insular population's fate is influenced by all these mechanisms and their respective impacts can vary through time as a result of fluctuating environmental conditions. For instance, changes in sea level can alter the distance between neighbouring islands and the mainland which directly impacts levels of gene flow. It is also important to note that, even under similar circumstances and island characteristics, all these processes have considerable stochasticity and thus the outcome each colonisation can be unique.

Island birds

The study of insular populations was indeed key to the development of evolutionary theory, starting with Darwin's first visit to the Galapagos in 1835 and, in particular, his work with birds that ensued. Birds are chief island colonizers and often assist other organisms in doing the same, such as plants and invertebrates[45]. Yet, there is a large variation in dispersal capacity among them, from sedentary to long-range migrators, rendering some avian taxa more likely to colonize islands than others. Moreover, winds directly influence flight routes[46], storm tracks provoke accidental displacement[45] and humans are responsible for a large amount of introductions[47], occasionally producing unexpected community assemblies[32,34,45].

Empirical studies of insular birds have provided evidence of both neutral and adaptive mechanisms driving phenotypical divergence (e.g.[48,49]). Although early studies denied its validity for birds[50], the island rule (convergence to intermediate body sizes) has since been shown to apply to numerous avian species[51,52]. Indeed, the examples of convergent trait evolution in insular birds are actually plentiful and go far beyond mere size. For example, changes have been observed in life-history traits such as higher investment in parental care[53], and in physical traits with trends towards larger brains[54], loss of flight[55] and reduced plumage color intensity and brightness[56,57]. Birds are also the protagonists of some of the most spectacular insular adaptive radiations[34,45,58], with the classical examples of beak adaptation in Hawaiian honeycreepers[59,60] and Darwin's finches[61,62].

Avian genomics

From allozymes[1] to the first complete human genome[63], the second half of the 20th century saw a staggering advance in sequencing technologies. In parallel, the field of molecular population genetics evolved to analyse the ever-increasing amount of markers[4], transitioning into the genomics era in the 21st century. Birds were no exception to the boom of genomics, with the chicken being among the first whole genome sequences of vertebrates in 2004[64], followed by the turkey[65] and the zebra finch[66] in 2010. The intrinsic characteristics of bird genomes may have facilitated this task along with that of subsequent whole-genome resequencing studies. Compared to mammals, birds have considerably less repetitive regions as well as shorter introns and intergenic regions[67]. As a result, bird genomes are on average one third the length of that of mammals, which reduces their sequencing costs proportionally.

The early days of avian genomics were heavily skewed towards model species, particularly those with an agroeconomic interest (chicken, turkey and domestic duck), however it has since expanded nearly-exponentially into non-model species. This is perhaps best illustrated by the National Center for Biotechnology Information (NCBI) which, in eight short years, went from 11 to 667 reported bird reference genomes (April 2013[68] to April 2021). Despite only a tenth of these being at chromosome-level, it speaks to the interest of bird genomes to a range of diverse research fields, such as evolutionary biology. The latter was kickstarted in 2012 with the sequences of the pied and collared flycatchers (*Ficedula sp.*) in the first study to look at signatures of speciation along the genome in wild birds[69]. Soon thereafter, the Avian Phylogenomics Project (http://avian.genomics.cn/en/index.html) generated the sequences of 45 bird species and produced their flagship papers on bird comparative genomics[70] and phylogenomics[71], among an extensive collection of 28 papers on the evolution of bird genomes in 2014 alone. From it, the Bird 10'000 Genomes (B10K) Project was born with the aim to

sequence all extant bird species[72], promising a wealth of publicly available data for future studies (e.g [73]).

Avian molecular population genomics

With the declining cost of producing genomic data, population-level data is also progressively becoming more available to a wider group of researchers. As such, molecular population genomics has flourished, and is nowadays far from limited to humans or cattle, being almost commonplace for non-model species (albeit rarely with whole-genome sequencing; WGS). The most immediate advantage of using genomic data in population studies is the massive increase in number of loci, from a few dozens or hundreds provided by microsatellites and Sanger sequencing, to tens of thousands (Restriction Associated DNA sequencing; RAD-seq) and up to millions (WGS). With the parallel development of analytical tools, this directly translates into higher resolution to detect fainter variations in allelic frequencies, and thus population structure, than ever before while requiring a less extensive sampling of individuals. Practically, this is a major benefit when working with threatened and/or hard-to-sample species as many non-model organisms are.

As the number of loci increases, so does the power to infer the past demographic histories of populations, namely changes in population size, population splits and admixture events. A number of methods have been developed or improved for this specific purpose (reviewed in [74]), three of which are most commonly used in non-model species. First, software such as *fastsimcoal2*[75] and *δaδi*[76] provide likelihood-based frameworks to analyse changes in the shape of the site frequency spectrum (SFS). In this approach, one simulates one or more demographic scenarios with varying parameters sampled from specified ranges, and then compares the simulated SFS to the observed one. This approach allows inferences at moderate speed for multiple populations (up to three in *δaδi*), and can be based either on coalescent (*fastsimcoal2*) or diffusion (*δaδi*) theories. Second, approximate Bayesian computation (ABC) follows a relatively similar workflow by comparing multiple summary statistics (F_{ST}, for example) of simulated and observed data. However, in comparison to SFS-based methods, it is still highly computationally intensive and can be inefficient for large datasets or wide parameter spaces. Thirdly, sequentially Markovian coalescent methods, like PSMC[77] or MSMC[78], infer changes in population effective size (N_e) from the coalescent times along the genome. Unlike the previous methods, it does not use simulations and is therefore quite fast, however it requires WGS data and not all implementations take into consideration population information. Crucially, this type of inference is known to be unreliable in recent time (less than about 800 generations). All three methods have been used to infer the history of bird populations (recently reviewed in [79]), for example, SFS in South American passerines[80], ABC in rosy-finches[81], and PSMC in penguins[82].

The final, and perhaps most exciting, opportunity offered by genomic data, and WGS in particular, is the study of the genomic landscape itself. Indeed, the sequencing of individual genomes from multiple populations brought with it the awareness that diversity and differentiation are not constant along the genome, sparking widespread interest on the role of microevolutionary processes in generating such variation. The genomic landscape was the topic of the first WGS study in non-model birds, looking at diverging populations and speciation in flycatchers[69,83]. Since then, it has been widely investigated under different contexts like hybridization[84], inbreeding[85] and local adaptation[48], and has been the focus of multiple reviews (e.g.[86,87]).

Study species

The barn owl (*Tyto alba* species complex) is a non-migratory nocturnal raptor present on all continents but Antarctica. It hunts mainly small rodents in a large range of open habitats in temperate, subtropical and tropical zones. Despite being commonly described as one of the six bird species with a cosmopolitan distribution[88], recent molecular evidences[89,90] suggested it actually comprises three genetically distinct sister species: the Afro-European (or Western; *T. alba*), American (*T. furcata*) and Australasian (or Eastern; *T. javanica*) barn owls. Representatives of the barn owl group can be found on many islands around the globe (both naturally and introduced), and in some it has diversified into subspecies or even species[89]. Insular barn owls often display specific variation in phenotypic traits compared to their mainland counterparts[91], providing an assortment of interesting study systems with varying degrees of isolation and phenotypic traits.

The first barn owl genome published was of the American species[70] yet it was of poor quality. It was followed by a first European barn owl genome[92] which increased in quality despite still being quite fragmented. It allowed for the pilot analyses of this thesis, while its improved version was being produced[93].

Variation in plumage colouration

Bird plumage coloration has long been the object of intense interest from naturalists and scientists alike[94]. One of the most notorious features of barn owls is their extraordinary variation in ventral plumage colour, from white to dark rufous (Figure 1), a heritable trait[95] that is observed within all three species[96], and has often been used as a taxonomic criterium to split subspecies[89]. On a larger geographical scale, plumage coloration is strongly associated with latitude (and all its covariates), displaying steep clines in Europe and in North and South America[97]. The European colour cline, in particular, has been extensively studied, with rufous owls mostly in the north and northeast and white ones in the south. The classically held view for

its origin postulated that two parent populations with different colours – a white one from Iberia and a rufous one from near the Black Sea – evolved in allopatry, and then met in a secondary contact zone in central Europe[98,99]. This led to the classification of two European barn owl subspecies, the white *T. a. alba* (Scopoli 1769) and the rufous *T. a. guttata* (Brehm and CL 1831), with intermediate morphs assumed to be hybrids. However, with the advent of molecular markers, the first population genetics studies appeared to contradict the traditional theory. Genetic analyses with microsatellites suggested that, not only did barn owls colonize continental Europe from a single refugium in Iberia[100], but also that the colour cline was the result of local adaptation rather than a neutral by-product of recolonization[100,101]. The obvious question that ensues is what selective agents could be responsible for maintaining such a sharp cline, especially considering that neutral genetic differentiation across the continent is weak[102]. A definite answer is not yet available, and the correlation of multiple climatic variables with latitude makes the question difficult to address. Nonetheless, it has been shown that each morph has its own prey and habitat preferences, suggesting disruptive selection on alternative foraging strategies[103]. In addition, a global study on the three barn owl species suggested that convergent selection produces darker morphs in regions of higher rainfall and lower temperatures[96].

Plumage colour determination

To understand how a trait may respond to adaptive pressure or neutral processes, it is essential to understand how it is determined. Plumage colour in birds normally results from the deposition of pigments on the feathers, light scattering on structural components of the feather itself[104] or a combination of these two[105]. Pigmentary colour can be obtained by directly synthesising the pigment[106], ingesting it in food[107] or through deliberate staining[108]. Melanin, the most common pigment in vertebrates[109], occurs in two types: eumelanin which yields black, grey and brown colour, and pheomelanin that generates shades of yellow to rufous (reddish-brown). Its synthesis is regulated by the melanocortin-1 receptor (*MC1R*), part the melanocortin system, and mutations in this gene have been linked to melanic colour variation in multiple organisms[110–112], including birds[106,113,114].

The variation in colour of barn owl plumage from white to dark rufous (see above) is due to differential deposition of pheomelanin, yet its genetic basis has not been fully resolved. In Europe, a valine-to-isoleucine point substitution (V126I) in the *MC1R* gene explains roughly 30% of colour variation[102,115]. Owls homozygous for the ancestral valine allele (white allele) tend to be whiter, while the derived isoleucine allele (rufous allele) is linked to darker morphs (Figure 1) and significantly increases the expression of several key genes involved in melanin synthesis[116]. This very same *MC1R* mutation has been linked to melanic colour variation in gyrfalcons[117] and domestic ducks[118]. Geographically, the distribution of *MC1R* alleles in European barn owls

follows the colouration cline, with high frequencies of the derived rufous allele in the north and northeast of Europe[100,102]. The rufous allele is estimated to have evolved from the ancestral white allele shortly before or early during the recolonization process and then have been gradually selected in the route north[102]. As discussed above, it remains difficult to identify the agents of selection on the European colour cline. This is further complicated by the pleiotropic nature of the melanocortin system, which regulates behaviour and physiology alongside the production of melanin, and associations between these traits are common among vertebrates[119,120]. As such, it is conceivable that colouration is not the actual target of selection but reflects instead the effect of local adaptation on other linked cryptic traits.

Figure 1 – Barn owl plumage colouration and its genetic determination. On the left, white and dark rufous barn owls from Switzerland, a central population in the European cline where all morphs occur. Photo ©Isabelle Henry. On the right, colour distribution per *MC1R* genotype (W- white allele; R- rufous allele). Darker birds have higher brown chroma. Boxplot adapted from Burri *et al.*[102].

Thesis Outline

The general aim of my thesis was to characterize the impact of landscape barriers on the widely-distributed barn owl using whole-genome sequences and modern population genomics tools (with one exception: chapter 5). Up to the start of this work, population genetic analyses on this species had been performed with microsatellites and short mitochondrial sequences[100–102]. While it provided a general picture of neutral genetic patterns around the Mediterranean Sea, it lacked the resolution to decipher processes occurring at finer geographical scales. Specifically, the history of island populations remained convoluted (for example, Crete and the Canary Islands) or even unstudied (for example, Cyprus and the British Isles), while the role of mountain ranges was overlooked.

During my PhD, I focused mainly on insular barn owl populations from the eastern Atlantic Ocean and the Mediterranean Sea. Over three chapters, I studied the roles of neutral and adaptive processes in generating phenotypic and genetic variation in these populations with different levels of isolation. Then, in the two last chapters I studied the impact of mountain ranges in creating and shaping genetic structure and demographic history to varying degrees.

In detail, my thesis was composed of the following chapters:

Chapter 1: Barn owls from the British Isles are conspicuously white in comparison to their dark rufous mainland counterparts at similar latitudes. This is puzzling as owls are assumed to have colonised Britain from northern Europe and should thus be rufous, not white. Contrasting phenotypes such as this are often assumed to be due to selection, a theory we aimed to test. In this chapter, we first confirmed this discrepancy with phenotypical data and then addressed it by investigating how barn owls colonised the British Isles after the last glaciation. To do so, we coupled demographic modelling based on genomic data with species distribution models. Subsequently, we investigated whether northern (i.e. rufous) owls are frequent migrants into Britain by analysing patterns of neutral genetic structure supported by ringing data. Lastly, we probed genes associated with colouration, other than *MC1R*, for signals of selection on British and Irish owls.

Chapter 2: The fate of an insular population is shaped by the timing of colonisation, fluctuations in population size as well as connectivity to neighbouring populations, is tightly linked to island characteristics and the circumstances of colonization. Although the colonization of a new and environment can be an opportunity for local adaptation to act and even facilitate speciation, small and isolated populations can also suffer the detrimental effects of inbreeding. In this chapter, we used 65 whole genome sequences to contrast two insular barn owl populations in the eastern Mediterranean. Namely, we focused on Crete and Cyprus, the two largest islands in the region, with similar geological age, climate, surface and distance to the mainland. We

investigated the mainland origin of each of these populations and attempted to discern their past demographic histories and how they impacted the current patterns of relatedness and inbreeding on each island.

Chapter 3: Islands are often the setting of ecological speciation as they provide a new isolated environment to adapt to, insulated from encumbering gene flow. In the Canary Islands, the two easternmost islands out of the seven that compose the archipelago are home to a subspecies of barn owl, *T. a. gracilirostris*, characterized by its small size. Its presence on the islands closest to the continent, while the western islands hold the "standard" *T. alba*, is suggestive of strong local adaptation leading to early onset speciation, supposedly fuelled by the particularly arid local climate. In this chapter, we sampled both the subspecies and standard barn owls from the Canarias and inspected their genomes for signals of local adaptation, both to the common insular environment and to their specific islands.

Chapter 4: The joint impacts of climatic fluctuations and landscape barriers influence the distribution of species by periodically isolating and reuniting populations. With the advent of sequencing technology, one can now disentangle the signatures of such events on the genetic makeup of individuals. Previous work with traditional genetic markers proposed that barn owls colonised Europe from a single glacial refugium and could have created a secondary contact zone in the Balkans where it met the Middle Eastern lineage. Using 94 whole-genome sequences, we revisited in this chapter the postglacial demographic history of barn owls in the Western Palearctic. Specifically, we tested the possibility of cryptic glacial refugia by combining demographic and species distribution modelling which allowed us to address the taxonomic relevance of the supposed rufous subspecies *T. a. guttata*. Lastly, we focused on the decisive role of mountain ranges in parting genetic lineages of barn owls at a continental scale.

Chapter 5: In this final chapter, we were also interested in the impact of mountains as barriers but focused on the American barn owl, the largest of the sister taxa. We first assessed how North American and European owls differ in terms of overland connectivity using both genetic and ringing data from both continents. Then, we concentrated on the role of the biggest mountain barrier in this geographic range – the Rocky Mountains – and tested its impact on barn owl dispersal and gene flow. Chronologically, this chapter was published early in my thesis while we were still in the process of obtaining the genomic data for the remainder of the chapters. As such, it was not based on whole-genome sequences like the rest, but rather microsatellites and mitochondrial sequences from approximately 300 American museum specimens, and combined with European data from Burri *et al.*[102].

Chapter 1

Unexpected post-glacial colonisation route explains the white colour of barn owls (*Tyto alba*) from the British Isles

Ana Paula Machado[a*], Tristan Cumer[a], Christian Iseli[b], Emmanuel Beaudoing[c], Anne-Lyse Ducrest[a], Melanie Dupasquier[c], Nicolas Guex[b], Klaus Dichmann[d], Rui Lourenço[e], John Lusby[f], Hans-Dieter Martens[g], Laure Prévost[h], David Ramsden[i], Alexandre Roulin[a†], Jérôme Goudet[a,j†]

[†] co-senior authors

Status

Under review at *Molecular Ecology* (submitted April 2021)

See Appendices I-III on bioRxiv: https://doi.org/10.1101/2021.04.23.441058

Author contributions

APM, TC, AR, JG designed this study; APM produced whole-genome resequencing libraries; APM, TC conducted the analyses; ALD, MD produced the new reference genome; CI, EB, NG assembled it; TC identified coding regions; KD, RL, JL, HDM, LP and DR provided samples; APM led the writing of the manuscript with input from all authors.

Abstract

The climate fluctuations of the Quaternary shaped the movement of species in and out of glacial refugia. In Europe, the majority of species followed one of the described traditional postglacial recolonization routes from the southern peninsulas towards the north. Like most organisms, barn owls are assumed to have colonized the British Isles by crossing over Doggerland, a land bridge that connected Britain to northern Europe. However, while they are dark rufous in northern Europe, barn owls in the British Isles are conspicuously white, a contrast that could suggest selective forces are at play on the islands. However, analysis of known candidate genes involved in colouration found no signature of selection. Instead, using whole genome sequences and species distribution modelling, we found that owls colonised the British Isles soon after the last glaciation, directly from a white coloured refugium in the Iberian Peninsula, before colonising northern Europe. They would have followed a yet unknown post-glacial colonization route to the Isles over a westwards path of suitable habitat in now submerged land in the Bay of Biscay, thus not crossing Doggerland. As such, they inherited the white colour of their Iberian founders and maintained it through low gene flow with the mainland that prevents the import of rufous alleles. Thus, we contend that neutral processes likely explain this contrasting white colour compared to continental owls. With the barn owl being a top predator, we expect future research will show this unanticipated route was used by other species from its paleo community.

Keywords

Demographic inference; MC1R; Plumage colouration; Reference genome; Species distribution modelling; Whole-genome resequencing.

Introduction

The dramatic climate fluctuations of the Quaternary were key in shaping the global distribution of species and communities observed today[16,121]. During the last glaciation, northern Europe was largely covered by ice caps, and the resulting lower sea levels unveiled an expanded coastline widely different from that of today. The inhospitable conditions throughout the continent forced many temperate species into warmer refugia, most commonly the southern peninsulas of Iberia, Italy and Balkans[122,123]. Once temperatures started increasing about 18 thousand years BP, ice sheets melted, the sea rose and these species re-expanded northwards into central and northern Europe, a key step in determining their modern distribution and genetic structure across the continent. Early comparative phylogeography studies described differences in the route and timing of colonisation from each refuge population and identified the main post-glacial recolonization patterns from the south [121,123,124]. However, advances in sequencing technology and the consequent increase in studies with high representation molecular markers have since provided numerous examples of alternative routes and cryptic refugia for different taxa in mainland Europe as well as on islands[125–129].

The colonisation of the British Isles by terrestrial organisms has often been described in the context of the main phylogeographic patterns, with mainland north-western Europe as its origin [121,123,130]. Such a route would have been facilitated by Doggerland, a large land bridge of alluvial plains that connected Great Britain (GB) to mainland northern Europe before submerging under the north Sea 8'000 years BP[131,132]. Most terrestrial vertebrates of GB do appear to have arrived via Doggerland, as evidenced by the similarity between its mammal fauna and that of northern rather than southern Europe[130]. Nonetheless, some species believed to have followed this path were found to have had glacial refugia on the islands themselves[125], including plants[133], amphibians[134,135] and mammals[136,137]. Some taxa revealed other surprising post-glacial patterns such as colonization of the British Isles from multiple refugia in independent waves (badger[138], water vole[139]) and even separate colonisation of Ireland and GB (stoat[140]).

Barn owls (*Tyto alba*) recolonised western Europe following the last glaciation from a refugium in the Iberian Peninsula[100,102]. On the mainland, barn owl ventral plumage colouration follows a latitudinal cline ranging from mostly white in the southern populations to dark rufous in the north[100,101]. Despite their post-glacial expansion route, the clinal variation in colour was not a neutral by-product of range expansion, but was rather created and maintained by an independent post-glacial selective process[100]. The genetic basis of this pheomelanin-based trait is not fully understood, but a specific non-synonymous variant (V126I) in the melanocortin-1 receptor (*MC1R*) gene has been found to explain roughly 30% of its variation in Europe[115]. The derived *MC1R* rufous allele produces the darkest owl phenotypes and follows the European colour cline of increasing frequency with latitude[102].

It is hypothesised that, given their aversion to crossing large water bodies, barn owls recolonized Great Britain following the traditional route by crossing over Doggerland[141]. However, barn owls from the British Isles are famously white[141,142] in stark contrast to their darker mainland counterparts at similar latitudes. Over-land expansion from a north-western European population, inhabited mostly by rufous owls with 10% - 45% rufous *MC1R* allele, would be at odds with the whiteness of the GB population. This disparity is especially startling, given that rufous individuals disperse further than white ones[143,144], and would thus be more likely to colonise the islands in the first place. Finally, with GB being a recently isolated island, its avifauna is very similar to that of continental Europe (albeit less species rich), and examples of such phenotypic divergence from the mainland are rare; the barn owl is thus an intriguing exception. Being sensitive to extreme cold[145], a northern refugium seems unlikely. However, such phenotypic disparity suggests that, unless strong selective pressure is involved, the colonisation timing and route of barn owls of the British Isles might have been less straightforward than has been assumed.

Here, we address the post-glacial colonisation history of barn owls in the British Isles in light of the puzzling whiteness of their plumage. First, with a new broad sampling of 147 individuals from western Europe, we confirm that owls from the British Isles do not fit into the expected colouration and *MC1R* pattern of the mainland, with darker individuals at higher latitudes. Taking advantage of a highly contiguous newly-assembled reference genome and using the whole-genome sequences of 61 individuals, we use the neutral genetic structure to model the demographic history of barn owl colonisation of the northern part of Europe and the British Isles from a glacial refugium in Iberia. Then, we use ringing data to support estimations of current gene flow. Lastly, we investigate the potential role of other colour-linked genes in maintaining the phenotypic disparity in plumage colour between the British Isles and mainland Europe.

Materials & Methods

Tissue sampling, MC1R genotyping and colour measurement

In total, 147 individual barn owls were sampled for this study from six European populations (Sup. Table 1): Ireland (IR), Great Britain (GB), France (FR), Switzerland (CH), Denmark (DK) and Portugal (PT). A denser sampling was performed in the British Isles (n=113) as this was the first time these populations were studied, while for the mainland populations data was already available[102]. Genomic DNA was extracted from blood, feathers or soft tissue using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions, including RNA digestion with RNase A. A previously established allelic discrimination assay[115] was used to molecularly determine individual genotypes at the amino acid position 126 of the Melanocortin 1

receptor (*MC1R*) gene of the 147 individuals (Sup. Table 1). Additional allelic frequencies at this locus published in Burri *et al.*[102] from the mainland populations of interest to this study were used for context (N=247 individuals; Appendix 1).

For all individuals with available breast feathers (*n*=145), pheomelanin-based colour was estimated as the brown chroma of the reflectance spectra (for detailed description see Antoniazza *et al.*[101]. Briefly, the brown chroma represents the ratio of the red part of the spectrum (600–700 nm) to the complete visible spectrum (300–700 nm). The reflectance of four points of the top of three overlapping breast feathers was measured using a S2000 spectrophotometer (Ocean Optics, Dunedin, FL) and a dual deuterium and halogen 2000 light source (Mikropackan, Mikropack, Ostfildern, Germany). An individual's brown chroma score was obtained as the average of these four points. Brown chroma data from Burri *et. al.*[102] were used to complete the dataset, using the same individuals as for the *MC1R* analysis (Appendix 1). Given the marked non-normality of the data, a non-parametric Kruskal-Wallis test was performed to detect differences in coloration between the six populations. Further, a Pairwise Wilcoxon Rank Sum test was used to identify significant differences between pairs of populations using a Bonferroni correction.

New reference genome

As the available reference genome for the European *Tyto alba* was fragmented[92], a new reference was produced in order to achieve a near chromosome-level assembly. A full description of the process and its detailed results are given in Appendix 2. Briefly, a long-read PacBio library was produced from a blood sample of a Swiss individual at an expected coverage of 100x for the barn owl's 1.3Gb genome. FALCON and FALCON-Unzip v.3[146] were used to assemble PacBio reads. Then, a high molecular weight DNA Bionano optical mapping library was used to assemble PacBio contigs into scaffolds. Finally, repeated regions were identified using RepeatModeler v.1.0.11[147] and masked with RepeatMasker v.4.0.7[148]. Coding regions were identified using the Braker2 pipeline v.2.0.1[149–153].

Whole-genome resequencing and SNP calling

For the population genomics analyses of this study, the whole genomes of 61 out of the 147 individual barn owls were sequenced (Sup. Table 1). In addition, one eastern (*T. javanica* from Singapore) and one American barn owl (*T. furcata* from California, USA) were used as outgroups. See Supplementary Methods for a complete description of the library preparation, sequencing, SNP calling and filtering. Briefly, individual 100bp TruSeq DNA PCR-free libraries (Illumina) were sequenced with Illumina HiSeq 2500 high-throughput paired-end sequencing technology at the

Lausanne Genomic Technologies Facility (GTF, University of Lausanne, Switzerland). The bioinformatics pipeline used to obtain analysis-ready SNPs was adapted from the Genome Analysis Toolkit (GATK) Best Practices[154] to a non-model organism following the developers' recommendations, producing a full dataset of 6'721'999 SNP for the 61 European individuals with an average coverage of 21.1x (3.36 SD).

Population structure and genetic diversity

To investigate population structure among our samples, sNMF v.1.2[155] was run for K 2 to 6 in 25 replicates to infer individual clustering and admixture proportions. For this analysis, singletons were excluded and the remaining SNPs were pruned for linkage disequilibrium (LD) with PLINK v1.946[156] (parameters -indep-pairwise 50 10 0.1) as recommended by the authors, yielding 319'801 SNP. The same dataset was used to perform a Principal Component Analysis (PCA) with the R package SNPRelate[157]. Treemix[158] was used to infer population splits in our data, using the LD-pruned dataset further filtered to include no missing data (180'764 SNP). To detect meaningful admixture between populations, 10 replicates were run for 0 to 8 migration events, with the tree rooted on the PT population, representative of the glacial refugium. An extra run without migration events was conducted with a north-American owl as an outgroup in the dataset to verify that the root did not affect the topology of the tree.

To estimate population statistics, individuals found to be mis-assigned to their given population based on genetic structure analyses (PCA and sNMF) were removed so as not to bias allelic frequencies (N=3 individuals from Ireland). Individual expected and observed heterozygosity and population-specific private alleles were estimated using custom R scripts for each genetic lineage identified by sNMF with K=4. To account for differences in sample sizes, private alleles were calculated by randomly sampling 9 individuals from the larger populations (GB and central Europe) 10 times in a bootstrap-fashion and estimating the mean. Individual-based relatedness (β[159]), inbreeding coefficient for SNP data, overall and population pairwise F_{ST}[160] were calculated with SNPRelate.

Gene flow and migration analyses

Migration surface estimate

The Estimated Effective Migration Surface (EEMS) v.0.0.9 software[161] was used to visualize geographic regions with higher or lower than average levels of gene flow within our dataset. The provided tool *bed2diff* was used to compute the matrix of genetic dissimilarities, from the dataset pruned for LD produced above. The free Google Maps api v.3 tool

(http://www.birdtheme.org/useful/v3tool.html) was used to draw the polygon outlining the study area in western Europe. EEMS was run with 750 demes in three independent chains of 5 million MCMC iterations with a 1 million iterations burn-in. Results were checked for MCMC chain convergence visually and through the linear relation between the observed and fitted values for within- and between-demes estimates using the accompanying R package rEEMSplots v.0.0.1[161]. The three MCMC chains were combined to produce maps of effective migration and diversity surfaces with the provided functions in rEEMSplots.

Treatment and analyses of capture-recapture data

In addition to genomic data, recapture data of ringed barn owls across Europe were obtained from the EURING database (obtained in March 2020[162,163]). Specifically, we estimated the frequency of crosses over open water between GB and central and western Europe, as well as between GB and Ireland. To do so, we kept records of birds that had been recaptured at least once after ringing (n=94'797 recaptures, n=80'083 individuals, from 1910 to 2019) and filtered the accuracy of the "time of capture" parameter to a period of within 6 weeks of the reported date to exclude potentially unreliable data points. We extracted the number of birds ringed and recaptured in GB and Ireland, as well as in the countries that produced or received migrant birds from these islands and central Europe (Belgium, Denmark, France, Spain, Germany, Switzerland and The Netherlands). Crosses between islands and to/from the mainland are reported and include birds that were found dead in the sea (n=8). All counts and percentages reported are relative to the number of individual birds recaptured (rather than number of recapture events, as a single bird can be recaptured multiple times).

Post-glacial species distribution

To support the demographic scenarios tested in the following section, we modelled the past spatial distribution of barn owls in western Europe, in order to identify the regions of high habitat suitability at the last glacial maximum (LGM, 20'000 years BP). A complete description of the models can be found in Supplementary Methods. Briefly, using Maximum Entropy Modelling (MaxEnt), a presence-only based modelling tool, we built species distribution models (SDM) for the Western Palearctic (Sup. Fig. 1) based on climatic variables extracted from the WorldClim database[164] at 5 arc min resolution. The best combination of feature and regularization multiplier based on the corrected AIC (as recommended by Warren & Seifert, 2011) was achieved with a quadratic model with 1 as regularization multiplier (Sup. Table 5). Then, the output of the models was transformed into a binary map of suitability in which only cells suitable in 90% of the models are presented as such in the map. All models were then projected to the mid-Holocene (6'000

years BP) and LGM (20'000 years BP) conditions extracted from WorldClim at the same resolution as current data. For each timepoint, the results of the models were merged and transformed into a binary map as for the current data.

Maximum-likelihood demographic inference

Data preparation

To discriminate between different demographic scenarios for the colonisation of the British Isles by barn owls we used the software *fastsimcoal2*[75,166]. Individuals and variants in the dataset used here as input went through additional filtering steps in an attempt to ensure neutrality and homogeneity between samples (Sup. Methods). Given their similarity (Fig. 1b&c), the original populations of France, Denmark and Switzerland were combined into a central European population (EU). The remaining populations were Portugal (PT), Great Britain (GB) and Ireland (IR), with 8 individuals each (Sup. Table 1). Population pairwise SFS were produced from the filtered dataset of 739'168 SNP.

Demographic scenarios and parameters

Three different scenarios of colonization of central Europe and the British Isles from the Iberian Peninsula were simulated (Figure 3), distinguishable by the difference in timing and origin of the insular populations: north-western (NW) European origin, Iberian origin and insular refugium. Each scenario was further split in two versions (A and B) to accommodate small changes in topology. For all scenarios, wide search ranges for initial simulation parameters were allowed for population sizes, divergence times and migration rates while accounting for census and geological data (Sup. Table 7). Splits were preceded by instantaneous bottlenecks, in which the founding population size was drawn from a log-uniform distribution between 0.01 and 0.5 of current population sizes. All times were relative to the end of the last glaciation (18'000 years BP, rounded to 6000 generations ago), bounded between the present and the previous demographic event in the model.

In scenario NW European origin A, after an initial post-glaciation size expansion, the ancestral PT population colonized central Europe. From here, barn owls sequentially reached Great Britain and Ireland, potentially across the Doggerland land bridge. In version B, a smaller second glacial refugium is hypothesized to have existed in southern France, above the Pyrenees, as the founder of the central European population after the glaciation. In both versions, barn owls reached the British Isles from central Europe. In the Iberian origin scenarios, the insular populations originated directly from PT. Spatially, this could have taken place across now-submerged land in

the Bay of Biscay, west of current-day France and north of Spain. Genetically, the insular birds would have been derived from the initial genetic pool in Iberia rather than from the subset in central Europe. Versions A and B of this scenario differ in the timing of colonization, with Europe being colonized before the islands in A and after in B. Lastly, the insular refugium scenarios hypothesize a separate and smaller glacial refugium in the south of the British Isles that would have been the origin of today's populations on the islands. Such refugia have been described for some terrestrial organisms albeit not birds[125,133,134,167]. Central Europe would be colonized post-glacially from PT. In version A and B of this scenario, the second glacial refugium would be part of an ancestral GB or IR population, respectively.

In summary, the NW European origin scenario reflects the shortest overland path based on current geography, whereas the remaining scenarios attempt to address the whiteness in the British Isles by avoiding shared ancestry with darker-coloured populations at different time scales, as well as the changes in the coastline during and after the last glaciation. For all scenarios, migration was allowed between neighbouring populations (Figure 3; Sup. Table 7).

Demographic inference

Demographic simulations and parameter inference were performed under a composite-likelihood approach based on the joint site frequency spectrum (SFS) as implemented in *fastsimcoal2*[75,166]. For each scenario, 100 independent estimations with different initial values were run (Sup. Methods). The best-fitting scenario was determined based on Akaike's information criterion (AIC; Akaike 1974) and confirmed through the examination of the likelihood ranges of each scenario as proposed in Kocher *et al.*[169]. For the best-fitting scenario, non-parametric bootstrapping was performed to estimate 95% confidence intervals (CI) of the inferred parameters. For each block-bootstrapped SFS, 50 independent parameter inferences were run for the best-fitting scenario (see Sup. Methods for a detailed explanation).

Genome scans of colour-linked genes

Genome-wide scans were used to compare patterns of divergence and diversity between populations. SNPs were filtered to a minimum derived allelic frequency of 5%, and VCFtools was used to calculate nucleotide diversity (π) for each population and to estimate F_{ST}[160] between pairs of populations in 20kb sliding windows with 5kb steps across the whole genome. For our comparisons, Great Britain and Ireland were combined as British Isles; France and Switzerland as central Europe. Denmark was not included in the latter due to its markedly darker phenotype (Fig. 1a). The British Isles were compared to all other groups of individuals: white in Portugal,

intermediate in central Europe and dark rufous in Denmark. Further, Portugal and Denmark were also compared.

In our genomic dataset, owls from the British Isles and Portugal carried the same genotypes at the *MC1R* mutation (100% V allele) despite there being considerably more variation in colour among Portuguese individuals (Fig. 1a). As such, we first investigated whether insular individuals showed particular diversity or divergence at the surrounding positions within the *MC1R* gene that could relate to their pure white colour. Since the *MC1R* gene in barn owls is particularly GC rich[115] and is embedded in a region with a lot of homopolymeric sequences, the sequencing in this region has a considerably lower coverage than the average of the genome. To account for this, the scaffold containing this gene was extracted from the raw SNP set and re-filtered with similar site thresholds as described above, except for allowing 25% overall missing data (instead of 5%), limiting the minimum individual DP to 5 (instead of 10) and the minimum minor allelic count to 3. VCFtools was used to calculate nucleotide diversity for each population and to estimate F_{ST}[160] between pairs of populations in 5kb sliding windows with 1kb steps along this scaffold.

Second, to widen our search to other colour-linked genes besides *MC1R*, we mapped 22 autosomal candidate genes (Appendix 3) onto the reference genome using Blast v.2.9.0[170]. Windows including the candidate genes were plotted onto genomic scans (5kb windows with 1kb step) to check for overlap with peaks or drops in diversity and/or differentiation.

Results

MC1R genotyping and colour measurements

Plumage colour comparisons showed that the British Isles have the whitest owls of all measured European populations (Fig. 1a; $X^2 = 243.28$, $p < 0.001$). Most pairwise comparisons were significantly different after correction, with the exception of between GB and IR owls, and between CH and FR. As for *MC1R* genotyping, notably no I allele was found among the 113 genotyped individuals of the British Isles indicating it is absent from these populations or at very low frequency.

New reference genome

The new reference genome produced for European barn owl was a near chromosome level assembly, and has been deposited at DDBJ/ENA/GenBank under the accession JAEUGV000000000. Sequencing of the new reference genome's PacBio library yielded 7.3 million long reads with a total sum length of unique single molecules of 135 Gbp (N50 > 31Kb) yielding a realized coverage of 108x. Its assembly with FALCON and FALCON-Unzip resulted in

478 primary contigs partially phased, and 1736 fully phased haplotigs which represented divergent haplotypes. Optical mapping with Bionano produced a final assembly of 70 scaffolds, slightly more than the barn owl's karyotype of 46 chromosomes [92]. The final assembly was 1.25 Gbp long, with an N50 of 36 Mbp and BUSCO score of 96.9% (see Appendix 2 Table 1 for full assembly metrics). In comparison, the previous reference assembly [92] had 21,509 scaffolds, with an N50 of 4.6 Mbp.

Population structure and genetic diversity

Our dataset was composed of four main genetic clusters identified by individual ancestry analyses (sNMF) and PCA clustering. Individuals from Portugal (PT), Great Britain (GB) and Ireland (IR) belonged to their specific population ancestry, while individuals from France (FR), Denmark (DK) and Switzerland (CH) formed a single central European cluster (Fig. 1b,c; Sup. Fig. 3). Consistently, the first axis of the PCA opposed PT to GB & IR, as seen with sNMF K=2 (Sup. Fig. 3). The second axis clustered the central European individuals together and opposed them to PT (Fig. 1b). GB and IR segregate in both the first and second axes. Three barn owls sampled in Ireland showed a clear genetic signal of belonging to the Great Britain genetic cluster (Fig. 1b,c; Sup. Fig. 3). To avoid their interference in estimating allelic frequencies, they were omitted when estimating diversity and differentiation statistics.

Analyses of genetic drift with Treemix yielded a population tree with two branches splitting from PT. The first is a long branch of drift that divides into GB and IR, while the second, shorter branch, diversified into the three central European populations (Fig. 4a). Plotting the likelihood of runs and the standard error (SE) of each tree showed that including one migration event from PT to CH (migration edge weight = 0.27) considerably increased the fit of the tree to the data (Sup. Fig. 5).

The overall F_{ST} was 0.035. Population pairwise F_{ST} were the highest between Ireland and central Europe (Sup. Table 3). Overall, populations within central Europe showed the smallest differentiation (F_{ST} below 0.02) and the British Isles had the highest values in comparison to all mainland populations (Sup. Table 3). Diversity estimates showed higher levels in PT than in any other population and the British Isles had the lowest (Sup. Table 2). Individual relatedness was highest within IR, followed by GB (Sup. Fig. 4). On the opposite end, PT had the lowest within-population relatedness as well as with the other populations, consistent with its higher diversity.

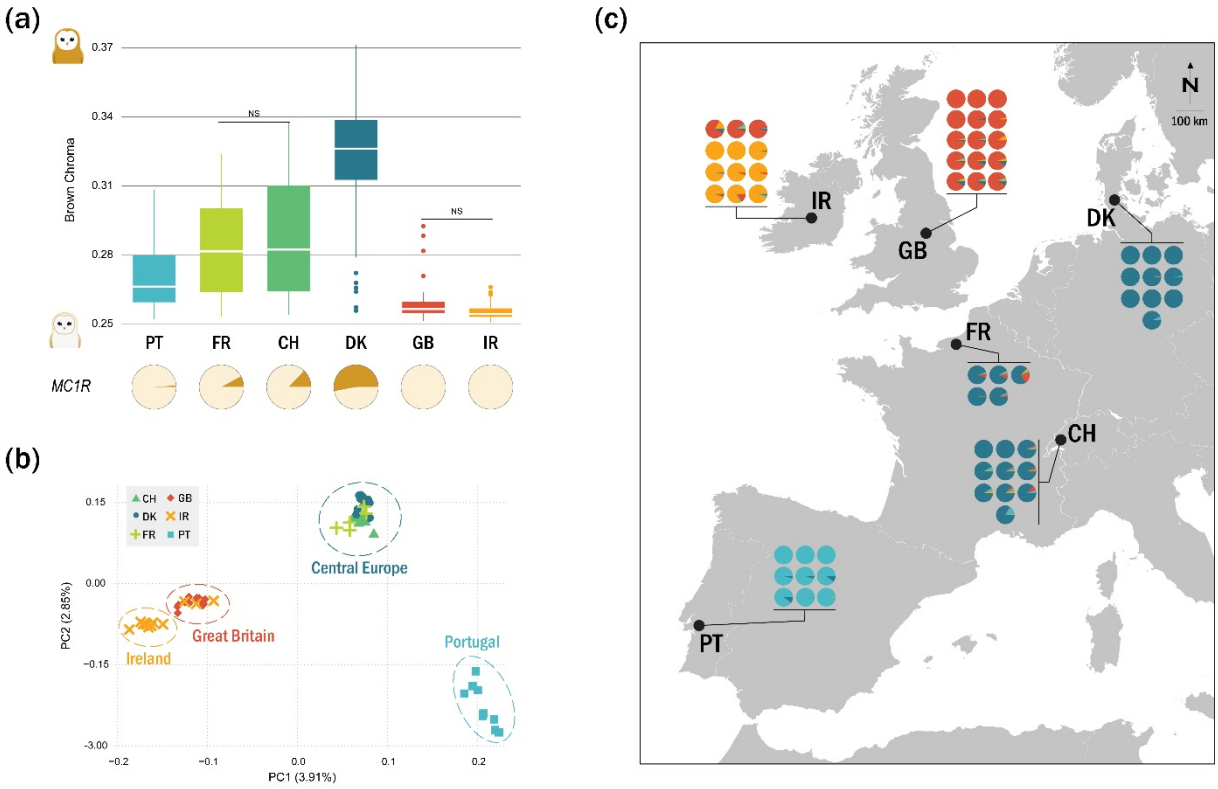


Figure 1 – Colouration and genetic structure of barn owl populations in Western Europe. **(a)** Brown chroma distribution and MC1R allelic frequencies of each studied population (total N=145). Higher brown chroma indicates redder owls. NS denotes the non-significant pairwise comparisons. The pies below the plot indicate the populations' allelic frequencies at the MC1R mutation: the rufous allele in brown and the white in beige. **(b)** PCA based on the pruned SNP set of the 61 individuals whose whole genome was re-sequenced. Point shape and colour denote populations according to the legend. Dashed circles enclose sample clusters observed in sNMF. Values in parenthesis indicate the percentage of variance explained by each axis. **(c)** Population structure. Small pie charts denote the individual proportion of each of K=4 lineages as determined by sNMF. Black dots are located at the approximate centroid of each sampled population.

Migration and gene flow

The English Channel – including the strait of Dover and the southernmost part of the North Sea – was identified by Estimated Effective Migration Surface (EEMS) as a region with lower than average gene flow between populations (Fig. 2a). This corridor extended west to the Atlantic. Furthermore, this analysis highlighted a region of low gene flow between the British and Irish populations. It put a barrier in Ireland by separating the north from the rest of the island, effectively isolating the three individuals sampled in Ireland that genetically resemble the British and clustering them with GB.

Analyses of capture-recapture data of ringed owls (N=80'083 individuals, from 1910 to 2019) revealed that all individuals ringed in Ireland (N=81 individuals) were recaptured in Ireland. As for GB, the vast majority (99.92%) of its ringed individuals (N=17'903) were also recaptured in GB and only 14 migrated out of the island: seven to Ireland (100% of this island's immigrants) and

seven to mainland Europe (Fig. 2b – Emigrants; Sup. Table 4a). In the opposite direction, GB received 21 individuals from the mainland (Fig. 2b - Immigrants), specifically from Belgium, the Netherlands and northern Germany (Sup. Table 4b). Of the immigrant birds, 19 were found dead, one severely injured with unknown fate, and one breeding. The latter was a female from the Netherlands, but the fate of its brood is not known. In the mainland, central European countries show considerably higher exchanges of individuals with each other (Sup. Table 4c) than with GB (Sup. Table 4b).

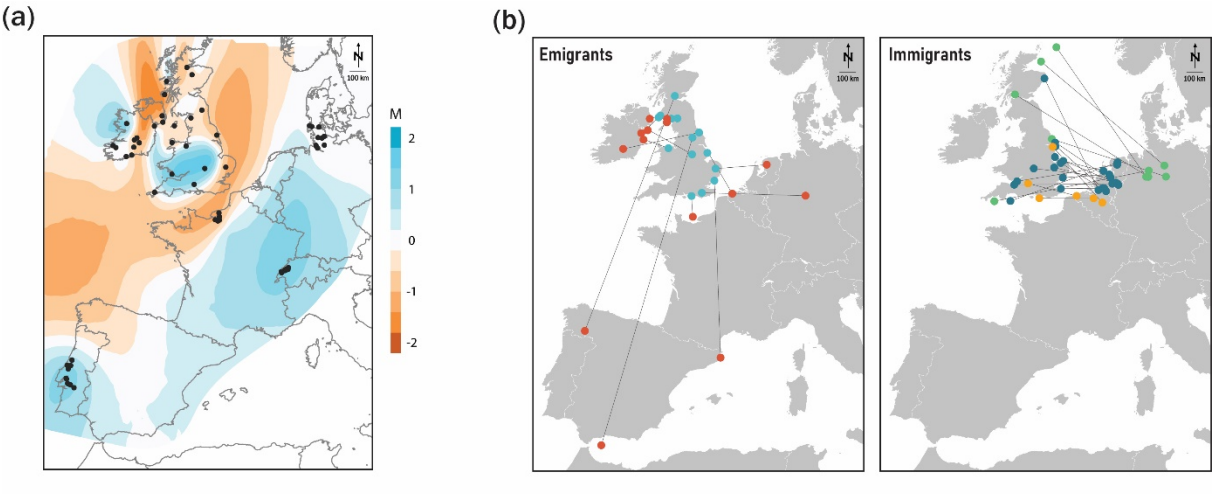


Figure 2 – Barn owl gene flow and dispersal between the British Isles and mainland Western Europe. **(a)** Estimated effective migration surface (EEMS) based on whole-genome data. Blue and brown shading denote regions of higher and lower than average gene flow, respectively. Black dots indicate individual sampling location. **(b)** Ringing and recapture locations of barn owls known to have flown out of (Emigrants) or into (Immigrants) Great Britain from 1910 to 2019, based on data courtesy of EURING. Lines simply connect two capture points and do not represent the actual path travelled by birds. Emigrant ringing locations in GB are coloured in blue, and recaptures in red. Immigrants into GB are coloured according to country of origin (orange – Belgium; green – Germany; blue – The Netherlands).

Post-glacial species distribution

Habitat suitability projections for barn owls in the past showed that, at the time of the last glaciation, there was suitable land for barn owls outside of the known refugium of Iberia from a climatic perspective (Fig. 4c). Specifically, south of today's British Isles there was a corridor of suitable land submerged nowadays, as well as along the south and western coasts of France, and a small cluster inland southern France. At the mid-Holocene (6'000 years BP), the coastline resembled that of present day, and the distribution of suitable habitat for barn owls resembled that of nowadays (Fig. 4c).

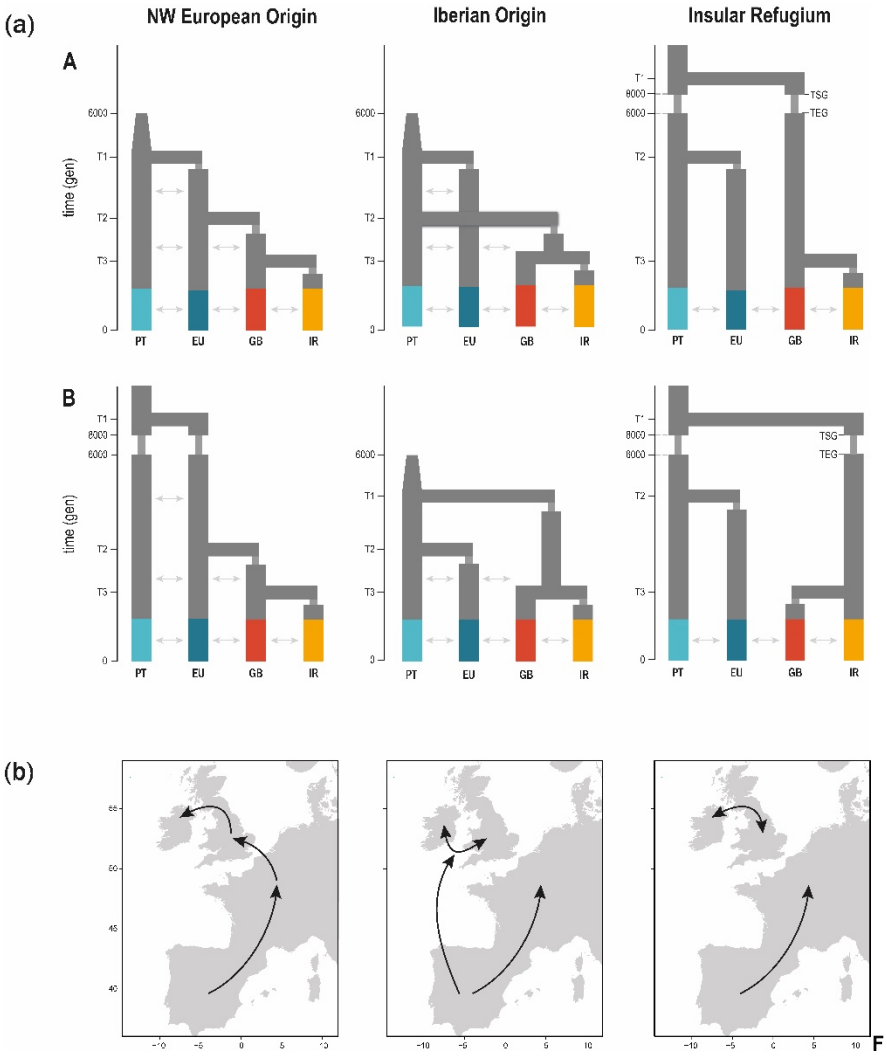


Figure 3 – Hypothesized demographic scenarios for the colonization of the British Isles by barn owls. **(a)** Tested demographic scenarios for the colonization of the British Isles by barn owls. There are three main topologies – NW European Origin, Iberian Origin and Insular Refugium – each with two version (A & B; first and second line respectively). The four main genetic clusters in our dataset were used: Portugal (PT), Central Europe (EU), Great Britain (GB) and Ireland (IR). Population EU in this analysis is composed of individuals from FR and DK. Indicated times were fixed in the models (6'000 and 8'000 generations ago), and the remaining time parameters were inferred relative to them or to the event immediately before (for example, T3 was bound between the present and T2). Cones depict post-glacial size increase and arrows gene flow between adjacent populations. In Insular Refugium topologies, TSG= time of start of glaciation in the insular lineage, TEG= time of end of glaciation in the insular lineage. **(b)** Schematic representation of the colonisation route to the British Isles for each scenario.

Demographic inference

AIC and raw likelihood comparisons showed the Iberian origin B model to be the best at explaining the SFS of our dataset (Sup. Table 6; Fig. 4b). In this model, an ancestral insular lineage split from the mainland refugium lineage in Iberia fairly soon after the end of the glaciation, estimated at approximately 13'000 years ago (95% CI: 7'000-17'000 years BP; calculated with 3-year generation time). Only much later, the model predicted the split of the central EU population from PT at 4'000 years BP (95% CI: 1'000-5'000 years BP) and the

separation between GB and IR at 1'200 years BP (95% CI: 220-2'200 years BP). Estimated effective population size was the largest in the PT population, followed by EU, GB and IR (Fig. 4b). Migration between populations was higher before these split than in recent times (Sup. Table 8; Ancestral vs Recent migration). Highest recent gene flow was observed from PT to EU, agreeing with Treemix's first migration event (Sup. Fig. 5). Migration levels between the two islands and with the mainland were of a similar order of magnitude and less than half of that between mainland populations, consistent with the two barriers to gene flow identified by EEMS (Fig. 2a). Point estimates with 95% confidence intervals for all parameters of the best model are provided (Sup. Table 8), as well as single point estimates for the rest of the models (Sup. Table 7).

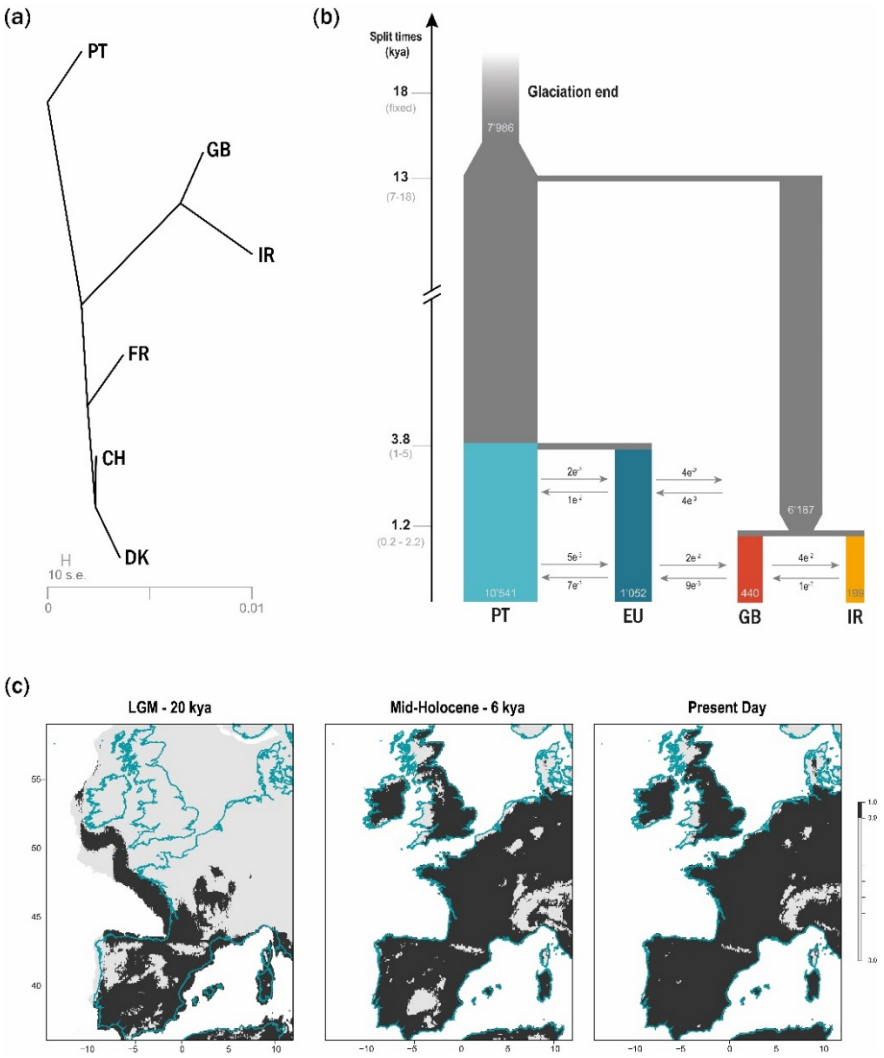


Figure 4 – Demographic history of barn owls in the British Isles. **(a)** Treemix analysis with zero migration events. **(b)** Best supported demographic model for the colonisation of the British Isles as determined by fastsimcoal2. Time is indicated in thousands of years, determined using a 3-year generation time, confidence intervals at 95% are given between brackets. Population sizes (haploid) are shown inside each population bar; arrows indicate forward-in-time migration rate and direction. Population EU in this analysis is composed of individuals from FR and DK. **(c)** Species distribution model of barn owls projected into past conditions – last glacial maximum (20'000 years BP) and mid-Holocene (6'000 years BP) – compared to today's distribution. Only locations with high suitability in at least 90% model averaging are coloured in dark grey. Below that threshold cells were considered as unsuitable (lightest grey shade on the graph). Modern coastline is shown in blue.

Genome scans of colour-linked genes

Genome-wide scans revealed some high peaks of differentiation between populations, but none overlapped with the colour-linked candidate genes tested (Appendix 3). In particular, the *MC1R* region showed no particular sign of increased differentiation between pairs of populations, nor drop in diversity, with the exception of the known causal SNP between populations with different genotypes (Fig. 5b; Appendix 3).

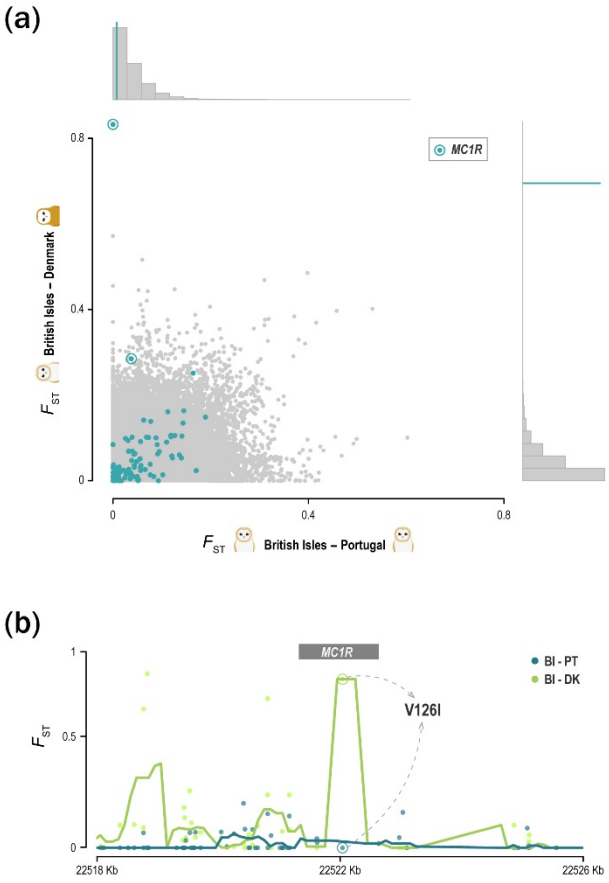


Figure 5– Differentiation at the colour-linked locus V126I of the *MC1R* gene between differently coloured barn owl populations in Europe. **(a)** Genome-wide F_{ST} values per window (in grey, 20Kbp windows with 5Kbp steps), between two white barn owl populations on the horizontal axis – British Isles (BI) and Portugal (PT) – and between one white and one rufous on the vertical axis – British Isles and Denmark (DK). The distribution of each variable is shown on the histograms. Blue dots indicate the F_{ST} at windows containing the tested colour-linked genes. Windows containing the *MC1R* are encircled in blue, and their mean is shown with the blue line on the histograms. **(b)** F_{ST} along the genome around the *MC1R* gene (grey box). Dots are per site values, and lines the mean over sliding windows (500bp with 100bp step), for the same comparisons as above: British Isles and Portugal in blue; British Isles and Denmark in green. Circled dots indicate the V126I locus in both comparisons.

Discussion

Like most terrestrial species, barn owls are assumed to have colonized the British Isles after the last glaciation by crossing over Doggerland, a land bridge that connected GB to northern Europe. In continental Europe, barn owls display a marked latitudinal colour cline maintained through local adaptation[101]. However, in the British Isles they are conspicuously white in comparison to their nearest mainland counterparts questioning whether this is their source population. The currently held interpretation for their whiteness is a strong selection on this trait after

colonisation. Here we provide evidence for a simpler explanation that does not require selection. Using whole-genome sequences and demographic simulations, we show that the colour disparity can be explained by the patterns in neutral genetic differentiation, resulting from an unexpected colonization route to the British Isles. We provide evidence for an early split of the insular lineage and low levels of gene flow with the mainland. Having found no evidence of selection on colour in the British Isles, it is plausible that this population has simply remained the white colour of its founders.

Genetic isolation from the mainland

Our results based on whole genomes revealed genetic structure among western European barn owls despite shallow differentiation for a cosmopolitan bird (overall F_{ST} 0.035) and showed genome wide genetic isolation between the islands and the mainland, accompanied by low levels of gene flow and migration. On the mainland, Portugal displayed the highest levels of genetic diversity (Sup. Table 2) and the largest estimated population size (Fig. 4b; Sup. Table 8), in accordance with its known role as a glacial refugium [100]. While forming its own population cluster (Fig. 1b,c), we found evidence of considerable gene flow towards central Europe (Fig. 2a, 4a,b; Sup. Table 8), consistent with a recent split between the two populations (less than 5'000 years BP; Fig. 4a) and the relatively low differentiation between them (Sup. Table 3). This suggests that the Pyrenees are permeable to barn owl migration, unlike other higher and larger mountain ranges [171]. In central Europe, barn owl populations appear to be remarkably homogenous genetically, despite covering a large geographical and colour range (Fig. 1, Sup. Table 3), in accordance with previous studies of continental Europe with traditional markers[101], and supported by capture-recapture data that revealed high amounts of exchanges in central Europe (Sup. Table 4c).

Ireland and GB showed the lowest diversity and estimated effective population sizes in our study (Fig.4; Sup. Tables 2, 8). Barn owl populations of each island are genetically distinct from each other as well as from the mainland (Fig. 1, 4a; Sup. Table 3). Genomic differentiation (Fig. 1, 2a, 4a,b; Sup. Table 3) and capture-recapture data with only a handful of exchanges recorded in the last century (Fig. 2b; Sup. Table 4a&b), suggest gene flow with the mainland is low. Specific analyses highlighted a barrier to gene flow extending from the Celtic Sea, through the English Channel to the North Sea (Fig. 2a), effectively isolating the British Isles from the mainland. Between the two islands, isolation appears to be recent (less than 2230 years BP; Fig. 4a,b; Sup. Table 8), despite relatively high genetic differentiation (Sup. Table 3) likely exacerbated by an important effect of genetic drift in such small populations. There is little sign of current pervasive admixture in either direction (Fig. 1c), consistent with the role of the Irish Sea as a strong barrier. However, there are records of owls from GB migrating into northern parts of Ireland (Fig. 2b –

Emigrants), the most easily accessible part of the island, while avoiding major water bodies by island-hopping from Scotland. Curiously, three of the individuals we sampled in Ireland for whole-genome sequencing (all sampled from found carcasses) appeared to be genetically from GB (Fig. 1b,c), driving EEMS to place a gene flow barrier nearly along the political border between the two countries of Ireland instead of the sea (Fig. 2a). Northern Ireland appears to be inhospitable for barn owls, at least in modern times, with only 1 to 3 pairs recorded per year in the whole country[172]. It could be acting as an extension of the sea barrier with the birds that fly in from GB being unable to find mates and thus not contributing to the genetic pool of the southern population, accentuating the differentiation between the two islands.

Disparity in plumage colouration

Plumage colouration in barn owls, and the linked *MC1R* locus, follow a clinal distribution in continental Europe maintained by local adaptation[101,102]. Here, we formally establish that barn owls from the British Isles do not follow the continental latitudinal cline and are whiter than any continental population in Europe, including even Portugal (Fig. 1a), confirming what was previously untested common knowledge among ornithologists. The rufous *MC1R* allele appears to be virtually absent in these populations in contrast to its close to 50% frequency at similar latitudes on the mainland, where dark morphs are positively selected (Fig. 1a [100,102]). While genome-wide scans confirmed the important role of the known *MC1R* mutation in determining rufous colouration (Fig. 5a), it appears to be restricted to the SNP variant itself and not the adjacent genomic regions (Fig. 5b). Our results are consistent with previous studies that showed that carrying a single copy of the rufous allele is sufficient to ensure a darker phenotype, while individuals homozygous for the white allele can have a wide range of colouration[102,115].

This colour trait is likely polygenic, given that the known *MC1R* mutation explains only 30% of its variation[102,115] and its high heritability[95]. Other loci could act in conjunction with a homozygous white *MC1R* to either produce whiter birds in GB or slightly darker morphs in Iberia. However, none of the other known colour-linked genes tested here explain how white owls homozygous for the white *MC1R* allele from Portugal reach darker phenotypes than those of the British Isles (Fig. 1a, 5a; Appendix 3). Alternatively, it is conceivable that the phenotype we observe – colouration – simply reflects the pleiotropic effect of insular local adaptation on other linked cryptic traits. The melanocortin system regulates behaviour and physiology alongside the production of melanin, and associations between these traits are common among vertebrates[119,120]. Further work, potentially focusing on colour-varied populations to avoid the confounding factor of population structure could help elucidate the genetic basis of barn owl plumage colouration. If such other loci are found, it would be fascinating to investigate their distributions and interaction with *MC1R* along the continental colour cline as well as on the British Isles.

Colonisation of the British Isles

Demographic simulations based on neutral sites showed that the British Isles were colonized from the glacial refugium in the Iberian Peninsula soon after the end of the glaciation (Fig. 3b). This would have occurred while the British Isles were still connected to the mainland and the landmass extended considerably further south than today's islands, following a corridor of suitable climatic conditions along the coast leading west (Fig. 4c) completely separate from Doggerland. It is also possible that this corridor was already occupied by barn owls in a continuous population with Iberia before becoming isolated, as this species easily maintains high over-land gene flow (Fig. 1b&c, 2a; Sup. Table 4c). Our wide confidence intervals make it hard to pin-point exactly the time of the actual split between the insular lineage from that of Iberia, but with the fast rise of sea levels and opening of the delta in the English Channel, the southern route to the islands would have been closed by 10'000 years BP[173,174]. Crucially, at this time prey would already be available in the form of voles, shrews, lemmings and bats[130]. Once separated, the insular lineage underwent a long period of genetic drift, isolated from the mainland population in Iberia but homogenous within itself before splitting between the two islands (Fig. 4a,b).

On the mainland, central European populations split genetically from the Iberian refugium much later (less than 5'000 years BP). Large population sizes and high overland gene flow (Fig. 4b; Sup. Table 8) might thus have maintained low differentiation for a long period of time, but also climatic conditions north of the Pyrenees may have taken longer to become favourable. The latter hypothesis would further counter the traditional point of view of Doggerland as the point of arrival for barn owls, as they could have not yet reached such high latitudes before Doggerland submerged 8'000 years BP. Intriguingly, our demographic model predicts high migration from GB into central Europe between the splits of the latter with Iberia and between the two islands (Fig. 4b), which appears unlikely with all land bridges submerged at this point (less than 5'000 years BP). It is possible that the migration rate was inflated as the model did not allow for gene flow between the ancestral insular and mainland populations before the first split and thus forced all migration to occur in a short time interval (Fig. 3).

In light of the inferred demographic history, barn owls of the British Isles would have inherited their whiteness from their source mainland population, the refugium in the Iberian Peninsula, and kept it through small population size, genetic drift and low gene flow. Although it is conceivable that some copies of the rufous *MC1R* allele were present in the founding insular population, similar to its frequency in Iberia (1%; Fig. 1a), in the absence of strong positive selection in the insular environment, it could have disappeared through genetic drift given the small effective sizes (Fig. 4b; Sup. Table 8). Thus, the selective pressure that renders the rufous colour and allele adaptive in northern continental Europe[101,102], may be absent in the British

Isles. Still, we cannot rule out that gene flow with the mainland is too weak and over too short a period of time to offer selection sufficient variation in the British Isles to increase the frequency of imported rufous alleles. If, conversely, the white morph was positively selected on the islands – potentially explaining its purer shade – we would have expected to find extended haplotypic differentiation when comparing it to the white mainland birds, which we did not (Fig. 5; Appendix 3). Therefore, it appears the white insular morph can be most parsimoniously explained by relaxation or absence of selective pressure in contrast to the mainland. Such a pattern is actually common among insular birds which, due to relaxed selection, tend to display less colourful plumage than their mainland counterparts [57,175], as also observed in the barn owl worldwide[176].

This early history of colonisation of the British Isles inferred here from whole-genome sequences and supported by SDM projections on past climatic features is apparently unique among terrestrial vertebrates, but it is far from the first to deviate from the most common colonisation route over Doggerland (e.g.[125,133–136]) or to indicate an earlier colonisation than generally assumed[140,177]. The case of the stoat (*Mustela erminea*) is particularly interesting as it was found to have had an isolated glacial refugium also in now submerged land southwest of today's French coastline on the Bay of Biscay (Fig. 4c – LGM [140]). From there they reached Ireland very early as the temperatures started rising but, as the Celtic Sea opened 15'000 years BP, only colonized GB much later over Doggerland[140]. The key difference between the two cases lies in the fact that barn owls maintained a homogenous population between GB and Ireland through flight.

Conclusion

Our study demonstrates that barn owls followed a highly uncommon post-glacial colonisation route to the British Isles. Likely taking advantage of the since submerged suitable habitat on the Bay of Biscay, barn owls reached the islands much earlier than expected from this southern point. The inferred demographic history could explain the whiteness of these populations through a combination of founder effect and low gene flow, and without the need to invoke selective pressures. We contend high quality population genomic data associated with species distribution hindcasting will reveal an unusual demographic history and post-glacial colonization for many non-model species. We wonder how often an intuitive selective explanation for a conspicuous phenotype could turn out to be the result of purely neutral processes.

Chapter 1 - Supporting Information

Supporting Methods

Whole-genome resequencing

The whole genomes of 63 individuals (61 European and 2 outgroups) were sequenced in this study. Their sex was determined with molecular markers[178] prior to sequencing. DNA quality and fragmentation were assessed with Fragment Analyzer™ (Advanced Analytical Technologies Inc.). Sample Purification Beads (SPB; Illumina, California, USA) were used to remove fragments smaller than 500 bp prior to library preparation of samples that showed high spread of DNA fragment size. Further, the intensity of the initial mechanical fragmentation step (Covaris, Woburn, MA, USA) was adjusted based on Fragment Analyzer profiles to promote homogenous library sizes. Individually tagged 100bp TruSeq DNA PCR-free libraries (Illumina) were prepared according to manufacturer's instructions. Whole-genome resequencing was performed on multiplexed libraries with Illumina HiSeq 2500 high-throughput paired-end sequencing technology at the Lausanne Genomic Technologies Facility (GTF, University of Lausanne, Switzerland).

Data preparation and SNP calling

Raw reads were trimmed with Trimommatic v.0.36[179] for Illumina adapters, and minimum sequence length of 70 bp. BWA-MEM v.0.7.15[180] was used to map the trimmed reads to the newly generated reference barn owl genome. Despite our libraries being PCR-free, potentially duplicate reads were marked with Picard-tools v2.9.0 (http://broadinstitute.github.io/picard) MarkDuplicates per run and per library.

Base quality score recalibration (BQSR) was performed following the iterative approach recommended for non-model species for which a set of "true variants" is not available, using high-confidence calls on the un-calibrated calls in a bootstrap-fashion to achieve convergence of the quality of variant calls. Here, a first calling of high-confidence variants was done with a combination of GATK's HaplotypeCaller and GenotypeGVCF v.4.1.3 and ANGSD v.0.921[181]. Both sets of calls were filtered for a maximum of 5% missing data, individual depth (DP > 10 and DP < 30), mapping quality (MQ > 40) and minor allelic frequency (MAF > 0.02). The intersect of the two call sets was used as the known set of variants to run BQSR a first time with BaseRecalibrator and ApplyBQSR in GATK v.4.1.3. The calibrated output was used to recall variants with GATK and ANGSD as above, and the intersect was used on a second round of recalibration. The results were similar to the previous run suggesting convergence had been

achieved as observed in other bird genomics studies [83]. Thus, the recalibrated calls obtained in the first round were kept for the remainder of the pipeline. Following BQSR, sequence variants were called with GATK's HaplotypeCaller and GenotypeGVCFs v.4.1.3 from the recalibrated bam files.

Genotype calls were filtered for downstream analyses using a hard-filtering approach as proposed for non-model organisms, using GATK and VCFtools[182]. Calls were removed if they presented: low individual quality per depth (QD < 5), extreme coverage (800 > DP > 1800), mapping quality (MQ < 40 and MQ > 70), extreme hetero or homozygosity (ExcessHet > 20 and InbreedingCoeff > 0.9) and high read strand bias (FS > 60 and SOR > 3). We filtered further at the level of individual genotype by removing calls for which up to 5% of genotypes had low quality (GQ < 20) and extreme coverage (GenDP < 10 and GenDP > 40). Lastly, we kept only bi-allelic sites with less than 5% of missing calls across individuals yielding a dataset of 6'721'999 SNP. For analyses of neutral population structure and demography, an exact Hardy-Weinberg test was used to remove sites that significantly departed ($p<0.05$) from the expected equilibrium using the R[183] package HardyWeinberg[184,185].

Post-glacial species distribution

We built species distribution models (SDM) using Maximum Entropy Modelling (MaxEnt), a presence-only based modelling tool, to identify the regions of high habitat suitability for barn owls at the last glacial maximum (LGM, 20'000 years BP). Current climatic variables for the Western Palearctic (Sup. Fig. 2) were extracted from the WorldClim database[164] at 5 arc min resolution using the R package rbioclim[186]. The chosen set of variables represents the climatic conditions experienced by the species through the year and were filtered to be correlated at less than 0.8. Retained variables were: mean diurnal range (bio2), minimum temperature of coldest month (bio6), temperature annual range (bio7), mean temperature of wettest quarter (bio8), precipitation seasonality (bio15), precipitation of driest quarter (bio17) and precipitation of coldest quarter (bio19).

To determine which combination of feature and regularization multiplier use in the model to optimise the prediction without over complexifying the model, we built models with linear, quadratic and hinge features, and models with a range (1 to 5) of regularization multipliers. The best combination of feature and regularization multiplier based on the corrected AIC (as recommended[165]) was achieved with a quadratic model with 1 as regularization multiplier (Sup. Table 5). We ran 100 independent maxent models, omitting 25% of the data during training to test the model. To avoid geographic bias due to different sampling effort in the distribution area

of the species, we randomly extracted 1000 presence points within the IUCN distribution map[187] for each model run[188].

Predictive performances of the models were evaluated on the basis of the area under the curve of the receiver operator plot (AUC) of the test data. For all models with an AUC higher than 0.8 (considered a good model [189,190]), the output of maxent was transformed into a binary map of suitability by assuming that a cell was suitable when its mean suitability value was higher than the mean value of the 10% test presence threshold. This conservative threshold allows to omit all regions with habitat suitability lower than the suitability values for the lowest 10% of occurrence records. We averaged the values of the models for each cell, and only cells suitable in 90% of the models are presented as such in the map.

All models were then projected to the mid-Holocene (6'000 years BP) and LGM (20'000 years BP) conditions extracted from WorldClim at the same resolution as current data. When projecting to past climates, the multivariate environmental similarity surface (MESS) approach[191] was used to assess whether models were projected into climatic conditions different from those found in the calibration data. Cells with climatic conditions outside the distribution used to build the model were considered as unsuitable for barn owls (0 attributed to cell with negative MESS) as we intended to highlight only the highly suitable regions. For each timepoint, the results of the models were merged and transformed into a binary map as for the current data.

Maximum-likelihood demographic inference with fastsimcoal2

Data preparation

To build a set of variants approaching neutrality, we kept only autosomal SNPs found outside of genic regions and CpG mutations[192]. The sites were filtered to include no missing data and to within two thirds of the standard deviation of the mean coverage to ensure homogeneity. To determine the ancestral state of the SNP using a parsimony approach, the genomes of the two outgroups were used as outgroups based on the *Tytonidae* phylogenetic tree[89]. Sites for which it was impossible to attribute a state based on the available outgroups were discarded (868 sites).

Demographic inference

For each run of each of the six tested demographic scenarios, the following options were set: -n 500000 (number of coalescent simulations), -M and -L 50 (estimate the parameters from the SFS with 50 expectation-maximization (EM) cycles to estimate parameters). As we do not currently have a good estimation of the barn owl mutation rate, the end of the glaciation

(rounded to 6000 generations ago, 18'000 years BP with a 3-year generation time) was fixed and all other parameters were scaled relative to it using the -0 option (based solely on polymorphic sites).

In the non-parametric bootstrapping of the best-fitting model, a block-bootstrap approach was employed as suggested by the authors to account for LD[75,166]. As such, the SNPs were divided into 100 blocks of similar size and then 100 bootstrap datasets were generated by sampling with replacement 100 blocks each, so as to obtain the same number of SNPs as the real dataset. For each bootstrapped SFS, 50 independent parameter inferences were run under the best-fitting scenario out of the six tested. Due to computational constraints, bootstrap runs were performed with only 10 EM cycles, an approach that has been previously used and described as conservative[193]. The highest maximum-likelihood run of each scenario was used to estimate 95% CI of all parameters.

Supporting Tables

Supplementary table 1 – Summary of the sampling scheme for the different analyses in this study. Full sample detail available in Appendix 1.

Population	Abbrev.	N Colour	N MC1R	N WGS	N FSC
Ireland	IR	44	44	12	8
Great Britain	GB	67	68	15	8
France	FR	48	48	5	5
Denmark	DK	88	88	10	3
Switzerland	CH	65	65	10	0
Portugal	PT	70	71	9	8
Total		382	384	61	32

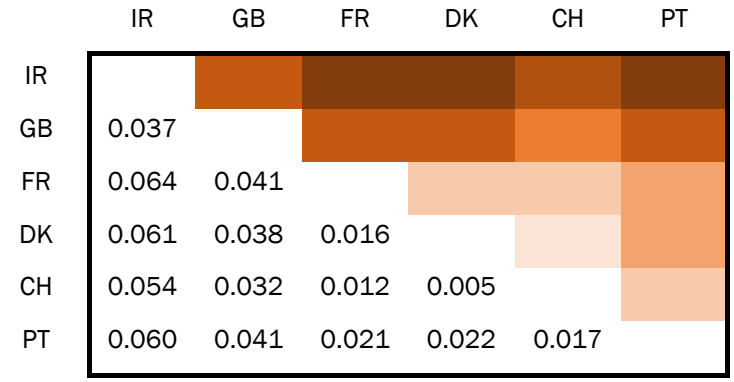
Number of samples for: N Colour – colour comparison analysis; N MC1R – MC1R genotyping; N WGS – whole genome re-sequencing and population genomics analyses; N FSC – demographic inference with *fastsimcoal2*.

Supplementary table 2 – Lineage genetic diversity for of 61 European barn owls. Individuals were grouped into the four main lineages identified in sNMF (Fig. 1 & Sup. Fig. 3).

Lineage	Abbrev.	N	Priv. Alleles	H_O (SD)	F_{IS} (SD)
Ireland	IR	9	84061	0.157 (0.004)	-0.023 (0.027)
Great Britain	GB	17	93144	0.158 (0.011)	0.029 (0.068)
Central Europe	EU	25	191666	0.174 (0.010)	-0.004 (0.057)
Portugal	PT	9	784099	0.188 (0.008)	-0.012 (0.041)

N – number of individuals; Priv. Alleles – private alleles accounting for different population sizes; H_O – observed heterozygosity and its standard deviation; F_{IS} – inbreeding coefficient.

Supplementary table 3 – Pairwise Weir & Cockerham's F_{ST} between barn owl populations in Western Europe. Above the diagonal, a heat map provides a visual representation of the F_{ST} values given below the diagonal.



Supplementary table 4 – Summary of barn owl capture-recapture events in Great Britain (GB) and Central Europe from 1910 to 2019, courtesy of EURING. Table shows number and percentages (%) of **(a)** emigrant and **(b)** immigrant owls to GB. Table **(c)** shows exchanges of individuals between countries in Central Europe; number of owls indicates migrants from the country on the row towards the country in column; background heatmap represents the values on top; two last columns give the number of ringed birds per country and how many of those emigrated elsewhere. BL= Belgium; CH=Switzerland; DE=Germany; DK=Denmark; FR=France; NL=Netherlands.

(a)

Country	Total immigrants to this country	Immigrants from GB	% immigrants from GB
Belgium	338	1	0.30
France	1506	1	0.07
Germany	1364	1	0.07
Netherlands	1107	1	0.09
Spain	39	3	7.69
Ireland	4	4	100
Northern Ireland	4	3	75

(b)

Country	Total ringed owls	Total emigrant owls	% emigrant owls	Emigrant owls to GB	% emigrant owls to GB
Belgium	6166	1070	0.28	3	0.049
Germany	20816	1493	0.35	5	0.024
Netherlands	25849	1177	1.19	13	0.050

(c)

→	BL	CH	DE	DK	FR	NL	N ringed	Emigrants
BL		3	80	3	342	616	7786	1118
CH	3		502	0	518	6	7862	1074
DE	61	66		136	486	424	24245	1522
DK	1	0	44		0	4	1003	57
FR	9	6	36	0		8	1318	67
NL	251	6	652	10	104		29487	1177

Supplementary table 5– Comparison of SDM model fit. AICc is reported for the multiple combinations of feature (linear, quadratic, hinge) and Beta multiplier (1 to 5).

	1	2	3	4	5
Linear	23953.4	23994.8	24037.1	24086.8	24125.8
Quadratic	23950.8	24002.6	24042.7	24095.2	24148.5
Hinge	24099.1	24185.4	24241.2	24304.4	24340.2

Supplementary table 6 – Likelihood and AIC of the demographic models tested with *fastsimcoal2*. Three main model topologies were tested, each with two versions (Figure 3). Models are sorted from best to worst according to the estimated likelihoods.

Model Type	Model	Est. Lhood	Δ Lhood	AIC	Δ AIC
Colour-based	B	-6410181	4342	29520023	0
Refugia	A	-6410377	4537	29520926	904
Colour-based	A	-6410439	4599	29521216	1193
Refugia	B	-6410674	4834	29522291	2269
Stepping-Stone	B	-6410811	4971	29522935	2912
Stepping-Stone	A	-6410865	5025	29523172	3150

Est. Lhood – Maximum-likelihood estimated for the simulated SFS per demographic model; Δ Lhood – difference between the likelihood of the simulated and observed SFS; Δ AIC – delta AIC

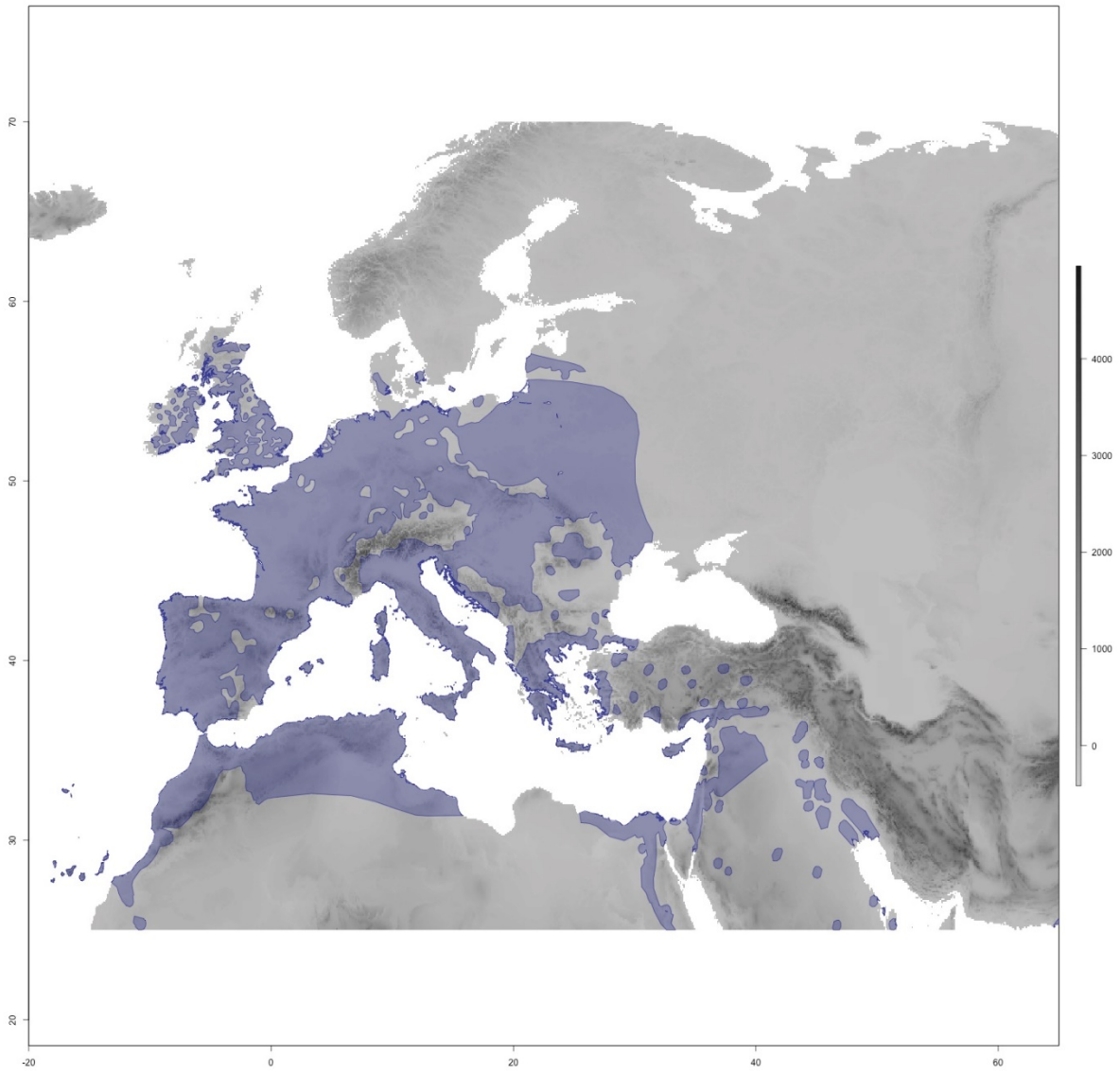
Supplementary table 7 – Parameter ranges and point estimate inferred for the demographic model tested with *fastsimcoal2*. Model "Iberian origin B" – identified as the best fitting model – is given in Sup. Table 8. All range distributions were uniform except for bottlenecks (†) which were log-uniform. When a parameter was absent in a model, the case was left blank.

		NW European		Iberian	Insular Refugium	
Parameter	Ranges	A	B	A	A	B
Current Population Sizes (haploid)						
PT	10000 – 4e5	58059	14700	10518	35058	61402
EU	1000 - 3.5e5	1018	4097	1011	1007	5314
GB	100 - 25000	1008	139	137	200	1712
IR	10 - 2500	421	221	232	179	390
Ancestral Population Sizes (haploid)						
PT in glac	200 – 2e5	39179		6637		
Ancestral island	200 – 2e5			11277		
PT before glac	1000 - 1e6		16066		8603	13846
GB before glac	1000 - 50000				41926	23204
EU before glac	1000 – 3.5e5		207200			
Times of Divergence (generations)						
T1		4875	12758	3656	11705	30398
T2		3635	256	3055	2843	2469
T3		304	204	288	1219	3318
TSG	8000 - 9000				8452	8148
TEG	3000 - 6000				3409	4482
Current Migration (flow is backwards in time)						
EU → PT	0 - 0.05	0.004	0.0001	0.005	0.007	0.001
PT → EU	0 - 0.05	0.011	0.005	0.011	0.004	0.001
GB → EU	0 - 0.05	0.004	0.058	0.054	0.039	0.005
EU → GB	0 - 0.05	0.011	0.002	0.013	0.005	0.0007
IR → GB	0 - 0.05	0.014	0.032	0.027	0.042	0.018
GB → IR	0 - 0.05	0.007	0.042	0.042	0.025	0.0001
Older Migration (2nd level from present in Sup. Fig. 1)						
EU → PT	0 - 0.05	0.002	0.018	0.0004		
PT → EU	0 - 0.05	0.035	0.014	0.001		
GB → EU	0 - 0.05	0.006	0.032	0.002		
EU → GB	0 - 0.05	0.021	0.049	0.030		
Ancestral Migration (oldest)						
EU → PT	0 - 0.05	0.002	0.004	0.008		
PT → EU	0 - 0.05	0.00002	0.022	0.011		
Instbot- bottleneck intensity at diverge †						
T1	0.01 - 0.5	0.052	0.006	0.006	0.009	0.003
T2	0.01 - 0.5	0.004	0.139	0.696	0.037	0.019
T3	0.01 - 0.5	0.010	0.016		0.044	0.110
Glacation Bottleneck size (haploid) †						
PT	0.01 - 0.5		295		2747	23223
EU	0.01 - 0.5		490			
GB	0.01 - 0.5				8	39

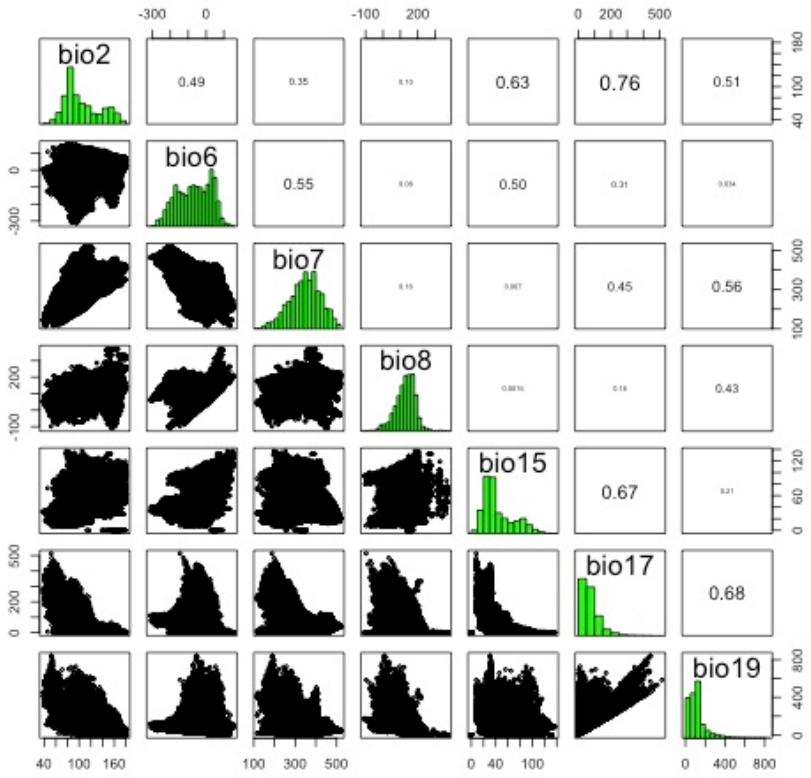
Supplementary table 8 – Parameter point estimates and 95% confidence interval for the best demographic model – Iberian origin B. Parameter names correspond to Figure 3. Times of divergence are in years calculated with a generation time of 3 years. Migration rates and number of individuals are given forward in time.

Parameter	Point Estimate	Lower Limit CI	Upper Limit CI
Current Population Sizes (haploid)			
PT	10541	10501	364413
EU	1052	1005	2382
GB	440	131	1838
IR	199	138	1071
Ancestral Population Sizes (haploid)			
Iberia glacial period	7986	5451	137520
Ancestral pop to GB & IR	6187	2112	20742
Times of Divergence (years)			
Split GB - PT	12940	7316	17812
Split EU - PT	3822	945	4984
Split IR - GB	1149	224	2229
Current Migration Rate (2Nm)			
PT → EU	0.005 (52)	0.0009 (10)	0.0089 (93)
EU → PT	0.0074 (8)	0.0038 (4)	0.0158 (17)
EU → GB	0.0152 (16)	0.0003 (0.3)	0.0548 (58)
GB → EU	0.0093 (4)	0.0051 (2)	0.0206 (9)
GB → IR	0.0361 (16)	0.0019 (1)	0.0473 (21)
IR → GB	0.0103 (2)	0.0011 (0.2)	0.0438 (9)
Ancestral Migration Rate (2Nm)			
PT → EU	0.0029 (262)	0.0002 (2)	0.0524 (552)
EU → PT	0.0248 (3)	0.0031 (3)	0.0585 (62)
EU → GB	0.0001 (0.1)	0.0004 (0.4)	0.0159 (17)
GB → EU	0.03 (186)	0.0121 (75)	0.0633 (392)
Instant Bottlenecks (N)			
Split GB - PT	0.015 (65)	0.288 (4)	0.01 (98)
Split EU - PT	0.018 (56)	0.032 (31)	0.002 (535)

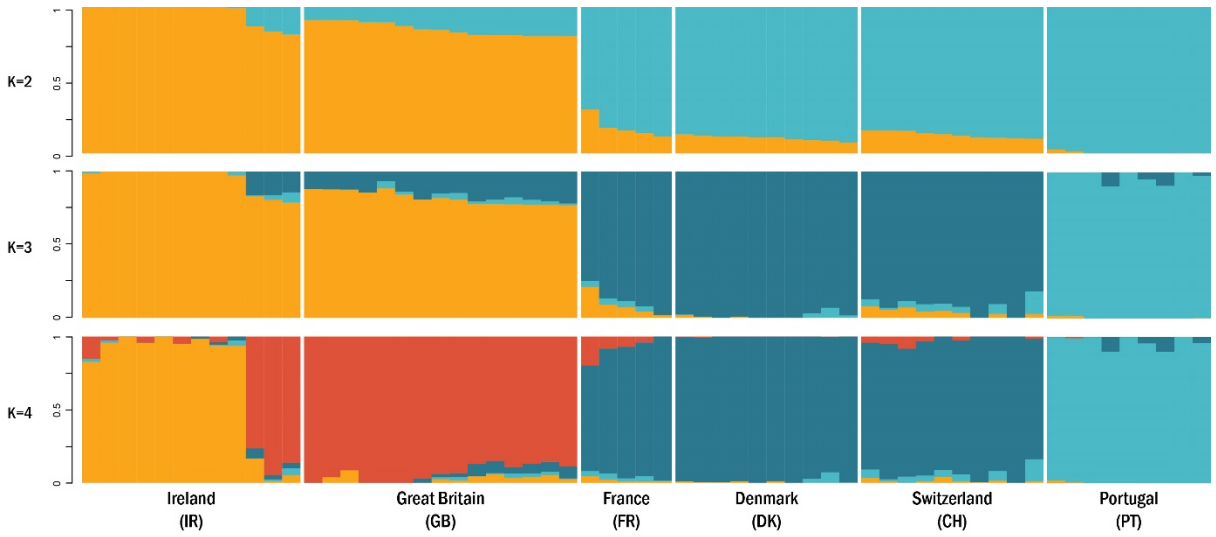
Supporting Figures



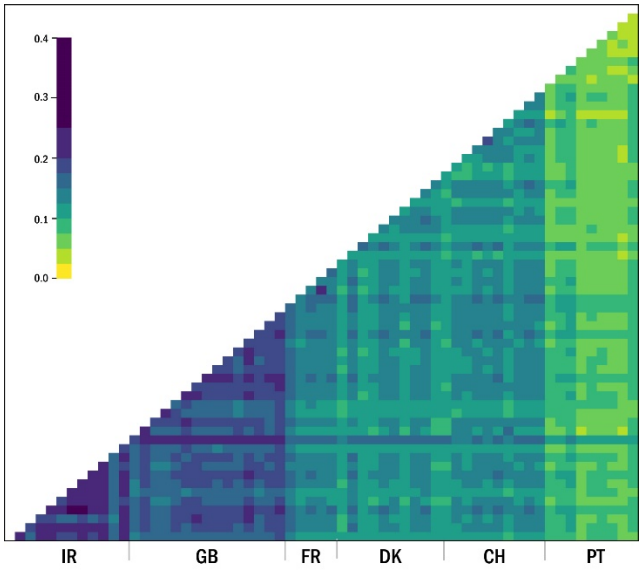
Supplementary Figure 1 – The map in grey represents the area considered for producing the Species Distribution Model (SDM) for the barn owl; shading denotes altitude according to the scale. The current distribution of barn owls is plotted atop the map in purple (data from IUCN: BirdLife International 2019). Random presence points were extracted within this distribution for the SDM.



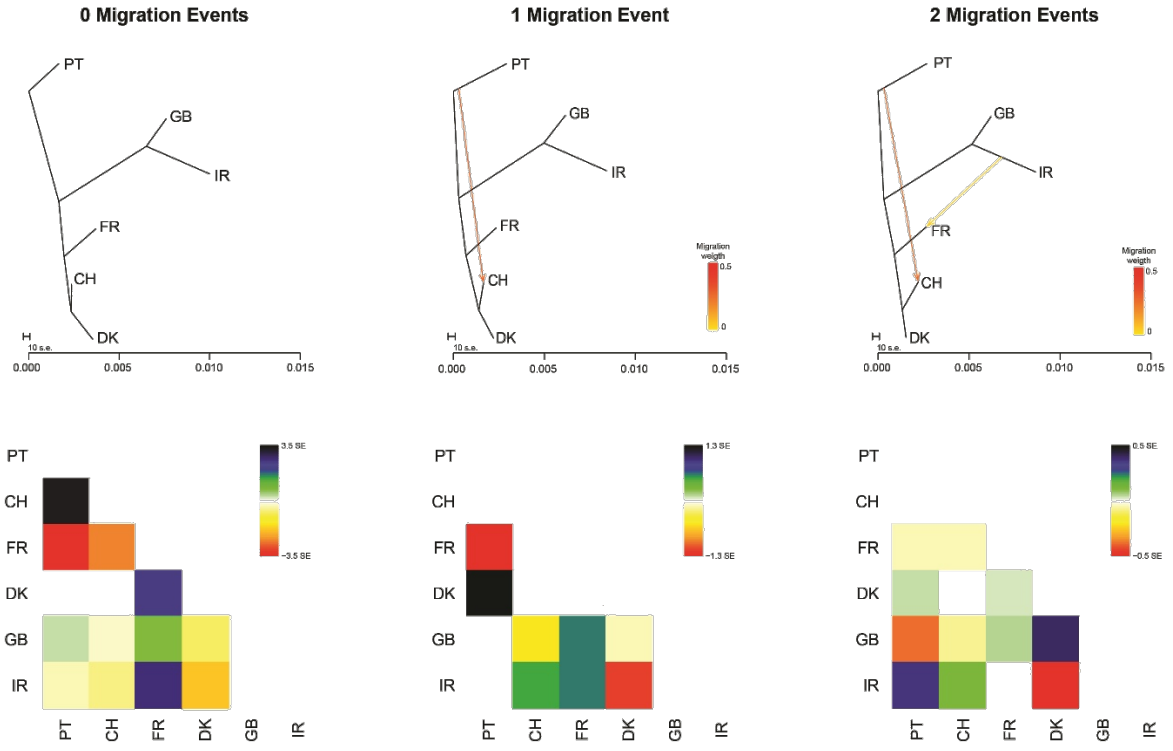
Supplementary Figure 2 – Pairwise correlation between retained climatic variables to produce the SDM. Only variables correlated at less than 0.8 were kept in the models, namely: Mean Diurnal Range (Bio2), Min Temperature of Coldest Month (Bio6), Temperature Annual Range (Bio7), Mean Temperature of Wettest Quarter (Bio8), Precipitation Seasonality (Bio15), Precipitation of Driest Quarter (Bio17) and Precipitation of Coldest Quarter (Bio19).



Supplementary Figure 3 – Individual clustering estimated by sNMF for K 2 to 4 lineages. Each vertical bar represents one individual, and the colours represent the relative contributions of each genetic lineage K.



Supplementary Figure 4 – Individual relatedness (β) matrix. Grey lines separate the populations. Population abbreviations follow Sup. Fig. 3.



Supplementary Figure 5 – Treemix results for 0 to 2 migration events. Highest likelihood runs are depicted, with the corresponding matrix of standard errors. With two migration events there was no topology convergence between replicates; the topology of the best run shown here was only present in 4 out of 10 runs.

Chapter 2

Genomic analyses reveal differential connectivity of barn owl insular populations in the eastern Mediterranean

Ana Paula Machado[a†], Alexandros Topaloudis[a†], Tristan Cumer[a], Eléonore Lavanchy[a], Vasileios Bontzorlos[k,l], Renato Ceccherelli[m], Motti Charter[n,o], Nikos Kassinis[p], Petros Lymberakis[q], Francesca Manzia[r], Anne-Lyse Ducrest[a], Mélanie Dupasquier[c], Nicolas Guex[b], Alexandre Roulin[a‡], Jérôme Goudet[a,j‡]

† first co-authors, sorted alphabetically

‡ co-senior authors

Status

Ready for submission, awaiting the submission of chapter 4 due to cross-referencing.

Author contributions

APM, AT, AR, JG designed this study; APM produced whole-genome resequencing libraries; AT, APM conducted the analyses with input from TC, EL; VB, MC, NK, PL provided samples; APM, AT led the writing of the manuscript with input from all authors.

Abstract

The study of insular populations was key in the development of evolutionary theory. The successful colonisation of an island depends on the geographic context, and specific characteristics of the organism and the island, but also on stochastic processes. As a result, apparently identical islands may harbour populations with contrasting histories. Here, we use whole genome sequences of 65 barn owls to investigate the patterns of inbreeding and genetic diversity of insular populations in the eastern Mediterranean Sea. We focus on Crete and Cyprus, islands with similar size, climate and distance to mainland, that provide natural replicates for a comparative analysis of the impacts of microevolutionary processes on isolated populations. We show that barn owl populations from each island have a separate origin, Crete being genetically more similar to other Greek islands and mainland Greece, and Cyprus more similar to the Levant. Further, our data show that their respective demographic histories following colonisation were also distinct. On the one hand, Crete harbours a small population and maintains very low levels of gene flow with neighbouring populations. This has resulted in low genetic diversity, strong genetic drift, increased relatedness in the population and remote inbreeding. Cyprus, on the other hand, appears to maintain enough gene flow with the mainland to avoid such an outcome. Our work provides a comparative population genomic analysis of the effects of neutral processes on a classical island-mainland model system. It provides empirical evidence for the role of stochastic processes in determining the fate of diverging isolated populations.

Keywords

Demographic inference; Inbreeding; Population genomics; *Tyto alba*; Whole genome sequencing

Introduction

Given their discrete borders, geographical isolation and abundance, islands are ideal systems to study patterns of genetic diversity in natural populations[34]. Due to the combination of biotic, abiotic, and stochastic forces, no two insular populations share the same demographic history[33]. Their fate is shaped by the timing of colonisation, fluctuations in population size and connectivity to neighbouring populations. These are directly impacted by the characteristics of the island, like carrying capacity and distance to the mainland, as well as the circumstances of colonisation such as bottlenecks and founder effects. The combined actions of reduced gene flow, *in situ* genetic drift, selection and potentially mutation influence the degree to which insular populations diverge[31,35,36]. Small populations are particularly sensitive to the effect of genetic drift, accelerating divergence from the surrounding populations. While high levels of gene flow can counter this effect, the lack of it can facilitate local adaptation by maintaining locally advantageous alleles [194] but can also lead to inbreeding with detrimental consequences[38].

In small isolated populations, without other sources for genetic diversity besides mutation and recombination, the relatedness among insular individuals increases over time under the effect of drift. As a result, levels of remote inbreeding may rise even with the avoidance of mating between close relatives. Although this is a common occurrence in island populations, mating between related individuals can lead to inbreeding depression[195] and, in extreme circumstances, local extinction[38,196]. As such, the study of the genetic makeup of insular populations can provide key information from a conservation perspective. Despite being widely used to estimate inbreeding and infer demographic histories, traditional genetic markers lack resolution to reconstruct particularly convoluted systems such as, for example, multiple islands or among modestly differentiated populations. Technological advances now provide more affordable high-representation genomic data such as the sequencing of whole genomes. Combined with increasingly sophisticated methods, it allows for more accurate inferences, even for non-model species[68].

The eastern Mediterranean offers an excellent setting to study insular demographic history. A biodiversity hotspot [197], the area is riddled with islands, the largest of which are Crete (CT) and Cyprus (CY). While fluctuating sea levels intermittently connected smaller islands to the mainland in the Quaternary, CT and CY have been isolated since the end of the Messinian salinity crisis (approx. 5 Mya[198]). They share many common features such as distance to mainland (95 and 75 km, respectively), surface area (8500 and 9200 km^2) and a Mediterranean-subtropical climate with mild winters and warm summers. Their strategic position makes them pivotal stop-overs in the seasonal migration of many bird species, and movements of bird populations are widely studied (e.g. [199,200]). However, thus far they have been the subject of only few genetic studies,

most on each island individually rather than comparatively, and typically focusing on human commensal small mammal species[201–203].

The Afro-European barn owl (*Tyto alba*) is a non-migratory bird of prey present across the African and European continents, as well as most of the surrounding islands and archipelagos[89]. In spite of being quite widespread and maintaining high gene flow overland[101,171], populations separated by water barriers appear to accumulate differentiation more quickly, with numerous insular subspecies[89,93,102]. In the eastern Mediterranean, the continental European barn owl lineage meets the eastern subspecies *T. a. erlangeri* (W. L. Sclater, 1921) from the Levant[102,204]. Although Crete and Cyprus populations supposedly belong to *T. a. erlangeri*[205], the low resolution genetic data previously available was insufficient to clarify the history of each island and how they relate to the mainland. Barn owls from Crete appeared to be quite distinct from all surrounding mainland, including the Levant[102], and the demographic history of the Cyprus owl population has never been studied.

Here, we investigate the genetic structure and past demographic history of insular and mainland barn owl populations in the eastern Mediterranean. We focus in particular on Crete and Cyprus, the two largest islands in the region, that have very similar intrinsic characteristics and are thought to harbour barn owls from the eastern subspecies found in the Levant (*T. a. erlangeri*). As such, the populations should have originated, independently or not, from the Levant. However, being closer to other Greek islands and the Greek mainland, Crete could have actually been colonised from there, which would be incompatible with it belonging to the same subspecies as Cyprus. Taking advantage of the whole genome sequences of 65 individuals and the recent publication of a high-quality reference genome[93], we address this by modelling the colonisation of both islands from the mainland. Lastly, we compare how their different demographic histories impacted their current genetic diversity and inbreeding levels.

Materials and Methods

Sampling, Sequencing and Genotyping

A total of 67 barn owl individuals from seven populations were used in this study (Table 1; Supporting Table 1): 10 in Italy (IT), 5 in islands of the Ionian Sea (IO), 10 in Greece (GR), 11 in islands of the Aegean Sea (AE), 11 in Crete (CT), 10 in Cyprus (CY) and 10 in Israel (IS). Of these, 47 were sequenced in Cumer *et al.*[204] (GenBank BioProject PRJNA727977; Sup. Table 1). One additional individual of the Eastern barn-owl species (*T. javanica* from Singapore[89]) was used as an outgroup for specific analyses. The outgroup was sequenced in Machado *et al.*[93] (GenBank BioProject PRJNA700797). The remaining 20 samples followed the same protocol described in [93,204]. In brief, we extracted genomic DNA using the DNeasy Blood & Tissue kit (Qiagen, Hilden,

Germany) and prepared individually tagged 100bp "TruSeq DNA PCR-free" libraries (Illumina) following the manufacturer's instructions. Then, whole-genome resequencing was performed on multiplexed libraries with Illumina HiSeq 2500 high-throughput paired-end sequencing technologies at the Lausanne Genomic Technologies Facility (GTF, University of Lausanne, Switzerland) with an expected sequence coverage of at least 15X.

The bioinformatics pipeline used to obtain analysis-ready SNPs from the raw sequenced of the 65 individuals plus the outgroup was the same as in Machado *et al.*[93] adapted from the Genome Analysis Toolkit (GATK) Best Practices [154] to a non-model organism following the developers' recommendations. Briefly, we trimmed the reads to 70bp length with Trimommatic v.0.36[179] and aligned them with BWA-MEM v.0.7.15[180] to the barn owl reference genome (GenBank accession JAEUGV000000000 [93]). Then, we performed base quality score recalibration (BQSR) following the iterative approach recommended for non-model species that lack a set of "true variants" in GATK v.4.1.3 using high-confidence calls obtained from two independent callers: GATK's HaplotypeCaller and GenotypeGVCF v.4.1.3 and ANGSD v.0.921[181]. Following BQSR, we called variants with GATK's HaplotypeCaller and GenotypeGVCFs v.4.1.3 from the recalibrated bam files.

For variant filtering we followed GATK hard filtering suggestions for non-model organisms, with values adapted to our dataset and expected coverage using GATK v4.1.3.0 and VCFtools v0.1.15[182]. A detailed documentation of the filters applied can be found in Sup. Table 2. We also removed scaffolds that belong to the Z chromosome due to it being hemizygous in females (Sup. Table 1). In preliminary analyses we corrected the origin of a sample, an injured owl found at sea and reported to a port in mainland Greece but that was genetically of Cretan origin and considered as such hereafter. We also removed one Italian (IT10) and one Israeli (IS10) individuals as relatedness analyses revealed they were each part of a sibling pair. The final dataset contained 5'493'583 biallelic SNPs with a mean coverage of 16.4X (4.38 SD) across 65 individuals (Sup. Table 1).

Mitochondrial DNA

Sequencing and assembly of mitochondrial genome

We produced a complete mitochondrial reference genome for the barn owl, from the same individual used for the reference nuclear genome recently published[93]. The mitochondrial genome was thus produced from the high molecular weight (HMW) DNA extraction described in detail in Machado *et al.*[93]. Briefly, HMW DNA was extracted from a fresh blood sample using the agarose plug method as described in Zhang *et al.* [206]. Then, 15-20 kb DNA fragments were

obtained with Megaruptor (Diagenode, Denville, NJ, USA) and checked on a Fragment Analyzer (Advanced Analytical Technologies, Ames, IA, USA). 5 µg of the sheared DNA was used to prepare a SMRTbell library with the PacBio SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations. The resulting library was size-selected on a BluePippin system (Sage Science, Inc. Beverly, MA, USA) for molecules larger than 13 kb. It was then sequenced on 1 SMRT cell 8M with v2.0/v2.0 chemistry on a PacBio Sequel II instrument (Pacific Biosciences, Menlo Park, CA, USA) at 30 hours movie length to produce HIFI reads.

After sequencing, we searched the circular consensus sequences (ccs) HIFI reads for sequences matching the 18128 bp mitochondrial genome of the previous assembly (NCBI Reference Sequence: NW_022670451.1[92] using minimap2[207] with the option -x asm5. We obtained twelve reads, which were reverse complemented as needed in order to be in the same orientation as our seed mitochondrial genome. No read was long enough to obtain a closed circular mitochondrial genome. Thus, we selected a css read of particularly high quality as an anchor and used two other overlapping reads to complete the circular sequence. From these three high quality reads, we manually assembled a full-length mitochondrial genome of 22461 bp. Mitochondrial css are provided in supplementary material and the reference sequence has been deposited at GenBank (currently awaiting the accession number).

We annotated the mitochondrial genome using MitoAnnotator v3.52[208] and removed the hyper-variable D-loop for the subsequent analyses, yielding a 15'571bp sequence.

Mitochondrial population structure and genetic diversity

To obtain the mitochondrial sequences of each individual, we mapped their trimmed whole-genome resequencing reads onto the newly assembled barn owl mitochondrial genome using the BWA-MEM v.0.7.15 algorithm[180]. We then called variants using the bcftools v1.8[182] mpileup (with mapping quality > 60, depth < 5000) and call (consensus calling, -c) for haploid data (ploidy=1). We then created a consensus fasta sequence with bcftools consensus, applying variants called above on the reference genome. We aligned individual fasta sequences using ClustalOmega v1.2.4[209] and manually checked the alignment for errors in MEGA X v10.1.7[210]. We generated a mitochondrial haplotype network using the R package pegas v0.14[211] and grouped similar haplotypes into haplogroups (Sup. Fig. 1). Finally, we quantified population diversity (nucleotide diversity, π) and divergence (Φ_{ST}) with Arlequin v3.5.2.2[212].

Population structure, diversity and inbreeding

To elucidate population structure in our dataset, we performed a principal component analysis (PCA) using the R-package SNPRelate v3.11[157] and inferred individual admixture proportions with the software sNMF v1.2[155]. sNMF was run for values of K ranging from 2 to 9, with 10 replicates for each K. Runs were checked visually for convergence within each K. For both analyses, we used a dataset of 603'496 biallelic SNPs obtained by pruning our SNP dataset for linkage disequilibrium (LD) using PLINK v1.9[156] (--indep-pairwise 50 10 0.1) as recommended by the authors. To investigate whether an island population was the product of admixture between two sampled populations, we used the f3 statistic[213] and TreeMix[158] both calculated with the TreeMix v1.13 software. TreeMix was run in 20 replicates, using a bootstrap per 500 SNP interval, with 0 to 3 migration events, using the same LD-pruned dataset as above, to which any sites with missing data were removed yielding a total of 598'599 SNPs.

We used SNPRelate to calculate an allele sharing matrix between individuals (β[159]) individual inbreeding coefficients relative to the total and then averaged per population (F_{IT}). We used the R package hierfstat v.0.5-9[214] to estimate population pairwise and population-specific F_{ST} as in Weir & Goudet[159]. Confidence intervals were obtained by bootstrapping 100 times 100 blocks of contiguous SNPs. We also used hierfstat to quantify individual inbreeding coefficients relative to their population of origin and then averaged per population (F_{IS}). For population genetic diversity, we calculated the observed individual observed heterozygosity and estimated the number of private alleles (i.e. alleles present in only one population) using custom made R scripts. To account for sample size differences in the estimation of private alleles, we subsampled 5 individuals (without replacement) from each population 100 times and calculated the mean number of private alleles in a population. When calculating the lineage-specific private alleles for K=5 from sNMF, we merged the populations of Greece, Ionian and Aegean islands and followed the same approach, this time sampling 9 individuals instead of 5 (corresponding to the new lowest sample size).

The Estimated Effective Migration Surface (EEMS) v.0.9 software[161] was used to visualize relative gene flow over the sampled region. First, we used the tool bed2diff to compute the matrix of genetic dissimilarities for the LD-pruned dataset mentioned above and utilized the Google Maps API v.3 tool (http://www.birdtheme.org/useful/v3tool.html) to draw a polygon outlining the study area. Then, EEMS was run with 700 demes in 3 independent chains of 2 million MCMC iterations with a 1 million iterations burn-in. We tested convergence of the results through a plot of observed-fitted values and the trace plot of the MCMC chain as suggested by the authors and plotted the results using the accompanying R package (rEEMSplots v.0.0.1).

We inferred runs of homozygosity (ROH) in the dataset by using the plink command --homozyg with default parameters (minimum 1 Mb length and 50 SNP). Only autosomal scaffolds of length

more than 1 Mb were considered in ROH inference (47/70 scaffolds) covering 92% of the total assembly length. Given that bird chromosomes are typically shorter than those of humans[70], for whom such methods were developed, we also called ROH with a minimum of 100Kb length. As the qualitative results were unchanged (data not shown), we kept the standard 1Mb threshold in a conservative approach to identify only identity by descent (IBD) segments and to facilitate potential comparisons with other studies. To estimate the index F_{ROH} we divided the sum of lengths of ROH in an individual with the length of the scaffolds[215] used after subtracting the number of 'N's (gaps) in the assembly. To visualize the distribution of ROH lengths per population, we divided ROH into five length classes: i) from 1Mb to under 2Mb, ii) from 2Mb to under 4Mb, iii) from 4Mb to under 6Mb, iv) from 6Mb to under 8Mb and finally, v) 8Mb or longer. We then calculated the number of base pairs falling within each ROH length class for every individual and averaged the values for each population.

To compare the levels of inbreeding, we tested whether F_{IT}, F_{ROH} and β differ significantly between populations using a non-parametric Kruskal-Wallis rank sum test since the normality assumption did not hold. Further, we performed a pairwise Wilcoxon rank sum exact test with a Bonferroni correction for multiple testing to assess significance in the differences between pairs of populations. Given the small sample sizes (Table 1), we excluded obvious hybrid individuals (AE01, CT06) to avoid biasing the average of their respective populations.

Demographic history

Demographic scenarios and parameters

To infer the origin and connectivity of the major insular barn owl populations (CT and CY), we used the software *fastsimcoal2*[75]. It uses coalescence simulations to estimate the composite likelihood of simulated demographic models under the observed site frequency spectrum (SFS). To model both island systems together, we would need to simulate the coalescence of the European and Levant lineages (sNMF K=2, Sup. Fig. 2) for which we have no time calibrating event and could be hundreds of thousands of generations in the past. Such inference would likely be unreliable as well as extremely consuming computationally. Thus, we inferred the demographic history of each island system separately, including their closest populations. For each island system, 'Crete' and 'Cyprus', we tested three demographic scenarios (Figure 2b).

To infer the history of 'Crete', we did not include IS in the simulated scenarios as population structure analyses show that CT's origin is not in the Levant, but rather from the European lineage (Figure 1). As such, we only considered the populations of AE and GR. The first two scenarios assume that both the Aegean islands and the island of Crete were colonized

independently from the Greek mainland population. In the first one, the colonization of Crete takes place after the colonization of the Aegean islands, while in the second scenario Crete is colonized first. The third demographic scenario assumes the islands are colonized in a stepping-stone fashion, with owls from mainland Greece reaching the Aegean islands first and from there colonizing Crete (Figure 2b). Due to the low sea levels at Last Glacial Maximum (LGM), the Aegean islands were part of a larger emerged land mass that allowed nearly continuous overland connectivity to the mainland[216]. As such, for every demographic scenario in 'Crete' we assumed that the colonization of the Aegean islands from Greece occurred at the LGM (rounded to 18'000 years BP, 6'000 generations with a 3-year generation time). While the exact date is an approximation, allowing for migration between all populations after they split should reduce potential biases.

For 'Cyprus', in addition to IS as a representative of the Levant origin, a ghost population was incorporated in an attempt to represent the unsampled Turkish coast north of Cyprus, where the distance from the island to the mainland is the shortest. Including this ghost population in the model served two purposes. First, to account for unsampled sources of migrants into CY. Second, to avoid inflating artificially the effective population size of the CY population to justify the non-negligible admixture signal from AE (Figure 1a) that the simulator might interpret as *in situ* mutations. In the first two scenarios, both the Ghost and Cyprus populations originate from Israel, with the difference being the order in which they are colonized (same topology as Figure 2b). For the third scenario, owls from Israel would give origin to the Ghost population first and from there reach Cyprus.

Data preparation

Population sizes were reduced to the number of the smallest population in each model, resulting in 10 individuals per population for 'Crete' and 9 for 'Cyprus' (Sup. Table 1). To calculate the observed SFS for both systems, we filtered the data to a homogenous set of neutral markers. Specifically, we only kept sites with no missing data and with a depth of coverage less than 2/3 standard deviation from the mean. We also excluded CpG mutations[192] and SNPs in genic regions. We inferred the ancestral state of the SNPs using the barn owl from Singapore, an outgroup to all our populations[89]. Where the outgroup was homozygous for an allele, we marked that allele as the ancestral under rules of parsimony, while any other sites were removed. Population pairwise SFS were produced from the filtered datasets, giving 479'244 and 477'987 SNPs for 'Crete' and 'Cyprus', respectively.

Demographic inference with fastsimcoal2

For each system and each scenario, we specified a range of parameters from which the software drew an initial number as input in the optimization cycle (Sup. Table 4, 6). We modelled population splits with an instantaneous bottleneck in which the founding population size is a fraction of the present size.

For each scenario and each island, we performed 100 software runs. For each run we set the number of coalescent simulations to 500'000 and estimated the parameters through 50 expectation-maximization (EM) cycles. As we do not currently have a good estimation of the barn owl mutation rate, the end of the glaciation (rounded to 6000 generations ago) was fixed and all other parameters were scaled relative to it using the -0 option (based solely on polymorphic sites).

The best-fitting scenario was determined using Akaike's information criterion (AIC[168]). For the best scenario of each system, we performed non-parametric bootstrapping to estimate the 95% confidence intervals of the inferred parameters. Specifically, we divided the SNP dataset in 100 blocks with an equal number of SNPs, from which we created 100 bootstrapped-SFS and performed 50 independent runs of the software for each, with 250'000 simulations. Due to computational constraints we reduced the number of EM cycles to 10, an approach used previously and characterized as conservative[193]. The highest likelihood run for each bootstrapped replicate was used to calculate the 95% CI of the inferred parameters.

Ancient population size inference

For inference of past effective population sizes, we used the Pairwise Sequential Markovian Coalescent (PSMC[77]). Specifically, we intended to estimate sizes in the distant past as this method is inaccurate for recent events. We ran the software on every individual of every population and calculated the median size for a population for each time interval. PSMC was executed with the same parameters as in Nadachowska-Brzyska *et al.*[217] (-N30 -t5 -r5 -p 4+30*2+4+6+10). For plotting we used a mutation rate of $8.28*10^{-9}$ mutations per site per generation as estimated for avian species by Smeds *et al.*[218] and a generation time of 3.6 years[145].

Results

Population structure and divergence in the eastern Mediterranean

Mitochondrial DNA exhibited an overall Φ_{ST} of 0.13 (AMOVA) across all sampled individuals and a range of nucleotide diversity (0.0013 – 0.0023; Table 1). The mitochondrial DNA analyses failed to show consistent population structure in the dataset. The first two haplogroups constructed from the haplotype network (Sup. Fig. 1) were present in all populations, while haplogroup 3 which was missing from Israel despite being predominant in nearby Cyprus, and haplogroup 4 which was found only on the mainland populations (Figure 1a). Cretan owls had the lowest haplogroup diversity with mostly haplogroup 3 present and the lowest nucleotide diversity (0.0013).

Principal component analysis based on whole nuclear genome SNP separated the populations approximately from West to East along the first axis with individuals for each population clustering together (Fig. 1b), similar to K=2 in sNMF (Sup. Fig. 2). The second axis separated the two islands (CT & CY) from the rest of the populations, with admixed individuals dispersing between sources of admixture. Admixture analyses with sNMF were consistent between runs up to K=5 (Sup. Fig. 2, Fig. 1a). For K=3, Crete separates from the European lineage and for K=4 CY separates from the Levant lineage (Sup. Fig. 2). For K=5 (Fig. 1a), Italy, Crete, Cyprus, and Israel formed separate clusters while owls from the Ionian islands, mainland Greece and the Aegean were grouped into a single population. Owls from the Aegean islands showed the highest proportion of admixture (mean=0.2, SD=0.1) with components from Crete, Cyprus, and Israel in addition to their majority Greek component (Fig. 1a). Some individuals from Crete and Cyprus appeared admixed between their respective island's and the Greek component (blue in Fig. 1a).

Tests for population admixture with *f3* yielded a single slightly but significantly negative value (f3=-0.00065, SE=6e-05, Z=-10.375), which showed the Greek population to be the product of admixture between the Aegean and the Ionian populations. None of the insular populations appeared to be the product of admixture between any population sampled in this study. The topology created by Treemix was rooted at IS, with CY splitting first. CT displayed the longest branch of genetic drift and split before AE and the rest of the European populations (Fig. 2a). The first migration event was from AE to GR (Sup. Fig. 3), and it was the only one consistent across runs.

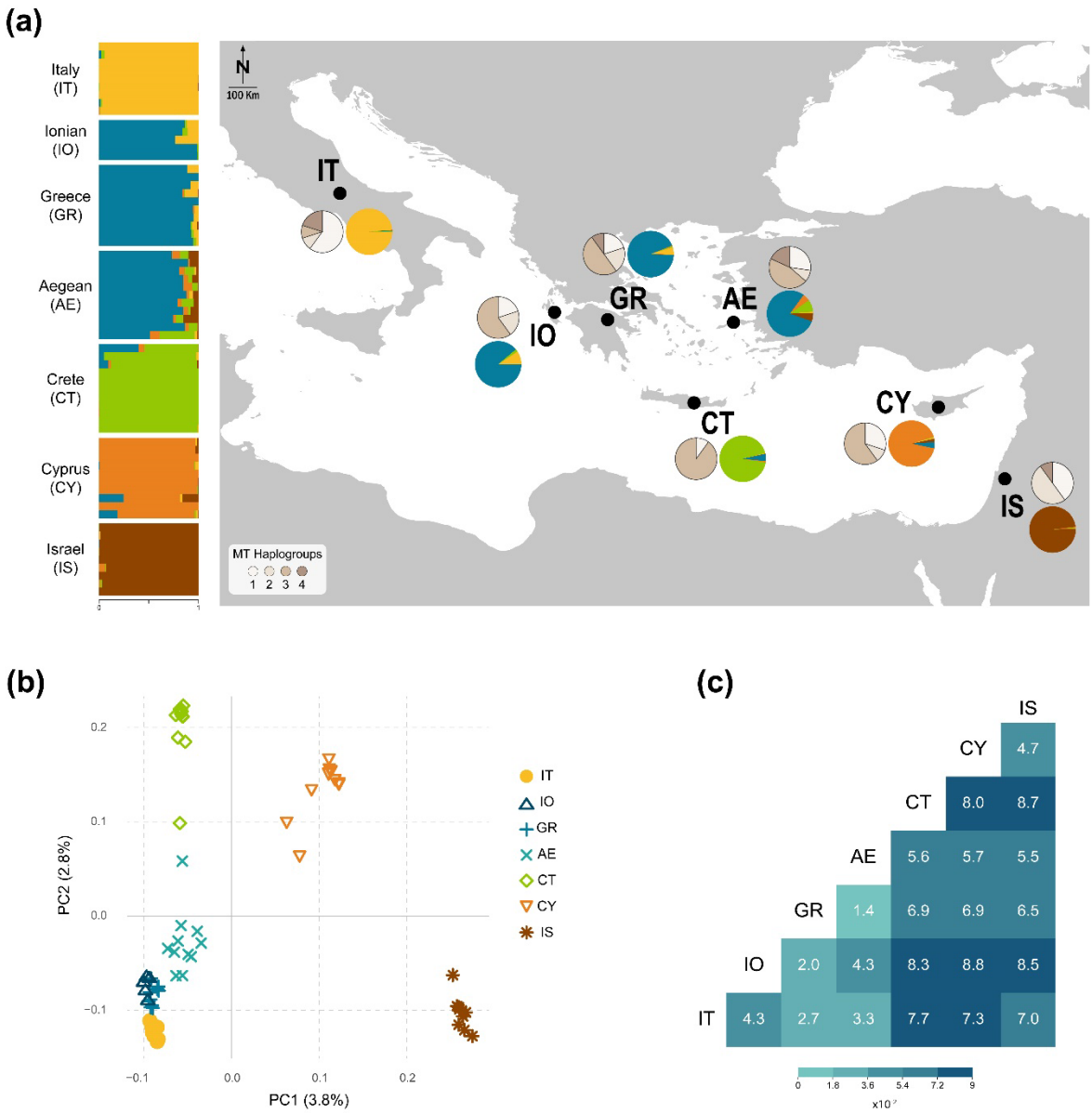


Figure 1 – Population structure of barn owls in the Eastern Mediterranean. **(a)** Nuclear and mitochondrial population structure. Horizontal bars indicate individual admixture proportions for K=5 as determined by sNMF. Black dots on map indicate the approximate centroid of each population; coloured pie charts represent the mean admixture proportions per population; pie charts in shades of beige represent mitochondrial haplogroup proportions per population. **(b)** PCA based on the pruned nuclear SNP set. Values in parenthesis indicate the percentage of variance explained by each axis. **(c)** Pairwise F_{ST} between sampled barn owl populations. Heat map illustrates the given values according to the legend.

Pairwise nuclear F_{ST} values ranged from 0.014 to 0.088 (Fig. 1c), with the highest found between Cyprus and Ionian (0.088), followed by between Crete and Israel (0.087). Crete exhibited overall the highest pairwise values with any population (all above 0.056). Matching population divergence, the quantitative depiction of gene flow through EEMS identified a strong barrier to migration around the island of Crete and regions of reduced migration around the southern Ionian islands and the island of Cyprus (Sup. Fig 5).

Genetic diversity and inbreeding

Genetic diversity based on nuclear SNP was generally highest in Israel and lowest in Crete, with Cyprus bearing comparable levels to any mainland population (Table 1). This was consistent for nuclear heterozygosity, population specific F_{ST} and gene diversity as well number of polymorphic sites in mtDNA. Private alleles were lowest among the closely related populations of Greece, the Ionian and Aegean Islands with Israel boasting the highest number. When considering GR, IO and AE as a genetic cluster (Fig. 1a), Crete actually had the lowest number of private alleles (Table 1).

Table 1 – Population genetic diversity, inbreeding and divergence estimates for barn owls of the eastern Mediterranean. The standard deviations of the values are provided between brackets for each parameter except for population specific F_{ST} where values are the standard error of the mean.

Pop	Abbr.	N	# PA	# PA lin.	#MT	π MT	H_O	Pop F_{ST}	F_{IS}	F_{IT}	F_{ROH}	β
Italy	**IT**	9	118’152 (202)	188’285	68 (23)	0.0021 (0.0011)	0.164 (0.002)	0.058 (0.005)	-0.024 (0.012)	0.014 (0.009)	0.028 (0.01)	0.112 (0.007)
Ionian Islands	**IO**	5	46’340		87	0.0021 (0.0013)	0.16 (0.004)	0.091 (0.004)	-0.039 (0.027)	0.019 (0.02)	0.067 (0.02)	0.143 (0.041)
Greece	**GR**	10	73’108 (220)	239’089 (294)	86 (19)	0.0021 (0.0012)	0.165 (0.005)	0.047 (0.003)	-0.02 (0.029)	0 (0.022)	0.038 (0.02)	0.101 (0.008)
Aegean Islands	**AE**	11	79’357 (198)		81 (18)	0.0023 (0.0012)	0.164 (0.01)	0.038 (0.002)	0 (0.059)	0.013 (0.049)	0.043 (0.03)	0.092 (0.014)
Crete	**CT**	11	82’202 (177)	124’440 (129)	51 (24)	0.0013 (0.0007)	0.153 (0.006)	0.115 (0.005)	-0.018 (0.037)	0.05 (0.034)	0.086 (0.04)	0.165 (0.024)
Cyprus	**CY**	10	121’550 (196)	177’675 (113)	72 (22)	0.002 (0.001)	0.165 (0.008)	0.061 (0.005)	-0.032 (0.05)	0.029 (0.039)	0.04 (0.03)	0.114 (0.024)
Israel	**IS**	9	271’400 (235)	413’375	43 (20)	0.0013 (0.0007)	0.175 (0.003)	0.007 (0.003)	-0.035 (0.016)	0.027 (0.008)	0.019 (0.01)	0.063 (0.012)

N: number of individuals in the population; #PA: private alleles per population, bootstrapped to the smallest N of 5 individuals; #PA lin.: private alleles per lineage of K=5 identified with sNMF, bootstrapped to the smallest N of 9 individuals; #MT: mitochondrial polymorphic sites per population, bootstrapped to the smallest N of 5 individuals; π MT: mitochondrial nucleotide diversity; H_O: observed heterozygosity; F_{ST}: population specific F_{ST} as in [159] bootstrapped over 100 blocks of contiguous SNP; F_{IS}: population level inbreeding coefficient; F_{IT}: mean individual inbreeding coefficient relative to the meta-population; F_{ROH}: mean inbreeding coefficient estimated from ROH; β: mean pairwise relatedness within population.

F_{IS}, the average inbreeding coefficients of individuals relative to their population, was slightly negative in all populations (Table 1), as expected with random mating in a species with separate sexes [219]. A Cretan individual had a local inbreeding coefficient below -0.1, likely due to it being a F1 hybrid between CT and AE (see individual bars in Fig. 1a and PC2 in Fig. 1b) and two samples

in the Aegean islands had a local inbreeding coefficient larger than 0.1 (Table 1; Fig. 3b). F_{IT} values, the average inbreeding of individuals relative to the total set, but averaged per population, were highest on the island of Crete followed by the Aegean and Cyprus. Israel had significantly lower F_{IT} than all other populations, whereas Crete's was higher than all but the Aegean (Fig. 3a; Table 1; X^2= 36.043, p < 0.001). Thus, Cyprus had higher F_{IT} than Israel, smaller than Crete and similar to every other population. Individual relatedness (β) was highest between two Ionian individuals found to be half-sibs (Sup. Fig. 4). Otherwise, individuals from Crete were more related to each other than any other pair of individuals in the dataset (Table 1; X^2= 195.77, p < 0.001). In its turn, individuals from Cyprus were only more related to each other on average than the populations of Israel and the Aegean.

Mean population F_{ROH} (i.e. proportion of the genome in runs of homozygosity) were also highest in Crete, followed by the Ionian, Aegean and Cyprus (Table 1). Individuals from Crete showed the highest proportion of ROH of all sizes (Fig. 3c; Sup. Fig. 6), while individuals from Israel had the lowest proportion in all categories. Individuals from Cyprus and the Aegean were also enriched in ROH segments compared to their mainland origin in most length classes, but much less so than Crete (Fig. 3c). Indeed, while F_{ROH} was significantly higher in Crete than in Greece, it was not the case between the Aegean and Greece nor between Cyprus and Israel (X^2= 11.862, p < 0.001).

Demographic history

We simulated three different demographic scenarios for each island system, two where the island was colonized from the mainland either before or after the other population in the model (AE for "Crete" and Ghost for "Cyprus") and one where the populations are colonized in a stepping-stone manner (Fig. 2b). The best demographic scenario inferred with *fastsimcoal2* for the island of Crete was the stepping-stone model (Fig. 2c; Sup. Table 3). Here, the Cretan population originates very recently from the Aegean islands 321 generations BP (68-1400 95% CI), itself colonized from mainland Greece at the fixed time of 6000 generations BP (Sup. Table 5). Estimated migration rates were higher towards the island from both GR and AE (6.7 and 3.7, respectively) and lower in the other direction (0.7 and 1.7). Inferred effective population sizes were highest for the Greek (1465 haploids; 509-7880 95% CI) mainland and lowest for Crete (373 haploids; 107-944 95% CI). Past instantaneous bottlenecks at colonisation were pronounced both for the Aegean and Cretan populations (48 [13-2922 95% CI] and 74 [6-243 95% CI] haploids, respectively).

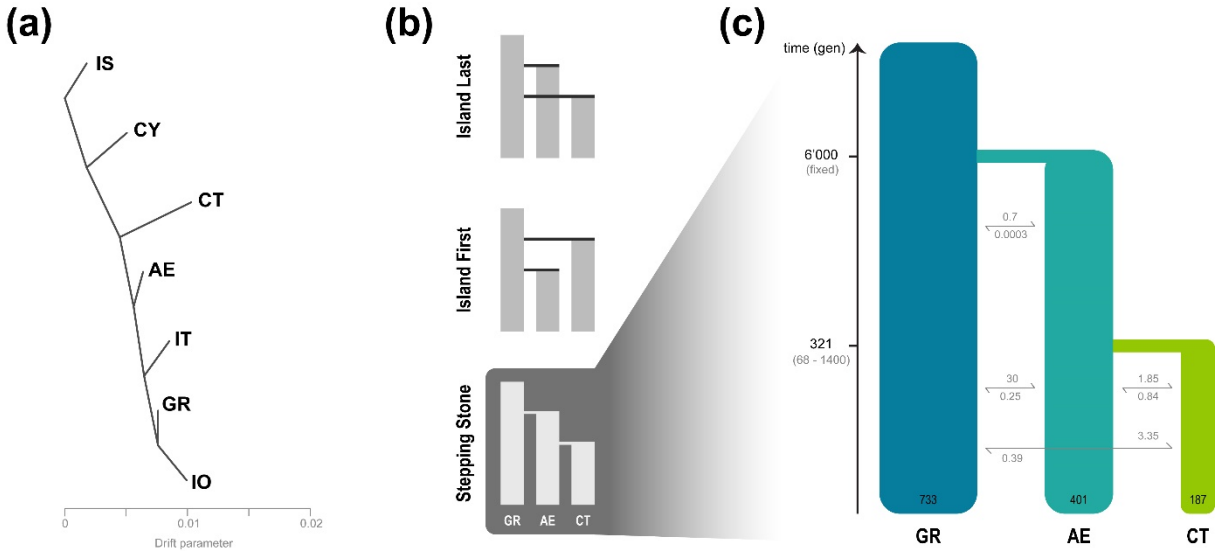


Figure 2 – Demographic history of barn owl insular populations in the Eastern Mediterranean. **(a)** Treemix analysis with zero migration events. Population abbreviations follow Figure 1. **(b)** Hypothesized demographic topologies for the colonisation of Crete. The same topologies were tested for Cyprus, with IS instead of GR, "Ghost" instead of AE and CY instead of CT. **(c)** Best supported demographic model for the colonisation of Crete as determined by fastsimcoal2. Time is indicated in generations, confidence intervals at 95% are given between brackets. Population sizes (diploid) are shown at the bottom of each population bar; arrows indicate forward-in-time number of migrants per generation.

For the island of Cyprus, the best-fitting scenario consisted of colonization from Israel after the colonization of the 'Ghost' population coinciding with a hypothesized mainland population residing on the southern coast of Turkey (similar topology as Fig. 2b; Sup. Table 3). Colonization time for CY was much more recent than the last glaciation (986 generations, less than 3'000 years BP; Sup. Table 7). However, the Ghost population was estimated to have an unrealistic large effective population size (65'310 diploids), and CY an extremely small one (61). The migration rates inferred indicate a complete replacement of CY each generation by the Ghost, suggesting this model is far from being an accurate representation of reality. As such we interpret its results with caution.

PSMC identified a pronounced bottleneck for all populations (around 20'000 years BP) but failed to show a clean split for the two Islands, particularly Crete, and any mainland population (Sup. Fig. 7).

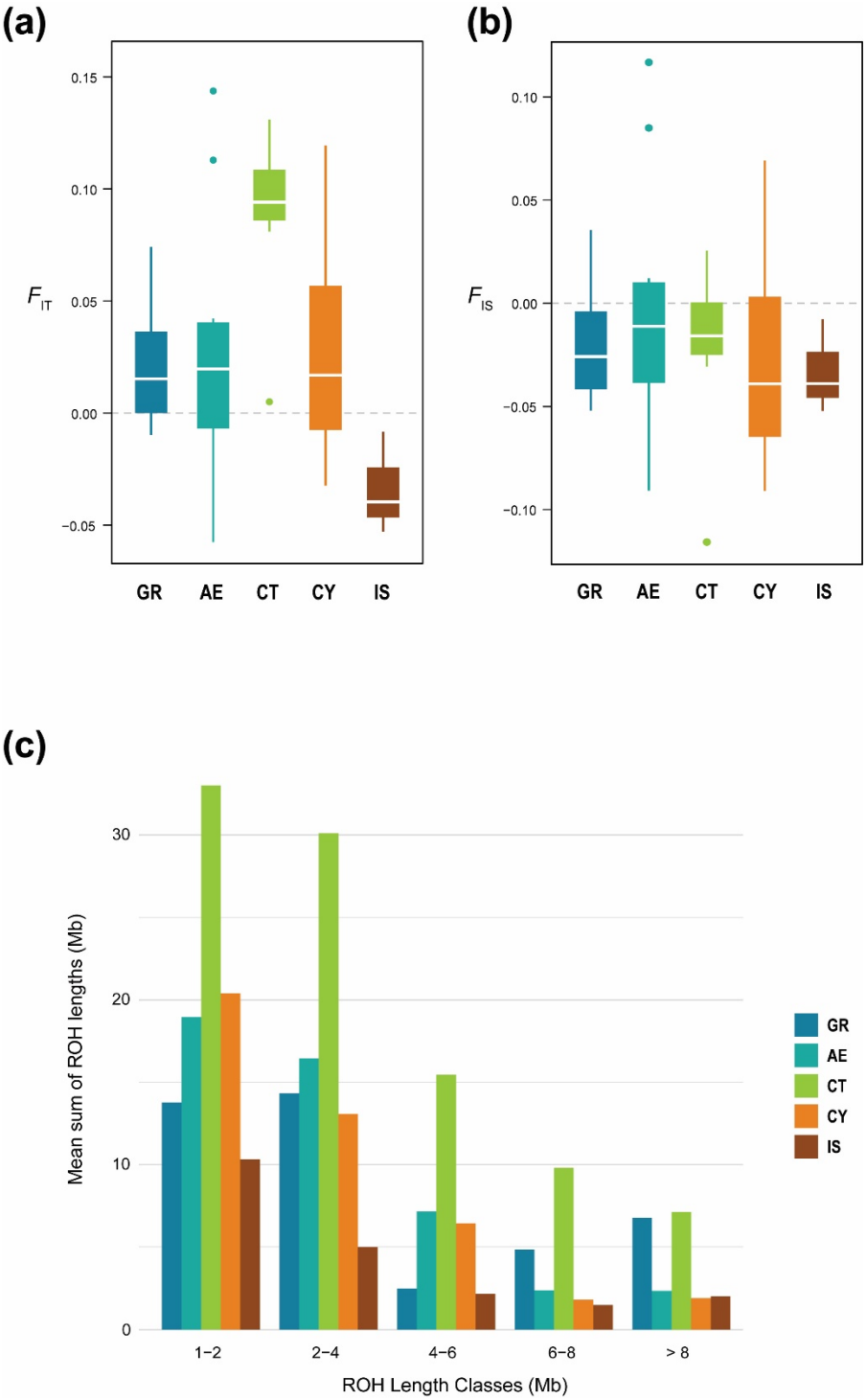


Figure 3 – Comparison of inbreeding in insular barn owls to their closest mainland counterparts in the Eastern Mediterranean. **(a)** F_{IT} measure of inbreeding calculated from individual allele matching proportions relative to the average in the dataset (dashed line is F_{IT}=0). **(b)** F_{IS} measure of inbreeding calculated from individual allele matching proportions relative to the average in the subpopulation (dashed line is F_{IS}=0). **(c)** Per population average length of ROH segments (in Mb) in each ROH length class.

Discussion

Although insular populations have greatly contributed to the development of evolutionary theory[31–33], the study potential of many of these remains untapped. The colonisation and settlement of an island by a given organism depend not only on the geographic context and specific island characteristics but also on stochastic events. As such, seemingly identical islands may yield populations with contrasting fates. Here, we investigate the demographic history and current patterns of inbreeding and genetic diversity of insular barn owls in the eastern Mediterranean Sea. In particular, we are interested in owls from Crete and Cyprus which, alongside the Levant region, were thought to form a subspecies *Tyto alba erlangeri*. These two similar islands in terms of size, climate and distance to mainland provide natural replicates for a comparative analysis of the colonisation and ensuing demographic processes. Using whole genome sequences, we show how each island and archipelago have unique histories and exhibit different degrees of isolation and the effect this has on the genomes of individuals. Specifically, Crete and Cyprus were colonized from distinct mainland locations, each from a different ancestral lineage, inconsistent with them belonging to the same subspecies. The population in Crete originated from the European lineage, more precisely from the Aegean islands, while the population in Cyprus came from the Levant in the east. Additionally, Crete underwent stronger genetic drift and inbreeding than Cyprus, resulting in a smaller and less diverse population.

Insular populations in the eastern Mediterranean

In the broader context of the Western Palearctic, our study targets two islands in the region where the European and eastern lineages of barn owls meet[204]. This is clearly shown in the genomic PCA, where the mainland populations of Italy and Greece in southern Europe were opposed to that of Israel in the Levant, with insular populations placed along this west-to-east genetic gradient roughly according to their geographic position (Fig. 1b, Sup Fig K=2). The main islands of Crete and Cyprus are the most genetically distinct populations (Fig. 1a,c), consistent with previous results for Crete[102]. Conversely, the Greek archipelagos – Ionian and Aegean – were genetically very similar to the Greek mainland population (Fig. 1a,c) suggesting they remain highly connected genetically. Such patterns of genetic differentiation reflect the geographical isolation of CT and CY, in contrast to the Aegean and Ionian archipelagos that are closer to the mainland through a network of adjacent islands and islets.

Overall, our results confirm that water bodies are strong barriers to barn owl movement [93,204]. For example, distant populations in the mainland, such as GR and IT, were much more similar to each other than any insular population, regardless of how distant each of them are (Fig. 1). Nonetheless, all insular populations showed small signals of admixture with their neighbouring

populations (Fig. 1a). This likely reflects the intricate geographic setting, as well as the overall low differentiation within this species (overall F_{ST} in our dataset 0.03, and 0.047 in the whole Western Palearctic[204], that mtDNA data lacked the resolution to detect (Fig. 1a, Table 1, Sup. Fig. 1; [102]). Insular populations had generally lower levels of population private diversity, while displaying similar levels of heterozygosity (Table 1), and higher within-population relatedness compared to the mainland (Sup. Fig. 4), reflecting their isolation. However, despite all populations appearing to mate randomly within localities (F_{IS} slightly negative as is expected from a dioecious species[219]; Fig. 3b, Table 1), the inbreeding levels of insular barn owls relative to the whole set of populations were quite large (F_{IT} and F_{ROH}; Fig. 3a,b, Table 1).

Crete and Cyprus

Despite the inherent physical similarities between the islands of Crete and Cyprus, their barn owl populations differ in many aspects. These natural replicates of island-mainland comparisons, with similar climatic conditions, supposedly harbour the subspecies *T. a. erlangeri* from the Levant (W. L. Sclater, 1921). However, while Cyprus' genetically closest mainland population is indeed Israel, Crete is actually most similar genetically to Greece (Fig. 1b). This demonstrates that Crete is not home to barn owls of the eastern subspecies, but rather from the European mainland lineage (*T. a. alba*). Not only they have separate geographic origins, but we also show that their *in situ* demographic histories are quite distinct.

Since its colonisation and founding bottleneck, Crete maintained a low population size with little gene flow with neighbouring populations (Fig. 2c; Sup. Fig. 5), generating background relatedness among individuals (Sup. Fig. 4). The low gene flow it maintains with the surrounding populations (Sup. Fig. 5) may be due to the very strong winds that surround the island[220] acting as a barrier by hindering flight. Thus, despite random mating within the island (low F_{IS}), remote inbreeding increased (high F_{IT} and F_{ROH}) due to high relatedness (high β), making CT the most inbred population by far in our dataset, as well as the least diverse (Table 1; Fig. 3a,b). Accordingly, it carried the highest proportion of ROH compared to any other populations (Table 1). Notably, CT was enriched in ROH of all sizes (Fig. 3c), suggesting a small effective size over a long time period until today[221]. This strong isolation coupled with small population size resulted in a very distinct genetic composition through the effect of genetic drift (Fig. 2a) as well as high individual relatedness and inbreeding.

In contrast, Cyprus appears to have maintained enough gene flow with the mainland preventing it from accumulating remote inbreeding, while allowing for differentiation. Winds in this region are weaker than around Crete[220], potentially facilitating the contact between Cyprus and Israel in the Levant, the most diverse population in our study. This could explain the surprisingly similar

patterns of genetic diversity in CY to that of mainland populations (Table 1), which suggest a higher effective population size in spite of the inference from fastsimcoal2 (Sup. Table 7). Furthermore, CY had considerably less runs of homozygosity (ROH) than CT, carrying only a slight enrichment in short length classes, similar to the Aegean and Ionian islands (Fig. 3c). Given the high inter-individual variability in relatedness and inbreeding coefficients (Table 1; Fig. 3a; Sup. Fig. 4), it appears that the gene flow with Israel and/or an unknown, unsampled population prevents the rise of population-wide inbreeding as observed in Crete. Interestingly, the most common mitochondrial haplogroup in CY was found in European populations but absent in IS (haplogroup 3, Fig. 1a). Although it could simply have been unsampled in the Levant, it may also be evidence of some gene flow between the European and eastern lineages as seen in the two admixed individuals of CY (Fig. 1a; see also paragraph after next)

Overall, the different levels of connectivity (i.e. levels of gene flow) of each island appear to be the main driver of their diverging histories. However, insular specificities may also contribute to this effect. The carrying capacity of CT and CY for barn owls could be different due to cryptic differences in nesting or roosting site availability, for example, in spite of their similar surface area. In addition, the mountainous landscape in CT could restrict dispersal movements as well as reduce the suitable surface for breeding. Finally, intrinsic characteristics of the colonisation of both islands may also have contributed to their diverging histories.

On the one hand, CY was colonised directly from the highly diverse and large mainland population of IS (Sup. Table 7). As such, both the settlers of the island and subsequent immigrants were likely unrelated and diverse, preventing the insular population from increasing steeply in relatedness. Our simulations suggest that colonisation occurred about 3000 years BP (1900 – 10000 years BP). However, this result should be interpreted cautiously as the modelling for this island system yielded unreasonable population size estimates (Sup. Table 7) likely due to our use of a ghost population to represent mainland Turkey. This is suspected to be a contact zone between the European and eastern barn owl lineages with sporadic gene flow[204]. Our observations support this hypothesis as islands on both sides of Turkey, namely CY and AE, carried some small genetic components from the other (Fig. 1a). In this context, our modelled ghost population would likely be admixed or even outbred which would explain its exaggerated population size. Sampling in Turkey will be key to clarify this hypothesis and fully describe the dynamics between barn owl populations in the eastern Mediterranean.

On the other hand, demographic simulations showed that CT was colonised from the AE archipelago rather than directly from mainland Greece (Fig. 2c; Sup. Table 5). This was supported by the second axis in the PCA which placed individuals in a gradient from GR to AE and then CT (Fig. 1b). Remarkably, one AE owl from the south-eastern island of Rhodes had approximately 50% Cretan origin hinting at how the patchwork of islands and islets in the region could have

been used as stepping stones during colonisation. Thus, CT was colonised from what is already a less diverse insular population, which in turn came from the GR mainland, itself less diverse than Israel (Table 1). This cumulative loss of diversity through recurrent bottlenecks and possible expansion could contribute to the quick increase in relatedness in the island given its small population size, despite its recent colonisation. Indeed, CT was inferred to have been colonised by barn owls around 1000 years BP (204 – 4200 years BP; Fig. 2c; Sup. Table 5). Accordingly, PSMC failed to uncover any signal of older divergence (Sup. Fig. 7). Nonetheless, considering the geological age of the island (5 million years BP) and that agricultural practices have been established there for millennia[222], this estimation appears extraordinarily recent. Absent any other source of evidence, one can only speculate as to why this population is so recent. It is possible that a massive migration led to population replacement at a time when sea levels were lower and the surrounding islands closer, masking any trace of an earlier settlement. Alternatively, earlier settlers could have been extinct due, for example, to a natural disaster such as the catastrophic Minoan volcanic eruption (3'500 years BP)[223].

Conclusion

Our work provides a comparative study on two natural replicates of island colonisation by the barn owl, a bird that despite being found in many islands avoids flying over open bodies of water. The use of whole genome sequences allowed us to demonstrate that Crete and Cyprus owls come from different genetic backgrounds, as each island originates from a distinct continental genetic lineage (Fig. 1). Further, their histories diverge resulting in noticeably different populations. Cyprus was colonised directly from the most diverse mainland population, accumulated differentiation but also remained sufficiently connected with it to maintain high levels of genetic diversity and prevent inbreeding (Table 1, Fig. 3). Crete was reached by island hopping in the Aegean from a less diverse mainland population. The small size and isolation of this island population facilitated the impact of genetic drift which, along with inbreeding, led to it diverging considerably from its founders despite the recent colonisation (Table 1, Fig. 2). Although further analyses would be necessary to study the functional consequences of inbreeding in Crete, this study shines a light on a real-life illustration of stochasticity in the classical island-mainland model systems.

Chapter 2 – Supporting Information

Supplementary Tables

Supplementary Table 1 – Description of samples used in this study. Individuals retained for the inference with fastsimcoal2 are indicated with †. Individuals marked with ‡ were removed from the analyses as they were each part of a sibling pair.

#	Pop	ID	Country	Location	Year	Tissue	Sex	Ref
1	Italy	IT01	Italy	Roma	2011	blood	Female	1
2	Italy	IT02	Italy	Roma	2015	blood	Female	1
3	Italy	IT03	Italy	Roma	2016	blood	Male	1
4	Italy	IT04	Italy	Roma	2016	blood	Male	1
5	Italy	IT05	Italy	Roma	2009	blood	Female	1
6	Italy	IT06	Italy	Grosseto	2014	blood	Female	1
7	Italy	IT07	Italy	Livorno	2001	blood	Female	1
8	Italy	IT08	Italy	Firenze	2011	blood	Female	1
9	Italy	IT09	Italy	Firenze	2016	blood	Male	1
10	Italy	IT10‡	Italy	Firenze	2017	blood	Female	2
11	Ionian	IO01	Greece	Kefalonia Island	2014	soft tissue	Male	2
12	Ionian	IO02	Greece	Kerkyra island	2013	soft tissue	Female	2
13	Ionian	IO03	Greece	Kerkyra island	2014	soft tissue	Female	2
14	Ionian	IO04	Greece	Zakynthos island	2015	blood	Male	2
15	Ionian	IO05	Greece	Zakynthos island	2015	blood	Female	2
16	Greece	GR01†	Greece	Agrinio	2014	blood	Male	1
17	Greece	GR02†	Greece	Athens	2014	blood	Male	1
18	Greece	GR03†	Greece	Chalandri	2012	blood	Male	1
19	Greece	GR04†	Greece	Corinth	2015	blood	Female	1
20	Greece	GR05†	Greece	Stoupa	2015	soft tissue	Female	2
21	Greece	GR06†	Greece	Lamia	2014	soft tissue	Female	1
22	Greece	GR07†	Greece	Mesolonghi	2014	soft tissue	Male	1
23	Greece	GR08†	Greece	Morfovouni	2015	blood	Female	1
24	Greece	GR09†	Greece	Panetolio	2014	blood	Female	1
25	Greece	GR10†	Greece	Spata	2015	blood	Male	1
26	Aegean	AE01	Greece	Rhodes island	2014	soft tissue	Female	1
27	Aegean	AE02†	Greece	Rhodes Island	2014	soft tissue	Female	1
28	Aegean	AE03†	Greece	Chios island	2015	blood	Male	1
29	Aegean	AE04†	Greece	Chios island	2012	blood	Female	1
30	Aegean	AE05†	Greece	Leros island	2012	blood	Male	1
31	Aegean	AE06†	Greece	Lesvos Island	2012	blood	Male	1
32	Aegean	AE07†	Greece	Lesvos Island	2013	soft tissue	Female	1
33	Aegean	AE08†	Greece	Rhodes island	2013	blood	Male	1
34	Aegean	AE09†	Greece	Leros island	2014	blood	Male	1
35	Aegean	AE10†	Greece	Rhodes Island	2015	blood	Female	1
36	Aegean	AE11†	Greece	Andros island	2015	blood	Male	2

37	Crete	CT01†	Greece	Crete	2015	soft tissue	Male	2
38	Crete	CT02†	Greece	Crete	2015	soft tissue	Male	2
39	Crete	CT03†	Greece	Crete	2015	blood	Male	2
40	Crete	CT04†	Greece	Crete	2015	soft tissue	Male	2
41	Crete	CT05†	Greece	Crete	2015	blood	Male	2
42	Crete	CT06	Greece	Crete	2015	blood	Female	2
43	Crete	CT07†	Greece	Crete	2013	soft tissue	Female	2
44	Crete	CT08†	Greece	Crete	2014	blood	Male	2
45	Crete	CT09†	Greece	Crete	2014	blood	Female	2
46	Crete	CT10†	Greece	Crete	2014	blood	Female	2
47	Crete	GR11†	Greece	Kalamata	2012	blood	Female	2
48	Cyprus	CY01†	Cyprus	Cyprus	NA	soft tissue	Female	1
49	Cyprus	CY02†	Cyprus	Cyprus	2016	soft tissue	Female	1
50	Cyprus	CY03†	Cyprus	Cyprus	2016	soft tissue	Male	1
51	Cyprus	CY04†	Cyprus	Cyprus	NA	soft tissue	Female	1
52	Cyprus	CY05†	Cyprus	Cyprus	NA	soft tissue	Male	1
53	Cyprus	CY06†	Cyprus	Cyprus	NA	soft tissue	Female	1
54	Cyprus	CY07†	Cyprus	Cyprus	2017	soft tissue	Female	1
55	Cyprus	CY08†	Cyprus	Cyprus	2015	soft tissue	Female	1
56	Cyprus	CY09†	Cyprus	Cyprus	2017	soft tissue	Female	1
57	Cyprus	CY10	Cyprus	Cyprus	2018	soft tissue	Female	1
58	Israel	IS01†	Israel	Lachish	2005	blood	Female	1
59	Israel	IS02†	Israel	Beit Shean	2005	blood	Male	1
60	Israel	IS03†	Israel	Hula	2005	blood	Female	1
61	Israel	IS04†	Israel	Beit Shean	2005	blood	Female	1
62	Israel	IS05†	Israel	Beit Shean	2005	blood	Male	1
63	Israel	IS06†	Israel	Beit Shean	2005	blood	Female	1
64	Israel	IS07†	Israel	Beit Shean	2005	blood	Female	1
65	Israel	IS08†	Israel	Hula	2005	blood	Female	1
66	Israel	IS09†	Israel	Hula	2005	blood	Female	1
67	Israel	IS10‡	Israel	Hula	2005	blood	Female	2
68	Outgroup	SGP	Singapore	Singapore	2015	soft tissue	Female	3

[1] Cumer *et al.* In Prep– GenBank BioProject PRJNA727977 [204]
[2] This study – GenBank BioProject PRJNA727915
[3] Machado *et al.* 2021 – GenBank BioProject PRJNA700797 [93]

Supplementary Table 2 – Filters applied to raw variant calls. In the last column x denotes the values of the sites kept in the dataset.

Parameter	Abbreviation in VCF file	Retained values
Quality by Depth	QD	5 < x
Root Mean Square Mapping Quality	MQ	40 < x < 70
Mapping Quality Rank Sum Test	MQRankSum	-12.5 < x
Read position Rank Sum Test	ReadPosRankSum	-8.0 < x
Strands Odds Ratio	SOR	x < 3.0
Fisher strand	FS	x < 60
Excess Heterozygosity	ExcessHet	x < 20
Inbreeding coefficient	InbreedingCoeff	x < 0.9
Site Depth	DP	500 < x < 1600
Genotype Depth	DP	10 < x < 40
Genotype Quality (Phred score)	GQ	20 < x
Percent of missing data per site		x < 5%
HWE exact test (p-value)		0.05 < x

Supplementary Table 3 – Likelihood and AIC comparison between simulated scenarios for systems Crete and Cyprus. Three topologies were tested for each system (Figure 2). The best scenario for each system is highlighted in bold.

System	Scenario	Likelihood	Δ likelihood	AIC	Δ AIC
CRETE	A	-2'506'864	5'219	5'013'755	603
	B	-2'508'348	6'703	5'016'724	3'572
	C	**-2'506'562**	**4'917**	**5'013'152**	**0**
CYPRUS	**A**	**-878'596**	**-1'025**	**1'757'220**	**0**
	B	-878'826	-1'277	1'757'679	460
	C	-878'794	-1'246	1'757'617	397

Likelihood – Maximum-likelihood estimated for the simulated SFS per demographic model; Δ likelihood – difference between the likelihood of the simulated and observed SFS; Δ AIC – delta AIC

Supplementary Table 4 – Parameter ranges and estimates of fastsimcoal2 for system "Crete". Initial range is the range from which the software draws initial the parameter values. The distributions are uniform for every parameter except *Bottleneck Intensity* which is log-uniform. Highest likelihood estimates are provided for every parameter. In scenarios A, C the Aegean islands diverge first from mainland Greece and therefore time of divergence for Crete is scaled to be less than 6'000 generations, and ancient migration is only between GR and AE. For scenario B, Crete is more ancient than AE and the time of split is ≥ 1 and ancient migration rates are between GR and CT. GR: Greece AE: Aegean CT: Crete.

Parameter	Initial Range	Point estimates		
		Crete A	Crete B	Crete C
Haploid population sizes				
GR	500 – 100'000	502	7156	1465
AE	50 – 100'000	283	638	802
CT	10 – 4'000	162	1468	373
Bottleneck Intensity (fraction of current population size)				
AE	0.01 – 0.5	0.187	0.064	0.060
CT	0.01 – 0.5	0.206	0.070	0.198
Time of divergence (proportion of 6000; generations)				
T^{CRETE} (scenario A, C)	0.001 – 0.999	0.039	-	0.053
T^{CRETE} (scenario B)	1 – 3	-	1.218	-
Ancestral migration rates (backwards in time)				
GR → AE (scenario A, C)	0 – 0.05	1e-06	-	1e-06
AE → GR (scenario A, C)	0 – 0.05	0.0018	-	0.0009
GR → CT (scenario B)	0 – 0.05	-	0.0029	-
CT → GR (scenario B)	0 – 0.05	-	0.0009	-
Current migration rates (backwards in time)				
GR → AE	0 – 0.05	0.0332	0.0002	0.0006
AE → GR	0 – 0.05	0.0678	0.0429	0.0417
GR → CT	0 – 0.05	0.0005	0.0003	0.0021
CT → GR	0 – 0.05	0.0198	0.0003	0.0046
AE → CT	0 – 0.05	0.0128	0.0074	0.0045
CT → AE	0 – 0.05	0.001	0.0022	0.0046

Supplementary Table 5 – System "Crete"; Point estimates and 95% confidence intervals from non-parametric bootstrapping for the best-fitting demographic scenario for system Crete – scenario C. Haploids/ generation in migration rates were estimated by multiplying each migration rate with the point estimate haploid size of the population of origin.

Parameter	95% Lower limit	Point estimate	95% Upper limit
Haploid population sizes			
GR	509	1'465	7'880
AE	383	802	60'962
CT	107	373	944
Bottleneck intensity (haploid size)			
AE	13	48	2922
CT	6	74	243
Time of divergence (generations ago)			
T^{CRETE}	68	321	1'400
Ancient Migration rates (forward migration in haploids/generation)			
GR → AE	0.1421	1.3185	61.823
AE → GR	0.0441	0.0577	36.09
Current Migration rates (forward migration in haploids/generation)			
GR → AE	0.082	61.09	57.867
AE → GR	7.218	0.481	56.942
GR → CT	0.063	6.739	10.987
CT → GR	0.0268	0.7833	6.0426
AE → CT	0.1604	3.6892	8.3408
CT → AE	0.0373	1.6785	9.698

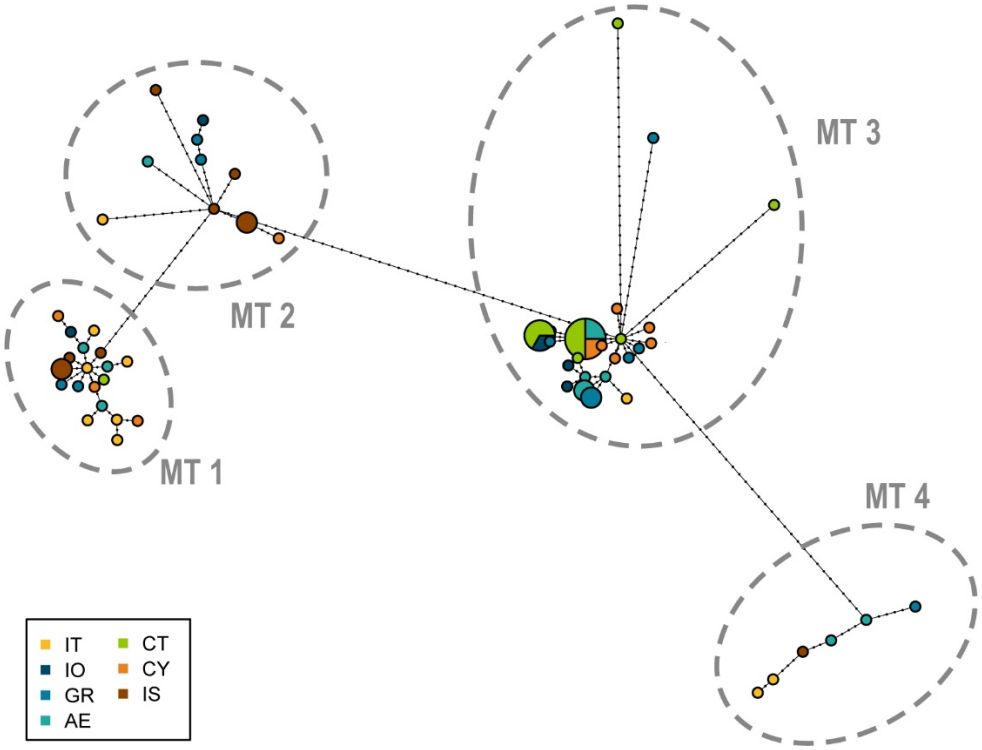
Supplementary Table 6 – Parameter ranges and estimates of fastsimcoal2 for system "Cyprus". Initial range is the range from which the software draws initial the parameter values. The distributions are uniform for every parameter except *Bottleneck Intensity* which is log-uniform. Highest likelihood estimates are provided for every parameter. In scenarios A, C the GH population diverges first from IS and therefore time of divergence for CY is scaled to be smaller and ancient migration is only between IS and GH. For scenario B, CY is more ancient than GH and the time of split is reversed, and ancient migration rates are between IS and CY. IS: Israel; GH: Ghost population; CY: Cyprus.

Parameter	Initial range	Point estimates		
		Cyprus A	Cyprus B	Cyprus C
Haploid population sizes				
IS	5'000 – 150'000	5'140	7'959	5'631
GH	100 – 150'000	130'619	487	109'265
CY	100 – 50'000	122	47'095	292
Bottleneck intensity (fraction of current population size)				
GH	0.01 – 0.5	0.110	0.296	0.025
CY	0.01 – 0.5	0.104	0.135	0.023
Time of divergence (proportion of 6000; generations)				
T^{CYPRUS} (scenario A, C)	0.001 – 0.999	0.164	-	0.075
T^{GHOST} (scenario A, C)	0.001 – 3	1.709	-	0.715
T^{CYPRUS} (scenario B)	0.001 – 3	-	0.646	-
T^{GHOST} (scenario B)	0.001 – 0.999	-	0.198	-
Ancestral migration rates (backwards in time)				
IS → GH (scenario A, C)	0 – 0.05	0.0001	-	0.0005
GH → IS (scenario A, C)	0 – 0.05	3e-05	-	0.0004
IS → CY (scenario B)	0 – 0.05	-	0.0431	-
CY → IS (scenario B)	0 – 0.05	-	0.0056	-
Current migration rates (backwards in time)				
IS → GH	0 – 0.05	0.001	0.0001	2e-05
GH → IS	0 – 0.05	0.0087	0.0086	0.0039
IS → CY	0 – 0.05	0.0001	0.0001	0.0001
CY → IS	0 – 0.05	0.0078	0.0005	0.0152
GH → CY	0 – 0.05	0.0018	0.0094	0.0108
CY → GH	0 – 0.05	0.0424	0.035	0.0051

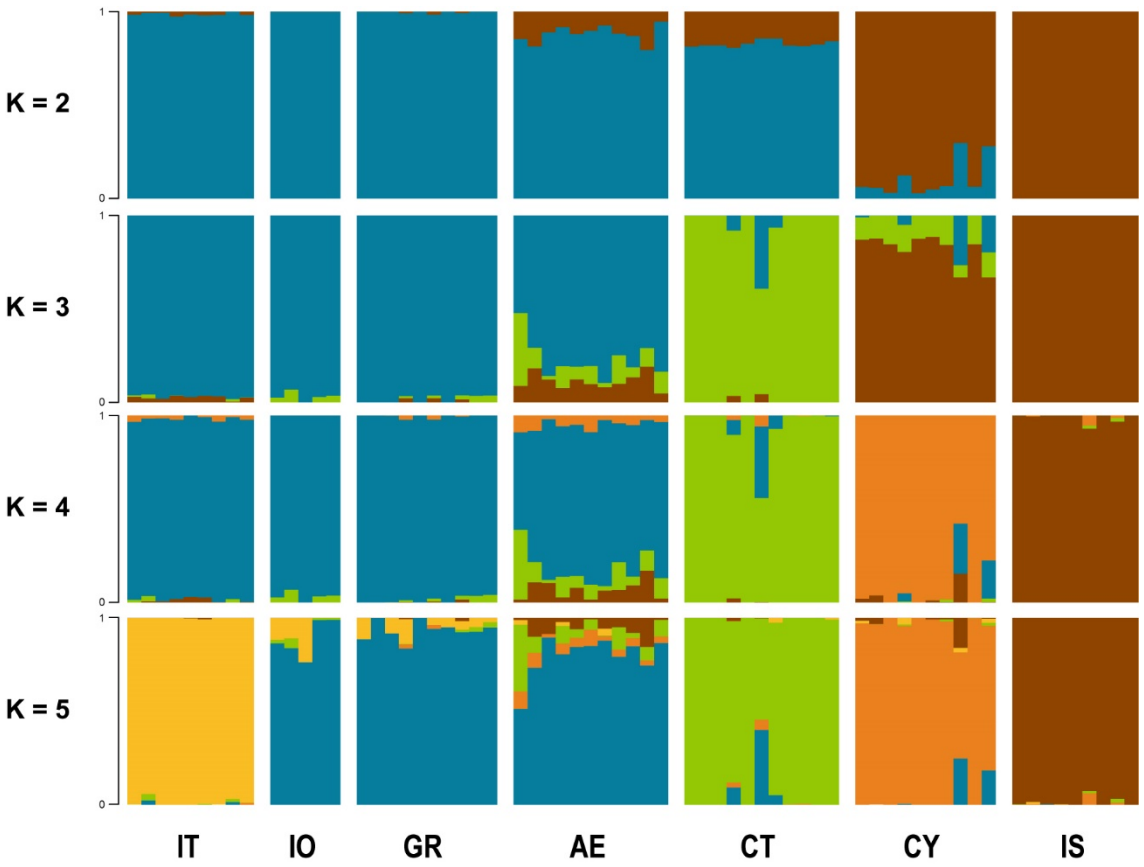
Supplementary Table 7 – System "Cyprus"; Point estimates and 95% confidence intervals from non-parametric bootstrapping for the best-fitting demographic scenario for system Cyprus – scenario A. Haploids/ generation in migration rates were estimated by multiplying each migration rate with the point estimate haploid size of the population of origin.

Parameter	95% Lower limit	Point estimate	95% Upper limit
Haploid population sizes			
IS	5'068	5'140	19'159
GH	34'780	130'619	172'130
CY	101	122	785
Bottleneck intensity (haploid size)			
GH	1'176	14'423	11'739
CY	3	12	145
Time of divergence (generations ago)			
T^{CYPRUS}	633	986.1	3'336
T^{GHOST}	2'392	10'250.9	11'739
Ancient Migration rates (forward migration in haploids/generation)			
IS → GH	4.8948	0.97146	160.38
GH → IS	231	13.1	6440
Current Migration rates (forward migration in haploids/generation)			
IS → GH	5.11	44.944	224
GH → IS	0.2069	133.022	13.4799
IS → CY	27.5	40.16	171
CY → IS	0.00026	0.0122	0.02353
GH → CY	145.9	5538.598	51'880
CY → GH	0.113	0.216	42.9

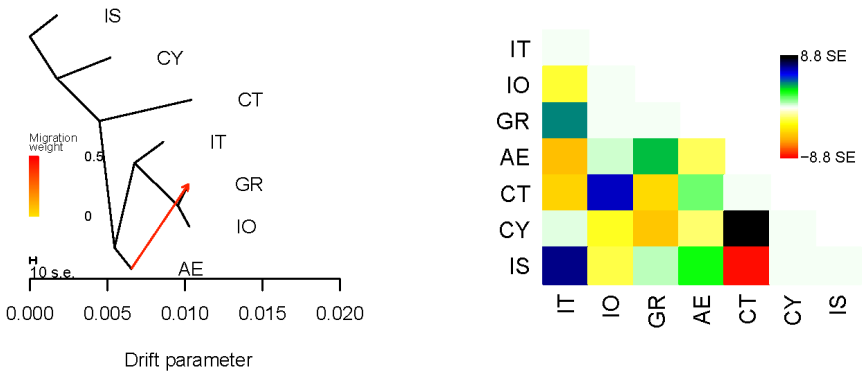
Supplementary Figures



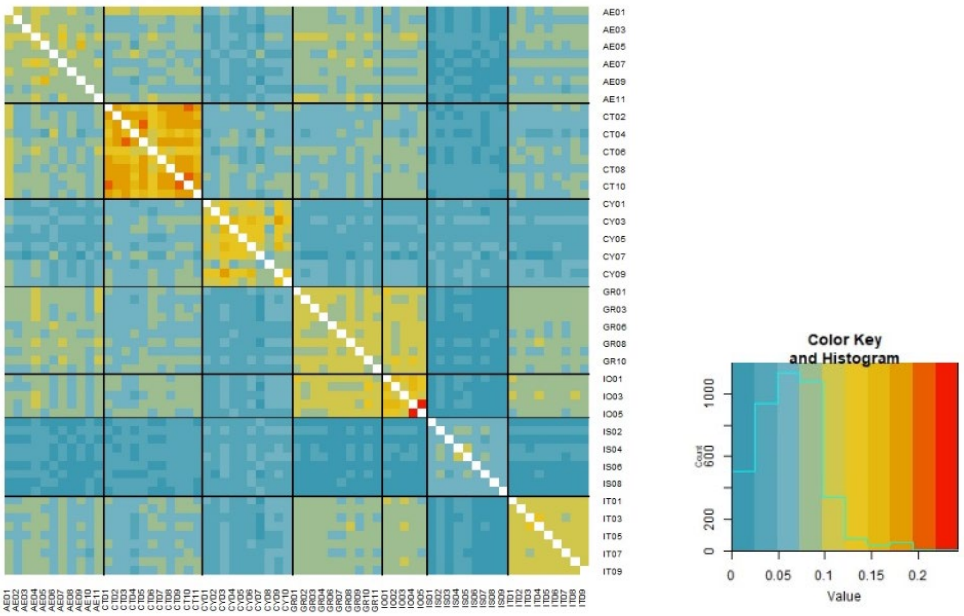
Supplementary Figure 1 – Haplotype network based on mtDNA of barn owls in the Mediterranean. The whole mitochondrial genome, except the D-loop, was used to construct the network. Small black dots represent individual mutation steps. Pie size is proportional to number of individuals carrying the haplotype. Large dashed circles indicate MT haplogroups 1 to 4, which are plotted in Figure 1a.



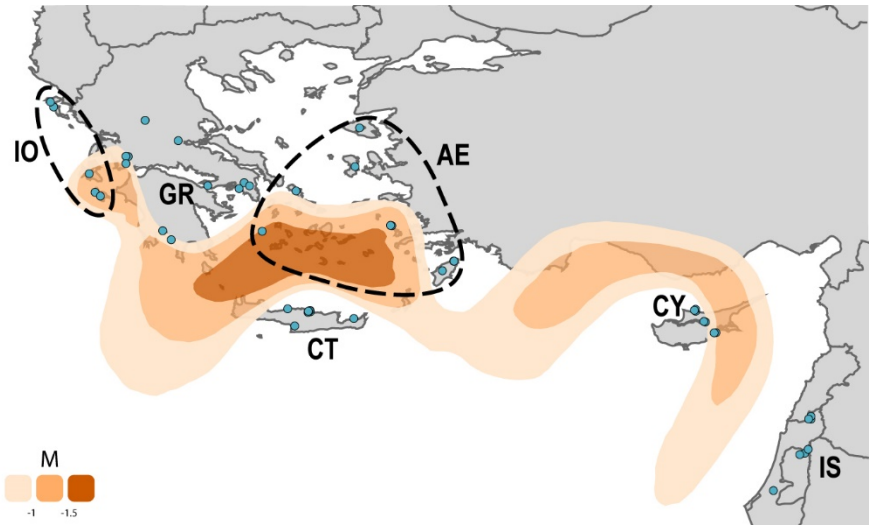
Supplementary Figure 2 – Individual clustering estimated by sNMF for K 2 to 5 lineages. Each vertical bar represents one individual, and the colours represent the relative contributions of each genetic lineage.



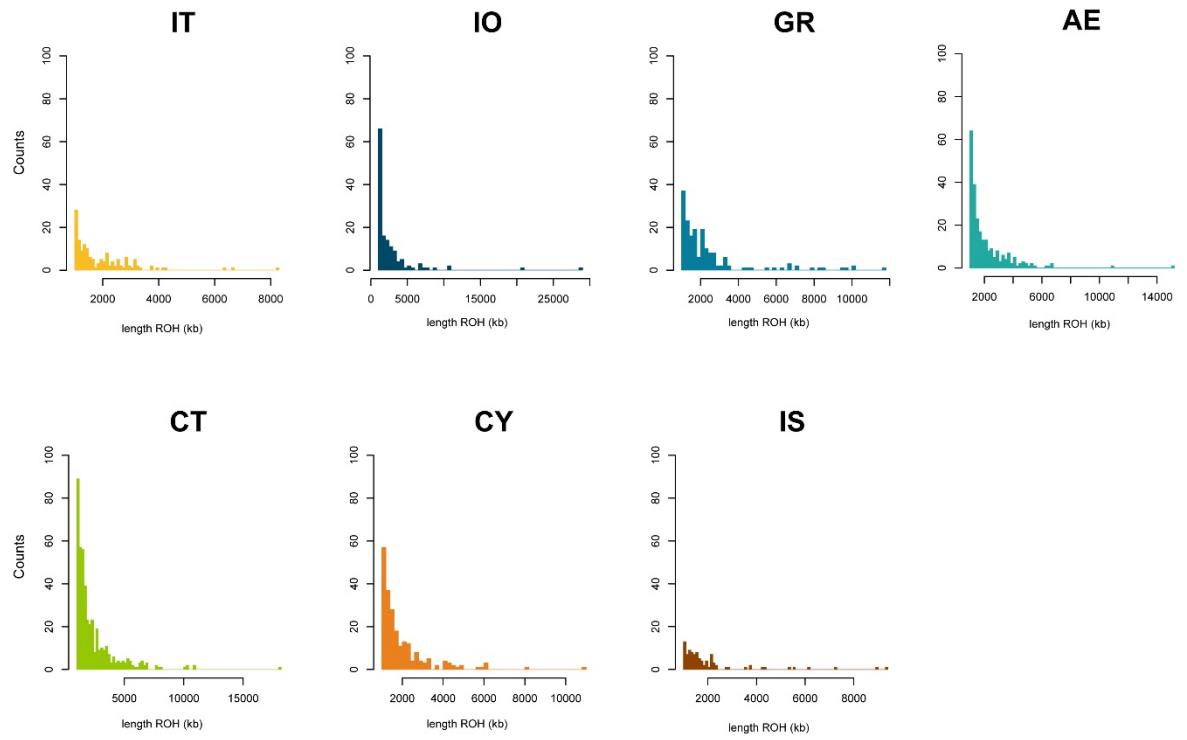
Supplementary Figure 3 – Treemix analysis for 1 migration events with its residual matrix.



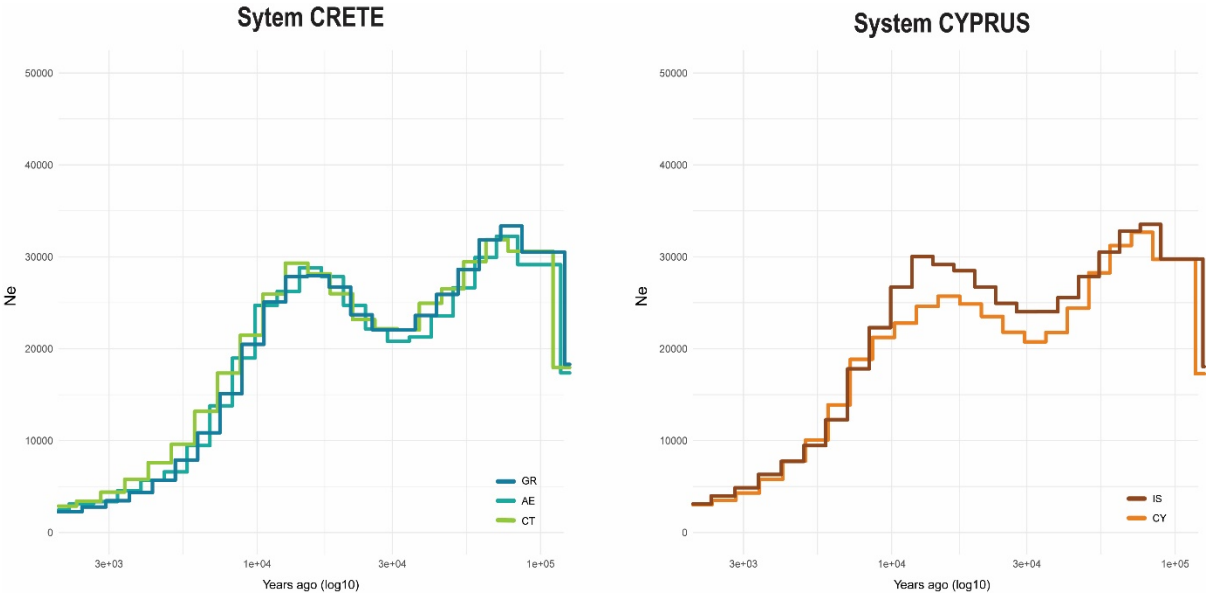
Supplementary Figure 4 – Pairwise individual relatedness (β) heatmap between all individuals.



Supplementary Figure 5 – Estimated effective migration surface (EEMS) based on whole-genome data. Orange shading denote regions of lower than average gene flow, darker shading indicates a stronger barrier. Dots indicate individual sampling location. Dashed lines delimit the Greek archipelagos sampled.



Supplementary Figure 6 – Distribution of ROH segments per population. Vertical axes are the same between plots but note the differences in the horizontal axes.



Supplementary Figure 7 – PSMC results for system Crete and Cyprus, separated for readability. Each line represents the median across individuals in each population. Plotted with 8.28x10^{-9} mutation rate and 3.6 years generation time.

Chapter 3

Insularity and ecological divergence in barn owls (*Tyto alba*) of the Canary Islands

Tristan Cumer[a†], Ana Paula Machado[a†], Felipe Siverio[s], Rui Lourenço[e], Motti Charter[l,m], Alexandre Roulin[a‡], Jérôme Goudet[a,j‡]

† first co-authors, sorted alphabetically

‡ co-senior authors

Status

In prep. We have received and are currently sequencing samples from Morocco that will be added to this project for publication.

Author contributions

APM, TC, AR, JG designed this study; APM produced whole-genome resequencing libraries; TC, APM conducted the analyses; FS, RL, MC provided samples; APM led the writing of the manuscript with input from all authors.

Abstract

Islands, and the at-times particular organisms that populate them, have long fascinated biologists, and were key to the formulation of modern evolutionary theory. Due to their isolation, islands offer unique opportunities to study the effect of neutral and adaptive mechanisms in determining genomic and phenotypical divergence. In the Canary Islands, an archipelago rich in endemics, the barn owl is thought to have diverged into a subspecies on the eastern islands of Fuerteventura and Lanzarote. Taking advantage of 37 whole-genomes and modern population genomics tools, we provide the first look at the genetic makeup of barn owls of this archipelago and contrast it to mainland populations in the Mediterranean Basin. We show that the Canarias hold surprisingly diverse, long-standing and monophyletic populations with a neat distinction of genes pools from the different islands. Then, we used a new method, less sensitive to structure than classical F_{ST}, to detect regions involved in local adaptation to the insular environment. We identified a haplotype-like region likely under positive selection in all Canarias barn owls, and genes in the region suggest morphological adaptations to insularity. In the eastern islands, where the subspecies is present, genomic differentiation pinpoint signs of locally adapted body proportions and blood pressure, consistent with the smaller size of this population living in a hot arid climate. In turn, genomic regions under selection in the western barn owls from Tenerife, showed an enrichment in genes linked to hypoxia, a potential response to inhabiting a small island dominated by a large mountain. Our results illustrate the interplay of neutral and adaptive forces in shaping divergence and early onset speciation.

Keywords

Local adaptation; Niche analysis; Population genomics; *T. a. gracilirostris*; Whole genome sequencing

Introduction

Due to the often-peculiar organisms that inhabit them, islands have always fascinated naturalists and scientists alike. Since Darwin's first visit to the Galapagos in 1835, the study of insular populations has been crucial to the development of evolutionary theory[31–33]. Labelled nature's test tubes, islands are home to a myriad of endemic (sub)species. Moreover, the combination of their relatively small size, discrete borders, geographical isolation and natural replication provides an excellent setting to study the evolutionary forces underlying population divergence and speciation[34]. The divergence of insular populations from their founders and surrounding islands is the result of neutral and selective forces. Disentangling the respective impacts of these two forces is a challenging task, as they are interconnected and can both contribute to genetic and phenotypic differentiation.

Isolated and small populations, such as those frequently found on islands, are under a markedly strong influence of genetic drift. It will alter the genetic makeup of the population by randomly removing rare alleles and fixating common ones, hence decreasing genetic diversity[196]. This is a common occurrence on islands which, coupled with low gene flow, can lead to inbreeding[195] and accelerate neutral divergence. Conversely, the absence of regular gene flow, and the often distinctive ecological conditions of the islands, can facilitate the action of local adaptation on beneficial alleles[194,224]. This process can lead to the emergence of ecomorphs via ecological divergence as populations adapt to unfilled insular niches[34], particularly so in remote islands that are colonized less often[225]. Ecomorphs can occur in different islands or in the same one[43,226,227], a concept mirrored in inland lakes[228], the aquatic homologous of islands. Eventually, ecomorphs can become new species or subspecies and, in extreme cases, result in adaptive radiations as illustrated by Darwin's finches[61,62].

The Afro-European barn owl (*Tyto alba*) is a non-migratory, nocturnal raptor present from Scandinavia to Southern Africa. It is also found on numerous islands and archipelagos where subspecies have often been described[89]. A recent study of the species' genetic structure in the Western Palearctic[204] described two main lineages occupying this region: the eastern lineage in the Levant and the western in Europe. In addition, Cumer *et al.* showed that barn owls from Tenerife (Canary Islands) were very distinct from both lineages. The Canarias are a volcanic archipelago that was formed several million years ago[229] about 100 km from the coast of north-western Africa. Unlike most other islands in the Western Palearctic, the Canarias' distance to the mainland has remained stable over time. This long-term isolation[230], along with its subtropical climate and elevation gradients[231], has resulted in high endemism, for example in plants[232], reptiles[233–235], mammals[236–238] and birds[239–241]. Among them, an endemic barn owl subspecies, *T. a. gracilirostris* (Hartert, E, 1905) based on its reportedly smaller size[205,242]. It is present in the

eastern Canarias (Lanzarote, Fuerteventura and surrounding islets), and is the only barn owl on these specific islands.

The presence of this subspecies on the eastern islands is surprising however, given that it is sandwiched between the western islands and the mainland which both harbour the nominal species *T. alba*. Lacking any evidence of different colonization origins or timing, this could suggest that local adaptation acting on the eastern population has accelerated divergence in comparison to the western. In both cases, neutral and adaptive microevolutionary processes promoting divergence of insular populations would leave traces on their genomic makeup. The advances in sequencing technology, and its decreasing costs, now allow to sequence the entire genomes of individuals. With the parallel development of sophisticated tools, it is possible to analyse changes in allelic frequencies at a high resolution to investigate the history of populations and inspect the genomic landscape for signals of local adaptation.

Here, we investigate the genomic bases of differentiation of barn owls from the Canary Islands. Making use of the whole-genome sequences from 37 individuals, we first describe the neutral genetic structure and diversity of the Canarias populations in contrast to the mainland, in order to retrace their history. Second, we employ a new relatedness-based method to probe the genomic landscape of these isolated populations for signals of local adaptation to the insular environment in regards to the mainland. Third, we characterize the climatic niches barn owls occupy on eastern and western islands, and explore how each population is diverging to adapt to their niches from a genomic perspective. Our results elucidate the genomic bases of the differentiation of insular populations, thus enlightening the classification of the Canarias barn owls.

Materials & Methods

Whole-genome sequencing, SNP calling and identification of coding regions

For this study, 37 individual barn owls were sampled from four populations (Figure 1; Sup. Table 1): Eastern Canaries (EC, Fuerteventura and Lanzarote), Western Canary (WC, Tenerife), Portugal (PT) and Israel (IS). All but the EC population have been published in previous work, including a North American owl that was used as an outgroup for some analyses[93,204]. For EC, we followed the same molecular and sequencing protocol as in the aforementioned publications. Succinctly, genomic DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) and individually tagged 100bp TruSeq DNA PCR-free libraries (Illumina) were prepared according to manufacturer's instructions. Whole-genome resequencing was performed on multiplexed

libraries with Illumina HiSeq 2500 high-throughput paired-end sequencing technologies at the Lausanne Genomic Technologies Facility (GTF, University of Lausanne, Switzerland).

The bioinformatics pipeline used to obtain analysis-ready SNPs from the 37 individuals plus the outgroup was the same as in Machado *et al.*[93], adapted from the Genome Analysis Toolkit (GATK) Best Practices[154] to a non-model organism following the developers' recommendations. Briefly, reads were trimmed with Trimommatic v.0.36[179] and aligned with BWA-MEM v.0.7.15[180] to the barn owl reference genome (GenBank accession JAEUGV000000000[93]). Base quality score recalibration (BQSR) was performed in GATK v.4.1.3 using high-confidence calls obtained from two independent callers: GATK's HaplotypeCaller and GenotypeGVCF v.4.1.3 and ANGSD v.0.921[181]. Following BQSR, variants were called with GATK's HaplotypeCaller and GenotypeGVCFs v.4.1.3 from the recalibrated bam files. Genotype calls were filtered using GATK and VCFtools[182] if they presented: low individual quality per depth (QD < 5), extreme coverage (600 > DP > 1000), mapping quality (MQ < 40 and MQ > 70), extreme hetero or homozygosity (ExcessHet > 20 and InbreedingCoeff > 0.9) and high read strand bias (FS > 60 and SOR > 3). We filtered further at the level of individual genotype for low quality (GQ < 20) and extreme coverage (GenDP < 10 and GenDP > 40). Lastly, we kept only bi-allelic sites with less than 5% of missing data across individuals resulting in 6'718'804 SNP. For analyses of neutral population structure and demography, an exact Hardy-Weinberg test was used to remove sites that significantly departed ($p<0.05$) from the expected equilibrium using the R[183] package HardyWeinberg[184,185] yielding 6'701'905 SNP with a mean coverage of 22.3X (4.53 SD).

To identify coding regions in the new reference genome, we extracted from NCBI all the annotated genes in the previous barn owl reference genome (GenBank assembly accession GCA_902150015.1[92]). Gene sequences where then mapped on the new assembly using minimap2 v2.8[207] with default parameters. Genes mapped with a quality equal or above 60 were kept for downstream analyses, in order to keep only high confidence coding regions.

Population structure and genetic diversity

To investigate population structure among our samples, sNMF v.1.2[155] was run for K 2 to 5 in 25 replicates to infer individual clustering and admixture proportions. For this analysis, singletons were excluded and the remaining SNPs were pruned for linkage disequilibrium in PLINK v1.946[156] (parameters -indep-pairwise 50 10 0.1) as recommended by the authors, retaining 288'775 SNP. The same dataset was used to perform a Principal Component Analysis (PCA) with the R package SNPRelate[157].

Individual observed heterozygosity and population-specific private alleles were estimated using custom R scripts for each population. Individual-based relatedness (β[159,243]), inbreeding

coefficients for whole genome SNP data (F_{IS} and F_{IT}), overall and population pairwise F_{ST}[159] were calculated with hierfstat v.0.5-9[214].

Demographic history

To investigate the demographic history of the insular populations and potential admixture events we used Treemix[158]. Using the LD-pruned dataset filtered further to include no missing data (228'980 SNP), Treemix was run for 10 replicates with 0 to 6 migration events, rooting the tree on the IS population, given its position on the PCA and the phylogeny (see below).

To explore the space of all possible admixture graphs between our populations, we used the heuristic search implemented in qpBrute[244,245] on the full set of SNPs, rooting the graph on the IS population. This *f4* based software runs a stepwise algorithm to add nodes to the graph, adding admixture events to prevent outliers. If a graph still yields f4 outliers (i.e. $|Z|<3$) it is discarded.

Phylogeny

To infer intraspecific phylogenies along the genome from SNP data, we constructed maximum likelihood phylogenetic trees using the Randomized Axelerated Maximum Likelihood (RAxML) v8.2.12[246], employing a generalized time-reversible (GTR) CAT model with Lewis ascertainment bias correction. First, we inferred genes trees in non-overlapping 100kb windows for all the individuals, including the American outgroup. Then, the produced gene trees were used to infer a concatenated maximum likelihood phylogeny with ASTRAL v5.7.5[247]. To verify that the consensus phylogeny was not driven by a specific region highly similar between the islands (see below and Fig. 2), we inferred a second tree with ASTRAL excluding RAxML windows in Super-Scaffold_1000006.

Detection of genomic regions under selection

Insular vs mainland barn owls

In this study, we aimed to identify genomic regions potentially under selection at two different levels. First, to detect signatures of selection specific to barn owls of the Canary Islands, we grouped insular individuals (EC and WC) and compared them to the mainland ones (PT and IS). A script (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py) by Simon Martin was used to estimate genome wide patterns of relative (F_{ST}) and absolute (d_{xy}) divergence between the insular and mainland groups, and to calculate nucleotide diversity (π) per group, in windows of 100kbp with 20kbp steps.

SNPrelate was used to calculate a pairwise matrix of linkage disequilibrium (r) from SNPs with over 5% minor allelic frequency (MAF), which was then squared to obtain r^2. For plotting, we estimated the mean of non-overlapping 100 SNP windows.

The topology weighting method implemented in Twisst[248], was used to quantify the relationships between the four populations in our dataset and visualize how they change along the genome. Twisst estimated the topology based on trees produced using RAxML in sliding windows (100kb of length, 20kb of slide; not including the outgroup) and estimated the weighting of each taxon topology (defined as the fraction of all unique population sub-trees) per window.

Finally, we calculated population specific F_{ST} from allele sharing matrices (β[159,243]) in sliding windows of 100kbp with 20kbp steps, using hierfstat. For each matrix, we calculated i) the mean allele sharing for pairs of individuals from islands and ii) the mean allele sharing for pairs of individuals one from an island and the other from the mainland from which we obtain the islands specific F_{ST} we shall refer to this estimate as F_{ST} island-specific (F_{ST}^{Can}; see Sup. Fig. 1 for a schematic representation). This method allowed to identify genomic regions of high differentiation exclusively on the islands with no confounding effect from the mainland (see also Weir *et al.*[249])

From the genome wide scans, we identified peaks of differentiation with at least two overlapping windows of F_{ST}^{Can} higher than 5 standard deviations (SD) from the mean (0.394), and extracted the genes in these regions from the list of coding regions of the reference genome identified in a previous section. The gene list was then fed to ShinyGo v0.61[250] to investigate potential enrichment of molecular pathways.

East vs West Canaries

On a second stage, to investigate potential genomic signals of differentiation, putatively linked with ecological adaptation to their distinct niches, we contrasted the two insular lineages (EC against WC). We estimated parwise F_{ST}, d_{xy} and π as for the island-mainland comparison described above.

Then, as for the island-mainland comparison above, we used hierfstat to calculate population specific F_{ST} in sliding windows of 100kbp with 20kbp steps based on a dataset including only the insular individuals. For each island, we consequently identified genomic regions with at least two overlapping windows of pariwise F_{ST} higher than 5 SD from the mean (0.157) and above the 99th quantile of each population's F (F_{ST}^{EC}= 0.399; F_{ST}^{WC} = 0.421) and extracted the genes in these regions as described above. This yielded two genes lists, one per island, which were input to ShinyGo as above.

Climatic niche analysis

To assess whether the barn owl populations of eastern and western Canarias occupy different climatic niches, we used the Outlying Mean Index (OMI) approach[251] as implemented in the R package ade4 v.1.7-16[252]. Observation points for barn owls were compiled from three different sources: Global Biodiversity Information Facility (GBIF), samples sequenced in this study and in [102]. We kept records only from the islands sampled in this study, namely Tenerife (*T. alba*; N=79) and Lanzarote and Fuerteventura (*T. a. g.;* N=34 and N=39, respectively).

Climatic variables for the Canary Islands were extracted from the WorldClim database[164] at 30 sec resolution (approximately 1 km^2) using the R package rbioclim[186]. Redundant variables (correlation of 1) were trimmed and the final model was run with the following 13 variables: Mean Diurnal Range (BIO2), Isothermality (BIO3), Temperature Seasonality (BIO4), Max Temperature of Warmest Month (BIO5), Min Temperature of Coldest Month (BIO6), Temperature Annual Range (BIO7), Mean Temperature of Driest Quarter (BIO9), Annual Precipitation (BIO12), Precipitation of Driest Month (BIO14), Precipitation Seasonality (BIO15), Precipitation of Warmest Quarter (BIO18), Precipitation of Coldest Quarter (BIO19).

Results

Population structure, phylogeny and genetic diversity

The overall F_{ST} in our dataset was 0.0698. Individual ancestry analyses with sNMF distinguished four genetic clusters in our dataset, separating each population into its own cluster (Fig. 1a). Similarly, PCA clustering clearly grouped individuals according to their population (Fig. 1c). The first axis opposed opposes the insular populations to the mainland, as did sNMF K=2 (Sup. Fig. 2), with EC and IS at each extreme. The second axis contrasted the two insular populations EC and WC, as in K=3 (Sup. Fig. 2), and finally the third axis segregated the two mainland populations PT and IS, as in K=4 (Fig. 1a, 1c). Three individuals from EC (one from Fuerteventura and two from Lanzarote) showed small ancestry levels from both WC and PT in sNMF and were placed intermediately on axes 1 and 2 of the PCA. In terms of differentiation, pairwise F_{ST} were the highest between both islands and IS (EC-IS 0.092; WC-IS 0.088) as well as between islands (WC-EC 0.084). PT had the lowest F_{ST} with all other populations in accordance with its central position on the PCA (all below 0.064), the lowest being with IS (PT-IS 0.043).

The consensus phylogenetic tree obtained with RAxML and ASTRAL grouped all individuals from each population together showing four well supported clusters (Fig.1b), despite signs of some incomplete lineage sorting (ILS; final normalized quartet score of 0.41). It positioned IS as the

basal population to all others of the Western Palearctic, followed by the split of PT and finally EC and WC. The tree topology remained the same when the Super-Scaffold_1000006 (see below and Fig. 2) was excluded (Sup. Fig. 3). Treemix yielded a population tree highly resembling the phylogeny and the first PCA axis, with PT splitting from the root IS first, followed by WC and finally EC (Sup. Fig 4). The first migration event was consistently from EC towards IS. Adding more than one migration events to the tree did not improve its fit to the data. The search for admixture graphs with qpBrute yielded three suitable graphs (i.e. no f4 outliers; Sup. Fig. 5). In all three, the root IS was connected directly to PT and EC, while WC was consistently positioned between them.

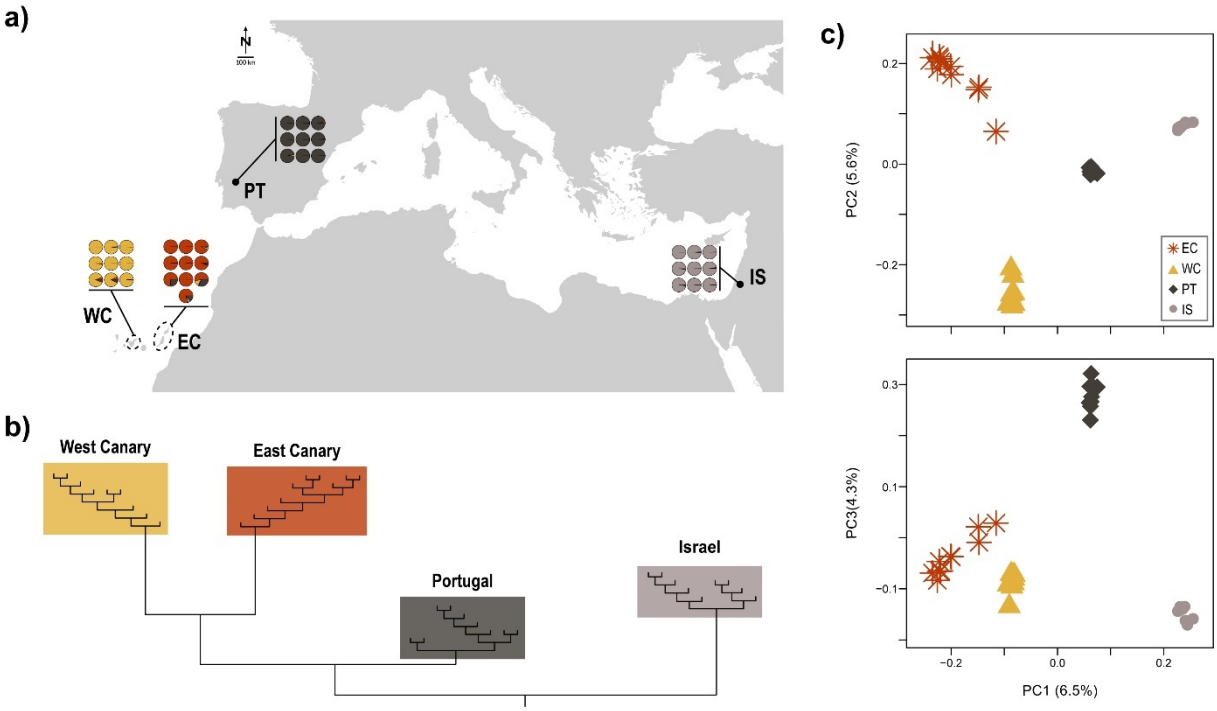


Figure 1 – Population structure of barn owls from the Mediterranean Basin and the Canary Islands. **a)** Individual admixture proportion of each of K=4 lineages as determined by sNMF. Black dots are located at the approximate centroid of each sampled population. Dashed lines encircle the island(s) sampled for each Canarias population. **b)** Maximum-likelihood phylogenetic tree inferred by RAxML and ASTRAL. Within-population branch lengths (inside shaded rectangles) were collapsed and are not to scale. Root position was obtained with an American barn owl as outgroup. **c)** PCA of the 37 individuals. Point shape and colour denote populations according to the legend. Axes one to three are shown and values in parenthesis indicate the percentage of variance explained by each axis.

Overall, mainland populations presented higher genetic diversity than the insular ones (Table 1). Nonetheless, both insular populations showed over 350'000 private polymorphic sites (Table 1). Accordingly, individual relatedness was higher within and between insular populations (Sup. Fig. 6), and PT had the lowest within population relatedness. All populations showed signs of random mating with F_{IS} close to zero but slightly negative as expected of dioecious species [219], while the inbreeding levels of insular barn owls relative to the whole set of populations (F_{IT}) were higher than those on the mainland (Table 1), a reflection of their higher relatedness.

Table 1 – Genetic diversity and population differentiation of barn owls from the Mediterranean Basin and the Canary Islands. Right-hand-side of the table shows the matrix of population pairwise F_{ST}.

Population	Abbrev.	N	#PA	H_O	F_{IS}	F_{IT}	F_{ST}	EC	WC	PT	IS
Eastern Canary	EC	10	405'394	0.155 (0.013)	-0.013 (0.09)	0.10 (0.08)	**EC**		0.084	0.063	0.092
Western Canary	WC	9	357'783	0.158 (0.008)	-0.038 (0.02)	0.08 (0.04)	**WC**	0.084		0.056	0.088
Portugal	PT	9	843'347	0.178 (0.007)	-0.014 (0.04)	-0.03 (0.04)	**PT**	0.063	0.056		0.043
Israel	IS	9	843'807	0.177 (0.003)	-0.025 (0.05)	-0.03 (0.02)	**IS**	0.092	0.088	0.043	

N – number of sampled individuals; #PA – private alleles in each population; H_O – mean observed heterozygosity (SD); F_{IS} - population level inbreeding coefficient (SD); F_{IT}: mean individual inbreeding coefficient relative to the meta-population (SD).

Detection of genomic regions under selection

Island vs mainland

Genomic comparisons of diversity and divergence between insular and mainland barn owls yielded 58 100kb windows of high differentiation including 77 genes (Fig. 2a). F_{ST}^{Can} allowed the detection of regions of high relatedness in the islands, whereas F_{ST} yielded less clear results (Sup. Fig. 7). ShinyGo analyses identified an enrichment of four functional categories related to morphogenesis in humans (Sup. Table 2). The largest enriched pathway – anatomical structure morphogenesis – included 22 of the genes in regions of high genomic differentiation. The remaining three categories – anatomical structure formation involved in morphogenesis, tube morphogenesis and blood vessel morphogenesis – were subsets of the longest, including 14, 11 and 9 of the 22 genes, respectively.

The largest of the peaks, including 21 consecutive windows, encompassed 15Mb of Super-Scaffold_1000006. Of the 77 genes found in peaks of differentiation, 69 were within this region. Here, insular owls showed a strong decrease in relative diversity compared to the mainland (F_{ST}^{Can}), as well as a drop of absolute divergence (d_{xy}) and nucleotide diversity (π). The region showed strong LD (r^2) in insular owls between neighbouring variants compared to the rest of the scaffold (Fig. 2b), which was not the case among continental ones (Sup. Fig. 8). In addition, F_{ST} was higher between island and mainland owls. Twisst showed roughly similar proportions of each tree along the genome, except for this region where there was a higher than average proportion of trees that joined EC and WC (Sup. Fig. 9). Of the 22 genes of the morphogenesis pathway, 17 were found in this closely-linked region of Super_Scaffold_1000006 (25% of the total 69 genes of the region).

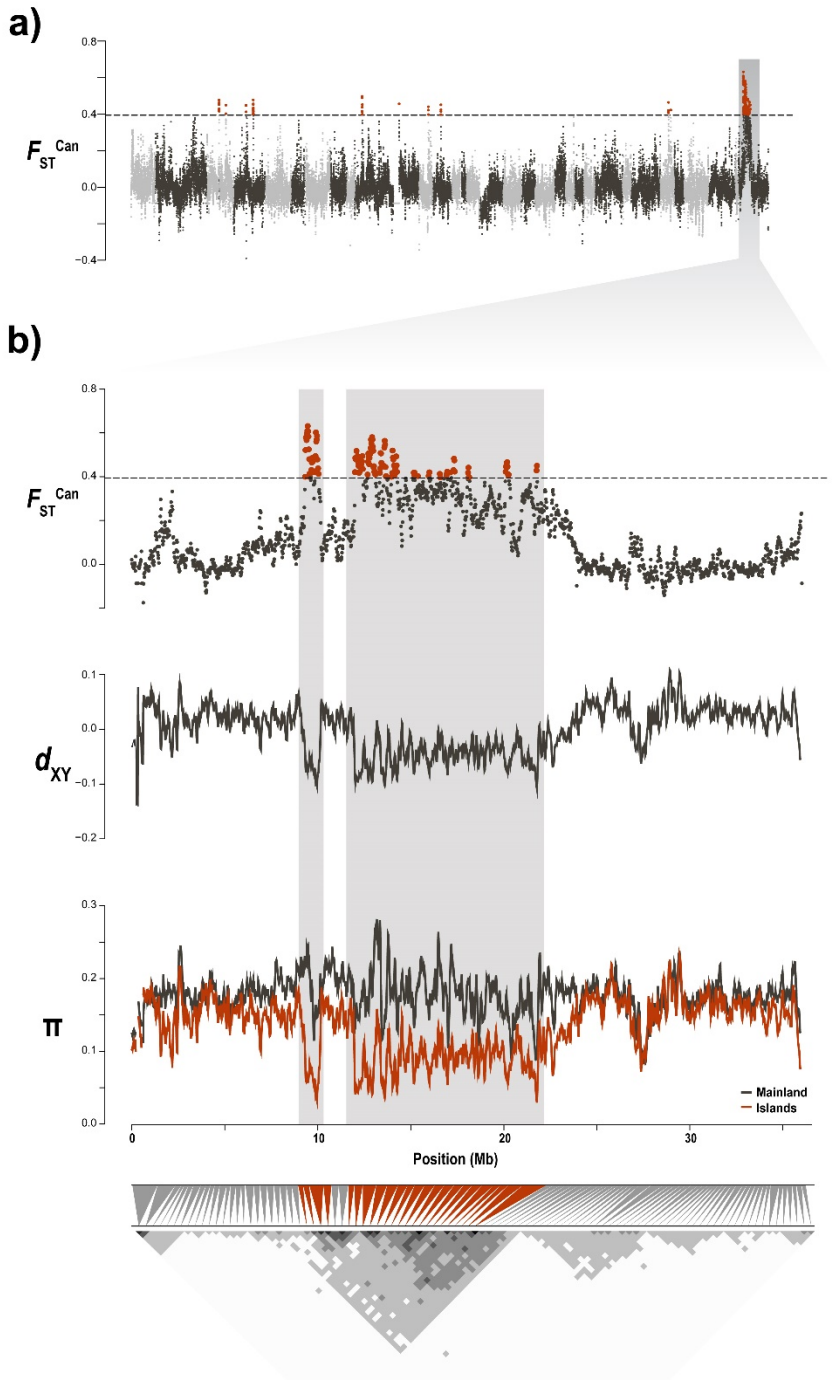


Figure 2 – Genomic landscape of differentiation between insular and mainland barn owls. **a)** Genome-wide F_{ST}^{Can} comparison between individuals from the Canary Islands (EC & WC) and from the mainland (PT & IS). Each dot represents a 100kb window. Dashed line indicates the 5 SD threshold used to identify genomic regions of high differentiation, emphasised in red. Alternating grey colours denote different scaffolds. Shaded vertical bar highlights Scaffold 100006. **b)** Zoom on Scaffold 100006 and, in particular, the ~15 Mb long highly differentiated genomic region (background shading) over windows of 100 kb. From top to bottom, we see in this region a high F_{ST}^{Can} between insular and mainland barn owls, low absolute distance (d_{XY}) between both islands and reduction of nucleotide diversity (π) among insular individuals (red line) compared to the mainland (black line). The bottom triangular matrix shows pairwise LD (r^2) in between groups of 100 SNP along the chromosome. Darker pixels show higher LD. Grey triangles match each pixel in the matrix diagonal to the region it spans on the chromosome above. Red triangles indicate pixels that overlap the region of high differentiation.

East vs West Canary Islands

We detected a total 46 putatively adaptive regions on the islands (21 in EC and 25 in WC), i.e. with high differentiation (F_{ST}) and increased relatedness within each island (F_{ST}^{EC} and F_{ST}^{WC}), from which we obtained two lists of genes. For EC, there were 29 such genes (Sup. Table 3). Enrichment analyses found no link to a specific GO pathway. For WC, we identified less regions, and its list contained 14 genes (Sup. Table 4). ShinyGo detected enrichment of multiple pathways linked to hypoxia and the cellular response to decreased oxygen by two genes (Sup. Table 5).

Climatic niche analysis

The OMI analysis yielded two axes that explained the climatic variability in our study area (Fig. 3b), with the first axis (OMI1) explaining nearly all of it (99.3%). OMI1 was positively correlated with temperature and negatively correlated with precipitation (Sup. Fig. 11). The eastern population *T. a. gracilirostris* occupied a narrow niche of high temperature, low precipitation and low seasonal and daily variability. *T. alba* in Tenerife, occupied a broader niche that covered most of OMI1, including some of the niche of *T. a. gracilirostris*. OMI2 explained little of the variability (0.7%), spreading slightly each subspecies' niche without segregating them.

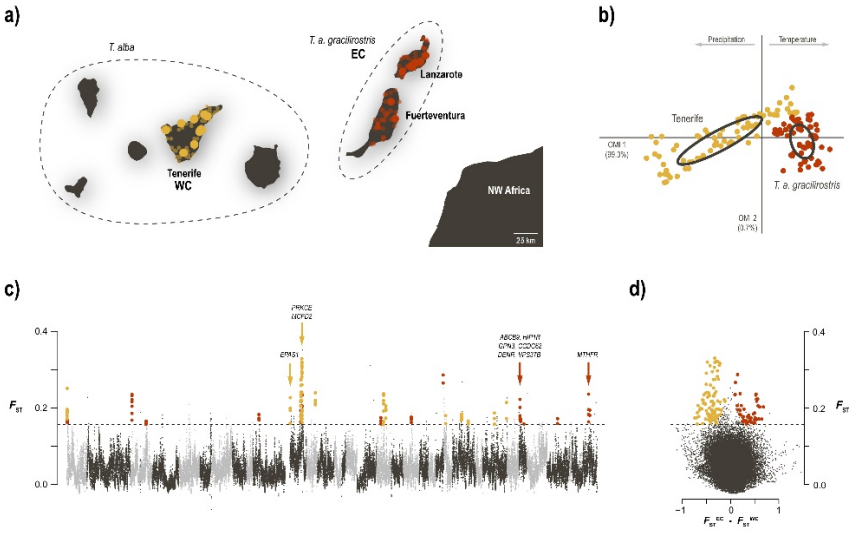


Figure 3 – Ecological divergence in barn owls from the Canary Islands. **a)** Sampling location of western (WC; yellow) and eastern (EC; red) Canary individuals, off the coast of north-western Africa. Sampled islands are named. Large dots indicate individuals sampled for WGS and population genomics analyses; small transparent dots are barn owl observations used in niche analysis. Dashed lines group islands according to barn owl taxonomy. **b)** Climatic niche occupied by the two Canarias populations. The first axis varies from cold and wet to hot arid environments. **c)** Genome-wide F_{ST} comparison between individuals from each Canarias population. Each dot represents a 100kb window. Alternating grey colours denote different chromosomes. Dashed line indicates the 5 SD F_{ST} threshold used to identify genomic regions of high differentiation in combination with F_{β}, highlighted in yellow and red for WC and EC, respectively. Arrows denote the location of genes linked with response to hypoxia in WC; and the cluster of 6 genes linked to morphological ratios and the main gene related to blood pressure in EC. **d)** Same F_{ST} windows and threshold as in **c** plotted against the difference between the two F_{β}, clearly highlight the two sets of windows putatively under local adaptation.

Discussion

Islands offer unique conditions for organisms to adapt and expand their niches. In the Canarias, an archipelago rich in endemics, the barn owl is one of the few raptors present and is thought to have diverged into a subspecies on the easternmost islands. Taking advantage of whole-genome sequences, we first describe the population structure of barn owls in the Mediterranean Basin and the Canary Islands, revealing that the latter are long-standing populations allowing us to address their taxonomic classification. Then, using a new, more sensitive method, we detect 58 putatively locally adapted genomic regions, many of them grouped in a haplotype-like region seemingly under positive selection in the islands. A quarter of the tightly linked genes in it enrich a pathway of anatomical morphogenesis suggesting morphological adaptations to insularity. We also identify genomic regions putatively locally adapted to either the eastern and western islands. For eastern Canarias, the identified genes in regions under positive selection belong to pathways linked to body proportions and blood pressure, consistent with the smaller owl size of this population living in a hot arid climate. For western Canarias, barn owls from Tenerife showed an enrichment of genes related to hypoxia, a potential response to inhabiting an island with a steep elevation gradient.

Barn owls from the Canary Islands and Mediterranean Basin

Our work shows that, while each population has its own unique genetic composition (Fig. 1), barn owls from the Canary Islands are distinct from those on the mainland surrounding the Mediterranean Basin. Indeed, clustering methods constantly opposed insular individuals to mainland ones (PCA axis 1 Fig. 1c; K=2 Sup. Fig. 2). Phylogenetic reconstruction grouped individuals per population, starting with IS, followed by PT and finishing with a monophyletic branch that split between both insular populations (Fig. 1b). Although there was considerable ILS throughout the genome (ASTRAL estimation 0.41; Sup. Fig. 9), as expected in a within-species context with overall low differentiation (F_{ST} 0.07), genomic data clearly provided the necessary resolution to resolve patterns that previous attempts based on a few genes had not[89]. The monophyly of the Canarias populations, which is not driven by selected genomic regions (Sup. Fig. 3), was also supported by the drift-based trees of Treemix. Thus, overall, our current results indicate that the two insular lineages – eastern and western Canarias – have a similar origin, although sampling north-western Africa would provide a more conclusive resolution.

In terms of genetic diversity, as expected the islands were less diverse than the mainland and showed slightly higher levels of inbreeding relative to the whole set[196]. However, they presented nearly 10-fold higher levels of private alleles than those reported for other islands in the Atlantic (British Isles[93]) and in the Mediterranean Sea (Cyprus and Greek islands[253]). Notably, the two

Canarias insular populations, about 165 Km apart, were more distant genetically from each other (in terms of pairwise F_{ST}) than either was from Portugal, over 1000 Km away. Further, they were more distant from each other even than WC was from any population in continental Europe in a recent genomic study[204], in agreement with previous observations with microsatellites[102]. Both the high private diversity and high differentiation can certainly be partially explained by our lack of samples from north-western Africa at the moment, which may be masking any shared polymorphisms between the islands and the nearest mainland as private to the former. However, as we did sample PT and IS, the most diverse populations in the Western Palearctic assumed to meet in northern Africa[204], much of this insular diversity is likely indeed private. Therefore, the Canarias appear to have been colonized much earlier than the other studied insular populations and have thus had the time for *in situ* mutations to accumulate in spite of genetic drift, which is supported by their higher divergence as well.

Insularity

We found multiple evidence of adaptation common to both barn owl lineages of the Canary Islands. To do so, we used a moment based estimator of population specific F_{ST}[159,243] to identify genomic regions with an excess of shared ancestry in all insular individuals relatively to the mean shared ancestry between islands and mainland individuals. Such regions, highly similar in insular individuals, are thus putatively under selection on the islands. This method provides a clearer result than classical pairwise F_{ST} scan (Sup. Fig. 7), as it focuses on the diversity in the target population rather than taking an average over the set of populations. Being a moment estimator, it is also very efficient compared to maximum likelihood or Bayesian estimators (e.g.[254]), and does not rely on the F-model, which assumes independent populations. Effectively, the island specific F_{ST}^{Can} identified regions of increased relatedness between insular individuals or, put more simply, regions in which all insular individuals resemble each other more, compared to the averaged relatedness along the genome. This population specific, moment based and model-free estimator of F_{ST} should be a useful addition to the population genomic toolbox to detect nested signals of local adaptation, especially when there is substructure in the groups one wishes to compare.

The genomic landscape of F_{ST}^{Can} differentiation between insular and mainland owls yielded 58 windows putatively under selection (Fig. 2a). Among these, a particularly large and clear peak of differentiation stood out. This region of approximately 15 Mb in length was highly similar among insular individuals (Sup. Fig. 10) as shown by the accompanying drop in d_{XY} and π (Fig. 2b). Furthermore, the increased linkage between alleles in this region suggests that it is transmitted in a haplotype-fashion. Crucially, the fact that we do not see the slightest surge of LD in mainland individuals confirms that it is not a by-product of a region of low-recombination in this species

(Sup. Fig. 8). Overall, we provide strong evidence of positive selection in this genomic region in Canary Islands owls, suggesting an adaptation to insularity (Fig. 2).

A quarter of the genes in this haplotype (17 out of 69), in conjunction with 5 other genes in potentially adaptative regions, significantly enriched the anatomical structure morphogenesis pathway (Sup. Table 2), a biological process related to the organisation and generation of anatomical structure during development. This suggests positive selection on some morphological trait on insular individuals. Given that there is evidence of gene flow into both islands (see admixed individuals in Fig. 1a, 1c), we propose two hypotheses to explain how selection might act on this haplotype. First, it could confer a significant advantage to individuals carrying it on the island and prevent those that do not carry it from reproducing or surviving. In this scenario, immigrants from the mainland not carrying this haplotype would not reproduce in both islands, while those that did carry it might. In the second scenario, selection would happen on the migrants themselves before reaching the island if, for example, the haplotype facilitates long flight over large spans of water. It is widely accepted that, given how dispersal capacity is highly variable across species, even among birds, some are more prone to colonizing islands than others. Therefore, it is also conceivable that, within a species or population, some individuals are morphologically more predisposed or have better dispersal abilities than others. Since the barn owl generally avoids flying over open water, as demonstrated by its consistently higher differentiation on islands[93,204,253], this appears like a plausible explanation. The absence of phenotypic measurements from the sequenced birds prevents us from establishing a link between phenotypes and genotypes on this data set, and we hence remain cautious on the speculation regarding the functional implications of this haplotype. In future work, a larger cohort should verify the frequencies of this haplotype as we only had 19 insular individuals in this study, and if possible include detailed morphometric measurements to allow a GWAS-like approach. In addition, the potential function of the 75% remaining genes in the region should be identified once a full genome annotation is available.

Ecological divergence

In the Canary archipelago, both the eastern islands and Tenerife have many specific endemic species across multiple taxa. This is generally attributed to their intrinsic characteristics driving ecological speciation namely, the arid and windy conditions of Lanzarote and Fuerteventura, and the elevation gradient of the Teide volcano (up to 3'715m tall) in Tenerife. We quantified the climatic differences between the two environments with a niche analysis based on reported barn owl observations, and show that indeed barn owls occupy significantly different niches on each group of islands (Fig. 3b). In the east, they are found on unvaryingly hot and dry locations, whereas Tenerife covers a wide range of temperature and precipitation (Fig. 3b).

From the genomic data, we found evidence of local adaptation on both insular populations (Fig. 3c, 3d). The eastern population had more genomic regions, and more genes, potentially under selection compared to the west (29 and 14 genes, respectively). Although no significant pathway enrichment was detected in the eastern population, there were two groups of genes with similar functions in the putatively adapted regions. The first group, composed of 12 genes, has significant links to body size and proportions in humans (see Sup. Table 3 and references therein). Among these, a specific set of 6 genes – *HIP1R*, *CCDC62*, *DENR*, *GPN3*, *VPS37B* and *ABCB9* – is tightly clustered in the barn owl genome (Fig. 3c), suggesting they might benefit from linked selection. These have been linked to body height, body-mass-index and other body measurement ratios[255–260], which could be evidence of a genetic determination of the smaller size of barn owls in the eastern population, thus suggesting it is indeed a morphological adaptation to living on these islands. The second group of genes in regions potentially under selection, includes 10 genes related to numerous blood parameters (Sup. Table 3), a similar signal to that seen in chickens adapted to hot arid environments[261]. In particular, the gene *MTHFR* has extensive connections to blood pressure[262–269], a trait known to vary with environmental temperature in mammals[270] and birds[271], potentially suggesting barn owls have circulatory systems adapted to the hot and dry conditions of the eastern Canarias.

In the western population, regions putatively under selection are enriched in genes involved in pathways related to the cellular response to hypoxia (i.e. low levels of oxygen; Sup. Table 5). While this analysis highlighted two genes, *EPAS1* and *PRKCE*, there were three other genes in highly differentiated regions with similar links to red blood cells and haemoglobin density (Sup. Table 4), namely *MCFD2*, *FZD8* and *ZNF512*[255,258,272–274]. Red blood cells, and the haemoglobin within, are responsible for transporting oxygen in the body and are direct targets of selection at high elevation[275]. The gene *EPAS1* in particular, is well known for being involved in adaptation to high altitude environments across vertebrates[276], hinting at an adaptation of barn owls to higher altitude in Tenerife. Indeed, Tenerife is a small island, made even smaller to barn owls by the presence of a colossal peak in the centre since mountains are limiting to this species[171,204]. However, being in a warm climate likely allowed the local population to expand their range by adapting to slightly higher altitudes than elsewhere.

Considering its wide range of distribution, even accounting for phenotypic plasticity, barn owls' capacity to adapt to a variety of prey and environments is unquestionable. As such, it is not surprising to detect signals of local adaption in the Canary Islands. With islands generally being species-poor, the species that do inhabit them adapt to different or broader niches via ecological divergence[34]. This is especially true of volcanic islands that arise isolated and uninhabited, in contrast to those intermittently connected to the mainland and more easily colonized. The community of birds of prey in the Canary archipelago includes less than half the species found in

the nearby mainland[205], likely due to the lack of suitable habitat and/or surface. A by-product of this is the reduction of inter-specific competition, which could have allowed the barn owl to maintain population sizes just large enough on the islands through time for selection to act and potentially expand its niche to better exploit the insular environment. Our results suggest this is happening in parallel on each island (Fig. 3c, 3d), consistent with their different niches (Fig. 3b) and relative genetic isolation, producing two ecomorphs.

Insular subspecies

The eastern islands of Lanzarote and Fuerteventura (Fig. 3a) are home to the barn owl subspecies *T. a. gracilirostris* (Hartert, E, 1905). This classification is based on its smaller size and potentially colouration patterns, although the latter is contested by ornithologists and inconsistent with reported phenotypical measurements[102]. The reduction in size is actually a common pattern in insular barn owls[176], and could be an adaptation to nesting in very small cavities (i.e. old lava hills) and/or to better navigate the strong winds in the eastern islands. The genomic data presented here is consistent with this population forming an endemic subspecies. It has diverged considerably from the mainland, with higher differentiation levels than barn owls from any other studied island in the Western Palearctic[93,253]. Moreover, we show it forms a monophyletic cluster (Fig. 1b) with high levels of private genetic diversity, and multiple genomic regions showing signs of local adaptation (Fig. 3c). While we did not find any specific pathway enrichment that could be linked to its small size, it is worth noting that our current genome annotation is imperfect and thus the extracted list of genes is likely incomplete.

In contrast, barn owls from Tenerife and the remaining islands are considered to belong to the nominal *T. alba* found also on the mainland surrounding the Mediterranean Basin (Fig. 3a). However, it too has considerably diverged from the mainland and shows signs of being locally adapted to, at least, Tenerife and its elevation gradient (see previous section). Furthermore, it clusters with the other insular Canary population rather than the mainland (Fig. 1). While it is not the aim of this study to evaluate what constitutes a subspecies, we provide evidence that the Tenerife population is diverging significantly from its founding population, both neutrally and adaptively, albeit at a slower pace than the eastern population.

The reasons why the eastern population is more divergent than the western, a puzzling fact considering it is closer to the mainland, are not yet fully resolved. Neutral divergence in F_{ST} between these two insular populations suggest they are the result of two independent colonisation events rather than a strict east-to-west progression as described for other taxa[277]. Although the islands themselves emerged from east to west, Tenerife is at least 11 million years old, twice the inferred time of formation of the *T. alba* species[89] and thus available at the time. Nonetheless, an earlier settlement of the eastern islands would have given more time for both

genetic drift and selection to promote divergence. A very small population size in the east, consistent with current census data[278], could account for the stronger drift in a scenario of simultaneous colonization. However, it would strongly hinder local adaptation, making it a less likely hypothesis since we identified more regions putatively under selection in this population. Finally, an earlier colonisation of the eastern islands could explain the link between EC and IS seen in Treemix (Sup. Fig. 4) and qpBrute (Sup. Fig. 5). The former suggested a migration event between these two populations, and the latter always linked EC, but not WC, directly to IS. This could suggest that EC has a trace of north or north-eastern African component in its ancestry, that WC does not. The inclusion of north-western African individuals in these analyses would be particularly useful to fully resolve the origin of the insular populations.

Conclusion

Due to their intrinsic characteristics, islands house numerous endemics making them ideal systems to study the bases of ecological divergence. We provide empirical evidence that both neutral and adaptive evolutionary mechanisms shaped divergence from the mainland in barn owls from the Canary archipelago. Our results show clear signs of genome-wide differentiation (i.e. neutral), a combination of mutations (high private diversity; Table 1) and drift (high F_{ST} and F_{IT}), consistent with theoretical expectations for populations established and isolated long ago despite some admixture. We also identify signals of local adaptation to common insular conditions (Fig. 2), as well as to each island's niche, creating ecomorphs (Fig. 3). While the history and functional effect of the putatively adapted genomic regions identified here deserve further investigation, these observations highlight how selection can still act on small isolated populations. This study illustrates the capacity of a widespread bird to adapt to the local ecological conditions of small islands, an adaptative capacity which may prove essential in facing a changing global climate.

Chapter 3 – Supporting Information

Supplementary Tables

Sup. Table 1 – Description of samples used in this study.

#	Pop	ID	Location	Year	Tissue	Sex	Ref
1	East Canary	EC01	Lanzarote	2010	muscle	Female	1
2	East Canary	EC02	Lanzarote	2010	muscle	Female	1
3	East Canary	EC03	Lanzarote	2012	muscle	Female	1
4	East Canary	EC04	Lanzarote	2012	muscle	Female	1
5	East Canary	EC05	Lanzarote	2012	muscle	Male	1
6	East Canary	EC06	Lanzarote	2012	muscle	Male	1
7	East Canary	EC07	Lanzarote	2007	feather	Male	1
8	East Canary	EC08	Lanzarote	2008	feather	Female	1
9	East Canary	EC09	Fuerteventura	2007	feather	Male	1
10	East Canary	EC10	Fuerteventura	2007	feather	Female	1
11	West Canary	WC01	Tenerife	1905	muscle	Male	2
12	West Canary	WC02	Tenerife	2003	muscle	Female	2
13	West Canary	WC03	Tenerife	2003	muscle	Female	2
14	West Canary	WC04	Tenerife	2003	muscle	Male	2
15	West Canary	WC05	Tenerife	2005	muscle	Male	2
16	West Canary	WC06	Tenerife	2005	muscle	Female	2
17	West Canary	WC07	Tenerife	2006	muscle	Male	2
18	West Canary	WC08	Tenerife	2006	muscle	Male	2
19	West Canary	WC09	Tenerife	2006	muscle	Male	2
20	Portugal	PT01	Pombal	2013	feather	Female	3
21	Portugal	PT02	Coruche	2013	feather	Male	3
22	Portugal	PT03	Évora	2013	feather	Male	3
23	Portugal	PT04	Coruche	2012	feather	Female	3
24	Portugal	PT05	Nazaré	2013	feather	Female	3
25	Portugal	PT06	Porto de Moós	2013	feather	Female	3
26	Portugal	PT07	Setúbal	2012	feather	Female	3
27	Portugal	PT08	Fátima	2013	feather	Female	3
28	Portugal	PT09	Santarém	2013	feather	Male	3
29	Israel	IS01	Lachish	2005	blood	Female	2
30	Israel	IS02	Beit Shean	2005	blood	Male	2
31	Israel	IS03	Hula	2005	blood	Female	2
32	Israel	IS04	Beit Shean	2005	blood	Female	2
33	Israel	IS05	Beit Shean	2005	blood	Male	2
34	Israel	IS06	Beit Shean	2005	blood	Female	2
35	Israel	IS07	Beit Shean	2005	blood	Female	2
36	Israel	IS08	Hula	2005	blood	Female	2
37	Israel	IS09	Hula	2005	blood	Female	2
38	Syngapore	SGP	Singapore	2013	soft tissue	Male	3
39	USA	USA	San Diego, California	2015	soft tissue	Female	3

[1] this study
[2] GenBank BioProject PRJNA727977 [204]
[3] GenBank BioProject PRJNA700797 [93]

Sup. Table 2 – ShinyGo pathway enrichment results for genes in putatively adapted genomics regions in the insular Canarias populations. All listed genes are located in the haplotype-like region in Figure 2, except those marked with †.

Enrichment FDR	Genes in list	Total genes	Functional Category	Genes
0.0271	14	1164	Anatomical structure formation involved in morphogenesis	GNG5 ANXA2† F3 RORA† VAV3 BCL10 CCN1 PRKACB DDAH1 WDR72† S1PR1 COL11A1 ADAM12† TGFBR3
0.0406	22	2785	Anatomical structure morphogenesis	PLPPR4 NEXN GNG5 ANXA2† TGFBR3 F3 CCN1 COL11A1 RORA† PALMD OLFM3 VAV3 BCAR3 BCL10 PRKACB BARHL2 DDAH1 NTNG1 WDR72† S1PR1 ADAM12† FGD5†
0.041	11	860	Tube morphogenesis	ANXA2† F3 VAV3 BCL10 CCN1 PRKACB DDAH1 S1PR1 RORA† ADAM12† TGFBR3
0.0419	9	603	Blood vessel morphogenesis	ANXA2† F3 VAV3 CCN1 DDAH1 S1PR1 RORA† ADAM12† TGFBR3

Sup. Table 3 – List of genes in putatively adapted genomic regions in the Eastern Canary population, grouped per location on the barn owl genome. Associated phenotypes are provided for each gene when available. Human-specific behavioural phenotypes are not reported (for example, alcohol consumption). Genes marked with † have phenotypes related to body size and proportions, and ‡ with blood parameters.

Super-Scaffold	Gene	Phenotype	References
1	MGMT†	Body height	255,260
3	LRTM1†	Body height	255,279
5	LAS1L	-	
	ZC3H12B	Alopecia	280
	HEPH	Alopecia	280
16	DAAM2†‡	Body height	255,281
		Platelet count	274
	MOCS1†‡	Body height	255
		Platelet count	274
23	MMAA‡	Hemoglobin measurement	258,282
	SMAD1‡	Hemoglobin measurement	258,282

44	PITPNM2‡	Platelet count	283
		Reticulocyte count	258
		BMI	255
	ARL6IP4	-	
	OGFOD2	-	
	ABCB9†	Waist-hip ratio	255
		BMI-adjusted waist-hip ratio	256
	VPS37B†	BMI-adjusted waist-hip ratio	256
	HIP1R†‡	BMI	257
		BMI-adjusted waist-hip ratio	256
		Waist-hip ratio	255
		Body height	255
		Platelet count	258
	CCDC62†‡	BMI-adjusted waist-hip ratio	256
		Platelet count	258
	DENR†‡	Mean corpuscular volume	258
		BMI-adjusted waist-hip ratio	256
		BMI	256
	GPN3†	BMI	255,256,259
		Body height	260
	ARPC3	-	
	ANAPC7‡	QT interval	284
20000042	P3H1	-	
	CLDN19	-	
	YBX1	-	
	PPIH	-	
	CCDC30‡	Systolic blood pressure	285
		Pulse pressure	285,286
	PPCS	-	
	KIAA2013†	Waist-hip ratio	279
	NPPC†	Body height	255,287–291
		BMI-adjusted waist-hip ratio	256
	MTHFR‡	Mean arterial pressure	267
		Diastolic blood pressure	262,265,266,268
		Systolic blood pressure	262–266
		Pulse pressure	265
		Platelet count	258,274
		Blood pressure	262
		Hypertension	269
		Mean corpuscular volume	255,258,274
		Erythrocyte count	255,258

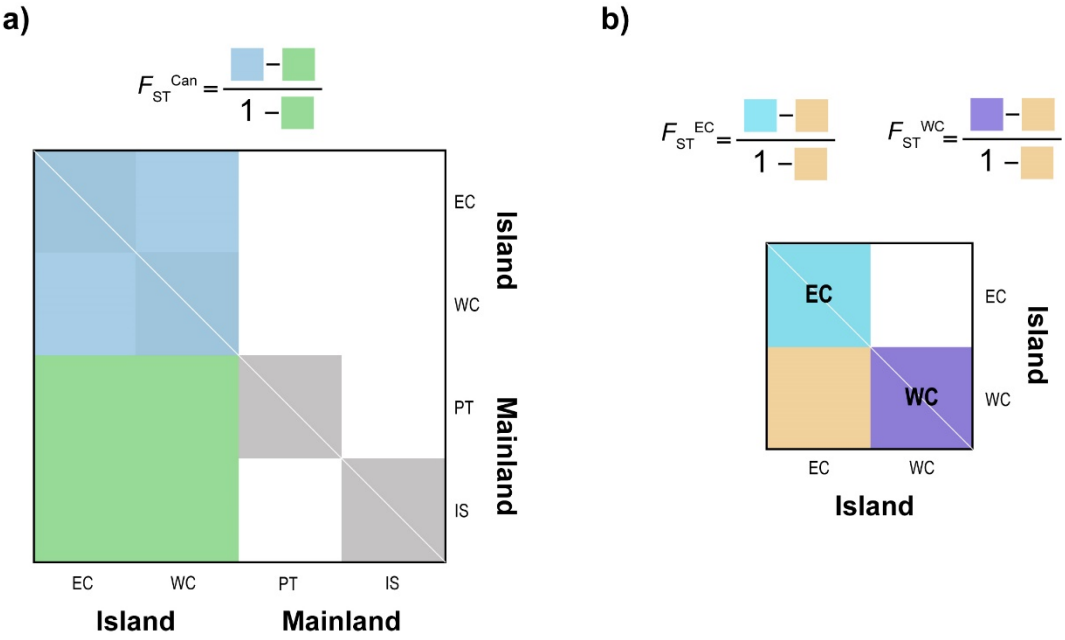
Sup. Table 4 – List of genes in putatively adapted genomic regions in the Western Canary population, grouped per location on the barn owl genome. Associated phenotypes are provided for each gene when available. Human-specific behavioural phenotypes are not reported (for example, alcohol consumption). Genes marked with † are linked to red blood cells and haemoglobin measurements, and with ‡ to other blood parameters.

Super-Scaffold	Gene	Phenotype	References
16	CAMKMT	Body height	255,288,290
	TMEM247	-	
	EPAS1†	Erythrocyte count	255
		Hematocrit	258
		Hemoglobin measurement	258,292
		High altitude adaptation	293,294
		PR interval	295
	PRKCE†	Hematocrit	258,272,274,296–300
		Erythrocyte count	255,258,272,274,297,298,300,301
		Hemoglobin measurement	258,272–274,296–300
	MCFD2†	Hematocrit	258,272
		Erythrocyte count	258,272
		Hemoglobin measurement	258,272,273
		Body height	255,279
	LRFN2	BMI	255,256,259
38	NXPH1	-	
	GJD4‡	PR interval	302
	FZD8†	Hematocrit	274
		Erythrocyte count	258,274
		Hemoglobin measurement	255,258
40	ZNF512†	Erythrocyte count	255
		Hair colour	255,303
	FNDC4	-	
	IFT172	Hair colour	255
	KRTCAP3‡	Systolic blood pressure	255
	NRBP1	Hair colour	255

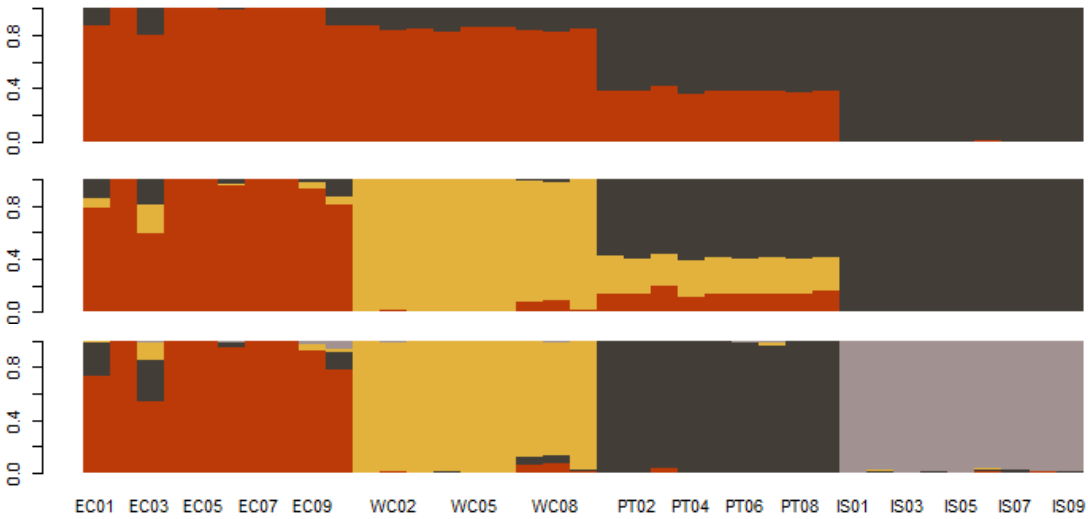
Sup. Table 5– ShinyGo pathway enrichment results for genes in putatively adapted genomics regions in the Western Canary population.

Enrichment FDR	Genes in list	Total genes	Functional Category	Genes
0.0394	2	66	Cellular response to hypoxia	EPAS1 PRKCE
0.0394	2	67	Cellular response to decreased oxygen levels	EPAS1 PRKCE
0.0394	2	78	Cellular response to oxygen levels	EPAS1 PRKCE

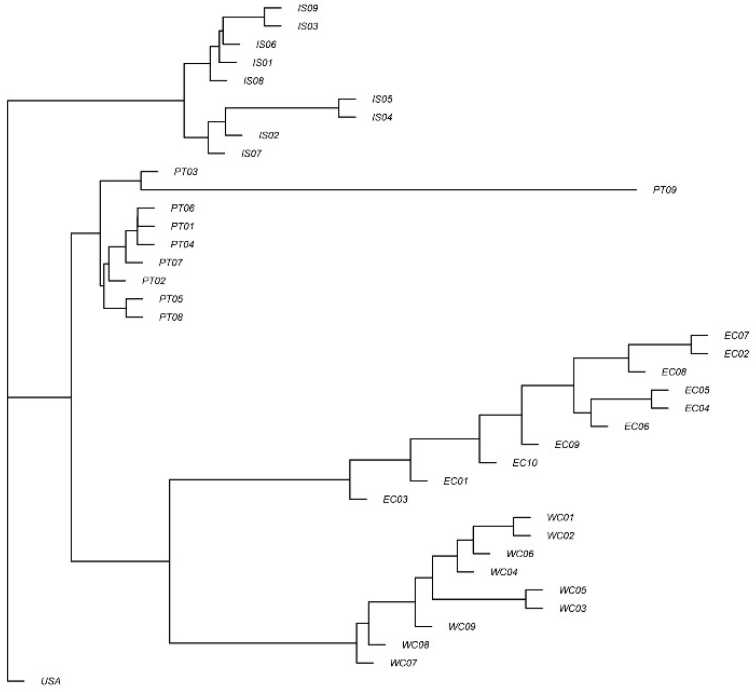
Supplementary Figures



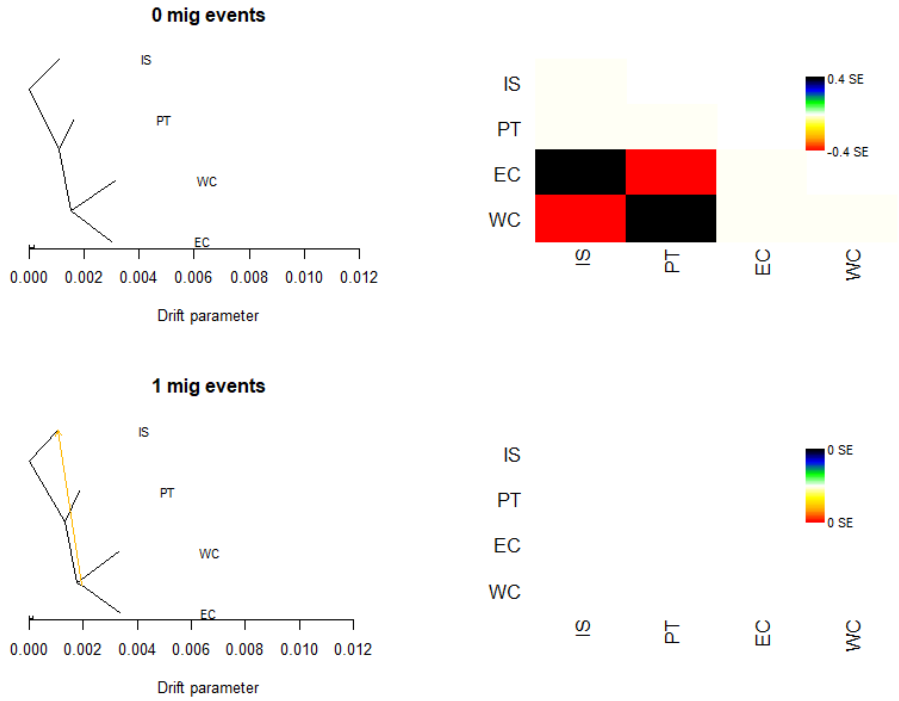
Sup. Fig. 1 – Graphical representation of the calculation of *F*β from individual relatedness matrices to compare **a)** insular and mainland individuals and **b)** individuals from each island. Coloured squares in formulas on top, represent the mean of the same-colour section of the matrix underneath, calculated in windows of 100kb along the genome. Note that in **b)** the matrix is smaller as it does not include any owl from the mainland. In both matrices, the diagonal, in white, is empty.



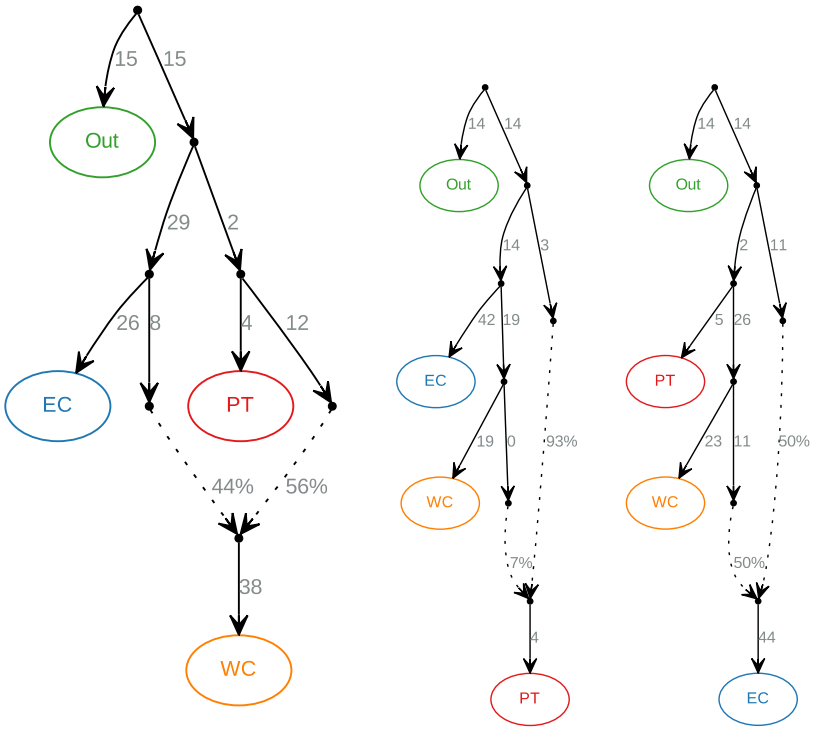
Sup. Fig. 2 – Individual clustering estimated by sNMF for K 2 to 4 lineages. Each vertical bar represents one individual, and the colours represent the relative contributions of each genetic lineage.



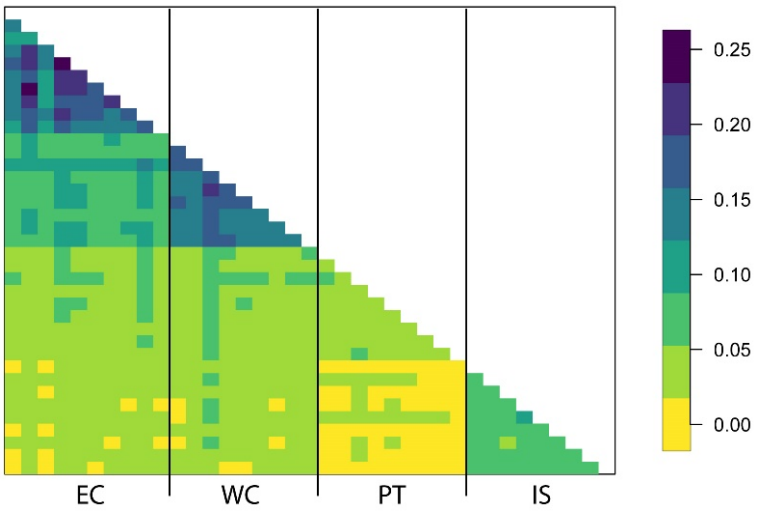
Sup. Fig. 3– Consensus phylogeny produced with ASTRAL, excluding the Super-Scaffold_10006. Note that the topology is the same as in Figure 1b with all the genome.



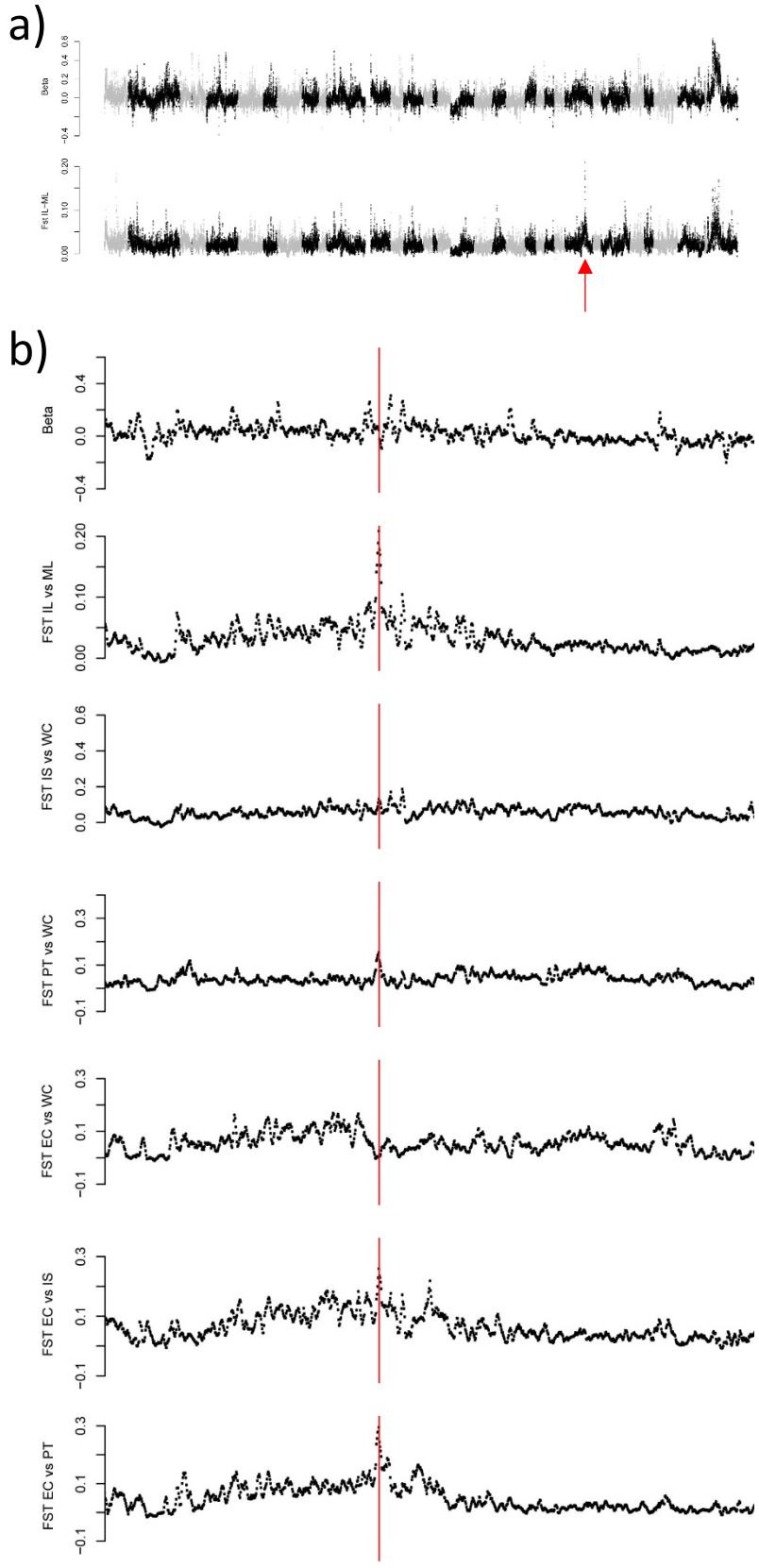
Sup. Fig. 4 – Treemix analysis with 0 and 1 migration events, and their residual matrices. Note that for 1 migration event, the residual error is 0 and thus treemix was unable to add more than one migration events to the tree.



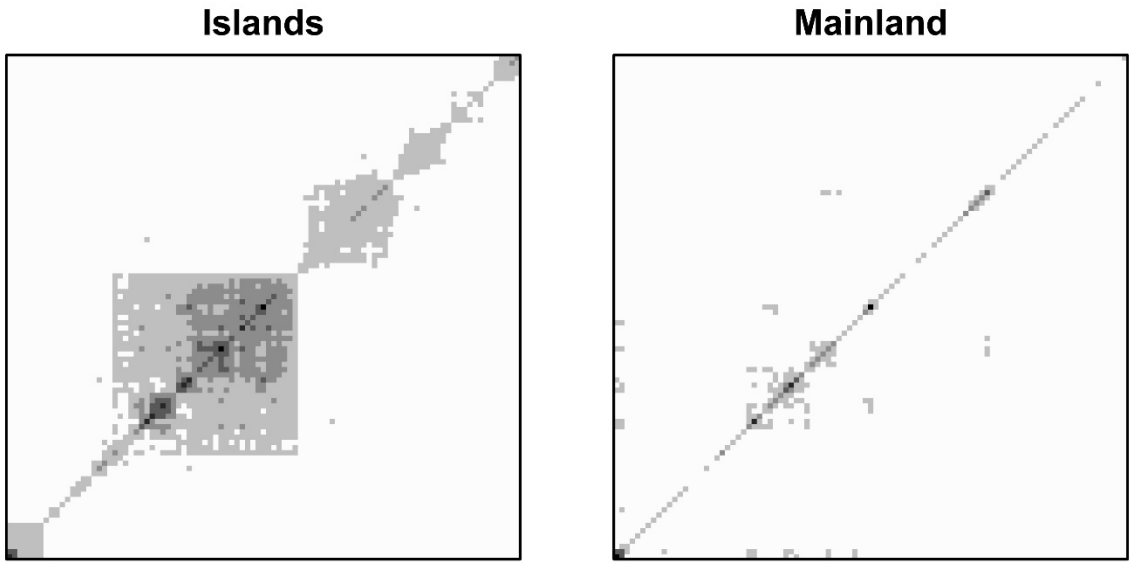
Sup. Fig. 5 – Admixture graphs produced by qpBrute that significantly explain the history of barn owls of the Canary archipelago. The first, bigger graph was the most likely. In all three, note that EC is directly linked to IS while WC is not.



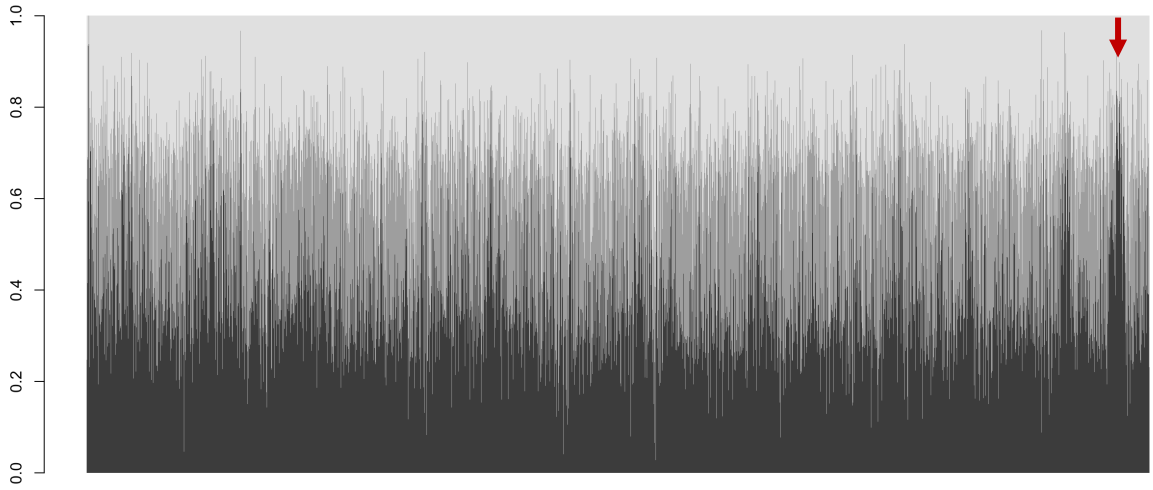
Sup. Fig. 6 – Pairwise individual relatedness (β) heatmap between all individuals.



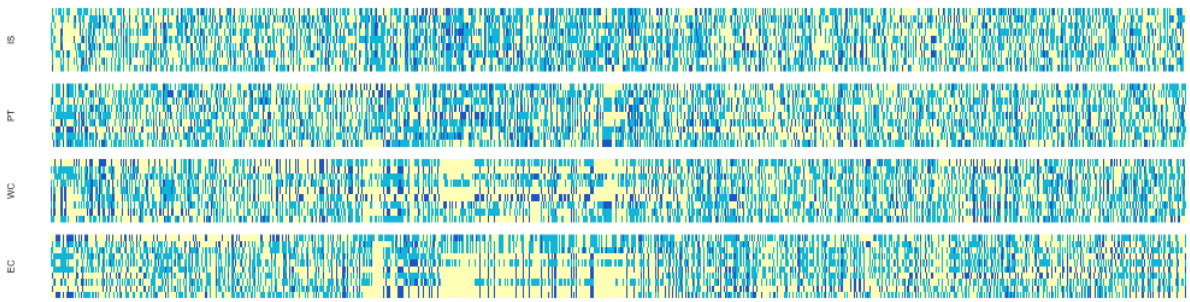
Sup. Fig. 7 – Comparison of F_{ST}^{Can} and F_{ST} scans for **a)** the whole genome and **b)** a zoom on the highest F_{ST} peak (indicated in red). Note that this peak is actually just a specificity of EC and not an overall difference between insular and mainland birds. F_{ST}^{Can} did not produce such a peak.



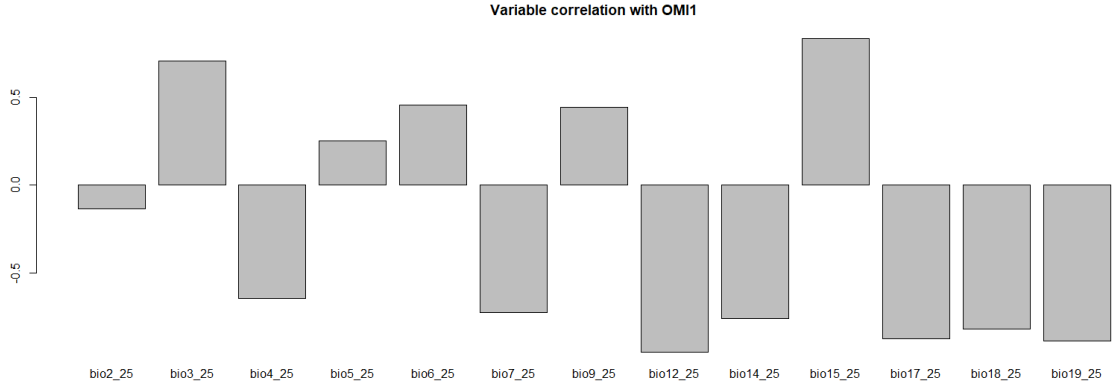
Sup. Fig. 8 – Matrices of pairwise Linkage Disequilibrium (LD; r^2) along Scaffold 10006 of barn owls from Islands and from Mainland. Each pixel shows the average r^2 of 100 SNP, with higher values being of darker colour. The Island matrix clearly shows a region of amplified LD that the Mainland does not. This region corresponds to the highly differentiated segment shown in Figure 2.



Sup. Fig. 9 – Twisst tree weighting output along the genome. At each 100kb window, the proportion explained by each tree topology in shown in shades of grey. Red arrow indicates the haplotype-like region selected on the islands (Figure 2). Note the increased proportion of dark grey trees, in which both island populations are placed in the terminal monophyletic branch, showing increased convergence among insular individuals in this genomic region.



Sup. Fig. 10 – Allele dosage representation along Scaffold 10006 of barn owls from the Canary Islands and the Mediterranean Basin. Each vertical bar represents the genotype at one biallelic SNP. In yellow, homozygote for one allele; light blue, heterozygote; dark blue, homozygote for the other allele. The highly differentiated region shown in Figure 2 is visible with in the EC and WC populations but not in the others, with clear long yellow segments.



Sup. Fig. 11 – Correlation of bioclim climatic variables with OMI1, the first axis of niche variance.

Chapter 4

Landscape and climatic variations of the Quaternary shaped secondary contacts among barn owls (*Tyto alba*) of the Western Palearctic

Tristan Cumer[a], Ana Paula Machado[a], Guillaume Dumont[a], Vasileios Bontzorlos[k,l], Renato Ceccherelli[m], Motti Charter[n,o], Klaus Dichmann[d], Hans-Dieter Martens[g], Nikos Kassinis[p], Rui Lourenço[e], Francesca Manzia[r], Kristijan Ovari[t], Laure Prévost[h], Marko Rakovic[u], Felipe Siverio[s], Alexandre Roulin[a†], Jérôme Goudet[a,j†]

† co-senior authors

Status

Ready for submission, awaiting on confirmation from some co-authors.

Author contributions

TC, APM, AR, JG designed this study; GD and APM produced whole-genome resequencing libraries and called the variants; TC and APM conducted the analyses; KD, RL, JL, HDM, PB, VB, MC, KD, HDM, NK, RL, FM, KO, LP, MR and FS provided samples to the study; TC led the writing with APM with input from all authors.

Abstract

The combined actions of climatic variations and landscape barriers shape the history of natural populations. When organisms follow their shifting niches, obstacles in the landscape can lead to the splitting of populations, on which evolution will then act independently. When two such populations are reunited, secondary contact occurs in a broad range of admixture patterns, from narrow hybrid zones to the complete dissolution of lineages. A previous study suggested that barn owls colonized the Western Palearctic after the last glaciation in a ring-like fashion around the Mediterranean Sea, and conjectured an admixture zone in the Balkans. Here, we take advantage of whole-genome sequences of 94 individuals across the Western Palearctic to reveal the complex history of the species in the region using observational and modelling approaches. Even though our results confirm that two distinct lineages colonized the region, one in Europe and one in the Levant, they suggest that it predates the last glaciation and identify a narrow secondary contact zone between the two in Anatolia. Nonetheless, we also show that barn owls re-colonized Europe after the glaciation from two distinct glacial refugia: a western one in Iberia and an eastern one in Italy. Both glacial lineages now communicate via eastern Europe, in a wide and permeable contact zone. This complex history of populations enlightens the taxonomy of *Tyto alba* in the region, highlights the key role played by mountain ranges and large water bodies as barriers and illustrates the power of population genomics in uncovering intricate demographic patterns.

Keywords

Demographic modelling; glacial refugium; Haplotypes; Population genomics; postglacial recolonization; Whole-genome resequencing

Introduction

Species distribution patterns fluctuate in response to climatic variations, as populations relocate to follow their shifting niches[304]. When organisms colonize new areas, obstacles in the landscape may lead populations to split with varying degrees of geographic isolation. Evolution, via mutation, drift, local adaptation and gene flow, will then act independently on each of the isolated populations. If two allopatric populations are later geographically reconnected, after a certain amount of time and divergence, it can create a secondary contact zone. For example, when populations on both sides of an obstacle meet at the end of it in a ring-like fashion, as described in birds[305], amphibians[306,307], or plants[308]. Likewise, climate oscillations can lead to cyclical isolation and secondary contacts between populations as regional suitability varies[309].

This complex interplay between climatic variations and landscape has been extensively studied in the Western Palearctic[121,123], specifically in light of the cycles of glacial and interglacial periods that characterized the Quaternary[310]. During the last glaciation, colder temperatures and the expansion of the ice sheets in the north rendered large areas unsuitable for many species, which led them to follow their niches southward. Species found refuge in the Mediterranean peninsulas and in northern Africa where climatic conditions were more amenable, forming isolated populations. At the end of the last glaciation maximum (i.e. approximatively 20k years ago[311]), this process was reversed as the climate warmed and the melting of the continental ice caps exposed free land that could be recolonized. An extensive literature addressing the post-glacial history of European organisms describes how the complex landscape of the continent, combined with the distribution of species during the glaciation, conditioned their recolonization processes[121,312]. However, the low-resolution of genetic data used before the genomic era was often insufficient to resolve the intricate and often fine-scale evolutionary processes that occurred during recolonization.

Rapid development of high-throughput sequencing technologies and corresponding methodological tools during the last decades has opened new avenues to study natural populations with high precision. In particular, it has allowed biologists to reconstruct the evolutionary history of species and highlight the diversity of processes acting when populations or subspecies interact in secondary contact. These processes have been found to result in a variety of situations. While the prolonged isolation of populations may lead to allopatric speciation (many examples in plants[313,314], amphibian[306], insects[315], mammals[316] and birds[317]), secondary contact tends to show a broad range of admixture patterns. When admixture occurs, it may vary from narrow hybrid zones between lineages[84], to the complete dissolution of a lineage[318], through a gradual level of admixture along a gradient of mixing populations[319].

Microsatellite and mitochondrial suggested that the barn owls (*Tyto alba*), a non-migratory raptor, colonized the Western Palearctic in a ring-like fashion around the Mediterranean Sea after the

last glaciation[102]. Under this scenario, a postglacial expansion from the glacial refugium in the Iberian Peninsula to northern Europe formed the western branch of the ring, with the eastern branch present across the Levant and Anatolia. While these observations led to conjecture of a potential admixture zone in the Balkans, the available data at the time combined with the overall low genetic differentiation in this species did not allow to fully resolve this question. Moreover, the peculiar genetic makeup of populations in the presumed contact zone brought into question the possibility of a cryptic glacial refugium in the eastern Mediterranean peninsulas.

The difficulty in resolving the post glacial expansion of barn owls is mirrored by their convoluted taxonomy in the Western Palearctic. In this region, *Tyto alba* is classified into different subspecies based on geography and plumage coloration. First, *T. a. erlangeri* (Sclater, WL, 1921) reported in Crete, Cyprus and Middle East, may match the Levant lineage. Second, *T. a. alba* (Scopoli, 1769) is white-coloured and supposedly present in western Europe and western Canary Islands, and could represent the western arm of the ring colonization. *T. a. guttata* (Brehm, CL, 1831), the third subspecies is a dark rufous morph allegedly found in Northern and Eastern Europe, in the Balkans and around the Aegean Sea. This taxonomy does not match any known genetic lineage identified so far and overlaps with the area where admixture between the two lineages that colonized Europe supposedly happens, making the presence of a subspecies in this area puzzling in light of the history known so far.

Here, taking advantage of whole genome sequences of 94 individuals from all around continental Europe and the Mediterranean Sea, we elucidate the demographic history of barn owls in the Western Palearctic. Combining descriptive and modelling approaches based on genomic and ecological data, we identify how the climatic variations and landscape of the region shaped the history of this species. We also investigate how previously isolated populations of barn owls interact at secondary contacts between different lineages, and discuss the convoluted taxonomy with regards to their history.

Material and methods

Samples and data preparation

Sampling, Molecular and sequencing methods

The whole genomes of 96 individual barn owls (*Tyto alba*) were used in this study (Sup. Table 1): 94 individuals were sampled in 11 Western Palearctic localities: Canary (Tenerife island - WC), Portugal (PT), France (FR), Switzerland (CH), Denmark (DK), Serbia (SB), Greece (GR), Italy (IT), Aegean islands (AE), Cyprus (CY) and Israel (IS). In addition, one Eastern (*Tyto javanica* from

Singapore) and one American barn owl (*Tyto furcata* from California, USA) were used as outgroups (Sup. Table 1). Illumina whole-genome sequences of individuals from PT, FR, CH, DK and the outgroups were obtained from the GenBank repository (BioProject PRJNA700797). For the remaining 61 individuals, we followed a similar library preparation and sequencing protocol as outlined in Machado *et al.*[93]. Briefly, genomic DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany), and individually tagged. 100bp TruSeq DNA PCR-free libraries (Illumina) were prepared according to manufacturer's instructions. Whole-genome resequencing was performed on multiplexed libraries with Illumina HiSeq 2500 PE high-throughput sequencing at the Lausanne Genomic Technologies Facility (GTF, University of Lausanne, Switzerland).

Data processing, SNP calling and technical filtering

The bioinformatics pipeline used to obtain analysis-ready SNPs was adapted from the Genome Analysis Toolkit (GATK) Best Practices[154] to a non-model organism following the developers' instructions, as in Machado *et al.*[93]. Raw reads were trimmed with Trimommatic v.0.36[179] and aligned to the reference barn owl genome[93] with BWA-MEM v.0.7.15[180]. Base quality score recalibration (BQSR) was performed using high-confidence calls obtained from two independent callers – GATK's HaplotypeCaller and GenotypeGVCF v.4.1.3 and ANGSD v.0.921[181] – as a set of "true variants" in GATK v.4.1.3.

Genotype calls were filtered for analyses using a hard-filtering approach as proposed for non-model organisms, using GATK and VCFtools[182]. Calls were removed if they presented: low individual quality per depth (QD < 5), extreme coverage (1100 > DP > 2500), mapping quality (MQ < 40 and MQ > 70), extreme hetero or homozygosity (ExcessHet > 20 and InbreedingCoeff > 0.9) and high read strand bias (FS > 60 and SOR > 3). Then, we removed calls for which up to 5% of genotypes had low quality (GQ < 20) and extreme coverage (GenDP < 10 and GenDP > 40). We kept only bi-allelic sites, excluded SNPs on the heterozome (Super scaffolds 13 and 42 [93]) and an exact Hardy-Weinberg test was used to remove sites that significantly departed (p=0.05) from the expected equilibrium using the package HardyWeinberg[184,185] in R[183], yielding a dataset of 6'448'521 SNP (mean individuals' coverage: 19.99X (sd: 4.38)). Lastly, we discarded singletons (minimum allelic count (mac) <2), yielding to a total of 5'151'169 SNP for the population genomic analyses.

SNP phasing and quality control

The set of 6'448'521 variants was phased in two steps. First, individual variants were phased using a read-based approach in which reads covering multiple heterozygous sites were used to resolve local haplotypes. To do so, WhatsHap v1.0[320] was run independently for each individual with default parameters. Secondly, variants were statistically phased with Shape-It v4.1.2[321]. This algorithm integrates local individual phase and applies an approach based on coalescence and recombination to statistically phase haplotypes and impute missing data. Shape-It was run following the manual instructions for a better accuracy: the number of conditioning neighbours in the PBWT was set to 8, and the MCMC chain was run with 10 burn-in generations, 5 pruning iterations, each separated by 1 burn-in iteration, and 10 main iterations.

To assess the quality of the phasing, we examined phase accuracy by using the switch-error-rate metric[322]. When comparing two phasing for an individual's variants, a switch error occurs when a heterozygous site has its phase switched relative to that of the previous heterozygous site. Thus, for each individual, we compared the true local phasing inferred form the read-based approach (WhatsHap) and the statistical phasing of this individual's variants statistically phased by Shape-It, with read-based phase information ignored only for the individual considered (same version and parameters than in the paragraph above). The final estimation of the switch error rate was done using the swithError code to compare both phasing sets (custom script by O. Delaneau https://github.com/SPG-group/switchError) (Sup. Fig. 1).

History of barn owls around the Mediterranean Sea

Population Structure and Genetic Diversity

In order to investigate population structure among our samples, sNMF[155] was run for a number of clusters K ranging from 1 to 10 with 25 replicates for each K to infer individual clustering and admixture proportions. For this analysis, SNPs were pruned for linkage disequilibrium in PLINK v1.946[156] (parameters --indep-pairwise 50 10 0.1) as recommended by the authors, yielding 594,355 SNP. Treemix[158] was used to calculate a drift-based tree of our populations, using this LD-pruned dataset. To detect admixture events between populations, 10 Treemix replicates were run for 0 to 10 migration events, with the tree rooted on the WC population, representative of a non-admixing population (see results).

Population expected and observed heterozygosity, population-specific private alleles, population-specific rare alleles (mac<5) and population-specific total number of polymorphic sites were estimated using custom R scripts on the 5'151'169 variants dataset. To account for differences in sample sizes, which ranges from 4 to 10, population-specific statistics were calculated by randomly sampling 5 individuals from the larger populations (all except FR and SB) 10 times in a

bootstrap-fashion and estimating the mean and standard deviation (SD). Individual-based relatedness (β)[159] and inbreeding coefficient for SNP data were calculated with the R package SNPRelate[157]. Overall F_{ST}, population pairwise F_{ST} and population specific F_{ST}[159] were computed with the hierfstat package v.0.5-9[323]. Confidence intervals for population specific F_{ST} were computed by dividing the SNPs into 100 blocs, and bootstrapping 100 times 100 blocks with replacement. Finally, Principal Component Analyses (PCA) were also performed with the R package SNPRelate, first with all individuals and second only with the 66 European ones (excluding WC, CY and IS).

Haplotype sharing

To measure shared ancestry in the recent past between individuals, we ran fineSTRUCTURE[324] on the phased dataset including all individuals. For this analysis we initially modelled haplotype sharing between individuals using ChromoPainter to generate a co-ancestry matrix, which records the expected number of haplotypes chunks each individual donates to another. For this ChromoPainter step, we converted phased haps files to chromopainter phase files using the impute2chromopainter.pl script provided at http://www.paintmychromosomes.com and generated a uniform recombination map with the makeuniformrecfile.pl script. Using the version of ChromoPainter built into fineSTRUCTURE v.2.0.8, we performed 10 EM iterations to estimate the Ne and Mu parameters (switch rate and mutation rate). The model was then run using the estimated values for these parameters (respectively 570.761 for Ne and 0.0074240 for Mu), and we used default settings to paint all individuals by all others (-a 0 0). We ran fineSTRUCTURE's MCMC model on the co-ancestry matrix for 500'000 burn-in and 500'000 sampling iterations, sampling every 10'000 iterations to determine the grouping of samples with the best posterior probability.

Modelling of history of European barn owl

Maximum-likelihood demographic inference

Data preparation

To describe the history of barn owls in Europe, we modelled five different demographic scenarios using fastsimcoal2[75,166]. Given the position of Italy on the PCA (Fig. 1e), its high F_{ST} and lower haplotype sharing with the rest of European populations (Fig. 2), we tested in particular whether it could have been a cryptic glacial refugium during the last glaciation. To focus on European populations and due to computational constraints, we simplified the dataset to model the history

of four populations of eight individuals (Sup. Table 1): PT as representatives of the known refugium in the Iberian Peninsula[100]; IT and GR representing the peninsulas of Italy and Balkans, respectively; and CH, a product of the recolonization of northern Europe from the Iberian refugium[100].

Autosomal SNPs were filtered to retain only neutrally evolving regions by excluding SNPs found in genic regions and CpG mutations[192]. To achieve homogeneity among SNPs, we removed all sites with missing data and excluded positions with a coverage outside two thirds of the standard deviation of the mean. We employed a parsimony approach based on the Tytonidae phylogenetic tree[89] to determine the ancestral state of the SNPs using the genomes of the two outgroups. Sites for which it was impossible to attribute a state based on the available outgroups were discarded. The remaining 770'718 SNPs were used to produce population pairwise site frequency spectra (SFS).

Demographic scenarios and parameters

Five different scenarios were tested to model the history of barn owls in continental Europe, with a special focus on the period since the last glaciation (Fig. 2; Sup. Fig. 8). Three models included only one refugium in the Iberian Peninsula and various possibilities of colonization scenarios, thus excluding the persistence of barn owls in a second refugium during the glaciation. The two last models included two refugia during the LGM, a western refugium in the Iberian Peninsula and an eastern refugium, in the Italian peninsula.

The models 1R-1, 1R-2 and 1R-3 included only one refugium in the Iberian Peninsula. 1R-1 model assumed only one colonization route around the north side of the Alps (forming the CH population), and from there move southeast to reach first the Balkans (GR) and then Italy (IT). The two other single-refugium models (1R-2 and 1R-3) assumed two distinct colonization routes, one north of the Alps to CH and the other south of the Alps along the Mediterranean coast to IT. 1R-2 assumes current Greece would have been colonized by owls from the Italian peninsula, following the route along the Mediterranean coast. In 1R-3, Greece would have been colonized via northern Europe, while the Mediterranean expansion would have stopped in Italy.

The last two models included a second, eastern, refugium (2R-1 and 2R-2). In these models, the western linage expansion from the Iberian population would have colonized Europe before the last glaciation, thus occupying all the Mediterranean peninsulas. During the last glaciation, two distinct populations would have survived, respectively in the Iberian (western refugium) and Italian (eastern refugium) Peninsulas. Both models assume that northern Europe was recolonized from the Iberian lineage after the glaciation, but they differ in the scenario of recolonization of southeastern Europe. In 2R-1, Greece was recolonized from the Italian refugium while in 2R-2 the expansion from the Iberian Peninsula would have recolonized all eastern Europe, including

current Greece. In this last scenario, the Italian population would be the only relic from the eastern refugium. For all scenarios, migrations between populations was allowed (see Fig. 2 and Sup. Table 3).

Wide search ranges for initial simulation parameters were allowed for population sizes, divergence times and migration rates (Sup. Table 3). Each population split was preceded by an instantaneous bottleneck, in which the founding population size was drawn from a log-uniform distribution between 0.01 and 0.5 proportion of current population sizes.

Demographic inference

Demographic simulations and parameter inference were performed under a composite-likelihood approach based on the joint SFS as implemented in fastsimcoal2[75,166]. For each of the five scenarios, 100 independent estimations with different initial values were run. For each run, there were 500'000 coalescent simulations (option -n), with 50 expectation-maximization (EM) cycles (-M and -L). As we do not have an accurate mutation rate for barn owls, we fixed the end of the glaciation to 6000 generations BP (approximately 18'000 years BP with a 3-year generation time) and scaled all other parameters relative to it using the -0 command option (using only polymorphic sites). The best-fitting scenario out of the five tested was determined based on Akaike's information criterion[168] (AIC) and confirmed through the examination of the likelihood ranges of each scenario as suggested in Kocher *et al.*[169]. Non-parametric bootstrapping was performed to estimate 95% confidence intervals (CI) of the inferred parameters under the best-fitting scenario. To account for LD, a block-bootstrap approach was employed as suggested by the authors[75,166]: the SNPs were divided into 100 same-size blocks, and then 100 bootstrap SFS were generated by sampling these blocks with replacement. Due to computational constraints, for bootstrapping we ran 50 independent parameter inferences per bootstrapped SFS with only 10 EM cycles each, instead of 50 cycles used for comparing scenarios above. This procedure has been defined as conservative[193], and is expected to produce quite large confidence intervals. We accepted this trade-off as our main goal was to determine the best demographic topology, accepting uncertainty on specific parameter values. The highest maximum-likelihood run for each bootstrapped SFS was used to estimate 95% CI of all parameters.

Niche modelling

In order to identify the regions of high habitat suitability for barn owls at the last glacial maximum (LGM, 20'000 years BP) and to support the demographic scenarios tested in the previous section, we modelled the past spatial distribution of the species in the Western Palearctic (Sup. Fig. 10). We built species distribution model (SDM) using Maximum Entropy Modelling (MaxEnt),

a presence-only based tool[325]. Current climatic variables for the Western Palearctic (Sup. Fig. 11) were extracted from the WorldClim database at 5 arc min resolution using the R package rbioclim[186], and filtered to remove variables with a correlation of 0.8 or higher. The variables retained were: Mean Diurnal Range (Bio2), Min Temperature of Coldest Month (Bio6), Temperature Annual Range (Bio7), Mean Temperature of Wettest Quarter (Bio8), Precipitation Seasonality (Bio15), Precipitation of Driest Quarter (Bio17) and Precipitation of Coldest Quarter (Bio19). We built models with linear, quadratic and hinge features, and with a range (1 to 5) of regularization multipliers to determine which combination optimized the model without over complexifying it. The best combination based on the corrected AIC (as recommended by Warren & Seifert[165]) was achieved with a quadratic model with 1 as regularization multiplier (Sup. Table 5). We ran 100 independent maxent models, omitting 25% of the data during training to test the model. To avoid geographic bias due to different sampling effort in the distribution area of the species, we randomly extracted 1000 presence points within the IUCN distribution map[187] for each model run[188].

Predictive performances of the models were evaluated on the basis of the area under the curve (AUC) of the receiver operator plot of the test data. For all models with an AUC higher than 0.8 (considered a good model[189,190]), we transformed the output of Maxent into binary maps of suitability. We assigned a cell as suitable when its mean suitability value was higher than the mean value of the 10% test presence threshold. This conservative threshold allows us to omit all regions with habitat suitability lower than the suitability values of the lowest 10% of occurrence records. Finally, we averaged the values of the models for each cell, and only cells suitable in 90% of the models were represented as such in the map.

We projected the models to the climatic conditions of the mid-Holocene (6'000 years BP) and the LGM (20'000 years BP), which we extracted from WorldClim at the same resolution as current data. When projecting to past climates, the Multivariate Environmental Similarity Surface (MESS) approach[325] was used to assess whether models were projected into climatic conditions different from those found in the calibration data. As our goal was to highlight only areas of high suitability for barn owls, cells with climatic conditions outside the distribution used to build the model were assigned as unsuitable (0 attributed to cell with negative MESS). For each timepoint, the results of the models were merged and transformed into a binary map as described for current data (Fig. 2c).

Barriers and corridors

Migration Surface Estimation in the Western Palearctic

The Estimated Effective Migration Surface (EEMS) v.0.0.9 software[161] was used to visualize geographic regions with higher or lower than average levels of gene flow between barn owl populations of the Western Palearctic. Using the SNP dataset pruned for LD produced above, we calculated the matrix of genetic dissimilarities with the tool bed2diff. The free Google Maps api v.3 application available at http://www.birdtheme.org/useful/v3tool.html was used to draw the polygon outlining the study area in the Western Palearctic. EEMS was run with 1000 demes in five independent chains of 5 million MCMC iterations with a burn-in of 1 million iterations. Results were visually checked for MCMC chain convergence (Sup. Fig. 12) and through the linear relation between the observed and fitted values for within- and between-demes estimates using the associated R package rEEMSplots v.0.0.1[161]. With the same package, we produced a map of effective migration surface by merging the five MCMC chains.

Isolation by distance in continental Europe

To investigate how population structure correlated with spatial distances between European populations and to detail the role of the Alps as a barrier to gene flow, we performed Mantel tests as implemented in the ade4 package v.1.7-15[252] for R. We compared the genetic distances (pairwise F_{ST} between populations, see section Population Structure and Genetic Diversity for details) with different measures of geographical distance between populations: the shortest distance over land via direct flight and the distance constrained by the presence of the Alps, forcing the connection of the Italian population to the other populations via the Greek peninsula (Sup. Fig. 13b). We also tested the linear regression between both variables in R.

Results

History of barn owls around the Mediterranean Sea

Genetic diversity and population structure in the Western palearctic

Despite an overall low differentiation (overall F_{ST}=0.047, comparable with the overall F_{ST}=0.045 estimated by Burri *et al.*[102]), the dataset revealed a structuration of the genetic diversity among barn owls of the Western Palearctic. The first axis of the genomic PCA (explaining 3.32% of the total variance) contrasted individuals from the Levant populations (IS and CY) to all other individuals (Fig. 1d), consistent with K=2 being the best estimate in sNMF (Sup. Fig. 4, 5). For K=3 (Fig. 1b), the Canary population (WC) formed an independent genetic cluster, and this was confirmed by the second axis of the PCA (explaining 2.6% of the variance) opposing it to all other

individuals (Fig. 1d, Sup. Fig. 2). This isolation of WC was also observable in Table 1, with a lot of private and rare alleles in the island, its higher F_{IT} and the highest population specific F_{ST} of all sampled populations. On the same PCA (Fig. 1d), individuals from European populations (FR, CH, DK, IT, SB, GR, AE) formed a third distinct cluster, matching their grouping in a single cluster at K=3 with sNMF (Fig. 1b). The Iberian individuals (PT) occupied a central position on the PCA (around 0 on both axes) and a mixed composition in sNMF (Fig. 1b). This central position of the Iberian population was also visible in the pairwise F_{ST}, where the highest value in the pairs involving PT is 0.055 (with both CY and WC) while all other pairwise comparisons involving populations from two of the distinct groups identified before (Levant, Canary and Europe) have values equal or higher (Fig. 1c). The Treemix analysis (Fig. 1f) was also consistent with these results, since it also identified these major linages, first isolating the Canary population in a specific lineage, then grouping the Levant populations (IS and CY) in a second lineage and finally all European populations in a third lineage. In Europe, the Iberian population was basal to all other populations.

Table 1 - Population genetic diversity, inbreeding and divergence estimates for 11 populations of the western Palearctic barn owls. Standard deviations of the mean are provided between brackets for each parameter, see methods section for details.

Pop	N	#Pl	#PA	#Rare	F_{IT}	F_{IS}	Pop F_{ST}
WC	9	2'217'235 (37'007)	123'016 (2'465)	407'403 (10'568)	0.067 (0.054)	-0.022 (0.057)	0.116 (0.006)
PT	9	2'639'343 (22'857)	102'252 (2'305)	539'157 (10'106)	-0.018 (0.042)	-0.008 (0.042)	0.003 (0.003)
FR	4	2'151'627 (0)	30'290 (581)	287'746 (1'468)	0.039 (0.124)	0.043 (0.131)	0.050 (0.005)
CH	10	2'494'462 (10'723)	47'532 (730)	386'654 (3'117)	0.025 (0.019)	-0.011 (0.019)	0.036 (0.002)
DK	10	2'410'615 (15'558)	40'650 (1'378)	349'800 (5'434)	0.026 (0.020)	-0.02 (0.021)	0.049 (0.002)
IT	9	2'404'069 (7'267)	66'297 (1'638)	401'842 (4'483)	0.035 (0.012)	-0.022 (0.012)	0.052 (0.005)
SB	5	2'336'060 (0)	32'096 (1'515)	326'906 (2'519)	0.025 (0.011)	-0.038 (0.011)	0.056 (0.004)
GR	9	2'454'653 (7'365)	44'996 (1'146)	378'422 (4'152)	0.018 (0.028)	-0.016 (0.027)	0.039 (0.002)
AE	10	2'460'422 (20'650)	56'260 (3'439)	403'259 (10'465)	0.018 (0.060)	-0.001 (0.062)	0.030 (0.002)
CY	10	2'338'318 (48'463)	113'377 (2'229)	480'021 (13'581)	0.022 (0.047)	-0.034 (0.049)	0.059 (0.004)
IS	9	2'509'099 (9'500)	172'624 (4'018)	608'944 (3'597)	-0.019 (0.015)	-0.036 (0.016)	0.021 (0.003)

N: number of individuals in the population; #Pl: number of polymorphic sites per populations; #PA: number of private alleles per population; #Rare: Number of rare alleles (5 or less) per population; F_{IT}: mean individual inbreeding coefficient relative to the meta-population; F_{IS}: population level inbreeding coefficient; Pop F_{ST}: population specific F_{ST} as in [159]. Populations: WC – Canary, PT – Portugal, FR – France, CH – Switzerland, DK – Denmark, IT – Italy, SB – Serbia, GR – Greece, AE – Aegean Islands, CY – Cyprus, IS – Israel.

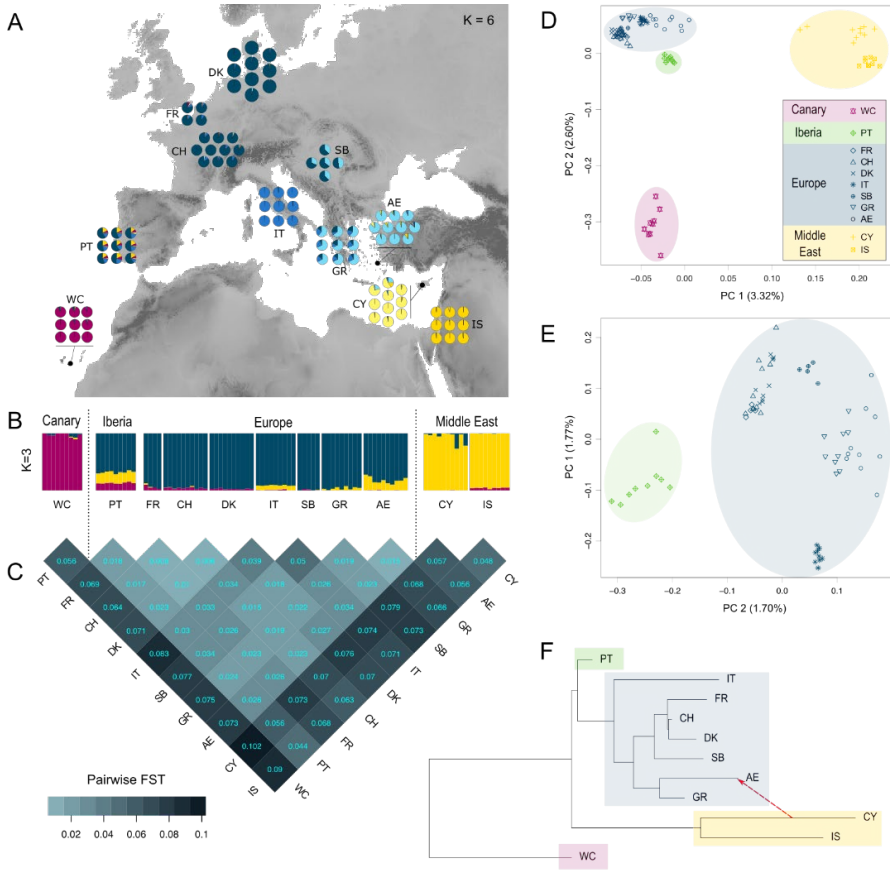


Figure 1 - Genetic structure of barn owl populations in Western Palearctic. **(a)** Population structure for K=6. Pie charts denote the individual proportion of each of lineages as determined by sNMF and are located at the approximate centroid of the sampled population. **(b)** Population structure for K=3. Each bar denotes the individual proportion of each of the 3 lineages as determined by sNMF. **(c)** matrix of pairwise F_{ST} between barn owl populations in Western Palearctic. The heatmap provides a visual representation of the F_{ST} values given in each cell. **(d)** PCA based on full set of 94 individuals. Point shape denote populations and colour circles enclose sample clusters observed in sNMF (K=3). Values in parenthesis indicate the percentage of variance explained by each axis. **(e)** PCA based on of the 66 European individuals. **(f)** Population tree and the first migration event in Western Palearctic populations inferred by Treemix.

Population structure in continental Europe

Focusing only on European samples, southern populations (PT, IT, GR and AE) harboured a higher genetic diversity than northern populations (CH, FR, DK, SB; Table 1). The first axis of the genomic PCA based on samples from European-only populations opposed the Italian individuals to all others (Fig. 1e; Sup. Fig. 3). This isolation of Italian samples was also apparent in the pairwise F_{ST} within Europe, with all the largest values involving IT. These results were consistent with sNMF on all samples for K higher than 3, where IT individuals formed an independent genetic cluster (Fig. 1a; Sup. Fig. 5), as well as Treemix, where the Italian population was the first to split and had the longest branch within European lineage (Fig. 1f). Consistently with the distribution of the individuals in the European PCA (Fig. 1e; Sup. Fig. 3), the ancestry coefficients

in the sNMF analysis of K=4 and K=5 (Sup. Fig. 5) revealed the genetic differentiation of northern populations (FR, CH and DK) compared the Italian one, also opposed in the first axis. Individual from GR and AE individuals shared ancestry with both Western and Italian population, in line with their central position along this first axis of the European PCA. For K=5, a third European component was distinguished in the Aegean individuals. This component was the majoritarian in Greek samples with contributions of both Northern and Italian component; and Serbian samples appeared as a mix of the northern and the Aegean component. The eastern lineage (CY and IS), grouped at previous K, were split into two distinct ancestry pools at K=6 (Fig. 1a). AE individuals harboured low amounts of CY ancestry, absent in all other European populations, and two CY individuals carried a large contribution of the Aegean component. Consistently, the first migration event detected by Treemix was from CY, a population from the Levant linage, to AE, a population from the European lineage (Fig. 1f; Sup. Fig. 6).

Fine Structure and Haplotype sharing

The clustering of individuals by FineStructure, based on shared haplotypes between individual was consistent with previous results (Sup. Fig. 7). Individuals from the different populations sampled were monophyletic, except for CH and FR individuals, mixed in the same population. Consistently with this grouping, haplotypes from any given population were more likely to be found in individuals from the same population, followed by its most related populations (Fig. 2). In the Levant lineage, IS haplotypes mostly painted IS Individuals but also CY individuals and vice versa. Iberian haplotypes mostly painted PT individuals, but also contributed greatly to the painting of all European individuals, decreasing with distance. Western European haplotypes (from FR and CH) mostly painted western European individuals, then northern individuals (from DK) and finally eastern individuals (from SB, GR and AE). The reverse pattern was observed for haplotypes from eastern Europe (GR, AE), with a gradient of contribution decreasing from east to west. Haplotypes from DK and SB mostly painted individuals from their own population, but also in their respective neighbours in both eastern and western European populations. Italian haplotypes were the most distinct haplotypes among European populations, mostly painting Italian individuals, followed by Greek individuals, and being painted by other populations at a lower rate than expected given its geographic position. Finally, AE haplotypes also painted more often CY individuals than IS individuals. This painting of levant individuals by AE haplotypes was higher than the contribution from any other European individual, and both CY and IS haplotypes painted more AE individuals than any other European individual.

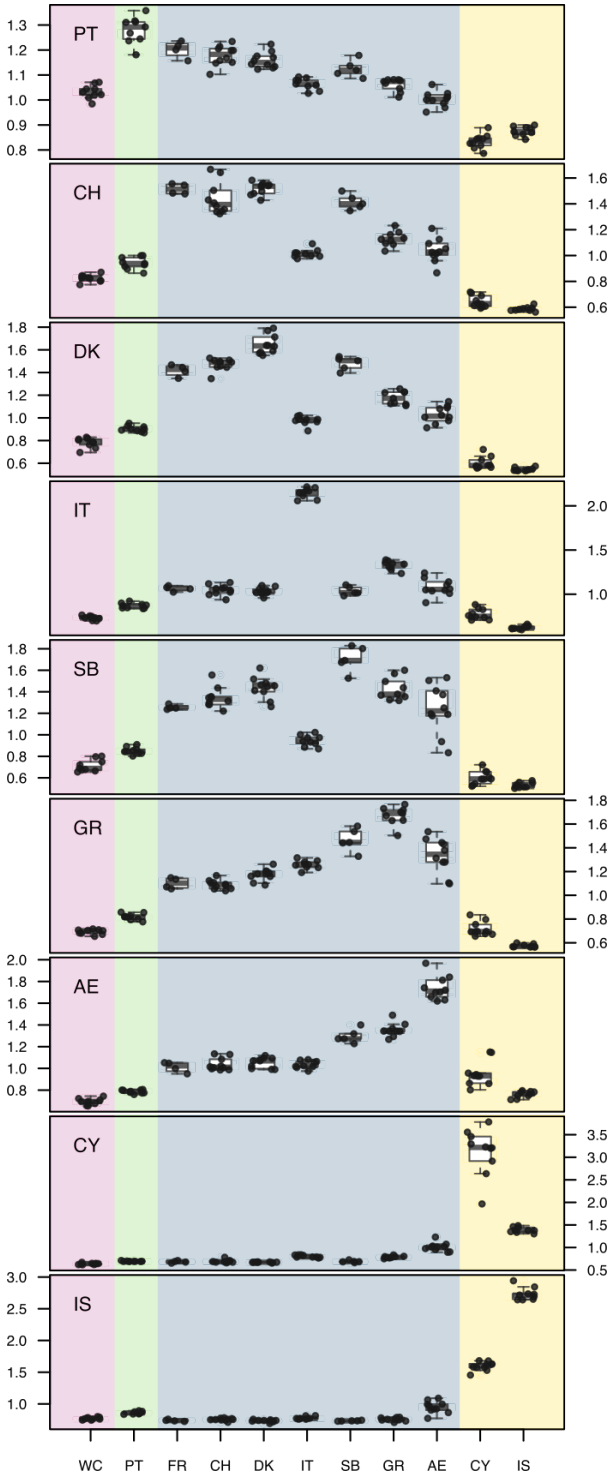


Figure 2 – Individual haplotype sharing between barn owl populations. Part of the total length of ChromoPainter chunks inherited from other genomes. Each graph summarizes the information of all the genomes from a given population. Background colours match the lineages identified in Fig. 1.

Modelling of the history of European barn owls

Species distribution modelling

Habitat suitability projections showed that, from a climatic point of view, there were suitable regions for barn owls all around the Mediterranean Sea during the glaciation (20'000 years BP; Fig. 3c). Large areas were suitable in northern Africa and the Iberian Peninsula, but also in the

two eastern Mediterranean peninsulas (current Italy and Greece). At this point, the sea levels were lower than today's and the two eastern peninsulas were more connected, allowing for a continuous region of suitable barn owl habitat. At the mid-Holocene (6'000 years BP), major changes in sea level revealed a coastline very similar to nowadays. Our projections revealed a reduction of habitat suitability in northern Africa at this time, while the suitability of western and northern Europe increased (Fig. 3c). Finally, today, nearly all continental Europe is suitable for the barn owls, with the notable exception of mountain areas (Fig. 3c).

Demographic inference

AIC and raw likelihood comparisons showed that the two refugia model 2R-1 explains best the SFS of our dataset (Sup. Table 2; Fig. 3b). In this model, an ancestral Italian lineage (IT) split from the Iberian lineage (PT) before the last glaciation, estimated at approximately 69'000 years BP (95% CI: 24'000-90'000 years BP; calculated with 3-year generation time; Sup. Table 4). After its initial expansion, the ancestral population is estimated to have been larger in the Italian peninsula that in Iberia (respectively 189K (11K-320k) and 11k (10K-73k) haploid individuals). During the glaciation (fixed between 24 and 18K years BP), both populations experienced a bottleneck, with a population size reduced to 6.8k (1.2k-141k) individuals in the Iberian lineage and 62 (36-116k) in the Italian. After the glaciation, the size of both populations increased to their current size, estimated at 44k (20K-380k) in the Iberian Peninsula and 1.3k (1k-326k) in the Italian Peninsula, both smaller but consistent with their estimated census size (55k-98k[326] and 6k-13k[327], respectively). The Greek population split from the Italian branch around 5'700 years BP while the Swiss (CH) population split slightly later from Iberia (5'000 years BP) and maintain a high level of gene flow (estimated to 90 (46-1.4k) from CH to PT and 8 (3-62) in the reverse direction). Current effective population sizes of the CH and GR populations are estimated to 3.4k (1k-205k) and 1.4k (1k-208k), respectively (Fig. 2b), in line with census results (1000-2500[328] and 3000-6000[329], respectively). Migration between these populations is estimated to be highest from IT and GR to CH (respectively 27 (0.2-157) migrants from IT and 42 (0.1-156) from GR) and lowest in the opposite direction (respectively 1.3 (0.1-96) migrants from CH to GR and 0.02 (0.2-38) from CH to IT) (Sup. Table 4). Point estimates with 95% confidence intervals for all parameters of the best model are given in (Sup. Table 4), as well as single point estimates for all models (Sup. Table 3).

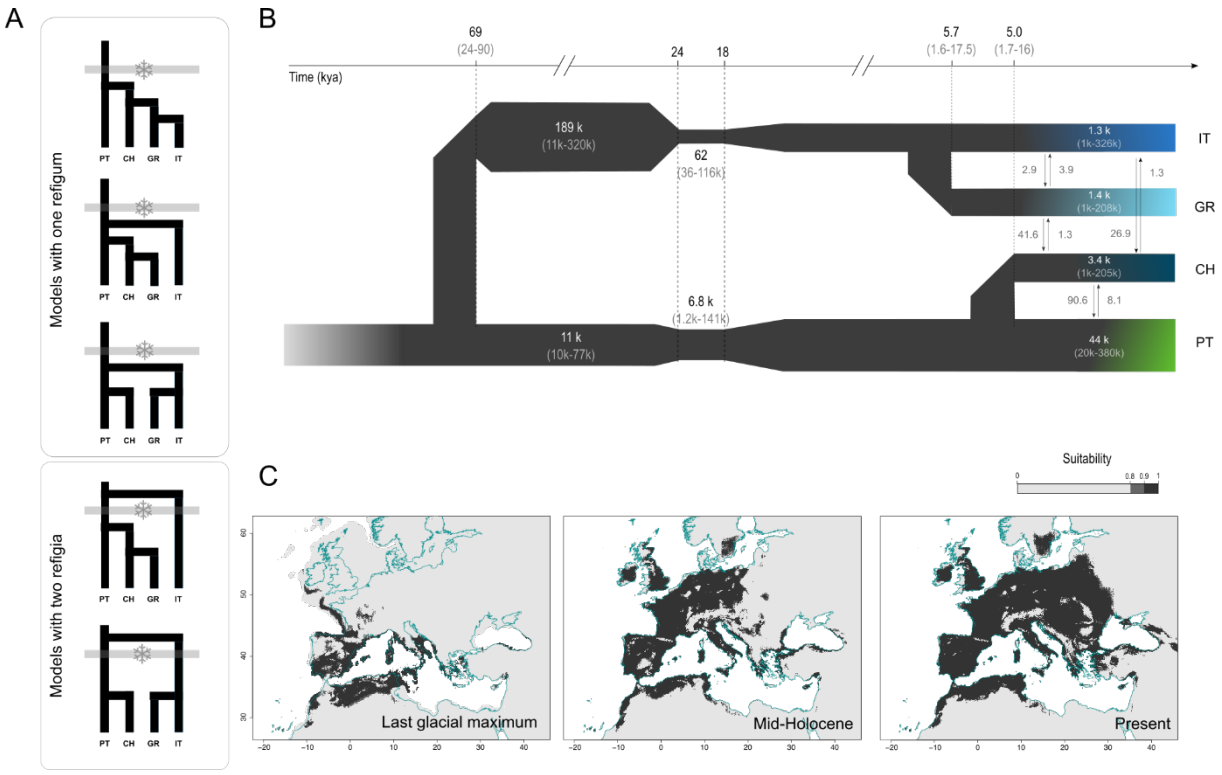


Figure 3 – Modelling of the history of the barn owl in Europe. **(a)** Schematic representation of the five demographic scenarios tested for the colonization of the Europe by barn owls. Three models included one refugium in the Iberia during the LGM while the last two included two refugia, one in Iberia and the second in Italy. Grey bars with snowflakes represent the last glaciation. **(b)** Best supported demographic model for the history of European barn owl populations as determined by fastsimcoal2. Time is indicated in thousands of years, determined using a 3-year generation time, confidence intervals at 95% are given between brackets. Population sizes (haploid) are shown inside each population bar; arrows indicate forward-in-time migration rate and direction. **(c)** Species distribution model of barn owls based on climatic variables, projected into the past (last glacial maximum - 20 kya; mid-Holocene - 6 kya) and today's conditions. Locations in dark grey were highly suitable in 90% of the models. Below that threshold cells were considered as unsuitable (lightest grey shade on the graph).

Barriers and corridors

Migration Surface Estimate in the Western Palearctic

Estimated Effective Migration Surface identified large water bodies, especially in the eastern Mediterranean and around Cyprus, as regions resisting to migration (Sup. Fig. 12). On the mainland, barriers to gene flow matched the main formations of the Alpide belt in the region, an orogenic formation spanning from western Europe to eastern Asia. From west to east, a light barrier overlapped with the Pyrenees, a strong barrier spanned the Alps to the Balkans and a third obstacle matched the Taurus mountains in Anatolia. A region with high gene flow was identified in continental Europe above the Alps, spanning from western Europe to the Balkans peninsula.

Isolation by distance in Europe

In continental Europe, the shortest path overland did not correlate significantly with genetic distance (Sup. Fig. 13) (mantel test, p-value = 0.193, R = 0.20 / linear model, p-value = 0.26, R2 = 0.012). On the contrary, when the geographic distance between populations included the barrier formed by the Alps (i.e. the Italian population was connected to other populations via the Greek peninsula, itself connected to western Europe via northern Europe), both tests were significant (mantel test, p-value = 0.002, R = 0.68 / linear model, p-value = 1.3x10-5, R2 = 0.507).

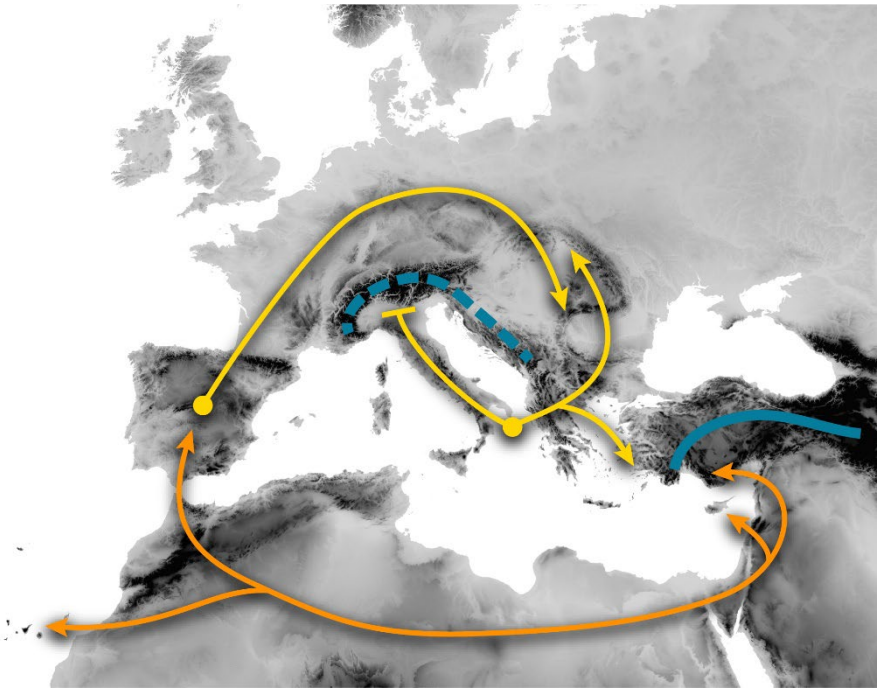


Figure 4 – Schematic representation of the history of barn owls in the Western Palearctic and the main barriers in the region. Orange arrows depict the colonization of the region by the three main lineages (Levantine, Canarias, European). Yellow arrows represent the modelled postglacial recolonization scheme of Europe, with two distinct refugia (yellow dots). Blue lines represent the main barriers identified in this work, namely the Alps in Europe (dashed line) and the Taurus and Zagros mountains in Anatolia (solid line).

Discussion

The history of natural populations is shaped by the combination of landscape barriers and climatic variations that isolate and mix lineages through their combined actions. Consistently with previous work[102], we show that barn owls colonized the Western Palearctic in a ring-like fashion around the Mediterranean Sea, with one arm around the Levant and the second throughout Europe. However, using whole genome sequences we found this colonization actually predates the last glaciation and pinpoint a narrow secondary contact zone between the two lineages in Anatolia rather than in the Balkans. In addition, we provide evidence that barn owls recolonized Europe after the LGM from two distinct glacial refugia – a western one in Iberia and an eastern in Italy – rather than a single one as it was previously thought. As temperatures started rising, western and northern Europe were colonized by owls from the Iberian Peninsula while, in the

meantime, the eastern refugium population of Italy had spread to the Balkans (Fig. 4). The western and eastern glacial populations finally met in eastern Europe. This complex history of populations questions the taxonomy of the multiple *Tyto alba* subspecies, highlights the key roles of mountain ranges and large water bodies as barriers to gene flow for a widespread bird and illustrates the power of population genomics in unravelling intricate patterns.

Colonization of the Western Palearctic and gene flow in Anatolia

Our results show that two distinct barn owl genetic lineages surround the Mediterranean Basin: one in the Levant, and a second in Europe (Fig. 1b-d, f), likely connected via northern Africa. Supported by the higher and specific diversity of the basal population of each arm (namely, IS and PT (Table 1), these observations are consistent with the ring colonization scenario hypothesized by Burri *et al.*[102]. However, we show that a barn owl population survived the last glaciation in Italy (see next section for details) and that its genetic makeup resembles the European lineage (Fig. 1b, d, f). Therefore, the ring colonization of Europe around the Mediterranean appears to have been pre-glacial, whereas the post-glacial history is more convoluted (see next section).

In previous studies, the ancestry of Greek and Aegean populations was unclear, with a hypothesized mixed origin between European and Levant lineage 20. This uncertainty was mostly likely due to the low resolution of genetic markers (mtDNA and microsatellites), as the genomic data reported here clearly show that Greek and Aegean owls are genetically much closer to European than to Levant ones (Fig. 1b.d, f). This observation indicates that the European lineage reached further east than previously assumed, allowing us to pinpoint the secondary contact zone between the European and Levant lineages to Anatolia, instead of the Balkans as it had been proposed. In Anatolia, the Taurus and Zargos mountain ranges form an imposing barrier that appears to have stopped the expansion of the Levant lineage both during the ring colonization and nowadays.

Despite the barrier, and although we do not see a complete admixture of the two lineages, there is evidence for some gene flow. Indeed, the first migration in Treemix (Fig. 1f) pointed to a secondary contact between CY and AE, consistent with the signals of admixture between those populations (Fig. 1a, b, 2). The admixture pattern however is restricted geographically to this narrow region and does not permeate further into either of the lineages, as surrounding populations (IS in the levant and GR and SB in Europe) do not show signals of admixture (Fig. 1, 2). Thus, the migration between populations on both sides seems limited and possibly only occurs along a narrow corridor along the Turkish coast where only a few barn owls have been recorded[330]. Further analyses with samples from Anatolia should allow to characterize in high resolution how and when admixture occurred in this region.

Glacial refugia and recolonization of Europe

Previous studies showed that barn owls survived the last glaciation by taking refuge in the Iberian Peninsula and maybe even in emerged land in the Bay of Biscay[93,100]. The observed distribution of diversity in Europe, and especially the specific makeup of the Italian populations, is best explained by a demographic model with two glacial refugia – one in Iberia and a second in Italy, derived from the Iberian population before the glaciation (Fig. 3a, b; Sup. Table 2). Environmental projections not only support this model, as Italy was highly suitable for the species at the time, but also show that, due to the low levels of the Adriatic Sea, the suitable surface extended to the west coast of the Balkans (LGM - Fig. 3c). Crucially, three of the barn owl's key prey also had glacial refugia in the Italian and Balkan peninsulas, namely the common vole (*Microtus sp.*)[331], the wood mouse (*Apodemus sp.*)[129] and shrews (*Crocidura sp.*)[332]. The inferred size of the pre-glacial population of barn owl inhabiting current Italy was larger than any other (189k [11K-320K]), prior to a strong bottleneck during the glaciation (population reduced to 62 individuals [36-116k]). These values are likely inflated by necessary simplification of the model (e.g. instant bottlenecks) or by gene flow from unmodeled/unsampled populations, for example from the Levant lineage to eastern European populations (GR), or from northern Africa to Italy via Sicily. The latter would have been facilitated by the increased connectivity between Italy and North Africa during the glaciation (Fig. 3c)[333–335].

With the warming following the LGM, Europe became gradually more suitable and, by the mid-Holocene (6000 years ago), most of western and northern Europe were appropriate for barn owls (Fig. 2b) as well as the common vole (*Microtus arvalis*)[336]. The genetic similarity between Iberian (PT) and north western populations (CH, FR, DK; Fig. 1a, e) indicates that barn owls colonized these newly available regions from the Iberian refugium as previously thought[100]. The contribution from the Italian refugium to northern populations appears to have been hindered by the Alps (see next sections), as suggested by the higher genetic distance between them (Fig. 1a, c, d and 2). Instead, at this time the Adriatic Sea had neared today's levels, isolating genetically and geographically the Italian refugium from its component in the Balkan Peninsula (Fig. 3b – IT-GR split ~6k, Fig. 3b, c). Only more recently did the rise of temperatures allow for areas in the east of Europe to become suitable, finally connecting the south-eastern populations near the Aegean Sea (GR and AE) with populations in the north-eastern part of Europe (Fig. 3c). In particular, the high heterozygosity and admixed ancestry of Serbian individuals (Table 1; Fig. 1a) suggest that the suture between the Iberian and the Italo-Greek glacial lineages took place in eastern Europe. This newly identified postglacial recolonization scheme of continental Europe by the barn owl matches the general pattern described for the brown bear (*Ursus arctos*) and a barn owl prey, the shrew (*Sorex sp.*)[123].

The isolation by distance pattern observed between European populations (Sup. Fig. 13) highlights a diffusion of alleles in the European populations rather than a narrow hybrid zone. Further, the inferred migration rates support high current gene flow in the region (Fig. 3b, CH and GR in Sup. Table 4), and we found signals of each ancestry in populations far from the suture zone (dark blue in GR and light blue in CH and DK, Fig. 1a). Finally, the measure of haplotype sharing decreases consistently with distance between populations around the northern side of the Alps with no mark of neat differentiation between populations from Iberia to Greece, excluding IT (Fig. 2). Surrounded by the sea and the Alps, Italy is the exception and appears to have avoided incoming gene flow (Sup. Fig. 12; isolation in Fig. 2), thus being a better-preserved relic of the refugium population (own cluster in sNMF K>4; Sup. Fig. 5). In contrast, the Balkan component admixes smoothly with the other European populations. Such seamless mixing of the two glacial lineages from southern refugia where barn owls are mostly white[101,102], brings further into question the subspecies *Tyto alba guttata*.

The case of *Tyto alba guttata*

Traditionally, in Europe the eastern barn owl (*Tyto alba guttata*, Brehm, CL, 1831) is defined by its dark rufous ventral plumage in contrast to the white western barn owl (*Tyto alba alba*, Scopoli, 1769)[337]. With a wide distribution, it is recorded from The Netherlands to Greece, including most of northern and eastern Europe[205]. However, this repartition does not match the history of any specific glacial lineage identified above, nor any genetically differentiated population (Fig. 4). The dark populations of northern Europe (DK) are genetically as similar to lighter western populations (FR, CH) than to the dark ones in the east (SB; Fig. 1). This colour variance within European populations has been shown to be maintained through local adaptation[101], and while the genomic basis and history of this trait remain worthy of future investigations, we suggest that all European barn owls form a single subspecies (*Tyto alba alba*), reflecting the entire European population, regardless of their colour.

Barriers and corridors shape the connectivity of the Western Palearctic meta-population

The partition of genetic diversity among barn owls in the Western Palearctic allowed us to identify barriers and corridors to gene flow. Populations isolated by large water bodies have accumulated substantial genetic differences as, for example, the higher F_{ST} in the Canary and Cyprus Islands (Fig. 1c; Table 1), and reflect the importance of water as a barrier to dispersion in this species 21. On the mainland, and as described for the American barn owl[171], major mountain ranges act as significant obstacles to migration for European barn owls and can generate genetic structure. First, the high mountain ranges of Taurus and Zagros coincide with the contact zone between the

Levant and the European lineages both nowadays and potentially at the time of the pre-glacial ring colonization of Europe (see above). Second, the Alps and the Balkan Mountains slowed the northward expansion of the glacial populations of Italy and Greece after the LGM and still constrain migration between populations on both sides of their ranges. If these results remain to be confirmed with observational data (i.e. ringing data not available for all countries), they emphasize that, despite its worldwide repartition and its presence on many islands, the connectivity of barn owl populations is heavily driven by biogeographical barriers.

Conclusion

The combination of whole genome sequencing and sophisticated modelling methods revealed the complex history of the barn owl in the Western Palearctic with a precision previously unachievable. It allowed the localization of a secondary contact zone as well as the discovery of a cryptic glacial refugium. However, several questions remain unanswered, awaiting for relevant samples to be collected and analysed: What role did northern African populations played in connecting the Levant and European lineages? Did they contribute to the diversity observed in Italy? How narrow is the contact zone between the Levant and European lineages in Anatolia? Lastly, the origin of barn owls from the Western Palearctic as a whole also deserves further investigation, as they are believed to have colonized the Western Palearctic from the east, given their supposed origin in south-eastern Asia approximately 4 million years ago. But this is at odds with the higher genetic diversity of the Iberian population compared to the Levant one. Such inconsistency points to the need for samples from around the world, to understand how this charismatic group of nocturnal predators conquered the entire planet.

Research on postglacial recolonization and the subsequent phylogeographic patterns peaked at the turn of the century, with many studies providing an overview of the history of a wide variety of organisms (reviewed by Hewitt[123]). The rise in availability of genomic data for non-model species, combined with the type of approaches used here, will rewrite the history of many of them. Furthermore, it will allow to detail the genomic consequences of such history both from a neutral and selective perspective. Applied to several species, these approaches will allow to redefine with greater clarity the broad phylogeographical patterns in the Western Palearctic and elsewhere, to re-think taxonomic classifications and to better understand how organisms might adapt to a changing environment in a complex, fragmented and rapidly changing landscape.

Chapter 4 – Supporting Information

Samples and data preparation

Supplementary Table 1 – Description of samples used in this study. Individuals retained for the inference with fastsimcoal2 are indicated with †.

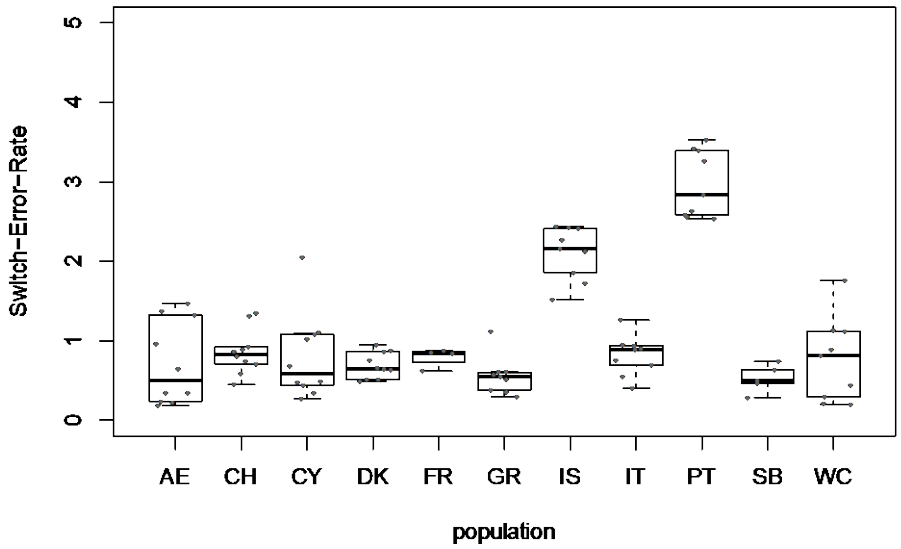
#	Pop	ID	Country	Location	Year	Tissue	Sex	Ref
1	AE	AE01	Greece	Rhodes island	2014	soft tissue	Female	1
2	AE	AE02	Greece	Rhodes Island	2014	soft tissue	Female	1
3	AE	AE03	Greece	Chios island	2015	blood	Male	1
4	AE	AE04	Greece	Chios island	2012	blood	Female	1
5	AE	AE05	Greece	Leros island	2012	blood	Male	1
6	AE	AE06	Greece	Lesvos Island	2012	blood	Male	1
7	AE	AE07	Greece	Lesvos Island	2013	soft tissue	Female	1
8	AE	AE08	Greece	Rhodes island	2013	blood	Male	1
9	AE	AE09	Greece	Leros island	2014	blood	Male	1
10	AE	AE10	Greece	Rhodes Island	2015	blood	Female	1
11	CH	CH01†	Switzerland	Estavayer-le-Lac	2004	blood	Female	2
12	CH	CH10†	Switzerland	Avenches	2008	blood	Male	2
13	CH	CH15†	Switzerland	Avenches	2010	blood	Male	2
14	CH	CH18†	Switzerland	GrangesMarnand	2001	blood	Male	2
15	CH	CH20	Switzerland	Payerne	2001	blood	Male	2
16	CH	CH21†	Switzerland	Chavornay	2011	blood	Male	2
17	CH	CH22	Switzerland	Estavayer-le-Lac	2001	blood	Male	2
18	CH	CH23†	Switzerland	Avenches	2004	blood	Male	2
19	CH	CH25†	Switzerland	Payerne	2003	blood	Male	2
20	CH	CH26†	Switzerland	Chavornay	2010	blood	Male	2
21	CY	CY01	Cyprus	NA	NA	muscle	Female	1
22	CY	CY02	Cyprus	Limasol	2016	muscle	Female	1
23	CY	CY03	Cyprus	Limasol	2016	muscle	Male	1
24	CY	CY04	Cyprus	Larnaca	NA	muscle	Female	1
25	CY	CY05	Cyprus	NA	NA	muscle	Male	1
26	CY	CY06	Cyprus	NA	NA	muscle	Female	1
27	CY	CY07	Cyprus	Larnaca	2017	muscle	Female	1
28	CY	CY08	Cyprus	Limasol	2015	muscle	Female	1
29	CY	CY09	Cyprus	Limasol	2017	muscle	Female	1
30	CY	CY10	Cyprus	Limasol	2018	muscle	Female	1
31	DK	DK01	Germany	Gettorf	2007	feather	Female	2
32	DK	DK02	Germany	Lindau	2007	feather	Male	2
33	DK	DK03	Germany	Rickling	2007	feather	Female	2
34	DK	DK04	Germany	Elskop	2007	feather	Female	2
35	DK	DK05	Germany	Itzehoe	2007	feather	Male	2
36	DK	DK06	Germany	Schwedeneck	2007	feather	Male	2
37	DK	DK07	Germany	Vaale	2007	feather	Female	2
38	DK	DK08	Denmark	Varnæs	2007	feather	Male	2
39	DK	DK09	Denmark	Bredebro	2007	feather	Male	2

40	DK	DK10	Denmark	Logumkloster	2007	feather	Female	2
41	FR	FR01	France	Rouen	2011	muscle	Male	2
42	FR	FR02	France	Rouen	2010	muscle	Male	2
43	FR	FR03	France	Rouen	2010	muscle	Male	2
44	FR	FR05	France	Rouen	2010	muscle	Male	2
45	GR	GR01†	Greece	Agrinio	2014	blood	Male	1
46	GR	GR02†	Greece	Athens	2014	blood	Male	1
47	GR	GR03†	Greece	Chalandri	2012	blood	Male	1
48	GR	GR04	Greece	Corinth	2015	blood	Female	1
49	GR	GR06†	Greece	Lamia	2014	soft tissue	Female	1
50	GR	GR07†	Greece	Mesolonghi	2014	soft tissue	Male	1
51	GR	GR08†	Greece	Morfovouni	2015	blood	Female	1
52	GR	GR09†	Greece	Panetolio	2014	blood	Female	1
53	GR	GR10†	Greece	Spata	2015	blood	Male	1
54	IS	IS01	Israel	Lachish	2005	blood	Female	1
55	IS	IS02	Israel	Beit Shean	2005	blood	Male	1
56	IS	IS03	Israel	Hula	2005	blood	Female	1
57	IS	IS04	Israel	Beit Shean	2005	blood	Female	1
58	IS	IS05	Israel	Beit Shean	2005	blood	Male	1
59	IS	IS06	Israel	Beit Shean	2005	blood	Female	1
60	IS	IS07	Israel	Beit Shean	2005	blood	Female	1
61	IS	IS08	Israel	Hula	2005	blood	Female	1
62	IS	IS09	Israel	Hula	2005	blood	Female	1
63	IT	IT01†	Italy	Roma	2011	blood	Female	1
64	IT	IT02†	Italy	Roma	2015	blood	Female	1
65	IT	IT03	Italy	Roma	2016	blood	Male	1
66	IT	IT04	Italy	Roma	2016	blood	Male	1
67	IT	IT05†	Italy	Roma	2009	blood	Female	1
68	IT	IT06†	Italy	Grosseto	2014	blood	Female	1
69	IT	IT07†	Italy	Livorno	2001	blood	Female	1
70	IT	IT08†	Italy	Firenze	2011	blood	Female	1
71	IT	IT09†	Italy	Firenze	2016	blood	Male	1
72	PT	PT01†	Portugal	Pombal	2013	feather	Female	2
73	PT	PT02†	Portugal	Coruche	2013	feather	Male	2
74	PT	PT03	Portugal	Évora	2013	feather	Male	2
75	PT	PT04†	Portugal	Coruche	2012	feather	Female	2
76	PT	PT05†	Portugal	Nazaré	2013	feather	Female	2
77	PT	PT06†	Portugal	Porto de Moós	2013	feather	Female	2
78	PT	PT07†	Portugal	Setúbal	2012	feather	Female	2
79	PT	PT08†	Portugal	Fátima	2013	feather	Female	2
80	PT	PT09†	Portugal	Santarém	2013	feather	Male	2
81	SB	SB01	Serbia	Kac	2012	blood	Male	1
82	SB	SB02	Serbia	Gudurica	2016	blood	Male	1
83	SB	SB03	Serbia	Gorobilje	2016	blood	Female	1
84	SB	SB04	Serbia	Lapovo	2015	soft tissue	Female	1
85	SB	SB05	Serbia	Striza	2006	soft tissue	Male	1
86	WC	WC01	Spain	Tenerife	1905	muscle	Male	1
87	WC	WC02	Spain	Tenerife	2003	muscle	Female	1

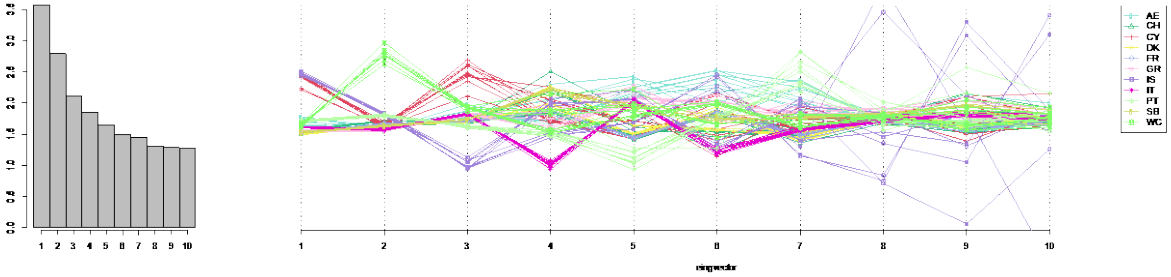
88	WC	WC03	Spain	Tenerife	2003	muscle	Female	1
89	WC	WC04	Spain	Tenerife	2003	muscle	Male	1
90	WC	WC05	Spain	Tenerife	2005	muscle	Male	1
91	WC	WC06	Spain	Tenerife	2005	muscle	Female	1
92	WC	WC07	Spain	Tenerife	2006	muscle	Male	1
93	WC	WC08	Spain	Tenerife	2006	muscle	Male	1
94	WC	WC09	Spain	Tenerife	2006	muscle	Male	1
95	Outgroup	SGP	SGP	Singapore	2013	muscle	Male	2
96	Outgroup	USA	USA	San Diego, California	2015	muscle	Female	2

[1] This study – GenBank BioProject PRJNA727977 [204]

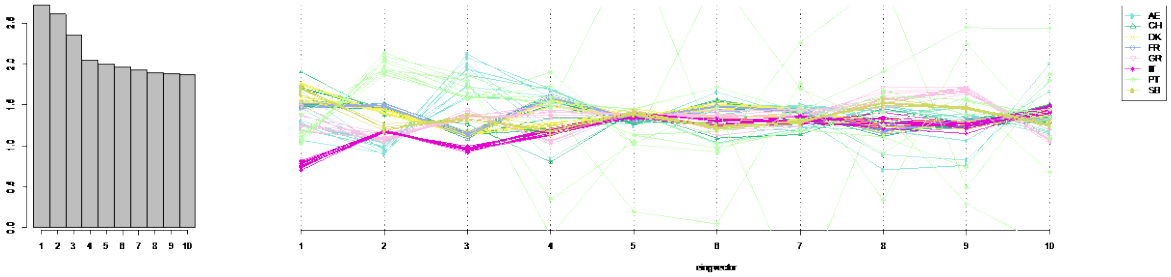
[2] Machado *et al.* 2021 – GenBank BioProject PRJNA700797 [93]



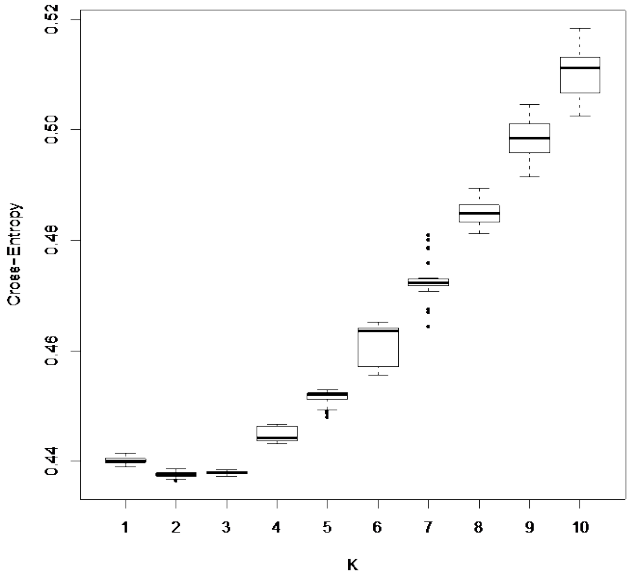
Supplementary Figure 1 – Individual phasing switch error rate (in %) grouped per population.



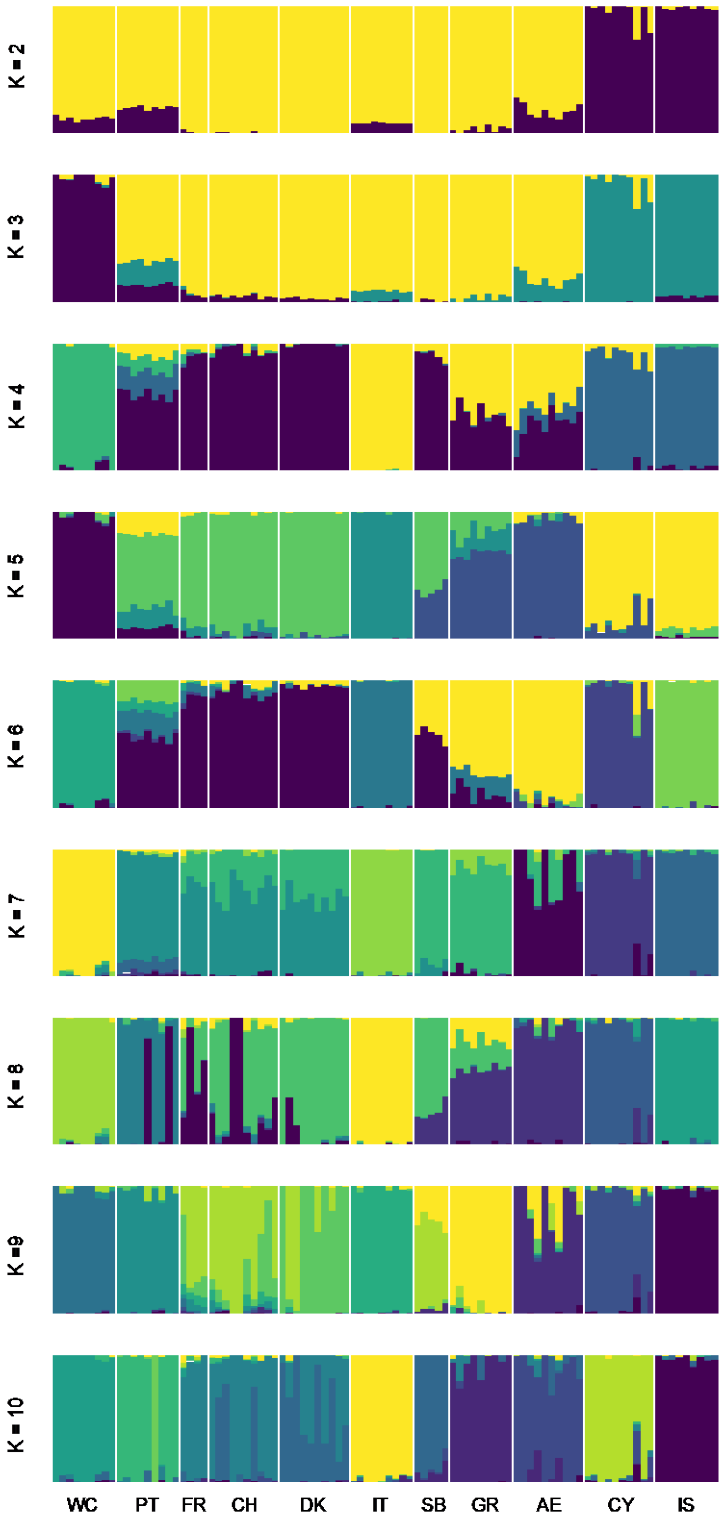
Supplementary Figure 2 – Screeplot of the 10 fist axes of the PCA with all the individuals (left). Position of the individuals on the 10 first axes of the PCA (right).



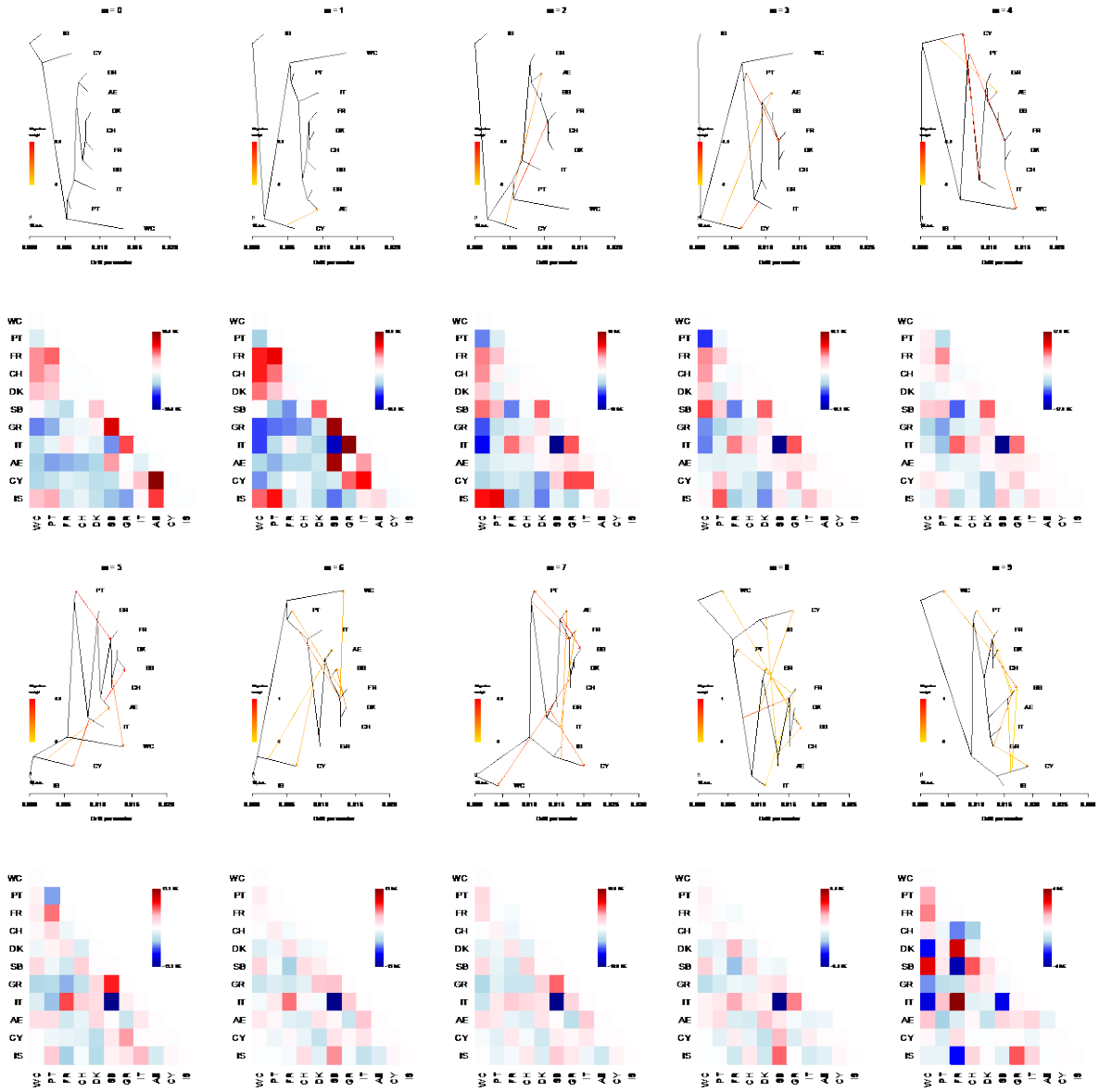
Supplementary Figure 3 - Screeplot of the 10 fist axes of the PCA with 66 European individuals (left). Position of the individuals on the 10 first axes of the PCA (right).



Supplementary Figure 4 – Values of the cross-entropy criterion for 25 sNMF runs per K. The number of clusters ranged from 1 to 10.

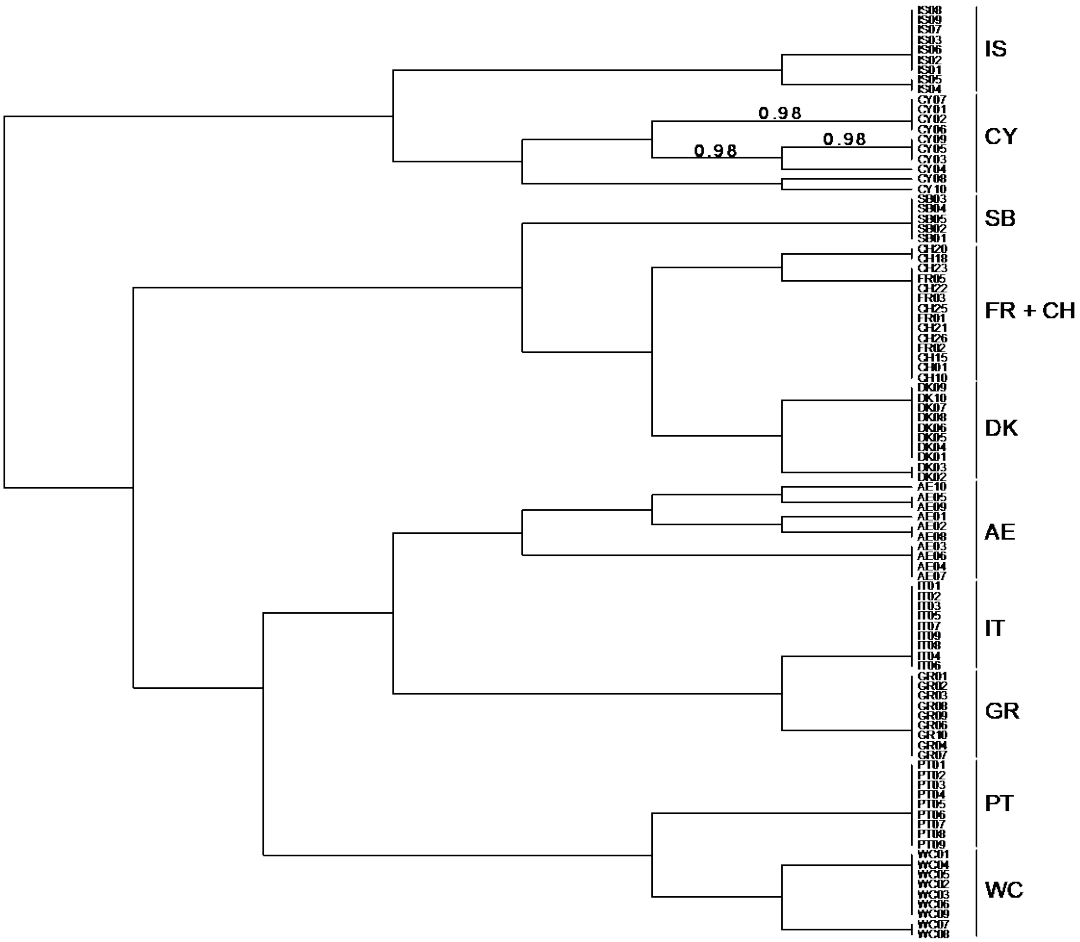


Supplementary Figure 5 – Individual ancestry estimated by sNMF for K ranging from 2 to 10. Each vertical bar represents one individual, and the colours represent the relative contributions of each genetic lineage K.



Supplementary Figure 7 – Results from Treemix for 0 to 9 migration events. Highest likelihood runs are depicted, with the corresponding matrix of standard errors.

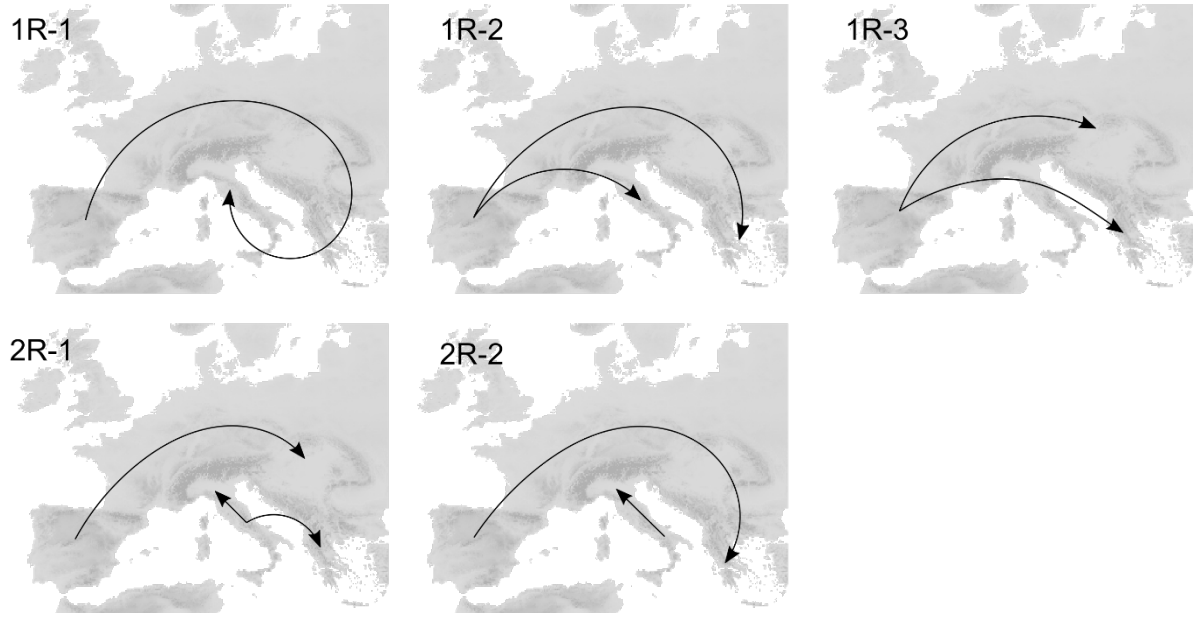
Haplotype sharing



Supplementary Figure 8 – Dendrogram of the individuals from FineStrcture. Sampled populations were monophyletic, except for Swiss (CH) and French (FR) individuals. Grouping is based on similarity and the tree is not based on any model of population differentiation[324].

Modelling the History of European barn owl

Maximum-likelihood demographic inference



Supplementary Figure 9 – Schematic representation of the recolonization of Europe for the five scenarios modelled in fastsimcoal2.

Supplementary Table 2 – Likelihood and AIC of the demographic models tested with *fastsimcoal2*. Five main model topologies were tested with one (1R) or two (2R) glacial refugia. Models are sorted from best to worst according to the estimated likelihoods.

Model Name	Est. Lhood	Δ Lhood	AIC	Δ AIC
2R-1	-6519338	3052.89	30022708.26	0
1R-1	-6522237	5951.772	30036060.1	13351.84496
1R-3	-6525492	9207.312	30051046.42	28338.16071
1R-2	-6529137	12852.437	30067830.84	45122.58168
2R-2	-6556984	40699.49	30196081.26	173372.9999

Est. Lhood – Maximum-likelihood estimated for the simulated SFS per demographic model; Δ Lhood – difference between the likelihood of the simulated and observed SFS; Δ AIC – delta AIC

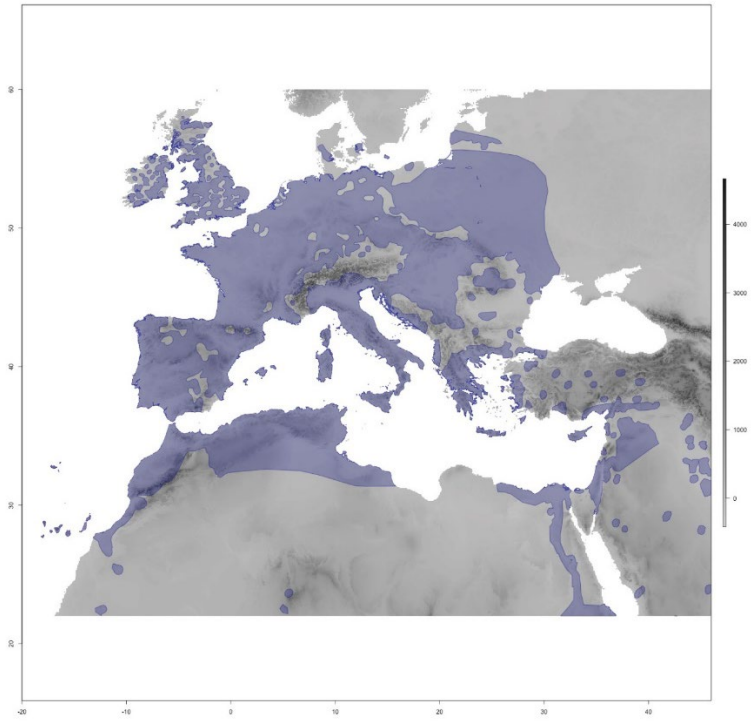
Supplementary Table 3 – Parameter ranges and point estimate inferred for the demographic model tested with fastsimcoal2. Model "2R-1" – identified as the best fitting model – is given in Sup. Table 7. All range distributions were uniform except for bottlenecks (*) which were log-uniform. When a parameter was absent in a model, the case was left blank.

		One Refugium			Two Refugia
Parameter	Ranges	1R -1	1R-2	1-R3	2R-2
Current Population Sizes (haploid)					
PT	10000 - 4e5	11713	11637	10218	386028
CH	1000 - 3.5e5	1009	194765	239033	261548
GR	1000 - 3.5e5	1031	1019	1037	4267
IT	1000 - 3.5e5	1022	5272	4692	296226
Ancestral Population Sizes (haploid)					
PT before glac	10000 - 4e5	-	-	-	244935
IT before glac	1000 - 3.5e5	-	-	-	115633
PT during glac	0.01 - 0.5 *	-	-	-	654486
IT during glac	0.01 - 0.5 *	-	-	-	83910
Times of Divergence (generations)					
Pre-glacial split PT-IT	8000 - 15000	-	-	-	8137
T1		5954	5946	5873	5830
T2		5937	5665	5853	0
T3		3956	1453	1331	-
Current Migration (flow is backwards in time)					
CH → PT	0 - 0.05	0.0074	0.0019	0.0001	0.0259
PT → CH	0 - 0.05	0.0029	0.0027	0.0037	0.0177
GR → CH	0 - 0.05	0.0212	0.0360	0.0172	0.0135
CH → GR	0 - 0.05	0.0038	0.0412	0.0182	0.0182
IT → GR	0 - 0.05	0.0164	0.0019	0.0016	0.0342
GR → IT	0 - 0.05	0.0005	0.0031	0.0057	0.0130
IT → CH	0 - 0.05	0.0013	0.0001	0.0001	0.0341
CH → IT	0 - 0.05	0.0002	0.0013	0.0001	0.0198
Older Migration					
CH → PT	0 - 0.05	0.0007	-	0.0017	0.00001
PT → CH	0 - 0.05	0.0209	-	0.0249	0.0001
GR → CH	0 - 0.05	0.0047	-	-	-
CH → GR	0 - 0.05	0.0709	-	-	-
*Instbot- instant founding population **					
Pre-glacial split PT-IT	0.01 - 0.5	-	-	-	69907
T1	0.01 - 0.5	84	189	99	12011
T2	0.01 - 0.5	19	6750	3957	58
T3	0.01 - 0.5	77	151	65	-

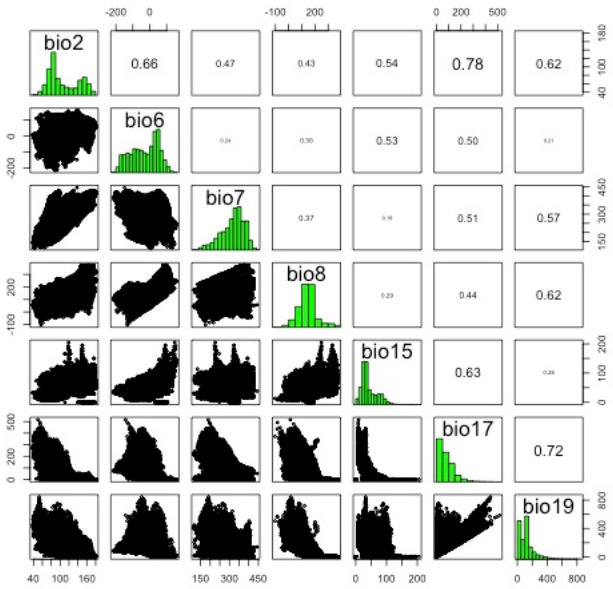
Supplementary Table 4 – Parameter point estimates and 95% confidence interval for the best demographic model: 2R -1. Parameter names correspond to Sup. Fig. 1. Times of divergence are in years calculated with a generation time of 3 years. Migration rates and number of individuals are given forward in time.

Parameter	Lower Limit CI	Point Estimate	Upper Limit CI
Current Population Sizes (haploid)			
PT	20405	44055	380159
CH	1006	3462	205409
GR	1007	1381	208303
IT	1039	1319	326428
Ancestral Population Sizes (haploid)			
PT before glac	10103	10859	73118
IT before glac	10775	189631	320792
PT during glac	700	6857	141005
IT during glac	36	62	116247
Times of Divergence			
Pre-glacial split PT-IT	24206	68907	90015
T1	1719	5095	16007
T2	1622	5717	17523
Current Migration (forward 2Nm)			
CH → PT	46.28	90.59	1413.66
PT → CH	2.65	8.05	62.01
GR → CH	0.13	41.66	156.06
CH → GR	0.11	1.26	95.91
IT → GR	0.21	2.88	135.06
GR → IT	0.16	3.89	60.28
IT → CH	0.17	26.94	157.00
CH → IT	0.20	0.02	37.80
*Instbot- instant founding population **			
Pre-glacial split PT-IT	66	80	73468
T1	18	119	16216
T2	23	29	19330

Niche modelling



Supplementary Figure 10 – The map in grey depicts the area considered for producing the Species Distribution Model (SDM) for the barn owl. Shading is relative to the altitude. The current distribution of barn owls is plotted atop the map in purple (data from IUCN: BirdLife International 2019). Random presence points were extracted within this distribution for the SDM.



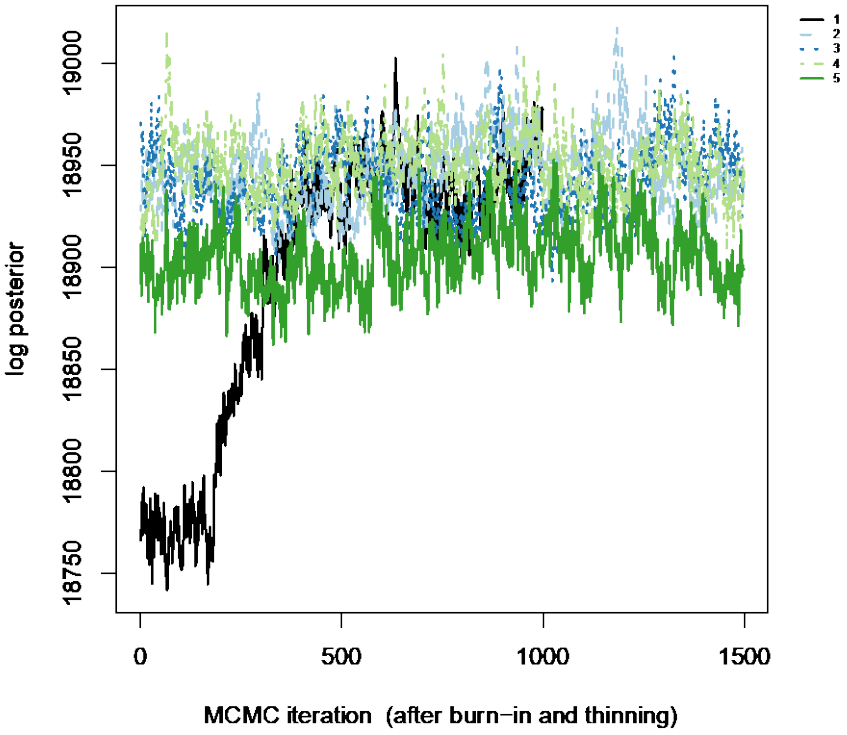
Supplementary Figure 11 – Pairwise correlation between climatic variables retained to produce the SDM. Only variables correlated at less than 0.8 were kept in the models, namely: Mean Diurnal Range (Bio2), Min Temperature of Coldest Month (Bio6), Temperature Annual Range (Bio7), Mean Temperature of Wettest Quarter (Bio8), Precipitation Seasonality (Bio15), Precipitation of Driest Quarter (Bio17) and Precipitation of Coldest Quarter (Bio19).

Supplementary Table 5 - Comparison of SDM model fit. AICc is reported for the multiple combinations of feature (linear, quadratic, hinge) and Beta multiplier (1 to 5).

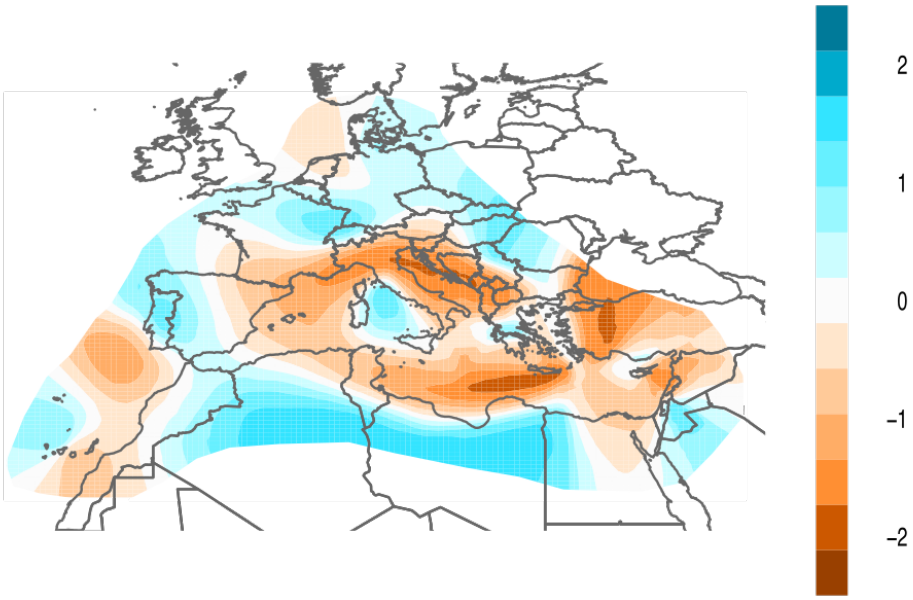
	1	2	3	4	5
Linear	23953.4	23994.8	24037.1	24086.8	24125.8
Quadratic	**23950.8**	24002.6	24042.7	24095.2	24148.5
Hinge	24099.1	24185.4	24241.2	24304.4	24340.2

Barriers and corridors

Migration Surface Estimate

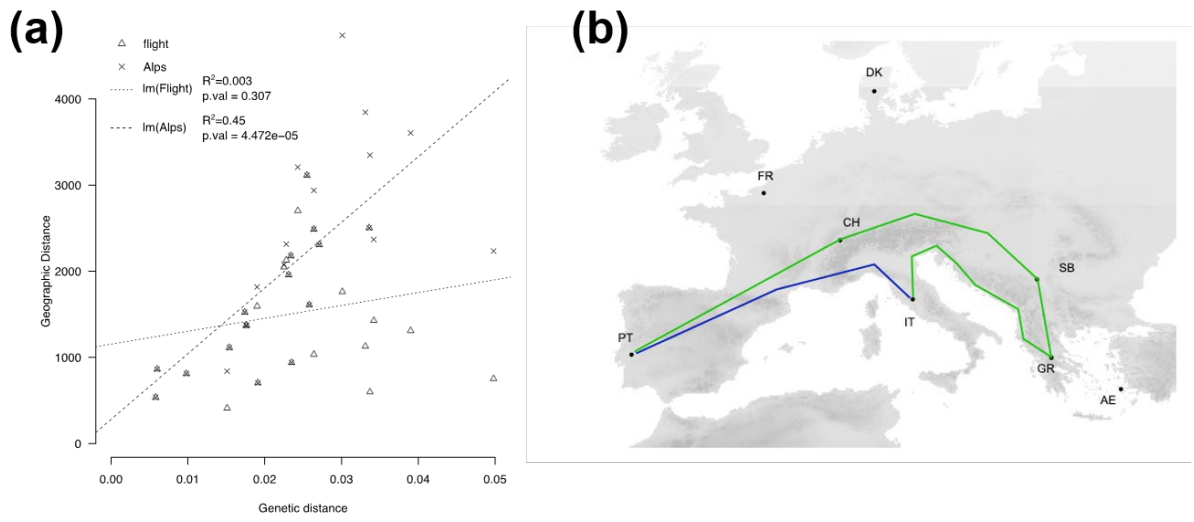


Supplementary Figure 12 - Convergence of MCMC chains for EEMS run in the Western Palearctic. 5 independently seeded MCMC chains reach approximate convergence.



Supplementary Figure 13 – Effective migration surface (EEMS) of barn owls in the Western Palearctic. The map depicts relative barn owl migration in western Palearctic. Blue indicates a greater migration rate over the average; and orange a lower migration than average. Map shows the mean of 5 independent EEMS posterior migration rate estimates between 1000 demes.

Isolation by distance



Supplementary Figure 14 – Isolation by distance in European barn owl. (a) Correlation of the pairwise genetic distance and two geographic distances between barn owl populations. Flight (Δ) refers to the straight-line flight distance over land between populations, while Alps (x) refers to distance between populations by considering the presence of the Alps. Lines depict the linear regression of both comparisons. R^2 and p-values of each model are reported in the legend. (b) Alternative distance models. The example illustrates the shortest overland distance (blue), and distance around the Alps (green) between the populations from Italy (IT) and the Iberian Peninsula (PT).

Chapter 5

The Rocky Mountains as a dispersal barrier between barn owl (*Tyto alba*) populations in North America

Ana Paula Machado[aΔ], Laura Clément[a], Vera Uva[a], Jérôme Goudet[a,j†], Alexandre Roulin[a†]

† co-senior authors

Status: Published in *Journal of Biogeography* (March 2018)

See Appendix S2: https://doi.org/10.1111/jbi.13219

Author contributions

APM, LC, VU, AR, JG designed this study; LC produced microsatellite data; VU produced and treated mitochondrial sequences; APM conducted the analyses of genetic and ringing data; APM led the writing of the manuscript with input from all authors.

DOI: 10.1111/jbi.13219

RESEARCH PAPER

WILEY Journal of Biogeography

The Rocky Mountains as a dispersal barrier between barn owl (*Tyto alba*) populations in North America

Ana Paula Machado[1] | Laura Clément[1] | Vera Uva[1] | Jérôme Goudet[1,2,†] | Alexandre Roulin[1,†]

[1]Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland

[2]Swiss Institute of Bioinformatics, Lausanne, Switzerland

Correspondence
Ana Paula Machado, Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland.
Email: anapaula.machado@unil.ch

Funding information
American Natural History Museum; Akademie der Naturwissenschaften; Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A-120517, 31003A_138180; Université de Lausanne; Basel Foundation for Biological Research

Editor: Camila Ribas

Abstract

Aim: Geological barriers within a species range play a key role in shaping patterns of genetic variation by restricting gene flow. Mountain ranges are particularly imposing barriers responsible for creating genetic differentiation across multiple taxa, from small amphibians to large mammals and birds. Here, we examined the population structure of North American barn owls (*Tyto alba*) and investigated whether the Rocky Mountains influence gene flow and dispersal at the continental scale.

Location: Continental North America.

Methods: We collected 292 museum samples covering the species range, genotyped them at 20 microsatellite markers and sequenced 410 bp of the mitochondrial gene *ND6*. Population and landscape genetics tools were used to study range-wide patterns of structure and identify gene flow barriers. Ring recapture data were also analysed to investigate individual movement patterns and frequency of exchanges between both sides of the Rocky Mountains.

Results: We found faint overall genetic structure, which is consistent with barn owl's high mobility across its continuous range. Nonetheless, we identified two distinct genetic groups on the western and eastern regions of the Rocky Mountains with a likely contact point through the narrow southern pass between them and the Sierra Madre Occidental in Mexico. Accordingly, most recaptured barn owls remain on the same side of the mountains. The Rockies appear to significantly isolate the populations in the west, which, as a consequence, display lower genetic diversity than their counterparts to the east.

Main conclusions: The Rocky Mountains appear to constrain barn owl dispersal and gene flow. Our study supports the hypothesis that regional landscape barriers can shape gene flow and population structure even in highly mobile organisms.

KEYWORDS
barn owl, barrier, dispersal, gene flow, landscape, microsatellites, mitochondrial DNA, population genetics, Rocky Mountains

1 | INTRODUCTION

Spatial patterns of genetic structure and differentiation have long been of interest to evolutionary biologists as they can ultimately

[†]Co-senior authors.

provide hints on the origin of speciation. Such patterns are shaped by the demographic history of populations and geography through the combined impacts of genetic drift and gene flow and selection. An organism's dispersal capacity is responsible for maintaining gene flow and subsequently determines the degree of genetic differentiation between populations. Low levels of dispersal facilitate the emergence of isolation by distance in homogenous or gradually differentiating environments. Intuitively, organisms with low-dispersal capacity differentiate at smaller geographic scales compared to those with higher dispersal capacity (Kisel & Barraclough, 2010). However, the presence of physical barriers can impede gene flow and quickly promote genetic and phenotypic differentiation. In the early days of modern speciation theory, Mayr (1942) posited that physical separation between populations (allopatry) was in fact crucial to initiate speciation.

Barriers to dispersal are often discrete geological or ecological features that create gaps between suitable patches of habitat. These can be of numerous types and act at different scales, from continental separation (Antonelli, 2017) to mountains (e.g. Boutilier, Taylor, Morris-Pocock, Lavoie, & Friesen, 2014), rivers (e.g. Hayes & Sewlal, 2016), forests (e.g. Keyghobadi, Roland, & Strobeck, 1999) and urban development (e.g. Frantz et al., 2012; Munshi-South, 2012). The balance between an organism's capacity to disperse and the magnitude of a given barrier determines the degree of differentiation it will generate. Thus, a barrier to dispersal for one species may have no effect on another species if the latter has the capacity to disperse across this barrier (e.g. Frantz et al., 2012). Mountain ranges have been identified as the third most important factor in biogeographical differentiation after tectonics and climate, mainly due to their role as a barrier to dispersal (Antonelli, 2017) that affects a large variety of taxa (e.g. DeChaine & Martin, 2005a; Huang et al., 2017; Smissen, Melville, Sumner, & Jessop, 2013; von Oheimb et al., 2013).

Here, we explore patterns of genetic structure and the potential impact of a landscape barrier on a nocturnal raptor, the barn owl (*Tyto alba*). This non-migratory cosmopolitan species is one of very few to have colonized all continents but Antarctica as well as many islands (del Hoyo, Elliott, Sargatal, & Christie, 1999; Taylor, 1994). It hunts across an extensive range of different environments, except regions with deep snow cover over large periods of the year. In addition, it has successfully adapted to anthropogenic environments as it often breeds in old buildings and is associated with traditional small-scale mixed farming (Taylor, 1994). Despite having colonized distant islands, both juvenile and adult barn owls disperse relatively short distances (Marti, 1999; van den Brink, Dreiss, & Roulin, 2012). Thus, it is capable of maintaining moderate levels of gene flow across suitable habitats within its range but is also susceptible to the presence of barriers. In Europe, there is a noteworthy pattern of isolation by distance (Antoniazza, Burri, Fumagalli, Goudet, & Roulin, 2010; Antoniazza et al., 2014) and populations isolated by large water barriers have accumulated substantial genetic differences as, for example, in the Canary Islands and Crete (Burri et al., 2016). However, European mountain ranges like the Pyrenees do not appear to affect population differentiation to the same degree (Burri et al., 2016) illustrating how barriers with different magnitudes can have varying impacts on the same organism.

In North America, the Rocky Mountains are a major high-elevation mountain range (up to 4,400 m) that extends over 4,800 km from north–west Canada to New Mexico (USA). The complexity of the landscape along the west coast is enhanced by smaller contiguous perpendicular ranges such as the Klamath Mountains, creating a patchwork of relatively isolated lowland. Examples of the Rockies' role as a landscape barrier to gene flow are numerous in plants (e.g. DeChaine & Martin, 2005b), terrestrial mammals (e.g. Geffen, Anderson, & Wayne, 2004; Schwartz et al., 2009), insects (e.g. DeChaine & Martin, 2005a) and passerines (e.g. Milot, Gibbs, & Hobson, 2000; Spellman, Riddle, & Klicka, 2007). There is even evidence of it impacting widespread and highly mobile birds such as red-tailed hawks (*Buteo jamaicensis*; Hull, Hull, Sacks, Smith, & Ernest, 2008). Unexpectedly, in a first genetic analysis of American barn owls, Huang et al. (2016) did not find any significant pattern of genetic differentiation across continental North America.

Using an established set of 20 microsatellite markers and a 410 bp sequence of the mitochondrial gene *ND6*, we examined neutral genetic variation across the range of the barn owl in North America. We employed approximately twice as many genetic markers and individuals as Huang et al. (2016) to increase resolution, and we compared our respective findings in detail. In our approach, we combine population genetics tools with analyses of recapture of ringed barn owls to characterize movement and gene flow at a range-wide scale. Furthermore, we specifically tested whether the Rocky Mountains are a physical barrier to gene flow and dispersal among populations.

2 | MATERIALS & METHODS

2.1 | Sampling and molecular analyses

We obtained barn owl samples across most of North America from natural history museums in USA and Canada. Samples were collected from either internal organs stored in 80% ethanol or muscle tissue from frozen carcasses dating from 1986 to 2011 (Table S1 in Supporting Information). Individuals were grouped into 10 populations by geographic proximity (Table 1; Figure 1; Table S1).

Genomic DNA was extracted using the BioSprint 96 DNA blood kit or the DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany) following manufacturer's instructions. All 292 individuals were genotyped at 20 microsatellite markers (Burri et al., 2008; Klein et al., 2009). Polymerase chain reactions (PCR) were carried out in a total volume of 8 μl, containing 2.5 μl of Qiagen Multiplex PCR Master Mix, 3 μl of DNA and varying concentrations of primers organized into four multiplexes (see Antoniazza et al., 2014; except Ta-206 and Oeo53). PCR thermal cycling was performed as follows: an initial denaturation step at 95°C for 15 min, 34 cycles at 94°C, 90 s at 57°C and 1 min at 72°C, with a final elongation step at 60°C for 30 min. The PCR products were analysed by capillary electrophoresis in a ABI 3100 sequencer (Applied Biosystems) with internal size standard GeneScan-500 LIZ. Fragment length analyses and scoring

TABLE 1 Diversity estimates per population for microsatellites and *ND6* in North American barn owls

			Microsatellites						*ND6*				
Population	**Abbrev.**	***N***	N_A	A_R	**PA**	H_o	H_e	F_{IS}	**S**	H_N	**π (SD)**	H_D **(SD)**	**Θ**
Florida	FLO	24	4.9	3.853	1	0.516	0.525	0.026	6	7	0.0021 (0.0005)	0.598 (0.107)	0.0039
New Jersey	NJ	11	4.2	4.006	3	0.513	0.540	0.036	5	5	0.0026 (0.0009)	0.618 (0.164)	0.0042
Louisiana	LOUI	19	4.9	3.954	4	0.500	0.509	0.023	6	7	0.0020 (0.0005)	0.608 (0.127)	0.0042
Ontario	ON	9	3.7	3.650	0	0.506	0.514	0.040	4	5	0.0026 (0.0008)	0.722 (0.159)	0.0036
New Mexico	NM	28	5.1	3.993	2	0.499	0.523	0.037	6	7	0.0018 (0.0005)	0.492 (0.114)	0.0038
California Los Angeles	CALA	41	5.2	3.814	2	0.504	0.509	0.001	5	6	0.0013 (0.0003)	0.485 (0.089)	0.0029
California San Francisco	CASF	41	5.4	3.862	1	0.481	0.507	0.036	3	4	0.0008 (0.0002)	0.304 (0.087)	0.0017
Oregon	OR	40	5.0	3.733	2	0.490	0.523	0.044	1	2	0.0004 (0.0002)	0.180 (0.075)	0.0006
Idaho	ID	38	5.0	3.711	1	0.474	0.505	0.050	2	3	0.0005 (0.0002)	0.203 (0.084)	0.0012
Washington	WA	41	5.0	3.709	3	0.502	0.522	0.027	2	3	0.0002 (0.0002)	0.096 (0.062)	0.0011
TOTAL		292				0.498	0.518	0.037	14	14	0.0011 (0.0001)	0.367 (0.036)	0.0055

N, number of individuals; N_A, mean number of alleles; A_R, allelic richness corrected for uneven sample sizes; PA, private alleles; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, inbreeding coefficient; S, number of polymorphic sites; H_N, number of haplotypes; π, nucleotide diversity; H_D, Haplotype diversity; Θ, Watterson's theta.

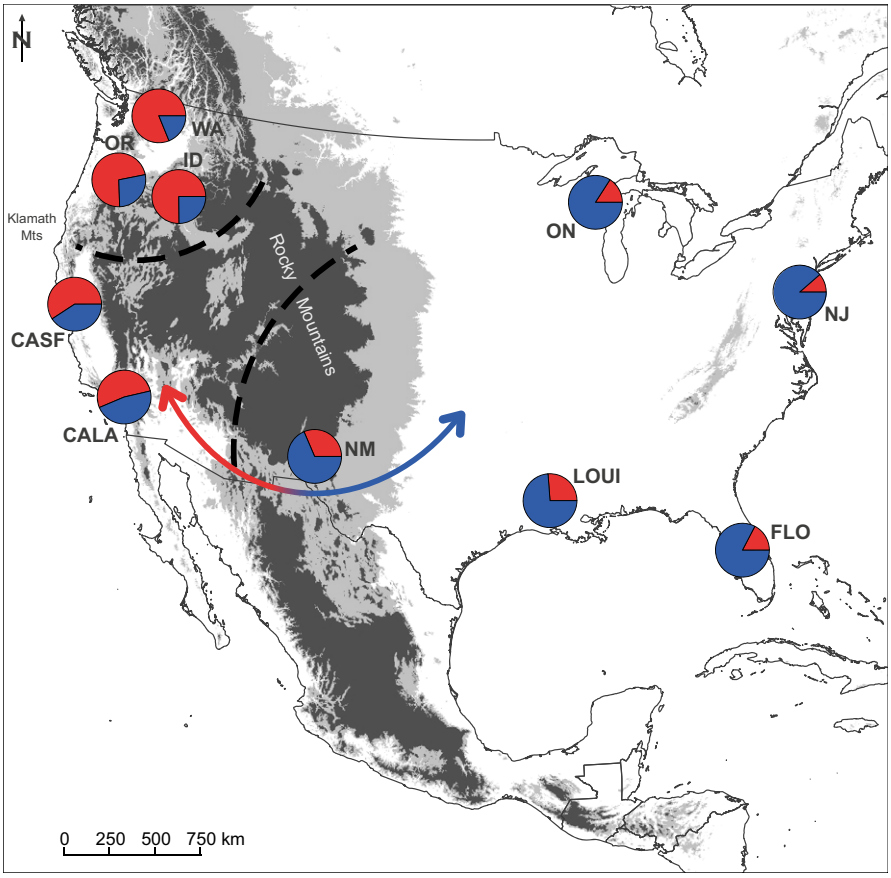


FIGURE 1 Population sampling location in North America. Abbreviations follow Table 1. Background colour denotes elevation: white below 750 m; light grey between 750 m and 1500 m; dark grey above 1500 m. Pie charts indicate admixture proportions per population between the west (red) and east (blue) lineages determined by STRUCTURE; the coloured arrow illustrates the contact route around South of the Rocky Mountains; and the dashed black lines indicate the barriers to gene flow identified by Monmonier's algorithm. Albers Equal Conic map projection

were performed in GeneMapper 4.0 (Applied Biosystems). Despite our samples being relatively recent, we tested for a relation between heterozygosity with year of sampling to control for the possibility of allelic dropout in older samples. Microsatellite data were tested for the presence of null alleles with Micro-Checker 2.2.3 (Van Oosterhout, Hutchinson, Wills, & Shipley, 2004) as well as for deviation from Hardy–Weinberg equilibrium and linkage disequilibrium between markers with Arlequin 3.5.2.2 (Excoffier & Lischer, 2010).

Additionally, a 410 bp fragment of the mitochondrial gene NADH dehydrogenase 6 (*ND6*) was sequenced for all individuals

using previously developed primers (Burri et al., 2016). PCR reactions were carried out in a total volume of 25 μl, containing 1× Buffer Gold, 2 mM MgCl2, 1× Q-Solution (Qiagen), 0.2 mM dNTP, 0.5 μM each primer, 1 U Taq Gold (Applied Biosystems) and 20 ng of genomic DNA. PCR thermal cycling was composed of an initial denaturation step for 7 min at 95°C, 35 cycles of 30 s at 95°C, 45 s at 62°C and 45 s at 72°C, and final elongation for 7 min at 72°C. Amplified DNA was sent for sequencing at Microsynth AG (Balgach, Switzerland). Mitochondrial DNA (mtDNA) sequences were aligned with MEGA 7 (Kumar, Stecher, & Tamura, 2016).

2.2 | Population genetic diversity and structure

To describe neutral variation, we calculated allelic richness (A_R; rarefied based on nine individuals), observed and expected heterozygosity (H_o and H_e) and F_{IS} per population as well as overall and pairwise F_{ST} values (Weir & Cockerham, 1984) between populations based on the 20 microsatellite data with the "hierfstat" package 4-22 (Goudet, 2005) in R 3.3.2 (R Development Core Team, 2016). Polymorphism data for mtDNA were obtained with DNASP 5.10 (Librado & Rozas, 2009) and global and population pairwise F_{ST} were calculated with ARLEQUIN. A median-joining haplotype network (Bandelt, Forster, & Röhl, 1999) was constructed in PopArt (Leigh & Bryant, 2015).

We examined patterns of isolation by distance as the correlation between pairwise genetic (F_{ST}) and Euclidean geographic distance matrices using a Mantel test for both microsatellites and *ND6*. The test was implemented in the R package "ecodist" 2.0.1 (Goslee & Urban, 2007) with 10^7 permutations to assess significance and 10^4 bootstrap iterations to determine the 95% confidence limits. Then, we contrasted the isolation by distance patterns in North America and in Europe (microsatellite data from Burri et al., 2016).

To infer the main axes of differentiation between populations, we performed principal coordinate analyses (PCoA) for microsatellite and *ND6* data as implemented in "hierfstat." At the individual level, we performed a principal component analysis (PCA) in "hierfstat" and used STRUCTURE 2.3.4 (Pritchard, Stephens, & Donnelly, 2000) to estimate the number of genetic clusters among our samples and individual admixture proportions based on microsatellite data. Analyses were run 10 times for each estimated *K* from 1 to 10, with 10^5 burn-in and 10^6 Markov chain Monte Carlo (MCMC) steps under an admixture model with correlated allele frequencies. To infer the most likely number of *K*, we followed the Δ*K* Evanno method (Evanno, Regnaut, & Goudet, 2005) as implemented in Structure Harvester (Earl & von Holdt, 2012).

2.3 | Barriers to gene flow

To determine the potential influence of the Rocky Mountains on gene flow and genetic differentiation among barn owls, we performed a series of analyses divided into two main steps. First, we aimed to determine whether the mountains have a measurable impact on gene flow employing (1) a partial Mantel test, (2) covariance analyses and (3) Monmonier's algorithm. Second, we compared (1) genetic diversity between populations on either side of the Rockies and (2) tested different routes of contact between them.

To investigate whether the Rocky Mountains act as a barrier to gene flow, we employed a partial Mantel test with the R package "ecodist" using the same parameters as for the Mantel test (see "Population genetic diversity and structure" above). The partial Mantel test allows us to test the correlation of a barrier with genetic differentiation (F_{ST}) while controlling for Euclidean geographic distance between populations. The barrier effect was encoded into a matrix as 0 or 1 if the populations were on the same or opposite sides of the Rockies.

Despite being commonly used in population genetics to detect ecological barriers to gene flow, partial Mantel test has been widely criticized (e.g. Legendre, Fortin, & Borcard, 2015). Indeed, this test performs poorly if the relationship between the distance variables is not strictly linear (Legendre & Fortin, 2010) and tends to produce high type I error rates when the variables are spatially autocorrelated (Bradburd, Ralph, & Coop, 2013; Guillot & Rousset, 2013). Additionally, Bradburd et al. (2013) noted that it does not allow us to quantify the relative effects of the ecological (barrier or gradient) and geographic distances. Thus, we employed a new method with the R package "Sunder" 0.0.4 (Botta, Eriksen, Fontaine, & Guillot, 2015), an extension of the original package "BEDASSLE" for biallelic markers developed by Bradburd et al. (2013). The rationale behind Bradburd's method lies on the expectation that allele frequencies in local populations covary in their deviation from the global mean allele frequency. The covariance among pairs of populations, which is summarized from raw allele counts into a matrix, is modelled as a decreasing function of geographic and ecological distance between the populations (Bradburd et al., 2013). Importantly, the package estimates model parameters such as the magnitude of the effects of the geographic and ecological variables in a Bayesian framework using a MCMC algorithm. "Sunder" further expands on this by (1) allowing the analysis of microsatellites through a multinomial approach (versus the previous binomial) and (2) incorporating a cross-validation step to select the most suitable model of genetic differentiation ("G": geographic, "E": ecological or "G+E": both), among others (Botta et al., 2015). Thus, we used "Sunder" to assess the contribution of the Rocky Mountains as an ecological barrier to genetic differentiation between barn owl populations using our microsatellite data. We launched 10 independent MCMC runs for 10^7 iterations, sampling every 1,000 steps, with equal testing of the three models—G, E and G+E. We set uniform priors with large upper bounds as recommended by the authors and updated all parameters in the MCMC iterations: α *e* [0, 1], indirect estimate of variance of allele frequencies; βG *e* $[0, 40{,}000]$ & βE *e* [0, 20], magnitudes of effect of geography and ecology, respectively; γ *e* [0, 1], smoothness of spatial variation in allele frequencies; and δ *e* [0, 1], degree of departure in allele frequency between neighbouring populations (named the nugget coefficient). We used 10% of the dataset (loci × locations) as validation set in cross-validation for model selection. Matrices of geography and barrier were the same as for the partial Mantel test above.

Finally, we employed Monmonier's maximum difference algorithm (Monmonier, 1973), as implemented in the R package "adegenet" 2.0.1 (Jombart, 2008), to detect the presence of barriers. This method developed by Manni, Guerard, and Heyer (2004) is based on Delaunay triangulation (Brassel & Reif, 1979) from the sampling coordinates and uses Monmonier's algorithm to find the path that displays the largest cumulative distances between connected vertices. We used the first axis of microsatellite PCoA as a measure of genetic distance between populations, using the third quartile of all distances between neighbours (default) as threshold.

On the second part of our analyses, we tested if populations on either side of the mountains have similar levels of genetic diversity —microsatellite allelic richness and observed heterozygosity and *ND6* nucleotide diversity and Watterson (1975)'s Θ estimator—using Student's *t* tests. Populations were divided into an east (Ontario, New Jersey, Louisiana, Florida and New Mexico) and a west (Washington, Oregon, Idaho and California) group.

Finally, we contrasted different routes of contact between populations on either side of the Rockies (see Figure S1 in Supporting Information) by analysing how different pairwise paths between populations explained genetic differentiation. We compared direct "flight distance" (Euclidean) independently of the presence of the mountains (orange line in Figure S1), with two distances accounting for the mountains as a barrier for dispersal, one through a northern pass (blue line in Figure S1) and the other around its southern tip (green line in Figure S1). The alternative routes were calculated by considering lower elevation points in the northern and southern parts of the Rocky Mountains within the natural distribution of barn owls as mandatory waypoints for individuals to move between opposite sides of the main range of the Rocky Mountains. These were approximated as (1) the cities of Yakima (Washington) and Helena (Montana) in the north and (2) Los Angeles (California) and Albuquerque (New Mexico) in the south. By simulating the three different routes, the relative distance between populations changes. For example, populations Ontario and Washington are fairly close in both the direct flight and northern routes, but very far in the southern route. Using these three sets of pairwise distances, we used Mantel tests to measure their relation to pairwise genetic differentiation between populations. Then, we used linear models, evaluated using the Akaike Information Criterion (AIC; Akaike, 1974), to determine which of these routes best explains the observed population genetic structure described by the first axis of the PCoA. This axis portrayed most of genetic variation among our populations thus eliminating population-specific noise, and the use of a single axis allows the testing of linear relations.

2.4 | Dispersal analyses

We analysed ring recapture data of North American and European barn owls to investigate (1) whether the Rocky Mountains shape the population structure by physically conditioning the movement of individuals across it and (2) the differences in dispersal distances between the two continents. North American data were obtained from USGS Bird Banding Laboratory (2017) and European from EURING (Du Feu, Clark, Schaub, Fiedler, & Baillie, 2016; Speek, Clark, Rohde, Wassenaar, & Noordwijk, 2001).

For all analyses, we filtered the raw data on date and coordinate accuracy as similarly as possible between the North American and European datasets based on their respective ringing system codes. Thus, we kept points with dates accurate up to the same month level in North America and to the 2-week range level in Europe. Additionally, we kept coordinates accurate up to a 10-min latitude × longitude block for North America and up to a 10-km radius for Europe.

To assess how often barn owls cross the Rocky Mountains, we kept all recapture points available following the date and coordinate precision filters, which yielded a dataset with a total of 3979 movements by 3734 individuals, between 1923 and 2017. We defined a crossing when an individual was captured on one side of the mountains and recaptured on the opposite side. The main ridge of the mountains was identified as the large dark grey patch labelled "Rocky Mountains" in Figure 1. Then, we compared movement distances and direction of barn owls ringed on either side of the Rockies. The Rocky Mountains are expected to be a barrier in a continent-wide context; thus, we anticipate that they will not necessarily restrict short-range movements but rather medium to long range. Thus, we compared distances larger than 60 km (4th quartile of the distribution of all recorded movements, $n = 967$) between both sides of the mountains with a nonparametric Mann–Whitney *U* test. Furthermore, we calculated the angle of these movements on a flat Earth (i.e. WGS84 projection) with the R package "adehabitatLT" 0.3.21 (Calenge, 2011) to investigate whether it may also be influenced by the barrier.

For comparisons between North America and Europe, we considered natal dispersal, the dispersal of juveniles from their birthplace to their first breeding site, as it is typically larger than between breeding years (Van Den Brink et al., 2012). Given recapture data do not provide information on breeding success, we based our analysis on sexually mature individuals in their first breeding season which approximates to 1-year-old barn owls. Thus, we calculated natal dispersal as the distance between the ringing location of an owl as nestling not capable of flight (up to 55 days old), and the location of its recapture the following breeding season. We set an age bracket for recapture between 6 and 18 months after ringing to account for variation in hatching date and ringing age. When individuals were recaptured multiple on occasions within this time range, we kept only one point as representative of the breeding date and location, selected as following: if the bird was found nesting, we kept the earliest date; if the bird was not found nesting, we kept the point closest to the peak of the breeding season (approximated to the 1 July). These filtering steps yielded datasets of 784 recaptured individuals in North America from 1925 to 2017 and 19641 individuals in Europe from 1910 to 2015. We used a Mann–Whitney *U* test to compare natal dispersal distances between both continents.

Recapture data are directly dependent on whether the general public reports ringed individuals, explaining in part the variation in the number of data for each region, which could bias our

comparison. Ringing and recapture locations in North America are heavily distributed along the coasts but considerably sparser in the interior of the continent (Figure 4). On the contrary, most locations in Europe are concentrated in the north-west and the United Kingdom with few data points along the eastern and southern peripheries. Therefore, medium distances could be underrepresented in North America and long distances in Europe simply due to different likelihoods of the public reporting ringed barn owls between regions. To account for this, we made a second set of Mann–Whitney *U* tests only on short-range distances which have a more comparable chance of being detected in both continents. We compared short-range natal dispersal between the two continents using distances below (1) 150 km, (2) 100 km and (3) 50 km. The thresholds were chosen so as to encompass the 3rd quartile of the distributions (52.18 km in Europe and 92.53 km in North America). Bonferroni correction for multiple testing was applied with four tests, setting the significance threshold at 1.25×10^{-2}.

2.5 | Comparison to previous results

In their study, Huang et al. (2016) used eight microsatellite markers and one mitochondrial gene (*ND2*) on 126 individuals, of which only 13 were not from the west coast. Our results pointed to the presence of neutral structure and isolation by distance, whereas Huang et al. (2016) found no such signal. To determine the cause of these dissimilarities, we ran specific clustering tests by varying (1) the number of markers used and (2) the geographical sampling.

First, we randomly chose 20 different sets of eight markers amongst our 20 markers in a bootstrap-like fashion to estimate how often we obtained similar genetic clustering results to those with the full dataset. Second, we replicated the geographical sampling of Huang et al. (2016) by removing from our dataset the populations of Florida, Louisiana and New Mexico (total 221 individuals). In this case, our populations of New Jersey and Ontario (20 individuals) stood for Huang et al. (2016)'s Pennsylvania (13 individuals) as north-east representatives. To avoid biasing the results because of the much heavier sampling in the west coast, we also randomly reduced the number of individuals in these populations (total 117 individuals) to match the proportions in Huang et al. (2016). We ran STRUCTURE for each test as described above under an admixture model, for *K* 1 to 5. For the bootstraps of eight loci, we included a location prior to account for the low number of markers (Hubisz, Falush, Stephens, & Pritchard, 2009) as done by Huang et al. (2016).

3 | RESULTS

3.1 | Genotyping

Genotyping was highly successful resulting in a final dataset with only 2.6% missing data. We found no relation between sample age and heterozygosity ($t = 0.891$, $p = 0.375$, $R^2 = 0.008$) indicating allele drop-out in older samples was unlikely.

Two loci showed signs of deviation from Hardy–Weinberg equilibrium as well as significant linkage disequilibrium for the Washington population. It is unlikely that these loci are physically linked on the same chromosome as they have been extensively used in the past with no evidence of linkage (e.g. Burri et al., 2008, 2016). Closer inspection revealed particularly low polymorphism for these markers in this specific population, and running STRUCTURE with and without them yielded similar results. We therefore kept all loci for subsequent analyses.

3.2 | Population genetic diversity and structure

Average allelic richness ranged from 3.65 to 4.01 among populations, and *ND6* haplotype diversity ranged from 0.10 to 0.72 (Table 1). We identified 14 *ND6* haplotypes with no more than four diverging sites between them. One haplotype was particularly frequent and present in all populations, resulting in a short star-like haplotype network centred around it (Figure 2a).

Overall Weir and Cockerham's (1984) F_{ST} was 0.018 (95% confidence interval 0.011–0.025) based on microsatellites and 0.021 (95% confidence interval 0.015–0.035) based on *ND6*. We found a significant pattern of isolation by distance (Mantel $r = 0.712$, $p = 0.001$) in North America, which was considerably weaker than that of Europe ($t = -10.889$, $p < 0.0001$; Figure 3). The highest pairwise F_{ST} detected with microsatellite data were between Washington and Idaho in the north-west and the populations of New Jersey and Florida on the east coast (Table 2). Similarly, the highest pairwise F_{ST} for

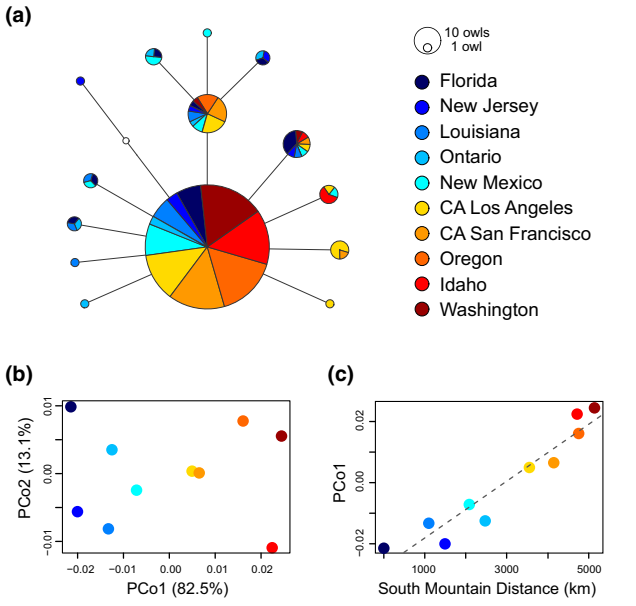


FIGURE 2 Population structure of North American barn owls. a) *ND6* median-joining haplotype network. Circle size is proportional to the number of individuals carrying that haplotype, and colours indicate their population of origin. A branch length denotes one mutation, and the empty circle a missing haplotype. b) Principle Coordinate Analysis (PCoA) based on microsatellite data of North American barn owls and **c)** correlation of PCo 1 and the distance around South of the Mountains

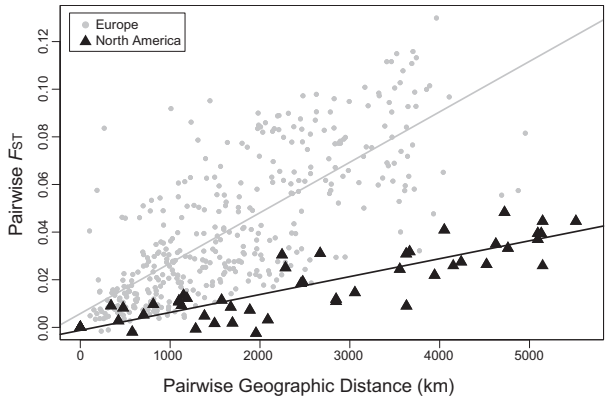


FIGURE 3 Comparison of microsatellite isolation by distance slopes of European and North American barn owls. Geographic distances represented are from the most likely demographic scenario of each continent: Ring distance in Europe (Burri et al., 2016), and around South of the Rockies in North America

mitochondrial data were between the north-western populations (Washington, Idaho and Oregon) and eastern populations (Ontario, New Jersey and Florida; Table 2).

The first PCoA axis described most of the variation in our genetic data (82.5%; Figure 2b) and clearly differentiated the populations from the east and west coasts. The same pattern was detected in the individual-based PCA (see Figure S2). Accordingly, admixture analyses in STRUCTURE found the highest support for two genetic clusters among our samples ($\Delta K(2) = 53.46$; Figure S3), separating populations from the east and the west (Figure 1). F_{ST} between the eastern and western groups was 0.022 for microsatellites and 0.060 for mtDNA. Interestingly, Californian populations (CA Los Angeles and CA San Francisco) showed high levels of admixture between the two lineages, consistent with their intermediate position along the first PCoA axis (Figures 1 and 2).

3.3 | Barriers to gene flow

The partial Mantel test detected a significant effect of barrier by the Rockies on genetic differentiation with microsatellite (Mantel $r = 0.522$, $p = 0.008$) and *ND6* data (Mantel $r = 0.415$, $p = 0.017$). Likewise, covariance analyses with "Sunder" identified the "geography + barrier" model as the most probable to influence genetic differentiation in all 10 runs (Table S3). Analyses with Monmonier's algorithm indicated the presence of two barriers (Figure 1). The first separated populations almost longitudinally, specifically along the axes between New Mexico and all populations to the west. This analysis identified a second barrier between California and the north-western populations (Washington, Oregon and Idaho).

Populations west of the Rocky Mountains had significantly lower nucleotide diversity ($t = 6.126$, $p < 0.001$) and Watterson's estimator Θ ($t = 6.084$, $p = 0.002$) of mitochondrial DNA than the populations east of the range (Table 1). Likewise, microsatellite observed heterozygosity was lower in the west ($t = 2.639$, $p = 0.034$). Despite showing the same tendency, the difference in allelic richness was not significant ($t = 1.725$, $p = 0.140$). Ontario showed particularly low levels of allelic richness, which skewed the eastern group's mean. Accordingly, if we excluded this population, the result was similar for allelic richness ($t = 3.951$, $p = 0.006$) as for the other diversity estimates.

The performed Mantel tests detected significant correlations of genetic differentiation with nearly all distances tested (Table 3; Figure S4). However, the distance around South of the Rocky Mountains had the best fit to both microsatellites (Mantel $r = 0.857$) and mtDNA (Mantel $r = 0.685$) compared to the other distances (Table 3; Figure S4). Similarly, AIC scores from linear models (Table S4) showed that the distance around the southern part of the Rocky Mountains best explained the first PCoA axis for microsatellite ($t = 10.003$, $p < 0.0001$, $R^2 = 0.926$; Table 3; Figure 2c) and mtDNA ($t = 6.038$, $p < 0.001$, $R^2 = 0.820$; Table 3; Figure S4f).

3.4 | Dispersal analyses

Of the 3734 ringed barn owls recaptured across North America, only five individuals (0.13%) crossed the Rocky Mountains (Figure 4; Table S5). When contrasting all movements between both sides of the Rockies, there was no difference in the distance travelled ($U = 1970900$, $p = 0.412$). However, when considering only

TABLE 2 Populations pairwise F_{ST} for microsatellite (below diagonal) and *ND6* (above diagonal) of North American barn owls

Population	FLO	NJ	LOUI	ON	NM	CALA	CASF	OR	ID	WA
FLO		−0.049	−0.030	−0.027	−0.008	0.015	0.056	0.122	0.103	0.171
NJ	0.002		−0.050	−0.069	−0.031	−0.021	0.023	0.108	0.101	0.202
LOUI	0.010	0.008		−0.049	−0.029	−0.011	0.026	0.092	0.094	0.165
ON	0.003	−0.001	0.001		−0.006	0.000	0.077	0.182	0.178	0.303
NM	0.019	0.010	0.005	−0.002		−0.009	0.003	0.044	0.038	0.092
CALA	0.024	0.022	0.020	0.012	0.010		0.003	0.048	0.059	0.101
CASF	0.026	0.026	0.016	0.009	0.008	−0.002		−0.010	0.024	0.037
OR	0.032	0.038	0.032	0.026	0.025	0.012	0.005		0.018	0.006
ID	0.049	0.041	0.032	0.038	0.029	0.013	0.009	0.008		0.011
WA	0.045	0.045	0.041	0.035	0.031	0.011	0.009	0.002	0.008	

Abbreviations follow Table 1.

TABLE 3 Mantel tests and PCo 1 linear model results comparing the three routes: Direct Flight (Euclidean), around North of the Rocky Mountains and around South of the Rocky Mountains

		Mantel test		PCo 1	
Route	Marker	*r*	*P*	R^2	*P*
Direct Flight (Euclidean)	Microsatellites	0.712	0.001	0.814	<0.001
	ND6	0.424	0.020	0.525	0.018
North of Mountain	Microsatellites	0.628	0.004	0.684	0.003
	ND6	0.279	0.055	0.370	0.062
South of Mountain	Microsatellites	0.857	<0.0001	0.926	<0.0001
	ND6	0.685	0.001	0.820	<0.001

distances above 60 km (4th quartile), eastern barn owls travelled further than those confined to the western side of the mountains (median east 193.95 km and west 114.60 km; $U = 150920$, $p < 1.0 \times 10^{-16}$; Figure S5). As for the direction of movements, these were mostly NW or SE in the western side of mountains (Figure 5). In the east, however, large-range movements displayed a clear trend south, whereas medium-range movements occurred in all directions (Figure 5).

Comparisons of natal dispersal distance between continents yielded consistent qualitative results across all tests (Table S6). When considering the complete dataset, North American barn owls travelled a median distance of 23.73 km from their birth place whereas the European only 16.61 km ($U = 8810300$, $p = 1.05 \times 10^{-12}$). Similarly, analyses of short-range distances (up to 50, 100 or 150 km) indicated individuals in North America disperse significantly larger distances than Europeans (Table S6).

3.5 | Comparison to previous results

Tests with eight microsatellite loci (as in Huang et al., 2017) yielded the same two-cluster signal as the 20 loci dataset in 100% of runs (Table S7). Indeed, the clustering was even more pronounced in some cases, likely due to the use of location as prior. This indicates that eight microsatellite markers should be sufficient to detect population structure across continental North America in the barn owl. As for the tests replicating the geographical sampling of Huang et al. (2017), STRUCTURE failed to capture any clustering signal with both all west coast samples (Ln*K*(1) = −9380.39) and with the reduced subset (Ln*K*(1) = −5015.22).

4 | DISCUSSION

We found clear genetic differentiation between barn owl populations along the east–west axis of their North American range, and we highlighted the potential role of the Rocky Mountains in reducing longitudinal gene flow and dispersal across it. Population structure and genetic diversity analyses revealed that north-western populations are

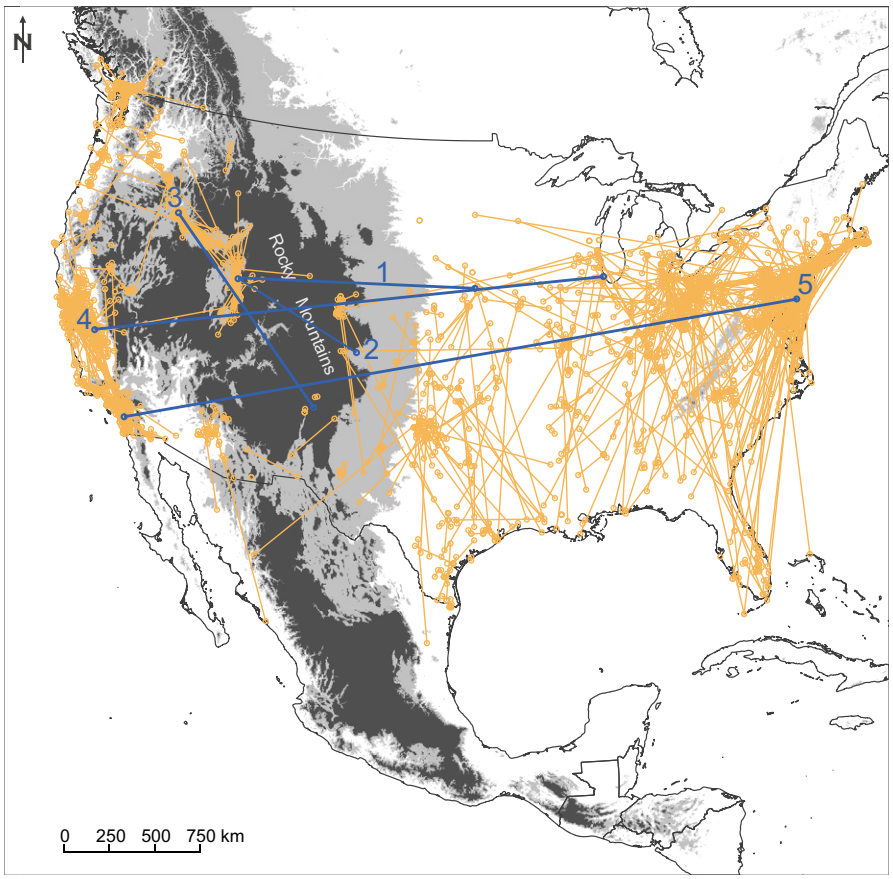


FIGURE 4 Barn owl movements across North America, based on ring recapture data courtesy of USGS Bird Banding Laboratory (2017). Lines simply connect two capture points and do not represent the actual path travelled by the individuals. Movements classified as crosses over the Rocky Mountains are highlighted in blue and numbered to match Table S5. Albers Equal Conic map projection

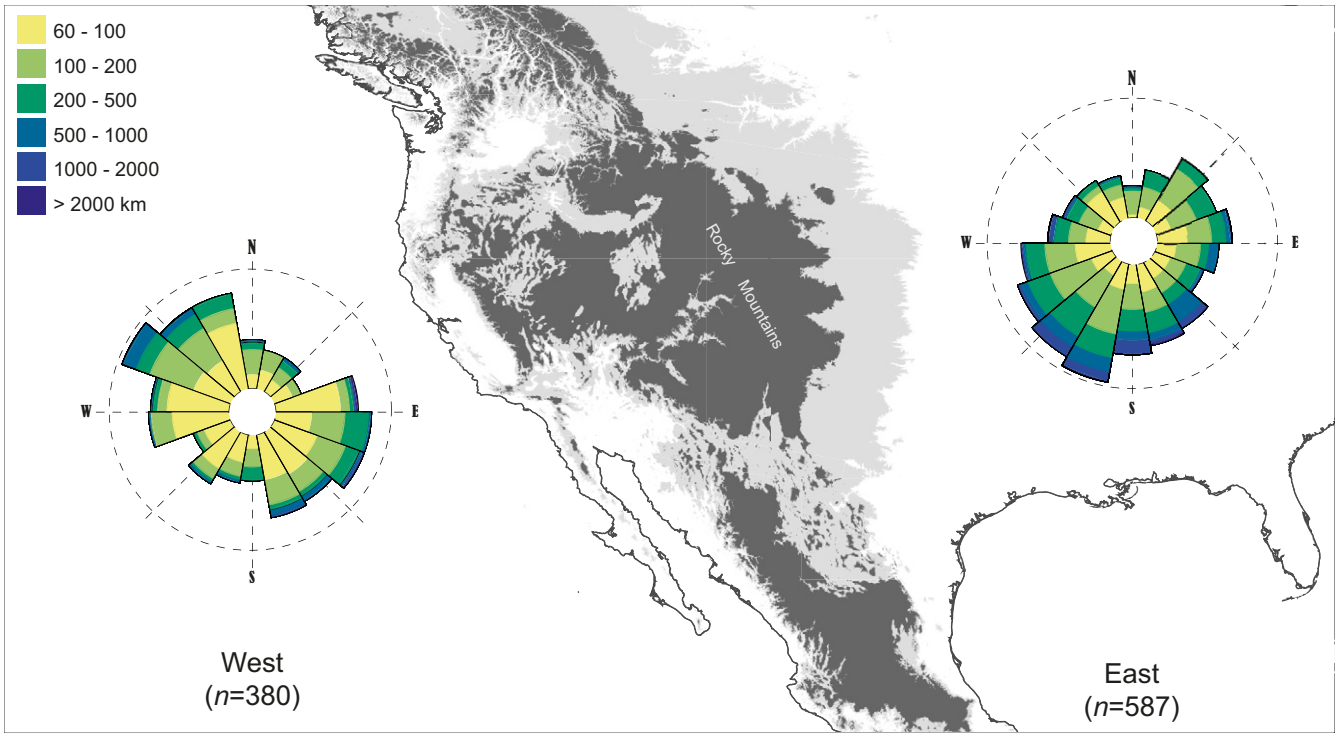


FIGURE 5 Direction and distance of barn owl movements on each side of the Rocky Mountains. Each circular plot is composed of 18 bins of 20 compass degrees; length of a bin indicates the amount of movements in that angle; colours indicate the proportion of each distance class in a bin according to the inset legend. Recapture data courtesy of USGS Bird Banding Laboratory (2017). Background colour represents elevation as in Figure 1. WGS84 map projection

particularly isolated from the remaining ones, a segregation that coincides with the main crest of the Klamath Mountains, contiguous to the Rockies. Our extensive sampling encompassing the distribution of the species in North America allowed us to uncover these results, which are not in line with previous work (Huang et al., 2016). In our study, we emphasize how large-sized landscape features can shape gene flow even in highly mobile and widespread organisms.

4.1 | Population structure

Overall, we found less differentiation and longer natal dispersal distances among barn owl populations in North America than in Europe. There was considerably lower neutral genetic structure (F_{ST} = 0.018 versus 0.045 for microsatellites and 0.021 versus 0.134 for *ND6*), weaker isolation by distance (Figure 3) and more pronounced natal dispersal in North America than in Europe (Table S6), reinforcing the image of more movement and exchange between populations. It is worth noting that North American barn owls are roughly 50% larger in body size than Europeans (Taylor, 1994), but it is unclear whether this might be related with dispersal capacity. Our results could also reflect better large-scale connectivity between suitable habitat, such as grassland, prairies and agricultural old-field in North America.

We revealed a clear differentiation along the east–west axis of continental North America (Figure 1), a pattern that has been observed in other avian taxa (Hull, Strobel et al., 2008; Klicka, Spellman, Winker, Chua, & Smith, 2011; Proudfoot, Gehlbach, & Honeycutt, 2007). Specifically, Bayesian analysis found two different genetic lineages, separating the west coast populations from those in the centre and the east coast. However, differentiation between the two lineages in space is not simply linear, as south-western populations from California displayed high levels of admixture between the two (Figures 1, S2, and S3a). Moreover, the highest differentiation (as measured by pairwise F_{ST}) was found between the north-western populations and essentially all east coast populations (Table 2). This suggests barn owl genetic structure is more complex than isolation by distance.

4.2 | Role of landscape: Rocky Mountains

The differentiation observed between the populations of North American barn owls appears to coincide with the Rocky Mountains (Figure 1), as supported by both genetic and ring recapture data analyses. This imposing assembly of mountains and plateaus is arguably the most notorious barrier in North America. It not only creates genetic differentiation at the intraspecific level (DeChaine & Martin, 2005b; Geffen et al., 2004; Hull, Hull et al., 2008; Milot et al., 2000; Schwartz et al., 2009; Spellman et al., 2007) but is also associated with species distribution breaks, illustrating its important phylogeographic role (Antonelli, 2017; DeChaine & Martin, 2005a).

The landscape between the major ridge of the mountains and the west coast is rugged and heterogeneous, in stark contrast to the extensive plains to the east (Figure 1). Western barn owls are thus

confined to pockets of somewhat isolated populations in low-elevation patches, namely the north–west, south–west and large valleys within the mountains (Figure 4). Restricted by the ocean to the one side and the mountains to the other, their long-range movements follow strict NW and SE directions (Figure 5). However, the exchange of individuals between suitable patches is reduced (Figure 4) which leaves a trace at the genetic level. There was a measurable degree of genetic differentiation between the populations of California and those in the north–west (Figures 1 and 2b), implying the presence of a latitudinal barrier to gene flow. Accordingly, Monmonier's algorithm identified a second barrier coinciding with the Klamath Mountains (Figure 1), a pattern that has been described in other taxa (e.g. Soltis, Gitzendanner, Strenge, & Soltis, 1997), including some owl species (Barrowclough, Groth, Mertz, & Gutiérrez, 2005; Haig, Mullins, & Forsman, 2004; Hull et al., 2010).

Populations in the north–west are therefore isolated by barriers in all compass directions with a gradient of magnitude: from the Pacific Ocean to the west, the species' ecological range limit to the north, the Rockies to the east and the Klamath Mountains to the south (Figure 1). Consequently, it is not surprising that they displayed the lowest genetic diversity (Table 1). Given this was a region of permafrost during the last glaciation (Hewitt, 2000), the northernmost populations are likely the result of post-glacial recolonization and therefore younger than the rest of the continent. On the other hand, their considerable amount of microsatellite private alleles (Table 1) suggests that isolation has also allowed for the accumulation of genetic differences in fast-mutating loci, contributing to the observed differentiation.

On the other side of the Rockies, the landscape is mostly composed of flat prairie plains. Unimpeded by any significant geological barrier, eastern barn owls cover longer distances (Figure S5) and travel mostly south in a wide angle (Figure 5). Hence, populations appear better connected with less genetic differentiation (Figure 1) and significantly higher genetic diversity than in the west (Table 1). Despite their 2,400 km of extent, the Appalachian Mountains in the north–east do not seem to restrict gene flow nor the movement of individuals among eastern barn owl populations (Figure 4). Although we recognize that our sampling in this region lacks fine-spatial resolution to detect genetic differentiation between populations on both sides, we found no such signal (Figure 1; Table 2), and recapture data show numerous individuals that cross the Appalachian Mountains (Figure 4). Indeed, the Appalachians are likely more suitable for barn owls as they occupy a considerably smaller area than the Rockies and rise to lower elevations (maximum 2,037 m, less than half of the Rockies), reducing their potential as a physical barrier to individuals' movements. Importantly, the snow cover on these mountains is not as deep and occurs over a shorter period.

Having established that the Rocky Mountains act as a barrier for gene flow and dispersal of barn owls, we attempted to identify the contact point between the populations on either side. We tested different routes and our analyses consistently pointed to the southern route as the most consistent with the observed genetic patterns (Table 3; Figure S4). The narrow pass between the Rocky Mountains and the Sierra Madre Occidental is a known biogeographical point at which multiple organisms cross the Western Continental Divide (Pyron & Burbrink, 2010). From the species' ecological point of view, the southern route would appear to be the most logical scenario as well, being the one at lowest elevation and with the least snow cover.

This is the first time a mountain range has been shown to significantly affect gene flow between barn owl populations. In Europe, the Pyrenees' short extent of 450 km does not seem to restrain exchanges between the Iberian Peninsula and France despite rising to 3,404 m (Burri et al., 2016). The potential role of the Alps (1,200 km in extent, up to 4,809 m elevation) has not yet been tested due to the lack of Italian barn owl samples available to previous studies. Additionally, investigating other major mountain ranges within the global distribution of barn owls, such as the Andes in South America or the Great Dividing Range in Australia, could reveal new mountainous barriers to gene flow and dispersal in this species.

4.3 | Comparison to previous results

Our findings do not concur with the previous report on barn owl's population structure in continental North America (Huang et al., 2016). We attempted to discern the causes at the origin of these differences, and our results concluded that it was our more extensive sampling in space that allowed the detection of the patterns presented above. Tests with the same number of markers as used by Huang et al. (2016) yielded similar results to ours (Table S7), suggesting the low number of markers was not responsible for the observed lack of structure in their study. However, our sampling covered most of continental North America (Figure 1), whereas Huang et al. (2016)'s only included one reduced population east of the Rocky Mountains (Pennsylvania, 13 individuals). Thus, when we replicated their geographical sampling, even with our 20 markers, Bayesian methods did not find significant genetic structure. Given the low level of overall genetic differentiation, it appears our extensive geographical sampling was crucial to detect it.

5 | CONCLUSION

Our study shows that large-sized landscape features can act as barriers to gene flow and dispersal for highly mobile and widespread organisms. We identified the Rocky Mountains as the main physical barrier to gene flow between barn owl populations along the east–west axis, followed by the Klamath Mountains that promote north–south differentiation along the west Coast. Thus, barn owl population structure seems to be strongly influenced by high-elevation, snow-covered barriers of large extent. Compared with previous work, our findings also highlight the importance of suitable geographical sampling when investigating patterns of genetic structure in space. Further work, including Central Plains and Mexican populations, would help elucidate the demographic history of North American barn owls and how it shaped current genetic structure and diversity patterns.

ACKNOWLEDGEMENTS

We are grateful to Raffael Winkler, Than J. Boves and James R. Belthoff as well as to the following institutions and their curators (in the USA, unless specified) for providing the tissue samples used in this study: American Museum of Natural History (New York), Academy of Natural Sciences of Philadelphia (Philadelphia), Bell Museum of Natural History (Minneapolis), Burke Museum of Natural History and Culture (Seattle), California Academy of Sciences (San Francisco), Cleveland Museum of Natural History (Cleveland), Washington State University Charles R. Conner Museum (Pullman), Cornell University Museum of Vertebrates (Lansing), Field Museum of Natural History (Chicago), University of Kansas Biodiversity Institute and Natural History Museum (Lawrence), Louisiana State University Museum of Natural Science (Baton Rouge), Museum of Southwestern Biology (Albuquerque), Museum of Vertebrate Zoology (Berkeley), Royal Ontario Museum (Ontario, Canada) and Smithsonian Institute (Washington DC). We thank Céline Simon for her valuable assistance with molecular work; Robin Séchaud for his advice on the analyses of recapture data and helpful discussion; and Arnaud Gaigher for comments on earlier versions of this manuscript. This study was funded by the Swiss National Science Foundation (31003A-120517 to AR and 31003A_138180 to JG), the American Natural History Museum in New York, the Swiss Academy of Sciences, the Basel Foundation for Biological Research and the "Fondation du 450eme anniversaire" and "Fondation Agassiz" of the University of Lausanne, to whom we are thankful.

ORCID

Ana Paula Machado http://orcid.org/0000-0003-2115-7237

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705

Antonelli, A. (2017). Biogeography: Drivers of bioregionalization. *Nature Ecology & Evolution*, *1*, 114. https://doi.org/10.1038/s41559-017-0114

Antoniazza, S., Burri, R., Fumagalli, L., Goudet, J., & Roulin, A. (2010). Local adaptation maintains clinal variation in melanin-based coloration of European barn owls (*Tyto alba*). *Evolution*, *64*, 1944–1954.

Antoniazza, S., Kanitz, R., Neuenschwander, S., Burri, R., Gaigher, A., Roulin, A., & Goudet, J. (2014). Natural selection in a postglacial range expansion: The case of the colour cline in the European barn owl. *Molecular Ecology*, *23*, 5508–5523. https://doi.org/10.1111/mec.12957

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*, 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Barrowclough, G. F., Groth, J. G., Mertz, L. A., & Gutiérrez, R. J. (2005). Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). *Molecular Ecology*, *14*, 1109–1120. https://doi.org/10.1111/j.1365-294X.2005.02465.x

Botta, F., Eriksen, C., Fontaine, M. C., & Guillot, G. (2015). Enhanced computational methods for quantifying the effect of geographic and environmental isolation on genetic differentiation. *Methods in Ecology and Evolution*, *6*, 1270–1277. https://doi.org/10.1111/2041-210X.12424

Boutilier, S. T., Taylor, S. A., Morris-Pocock, J. A., Lavoie, R. A., & Friesen, V. L. (2014). Evidence for genetic differentiation among Caspian tern (*Hydroprogne caspia*) populations in North America. *Conservation Genetics*, *15*, 275–281. https://doi.org/10.1007/s10592-013-0536-1

Bradburd, G. S., Ralph, P. L., & Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution*, *67*, 3258–3273. https://doi.org/10.1111/evo.12193

Brassel, K. E., & Reif, D. (1979). A procedure to generate Thiessen polygons. *Geographical Analysis*, *11*, 289–303.

Burri, R., Antoniazza, S., Gaigher, A., Ducrest, A.-L., Simon, C., Fumagalli, L., … Roulin, A. (2016). The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution*, *70*, 140–153. https://doi.org/10.1111/evo.12824

Burri, R., Antoniazza, S., Siverio, F., Klein, A., Roulin, A., & Fumagalli, L. (2008). Isolation and characterization of 21 microsatellite markers in the barn owl (*Tyto alba*). *Molecular Ecology Resources*, *8*, 977–979. https://doi.org/10.1111/j.1755-0998.2008.02121.x

Calenge, C. (2011). *Analysis of animal movements in R: The adehabitatLT package*. Vienna, Austria: R Foundation for Statistical Computing.

DeChaine, E. G., & Martin, A. P. (2005a). Historical biogeography of two alpine butterflies in the Rocky Mountains: Broad-scale concordance and local-scale discordance. *Journal of Biogeography*, *32*, 1943–1956. https://doi.org/10.1111/j.1365-2699.2005.01356.x

DeChaine, E. G., & Martin, A. P. (2005b). Marked genetic divergence among sky island populations of *Sedum lanceolatum* (Crassulaceae) in the Rocky Mountains. *American Journal of Botany*, *92*, 477–486. https://doi.org/10.3732/ajb.92.3.477

Del Hoyo, J., Elliott, A., Sargatal, J., & Christie, D. (1999). *Handbook of the birds of the world*. Barcelona: Lynx Editions.

Du Feu, C. R., Clark, J. A., Schaub, M., Fiedler, W., & Baillie, S. R. (2016). The EURING Data Bank – a critical tool for continental-scale studies of marked birds. *Ringing and Migration*, *31*, 1–18. https://doi.org/10.1080/03078698.2016.1195205

Earl, D. A., & von Holdt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*, 359–361. https://doi.org/10.1007/s12686-011-9548-7

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. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Frantz, A. C., Bertouille, S., Eloy, M. C., Licoppe, A., Chaumont, F., & Flamand, M. C. (2012). Comparative landscape genetic analyses show a Belgian motorway to be a gene flow barrier for red deer (*Cervus elaphus*), but not wild boars (*Sus scrofa*). *Molecular Ecology*, *21*, 3445–3457. https://doi.org/10.1111/j.1365-294X.2012.05623.x

Geffen, E., Anderson, M. J., & Wayne, R. K. (2004). Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology*, *13*, 2481–2490. https://doi.org/10.1111/j.1365-294X.2004.02244.x

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, 1–19.

Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*, 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, *4*, 336–344. https://doi.org/10.1111/2041-210x.12018

Haig, S. M., Mullins, T. D., & Forsman, E. D. (2004). Subspecific relationships and genetic structure in the spotted owl. *Conservation Genetics*, *5*, 683–705. https://doi.org/10.1007/s10592-004-1864-y

Hayes, F. E., & Sewlal, J. A. N. (2016). The Amazon River as a dispersal barrier to passerine birds: Effects of river width, habitat and taxonomy. *Journal of Biogeography*, *31*, 1809–1818.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, *405*, 907–913. https://doi.org/10.1038/35016000

Huang, A. C., Elliott, J. E., Cheng, K. M., Ritland, K., Ritland, C. E., Thomsen, S. K., … Martin, K. (2016). Barn owls (*Tyto alba*) in western North America: Phylogeographic structure, connectivity, and genetic diversity. *Conservation Genetics*, *17*, 357–367. https://doi.org/10.1007/s10592-015-0787-0

Huang, Z. S., Yu, F. L., Gong, H. S., Song, Y. L., Zeng, Z. G., & Zhang, Q. (2017). Phylogeographical structure and demographic expansion in the endemic alpine stream salamander (Hynobiidae: *Batrachuperus*) of the Qinling Mountains. *Scientific Reports*, *7*, 1871. https://doi.org/10.1038/s41598-017-01799-w

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, *9*, 1322–1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x

Hull, J. M., Hull, A. C., Sacks, B. N., Smith, J. P., & Ernest, H. B. (2008). Landscape characteristics influence morphological and genetic differentiation in a widespread raptor (*Buteo jamaicensis*). *Molecular Ecology*, *17*, 810–824. https://doi.org/10.1111/j.1365-294X.2007.03632.x

Hull, J. M., Keane, J. J., Savage, W. K., Godwin, S. A., Shafer, J. A., Jepsen, E. P., … Ernest, H. B. (2010). Range-wide genetic differentiation among North American great gray owls (*Strix nebulosa*) reveals a distinct lineage restricted to the Sierra Nevada, California. *Molecular Phylogenetics and Evolution*, *56*, 212–221. https://doi.org/10.1016/j.ympev.2010.02.027

Hull, J. M., Strobel, B. N., Boal, C. W., Hull, A. C., Dykstra, C. R., Irish, A. M., … Ernest, H. B. (2008). Comparative phylogeography and population genetics within *Buteo lineatus* reveals evidence of distinct evolutionary lineages. *Molecular Phylogenetics and Evolution*, *49*, 988–996. https://doi.org/10.1016/j.ympev.2008.09.010

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–140. https://doi.org/10.1093/bioinformatics/btn129

Keyghobadi, N., Roland, J., & Strobeck, C. (1999). Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Molecular Ecology*, *8*, 1481–1495. https://doi.org/10.1046/j.1365-294x.1999.00726.x

Kisel, Y., & Barraclough, T. G. (2010). Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist*, *175*, 316–34. https://doi.org/10.1086/650369

Klein, A., Horsburgh, G. J., Küpper, C., Major, Á., Lee, P. L. M., Hoffmann, G., … Dawson, D. A. (2009). Microsatellite markers characterized in the barn owl (*Tyto alba*) and of high utility in other owls (Strigiformes: AVES). *Molecular Ecology Resources*, *9*, 1512–1519. https://doi.org/10.1111/j.1755-0998.2009.02715.x

Klicka, J., Spellman, G. M., Winker, K., Chua, V., & Smith, B. T. (2011). A phylogeographic and population genetic analysis of a widespread, sedentary North American bird: The hairy woodpecker (*Picoides villosus*). *The Auk*, *128*, 346–362. https://doi.org/10.1525/auk.2011.10264

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*, 1870–1874. https://doi.org/10.1093/molbev/msw054

Legendre, P., & Fortin, M. J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, *10*, 831–844. https://doi.org/10.1111/j.1755-0998.2010.02866.x

Legendre, P., Fortin, M. J., & Borcard, D. (2015). Should the Mantel test be used in spatial analysis? *Methods in Ecology and Evolution*, *6*, 1239–1247. https://doi.org/10.1111/2041-210X.12425

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116. https://doi.org/10.1111/2041-210X.12410

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452. https://doi.org/10.1093/bioinformatics/btp187

Manni, F., Guerard, E., & Heyer, E. (2004). Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Human Biology*, *76*, 173–190. https://doi.org/10.1353/hub.2004.0034

Marti, C. D. (1999). Natal and breeding dispersal in barn owls. *Journal of Raptor Research*, *33*, 181–189.

Mayr, E. (1942). *Systematics and the origin of species*. New York, USA: Columbia University Press.

Milot, E., Gibbs, H. L., & Hobson, K. A. (2000). Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). *Molecular Ecology*, *9*, 667–681. https://doi.org/10.1046/j.1365-294x.2000.00897.x

Monmonier, M. S. (1973). Maximum-difference barriers: An alternative numerical regionalization method. *Geographical Analysis*, *5*, 245–261.

Munshi-South, J. (2012). Urban landscape genetics: Canopy cover predicts gene flow between white-footed mouse (*Peromyscus leucopus*) populations in New York City. *Molecular Ecology*, *21*, 1360–1378. https://doi.org/10.1111/j.1365-294X.2012.05476.x

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

Proudfoot, G. A., Gehlbach, F. R., & Honeycutt, R. L. (2007). Mitochondrial DNA variation and phylogeography of the eastern and western screech-owls. *The Condor*, *109*, 617–627. https://doi.org/10.1650/8262.1

Pyron, R. A., & Burbrink, F. T. (2010). Hard and soft allopatry: Physically and ecologically mediated modes of geographic speciation. *Journal of Biogeography*, *37*, 2005–2015.

R Development Core Team (2016) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/.

Schwartz, M. K., Copeland, J. P., Anderson, N. J., Squires, J. R., Inman, R. M., McKelvey, K. S., … Cushman, S. A. (2009). Wolverine gene flow across a narrow climatic niche. *Ecology*, *90*, 3222–3232. https://doi.org/10.1890/08-1287.1

Smissen, P. J., Melville, J., Sumner, J., & Jessop, T. S. (2013). Mountain barriers and river conduits: Phylogeographical structure in a large, mobile lizard (Varanidae: *Varanus varius*) from eastern Australia. *Journal of Biogeography*, *40*, 1729–1740. https://doi.org/10.1111/jbi.12128

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. (1997). Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. *Plant Systematics and Evolution*, *206*, 353–373. https://doi.org/10.1007/BF00987957

Speek, G., Clark, J. A., Rohde, Z., Wassenaar, R., & vanNoordwijk, A. J. (2001). *The EURING exchange-code 2000*. Heteren.

Spellman, G. M., Riddle, B., & Klicka, J. (2007). Phylogeography of the mountain chickadee (*Poecile gambeli*): Diversification, introgression, and expansion in response to Quaternary climate change. *Molecular Ecology*, *16*, 1055–1068. https://doi.org/10.1111/j.1365-294X.2007.03199.x

Taylor, I. (1994). *Barn owls: Predator-prey relationships and conservation*. Cambridge: Cambridge University Press.

USGS Bird Banding Laboratory (2017). *North American bird banding and band encounter data set*. Laurel, MD: Patuxent Wildlife Research Center.

Van Den Brink, V., Dreiss, A. N., & Roulin, A. (2012). Melanin-based coloration predicts natal dispersal in the barn owl, *Tyto alba*. *Animal Behaviour*, *84*, 805–812. https://doi.org/10.1016/j.anbehav.2012.07.001

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*, 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x

Von Oheimb, P. V., Albrecht, C., Riedel, F., Bössneck, U., Zhang, H., & Wilke, T. (2013). Testing the role of the Himalaya Mountains as a dispersal barrier in freshwater gastropods (*Gyraulus* spp.). *Biological Journal of the Linnean Society*, *109*, 526–534. https://doi.org/10.1111/bij.12068

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*, 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

DATA AVAILABILITY

Microsatellite and mtDNA sequence data are provided in Appendix S2.

BIOSKETCH

This publication is part of **Ana Paula Machado**'s PhD programme in the research groups of **Alexandre Roulin** (https://www.unil.ch/dee/roulin-group) and **Jérôme Goudet** (https://www.unil.ch/dee/goudet-group) at the University of Lausanne. Her current research interests include the population genetics and genomics of the cosmopolitan *Tyto alba*, as well as how this species interacts and adapts to different landscape features within its range.

Author contributions: A.P.M., J.G. and A.R. conceived the project; A.R. collected the samples from the different museums; L.C. and V.U. performed the molecular work; A.P.M. conducted the analyses and led the writing with contribution from all authors.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Machado AP, Clément L, Uva V, Goudet J, Roulin A. The Rocky Mountains as a dispersal barrier between barn owl (*Tyto alba*) populations in North America. *J Biogeogr*. 2018;00:1–13. https://doi.org/10.1111/jbi.13219

Chapter 5 – Supporting Information

Supporting Tables

Table S1 – Description of individuals used for the genetic analyses in this study.

Pop	Males	Females	U	Time range	States
FLO	8	16	-	1993 - 2008	Florida
NJ	3	8	-	1986 - 2004	New Jersey, New York, North Carolina, Pennsylvania
LOUI	10	9	-	1991 - 2007	Louisiana, Alabama, Florida, Mississippi
ON	4	5	-	1988 - 2011	Ontario (Canada), Minnesota, Illinois, Ohio
NM	10	17	1	1996 - 2009	New Mexico, Arizona, Kansas, Texas
CALA	22	17	2	1989 - 2007	California
CASF	19	20	2	1994 - 2005	California
ID	13	24	-	1993 - 2008	Idaho, Oregon, Washington
OR	20	19	2	1990 - 2005	Oregon, Washington
WA	16	25	-	1988 - 2008	Washington, British Columbia (Canada)
Total	125	160	7	1986 - 2011	

Pop – populations abbreviated following Table 1; U – sex unknown; States – sampling states in the USA or in Canada when indicated.

Table S2 – Microsatellite summary statistics per locus.

Locus	N	% NA	N_A	A_R	H_o	H_e	F_{ST}	F_{IS}
Ta-210	291	0.34	2	1.078	0.01	0.01	0.01	-0.01
Ta-216	291	0.34	17	6.047	0.77	0.77	0.02	0.00
Ta-306	289	1.03	2	1.930	0.24	0.22	-0.01	-0.08
Ta-218	290	0.68	5	3.314	0.41	0.43	0.02	0.04
Ta-220	283	3.08	8	4.828	0.62	0.71	0.02	0.12
Ta-204	291	0.34	3	2.957	0.61	0.60	0.01	0.00
Ta-214	285	2.40	4	2.304	0.24	0.27	0.03	0.13
Ta-305	292	0	6	3.418	0.61	0.64	0.05	0.04
Ta-310	289	1.03	4	1.407	0.05	0.05	-0.01	-0.01
Ta-413	291	0.34	18	8.169	0.80	0.87	0.02	0.08
Ta-202	289	1.03	6	4.438	0.62	0.71	0.00	0.13
Ta-212	290	0.68	7	3.574	0.67	0.64	0.01	-0.05
Ta-215	290	0.68	10	4.247	0.60	0.62	0.05	0.04
Ta-402	285	2.40	18	7.512	0.89	0.85	0.01	-0.05
Ta-408	266	8.90	13	4.541	0.53	0.64	0.00	0.17
Ta-FEP-42	275	5.82	2	1.991	0.32	0.34	0.03	0.06
Ta-54f2	272	6.85	9	6.260	0.76	0.81	0.00	0.06
Ta-Calex-05	266	8.90	3	1.338	0.05	0.04	0.04	-0.06
Ta-TGU-06	292	0	8	3.966	0.59	0.55	0.00	-0.08
Ta-RBG-18	270	7.53	6	3.251	0.59	0.57	0.02	-0.02

N – number of individuals successfully genotyped; % NA – percentage of missing genotypes; N_A –number of alleles.

Table S3 – Summary of 'Sunder' results. Likelihoods and effect size (β) estimates for each tested model across runs: G: geography; E: ecology (barrier); G+E: both. The highest likelihood (highlighted in bold) was consistently attributed to the G+E model.

# Run	G		E		G + E		
	Likelihood	β_G	Likelihood	β_E	Likelihood	β_G	β_E
1	-1392.07	2951	-1397.28	12.55	**-1389.17**	3473	11.19
2	-1181.17	3086	-1182.82	15.00	**-1178.63**	3578	11.50
3	-1339.94	3355	-1337.08	13.77	**-1332.42**	3490	15.12
4	-924.87	2862	-926.21	10.79	**-920.92**	3483	11.39
5	-1635.09	3007	-1640.24	10.23	**-1634.47**	3614	11.43
6	-1132.05	2934	-1130.43	11.71	**-1126.41**	3449	8.50
7	-1175.85	2887	-1178.62	13.15	**-1174.69**	3625	11.01
8	-1159.15	3261	-1158.68	9.48	**-1155.52**	3433	12.83
9	-1406.23	2749	-1409.58	12.86	**-1404.61**	3505	10.19
10	-1112.97	3300	-1117.56	10.44	**-1111.76**	3681	10.38

Table S4 – Complete results of linear models for the first PCoA axis comparing different contact routes, ordered by the most likely scenario according to AIC scores. Models for the same genetic markers were contrasted between the three possible routes: South of the Rocky Mountains, Direct Flight (Euclidean), or North of the Rocky Mountains.

Marker	Route	t	R^2	P	AIC	Δ AIC
Microsatellites	South	10.000	0.926	< 0.0001	-73.793	0
	Direct Flight	5.921	0.814	< 0.001	-64.593	9.200
	North	4.160	0.684	0.003	-59.277	14.516
ND6	South	6.038	0.820	< 0.001	-34.100	0
	Direct Flight	2.972	0.525	0.018	-24.385	9.714
	North	2.168	0.370	0.062	-21.569	12.531

Table S5 – Dispersal characteristics of the 5 individuals that crossed the Rocky Mountains, numbered to match the highlighted movements in Figure 4.

N	Ring no.	Capture		Recapture		Time (days)	Distance (km)
		Date	Location	Date	Location		
1	98772088	18-05-84	Salt Lake City, Utah	27-11-85	Omaha, Nebraska	558	1311
2	98791285	24-04-91	Salt Lake City, Utah	21-03-92	Pueblo, Colorado	332	768
3	180795278	05-05-09	Nampa, Idaho	25-05-12	Santa Fe, New Mexico	1116	1300
4	220647152	02-05-99	Merced, California	30-10-00	Chicago, Illinois	547	2836
5	220606110	27-05-90	Bridgeton, New Jersey	12-03-99	Temecula, California	3211	3776

Table S6 – Natal dispersal in North American and European barn owls. Mann-Whitney *U* tests results for the full dataset and subsets. All tests were significant following Bonferroni correction.

Test	North America		Europe		*U*	p
	N	Distance	N	Distance		
Complete dataset	784	23.73	19641	16.61	8810300	1.05×10^{-12}
Up to 150 km	651	18.65	17963	13.94	6293000	3.30×10^{-4}
Up to 100 km	602	18.51	16742	12.56	5440900	2.98×10^{-4}
Up to 50 km	525	17.61	14586	10.20	4228900	1.14×10^{-5}

N – number of individuals; Distance – median dispersal distance in km.

Table S7 – Summary of STRUCTURE bootstrap results comparison to Huang *et al.* (2016). All runs indicated K=2 as the most likely population structure.

# Run	Loci	Mean Ln(K)	SD	L'(K)	|L''(K)|	Delta K
1	Ta-210, Ta-306, Ta-218, Ta-204, Ta-310, Ta-202, Ta-402, Ta-408	-4224.30	3.61	49.39	97.06	26.89
2	Ta-210, Ta-216, Ta-214, Ta-305, Ta-310, Ta-215, Ta-402, FEP-42	-4433.18	4.08	124.78	168.23	41.27
3	Ta-210, Ta-216, Ta-214, Ta-305, Ta-413, Ta-402, Ta-408, FEP-42	-5777.86	5.4	106.11	163.83	30.33
4	Ta-216, Ta-306, Ta-220, Ta-204, Ta-310, Ta-212, Ta-215, RBG-18	-4696.35	2.98	85.86	112.46	37.74
5	Ta-210, Ta-204, Ta-305, Ta-413, Ta-202, Ta-212, Ta-215, Ta-408	-5624.21	4.15	123.48	167.54	40.4
6	Ta-216, Ta-220, Ta-204, Ta-214, Ta-305, Ta-413, Ta-402, FEP-42	-6403.00	4.24	110.49	110.12	25.98
7	Ta-306, Ta-220, Ta-204, Ta-305, Ta-310, Ta-202, Ta-402, TGU-06	-5132.36	6.34	63.39	90.33	14.25
8	Ta-306, Ta-204, Ta-214, Ta-310, Ta-202, Ta-212, Ta-402, RBG-18	-4545.51	5.92	62.28	104.72	17.7
9	Ta-216, Ta-306, Ta-218, Ta-204, Ta-413, Ta-202, Ta-215, Calex-05	-5276.06	4.756	107.24	154.28	32.44
10	Ta-216, Ta-306, Ta-218, Ta-310, Ta-202, Ta-215, Ta-402, Ta-408	-5350.93	1.55	112.94	144.37	93.15
11	Ta-216, Ta-218, Ta-220, Ta-214, Ta-305, Ta-408, Calex-05, TGU-06	-4820.90	5.21	75.81	105.73	20.29
12	Ta-306, Ta-220, Ta-214, Ta-305, Ta-215, Ta-408, FEP-42, RBG-18	-4343.71	4.52	112.69	144.43	31.92
13	Ta-210, Ta-216, Ta-204, Ta-305, Ta-310, Ta-413, Ta-212, FEP-42	-4816.5	2.2	73.47	115.75	52.55
14	Ta-210, Ta-306, Ta-218, Ta-220, Ta-214, Ta-305, Ta-202, TGU-06	-4031.33	4.72	78.13	120.93	25.65
15	Ta-210, Ta-216, Ta-220, Ta-212, Ta-408, 54f2, Calex-05, TGU-06	-5153.84	26.09	-45.75	24.65	0.945
16	Ta-216, Ta-306, Ta-218, Ta-305, a-215, FEP-42, 54f2, TGU-06	-5136.16	1.49	99.08	240.18	161.43
17	Ta-216, Ta-218, Ta-413, Ta-202, Ta-212, Ta-215, Ta-408, Calex-05	-5921.91	5.11	117.03	203.79	39.9
18	Ta-210, Ta-306, Ta-220, Ta-204, Ta-310, Ta-413, Ta-408, RBG-18	-4421.02	5.22	42.98	78.68	15.07
19	Ta-210, Ta-306, Ta-305, Ta-408, 54f2, Calex-05, TGU-06, RBG-18	-3977.20	10.11	29.2	112.22	11.1
20	Ta-210, Ta-306, Ta-218, Ta-413, Ta-212, Ta-215, Ta-408, RBG-18	-4824.81	4.65	100.3	165.68	35.59

Loci – set of 8 microsatellites used in each bootstrap run; Mean Ln(K) - Lean likelihood; SD – standard deviation; L'(K) – Absolute change of the likelihood distribution (mean); |L''(K)| - Rate of change of the likelihood distribution (mean).

General Discussion

In this thesis, I used molecular tools to study barn owl populations with varying degrees of isolation. From mountain ranges that distort otherwise smooth patterns of isolation by distance (chapters 4 and 5), to islands that harbour populations diverging both genetic and phenotypically (chapters 1 to 3). In the former, I focused on the tallest chains of Europe and North America, the Alps and the Rocky Mountains. In the latter, we span three island systems ranging in age, surface area and distance to mainland. Notably, in addition to genomic data, throughout this work I had the opportunity to work with phenotypical measurements, capture-recapture and climate data, in a complimentary approach that was often crucial to present well-supported results. In the following sections of this discussion, I summarize how the combined results of the presented chapters contribute to our knowledge of the species in light of what was previously known, discuss future avenues of research and comment on the importance of using genomic data.

Barn owls of the Western Palearctic

Up until the start of this thesis, classical genetic markers had provided a description of the broad patterns of genetic diversity and structure among barn owl populations of the Western Palearctic[100–102]. Briefly, it was known that most of continental Europe was derived from an Iberian glacial refugium, while an eastern lineage was present in the Levant. The colour cline was observable in the European lineage, with rufous individuals predominantly found in the north and northeast and white ones in the south and southwest. Due to the odd genetic makeup of the Greek populations, it was conjectured that the east European rufous birds and the Levant white ones might meet in a secondary contact zone between the lineages in the southern Balkans. This hypothesis was supported by the sharper changes in plumage colouration observed in the Balkans compared to the west, instigating the hypothesis that some degree of reproductive isolation might have arisen between owls of different colour. Using whole genome sequences, we were able to address the postulated scenario and rewrite the history of insular and mainland barn owls of the Western Palearctic.

Colonisation routes and glacial refugia

Genomic data confirmed that the Mediterranean Sea is surrounded by two main barn owl genetic lineages, the European (or western) and the Levant (or eastern), assumed to be connected over northern Africa. While this is concordant with previous results, we pinpointed their meeting point to Anatolia rather than the Balkans. Since there are virtually no rufous owls in Anatolia, this finding dissociates the secondary contact zone between lineages from the region of transition in colour and therefore refutes the hypothesised scenario and reproductive isolation. In the west, off

the northern African coast, the Canary archipelago was colonised by barn owls with fluctuating levels of both European and Levant ancestry, consistent with north Africa carrying a genetic gradient between the two lineages.

During the last glaciation, the known barn owl glacial refugium in the Iberian Peninsula probably extended into a contiguous narrow westwards corridor in now submerged land over the Bay of Biscay. Moreover, we showed the species survived in a second cryptic glacial refugium in continental Europe. Considering the lower sea levels during the last glaciation, the core of this refugium was most likely in southwest Italy and potentially spread over what is now the Adriatic Sea into the eastern Balkans.

As temperatures rose and the ice caps started melting, barn owls in the Bay of Biscay were soon cut off from the mainland and founded what would become the British Isles. At this time, Cyprus already boasted a suitable climate for the species, however we were unable to reliably determine its colonization time. In mainland Europe, the southwest became suitable first and, from the Iberian refugium, barn owls started colonizing central and northern Europe before turning east above the Alps. Previous work on colouration and *MC1R* indicate this lineage became increasingly rufous in the populations further north. In the meantime, from the Italo-Balkan refugium, they could only spread along the coast south to the Aegean Sea and its patchwork of small islands. Only much more recently, barn owls from the Aegean eventually reached Crete. Finally, as conditions became more favourable in the continental inland, the Iberian and Italo-Balkan glacial lineages ultimately met in eastern Europe. The admixture between the two is prevalent and traces of it reaches populations far from the assumed meeting point, corroborating the absence of reproductive isolation. The discovery of a second refugium in Italy explained the peculiarity of Greek owls which, along with other inland Balkan populations, have high proportions of the genetic signatures of both glacial refugia.

Among the traditional postglacial colonization patterns[121], the routes described here resemble that of the brown bear pattern the most (though far from perfectly), with a greater contribution from an Iberian refugium. It is a quite distinct pattern from that of the tawny owl[339] (*Strix aluco*) who used the central Balkan Peninsula as a glacial refugium. Despite the many similarities to barn owls – widespread, colour polymorphic, mid-sized nocturnal raptors – tawny owls are forest-dwellers, which probably allowed them to exploit this refugium while barn owls could not. As far as we were able to determine, the glacial history of the Eurasian kestrel (*Falco tinnunculus*), the diurnal homologous to the barn owl that also prefers open habitats, has unfortunately never been studied and therefore we cannot make a comparison to it. In conclusion, taking into consideration the changing coastline and regions of climatic suitability for our species was crucial in determining the location of glacial refugia and retrace colonization routes. This approach, in

combination with the power of genomic data, is likely to rewrite the history of other species, as in the recent case of yellow warblers (*Setophaga petechia*)[340].

Evolutionary change in insular populations

Genetic drift is commonly recognised as the main driver of neutral genetic divergence in insular populations, whereas variation in morphological traits is usually assumed to have a selective basis[58]. In this thesis, I analysed these suppositions in two insular systems, the British Isles and the Canary Islands. In the former, there is a clear divergent phenotype with insular owls being white instead of rufous as on the nearest mainland. We showed that this pattern can actually be parsimoniously explained by neutral processes, as British and Irish owls simply inherited the light colouration of their founders in the Iberian Peninsula, and kept it via reduced gene flow with the mainland after becoming geographically isolated by the rise of the sea levels. We are unable to completely discard the possibility that a rufous morph would be more advantageous on the islands, if it is better adapted to latitude-related factors, for example. However, even if that was the case, the small population size and low gene flow facilitate the loss of incoming rufous alleles by drift, making the white colour a direct product of neutral evolution.

In the Canary archipelago, barn owls of the eastern islands are shorter than those of the neighbouring islands and mainland. We found evidence of local adaptation in multiple genomic regions, including genes linked to body size and proportions, suggesting that its size might indeed be the result of selective pressure. Interestingly, we also found signals of adaptive forces acting on other, less obvious, traits like adaptations linked to blood-pressure in hot arid conditions in the eastern populations, red blood cells and haemoglobin in the high-elevation western population, as well as a supposed insular morphological adaptation with an unknown phenotype.

To summarize, our case studies provide evidence of a first conspicuous phenotype being the result of genetic drift and a second one due to local adaptation. In addition, we also found adaptation acting on cryptic traits. Ultimately, disentangling the effects of each evolutionary mechanisms on the genomic makeup of any given insular population remains a challenging task. Given the uniqueness of each island system and colonization circumstances, it is doubtful that a broad rule will ever emerge, except that all mechanisms likely interact simultaneously and their importance can vary with time and fluctuating local conditions.

The key role of landscape barriers

In the absence of landscape barriers, barn owls keep constant levels of gene flow with neat patterns of isolation by distance over thousands of kilometres, a common feature among raptors (reviewed in[341]). However, the impact of mountains, fluctuating sea levels and distance to the

mainland is striking when retracing the glacial and postglacial history described above. Mountains ranges force owls to contour them as, for example, around the north of the Alps in Europe (chapter 4) and around the south of the Rocky Mountains in North America (chapter 5).

As illustrated by Italy, the cumulative effect of barriers, like mountains and water, can isolate even more a population. Cornered in the north by the Alps and surrounded by the Mediterranean Sea in all other directions, the Italian population had little exchange with the rest of Europe and remained a well-preserved relic of the glacial refugium. A similar combination of large mountains and water bodies, resulted in the nearly complete discontinuation of gene flow in Anatolia, with only sporadic exchanges between the Levant and European lineages. In contrast, the meeting of the Iberian and Italo-Balkan glacial lineages in eastern Europe, in the absence of large barriers, is ubiquitous and diluted over long distances. On their own, large water bodies also significantly hinder exchanges of individuals with insular populations being consistently more similar genetically to their founding population than to other nearer populations, regardless of how old the colonisation is (chapters 1-3). This is outstandingly clear in the Canary archipelago, where islands 200 km apart are more distinct from each other than from the mainland that founded them over 1000km away.

Unexplored regions

Despite the vast advances reported here on our knowledge of the history of barn owls in the Western Palearctic, several questions remain unanswered and would require more extensive sampling. First, whether the populations of northern Africa form a continuum between the European and the Levant lineage, or the northern part of the Sahara Desert acts as yet another barrier between them. If the latter is true, the theory of a ring-like colonization around the Mediterranean Sea would be replaced by one of two lineages of unknown geographical origin with two current contact zones, one in Anatolia and the other in northern Africa, between Libya and Egypt. Similarly, the regions east of the Balkans and around the Black Sea, where the supposed *T. a. guttata* would have originated, merit investigation.

In terms of island populations, three insular subspecies remain unexplored at present: *T. a. ernesti* in Sardinia and Corsica, *T. a. schmitzi* in Madeira and *T. a. detorta* in Cape Verde. While all three are worth characterizing, the Madeiran subspecies is perhaps the most intriguing and one we had hoped to sample during this work. Being far (over 500km) yet nearly equidistant from the mainland and the Canary Islands, its study might contribute to the history of the Canarias population and potentially yield curious colonization patterns. Lastly, the European barn owl lineage spans a large variety of climates, habitats and prey species. It could be an indication that local adaptation occurs, in clinal form or not, on other traits besides colour.

Taxonomy of the Afro-European barn owl

Although it is not the goal of this thesis to contribute to the debate of what constitutes a subspecies, our work was intrinsically linked to such classifications and, at times, conflicted with them. Currently, there are seven described subspecies of barn owl in the Western Palearctic[205], four of which were part of this thesis. First, *T. a. gracilirostris* (Hartert, E, 1905) is found in the eastern islands of the Canary archipelago, namely Fuerteventura, Lanzarote and the small islets around them. This classification was based on its smaller size and supposedly darker plumage, although the latter is highly contested by ornithologists (F. Siverio, personal communication) and phenotypical measurements[102]. Our results concur with its subspecies status, in the extent to which it forms a long-term isolated genetic cluster, is monophyletic and displays genomic signs of being locally adapted to its environment. Moreover, it is by far the most differentiated insular population out of all the ones studied here. Second, *T. a. erlangeri* (Sclater, WL, 1921) is present in the Levant, Cyprus and Crete, and can easily be matched to the Levant (eastern) genetic lineage described above. Yet, we distinctly show that Crete does not belong to this lineage, but rather to the European one, and should therefore be excluded from the classification. Finally, *T. a. alba* (Scopoli 1769) and *T. a. guttata* (Brehm and CL 1831), are the, respectively, white and rufous subspecies of European barn owls (see General Introduction). The former is distributed in the south and west of Europe, and the latter in the north and east. However, as detailed above, our results do not support this split, as colouration does not reflect the neutral genetic patterns as it is the result of local adaptation acting within the same genetic lineage. Furthermore, though we identified a second glacial refugium, owls of southern Italy are white, thus refuting the hypothesis of differently coloured refugia (see General Introduction).

Plumage colour determination and maintenance

Traditional genetic markers had shown that the European barn owl colouration cline is due to local adaptation and not the result of neutral processes. Although we didn't specifically test this, the genomic data presented here concurs with this conclusion. Indeed, the genetic makeup of European populations reflects their respective demographic histories instead of their plumage colour. When it comes to the genetic determination and the maintenance of this trait, however, multiple aspects are still unclear.

The mutation detected at the *MC1R* gene of European barn owls only explains approximately 30% of colour variation[102,116]. The rufous derived allele appears to be decisive in generating rufous phenotypes, even in heterozygote state. Yet, among homozygous white owls there is a great deal of variation in shade, with some reaching intermediate phenotypes (see Figure 1 in General Introduction). This was particularly clear in our study of the British Isles, where barn owls are whiter than any mainland population, even in Iberia where being white is expected to be

advantageous. Besides both populations being unsurprisingly homozygous white at *MC1R*, we were unable to find other regions that could explain this difference through our candidate gene approach. Given its high heritability (h^2=0.81[95]), and that 70% of variation is still unaccounted for, it seems plausible that this trait has a polygenic basis. To probe this question further, we have begun a study using a GWAS (genome-wide association study) approach on both mixed and pure coloured-populations across Europe. We hope to detect other genomic regions that may compound on the effect of *MC1R* and explain the remaining phenotypical variance. Preliminary results indicate that a region in chromosome Z accounts for some of the variation amid homozygous white owls in Switzerland (unpublished data). Since female owls only have one copy of the Z chromosome, we removed it in our study of the British Isles and therefore could not have picked up this signal. It will be extremely interesting to see if this same region can explain the difference between British and Iberian white owls.

Finally, the selective forces maintaining the differential geographic distribution of plumage colouration still require additional research. The fact that it occurs on the three barn owl sister species, and are not linked to the same *MC1R* mutation (unpublished data), insinuates independent convergent evolution of the rufous phenotype in each taxon. Worldwide analyses of climatic data, found that darker barn owl morphs occur in colder and wetter regions[96], suggesting more melanic plumage might be better adapted to living in such conditions. The specific advantages they might have however, like better insulation, impermeability or camouflage, remain undistinguishable until further examination.

Population genomics in non-model species

This thesis adds to the growing body of literature of empiric population genomics studies in non-model species, and the added value of using whole genome sequences was clear at numerous points throughout it. First, the massive increase in number of loci resolved the convoluted structure patterns in and around the eastern Mediterranean, clarifying the origin of heavily diverged insular populations and pinpointing secondary contact zones. Second, it provided the necessary power to discover unexpected demographic scenarios, such as the colonisation of the British Isles, and even glacial refugia. Finally, accessing the genomic landscape allowed the identification of potential locally adaptative genomic regions in the Canary Islands. In parallel, the availability of a good reference genome assembly at near chromosome-level, allowed a better mapping of the reads and consequently contributed to increasing the number of SNPs, improving the resolution of the genomic landscape and facilitating the phasing of reads.

Notwithstanding, as for many non-model species, a few limitations remained and impacted this work, with perhaps the most obvious being the difficulty in sampling, not only in terms of quantity but also quality of the samples. Even though we were able to cover quite a wide geographic

range, and the barn owl is less complicated to sample than some other species, we were not able to sample all the locations we set out to (see above). A trade-off of using genomic data, is the need for fresh and good quality source tissue, especially when producing PCR-free genomic libraries, and thus museum samples, the typical source material in previous studies, were not usable. In model species, the wealth of samples and high-coverage data, allows the imputation of genotypes (or genotype likelihoods[342]) on extremely low coverage data (less than 1x), a technique seen most of all in humans[343] with some examples in other groups of interest as well, such as cattle[344,345]. This type of strategy was obviously out of our reach since most species do not have high quality reference panels available. The second, and also common, limitation was the lack of a complete annotation of the reference genome, particularly in chapter 3, where we investigated regions putatively under local adaptation in owls from the Canary Islands. The list of genes in these regions was probably incomplete due to the draft annotation available, thus reducing the power of detection in enrichment analyses. Thirdly, we only considerer bi-allelic loci in our studies, missing the signals that indels and structural variants could add. Finally, with our short-read sequencing data, we were restricted in the use of haplotype-based analyses. This was overcome in chapter 4 by harnessing the power of a large sample set (nearly 100 genomes) in a combined physical and statistical phasing approach, which is currently unfeasible for smaller sets. Still, the development of statistical tools is likely to overcome this limitation in the near future.

Demographic inference methods

Throughout this thesis, demographic inference was a frequent analysis step, and a particularly crucial one in verifying the hypotheses for the colonisation of the British Isles and the second glacial refugium in Italy. This type of analyses is gradually becoming more common in non-model organisms but still far from being widespread, especially from WGS data and with the complexity of the scenarios simulated here. We used *fastsimcoal2*, an SFS-based software that relies on coalescent simulations, and arguably the best available option to us at the start of this work in terms of simulating capacity, speed and time scale (see General Introduction).

Nonetheless, our approach was still imperfect, yielding often very large parameter confidence intervals, and at times insufficient, for example, preventing a reliable estimation of the colonisation of Cyprus. This was due to multiple factors, such as the necessary simplification of what are certainly more intricate demographic histories, not having a reliable mutation rate for our species, the lack of known events to calibrate time estimates and using non-parametric instead of parametric bootstrapping which likely contributed to the wide intervals. Lastly, it was still a quite long and computationally intensive process.

A recent study[318] used *δaδi* for population pairwise inferences, which were then used to calibrate the large multi-population modelling in *fastsimcoal2*. Though it is hard to estimate if there is a

gain in terms of speed, this two-step approach seems promising as it probably yields more accurate parameter inferences by reducing the initial search range and facilitating parameter optimization in *fastsimcoal2*. Finally, newer methods continue to be developed, such as *moments*, a software that provides two new tools for demographic inference. The first tool[346] improves upon, and is faster than, *δaδi* by circumventing the solving of diffusion equations to estimate the SFS. It allows for the inference of selection on top of neutral demographic events, can now model up to five populations and is being increasingly used in non-model species[347–350], including birds[317,351]. The second tool proposes a whole new method of neutral demographic inference based on patterns of LD between loci, that does not require phased data and can be applied to tens of populations[352,353]. Both approaches are promising for faster and scalable inference of demographic histories, and the next few years are likely to see other the emergence of other inference tools.

Conclusion

In the present thesis, we showed how a widespread bird is affected by landscape barriers, namely mountains and large bodies of water, and illustrate diverse outcomes of wild isolated populations. The use of the barn owl as study organism provided multiple study opportunities, from its extensive distribution and presence on many islands, to its conspicuous colour variation, the availability of a reference genome and the compiled knowledge from decades of study. We add to this knowledge using genomic data, with which we elucidated convoluted patterns that traditional markers could not, revealed unexpected histories and detected adaptive genomic signatures.

Finally, it is worth noting that not all groups, institutions or countries have access to the kind of funds and resources required for the work presented here, not only for sequencing but also for computing. Consortiums that generate large amounts of data, like B10K, and the standardization of open-access policy are crucial in addressing, at least, the sequencing hurdle and should therefore be encouraged. By allowing more researchers to exploit the existing data, a broader range of issues can be tackled thereby improving the total knowledge on the studied organism or communities and contributing to the overall advancement of molecular population genomics, particularly in non-model species.

Author affiliations

a. Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
b. Bioinformatics Competence Centre, University of Lausanne, Lausanne, Switzerland
c. Lausanne Genomic Technologies Facility, Lausanne, Switzerland
d. Hyldehegnet 27, 6400 Sønderborg, Denmark
e. Mediterranean Institute for Agriculture, Environment and Development, Laboratory of Ornithology, IIFA, University of Évora, Évora, Portugal
f. BirdWatch Ireland, Kilcoole, Co. Wicklow, Ireland
g. Gettorfer Weg 13, 24214 Neuwittenbek, Germany
h. Association CHENE, Centre d'Hébergement et d'Etude sur la Nature et l'Environnement, 76190 Allouville-Bellefosse, France
i. Barn Owl Trust, Devon, United Kingdom
j. Swiss Institute of Bioinformatics, Lausanne, Switzerland
k. Green Fund, Kifisia, Athens, Greece
l. "TYTO" - Organization for the Management and Conservation of Biodiversity in Agricultural Ecosystems, Larisa, Greece
m. Centro Recupero Rapaci del Mugello, Firenze, Italy
n. Shamir Research Institute, University of Haifa, Katzrin, Israel
o. Department of Geography and Environmental Sciences, University of Haifa, Haifa, Israel
p. Game and Fauna Service, Ministry of the Interior, Nicosia, Cyprus
q. Natural History Museum of Crete, University of Crete, Herakleio, Greece
r. Centro di Recupero per la Fauna Selvatica–LIPU, Rome, Italy
s. Los Barros 21, E-38410 Los Realejos, Tenerife, Canary Islands, Spain
t. Natural History Museum of Belgrade, Belgrade, Serbia
u. Palić Zoo, Palić, Serbia

Publication list

(Products of collaborations during the thesis)

1. Prost, S., Machado, A. P. *et al.* Genomic Analyses Show Extremely Perilous Conservation Status of African and Asiatic cheetahs (*Acinonyx jubatus*)– Under Review at *Mol. Ecol.*
2. Séchaud, R., Machado, A. P. *et al.* Behaviour-specific habitat selection patterns of breeding barn owls. *Mov. Ecol.* 9, (2021) doi: 10.1186/s40462-021-00258-6.
3. Séchaud, R., Machado, A. P., Schalcher, K., Simon, C. & Roulin, A. Communally breeding female Barn Owls Tyto alba are not related and do not invest similarly in the communal family. *Bird Study* 1–4 (2020) doi:10.1080/00063657.2020.1732291.

References

(Concerns the whole document, except chapter 5)

1. Harris, H. Enzyme polymorphisms in man. *Proc. R. Soc. B Biol. Sci.* **164**, 298–310 (1966).
2. Hubby, J. L. & Lewontin, R. C. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura. *Genetics* **54**, 577–594 (1966).
3. Hahn, M. W. *Molecular Population Genetics*. (Oxford University Press, 2019).
4. Charlesworth, B. & Charlesworth, D. Population genetics from 1966 to 2016. *Heredity (Edinb).* **118**, 2–9 (2017).
5. Wright, S. Evolution in mendelian populations. *Genetics* **16**, 97 (1931).
6. Kimura, M. & Weiss, G. H. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* **49**, 561–576 (1964).
7. Wright, S. Isolation by Distance. *Genetics* **28**, 114–138 (1943).
8. Twyford, A. D., Wong, E. L. Y. & Friedman, J. Multi-level patterns of genetic structure and isolation by distance in the widespread plant *Mimulus guttatus*. *Heredity (Edinb).* **125**, 227–239 (2020).
9. Pusadee, T., Jamjod, S., Chiang, Y. C., Rerkasem, B. & Schaal, B. A. Genetic structure and isolation by distance in a landrace of Thai rice. *Proc. Natl. Acad. Sci.* **106**, 13880–13885 (2009).
10. Darvill, B., Ellis, J. S., Lye, G. C. & Goulson, D. Population structure and inbreeding in a rare and declining bumblebee, *Bombus muscorum* (Hymenoptera: Apidae). *Mol. Ecol.* **15**, 601–611 (2006).
11. Blanquer, A. & Uriz, M. J. Population genetics at three spatial scales of a rare sponge living in fragmented habitats. *BMC Evol. Biol.* **10**, 1–9 (2010).
12. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
13. Pinhal, D. *et al.* Restricted connectivity and population genetic fragility in a globally endangered Hammerhead Shark. *Rev. Fish Biol. Fish.* **30**, 501–517 (2020).
14. Manel, S., Schwartz, M. K., Luikart, G. & Taberlet, P. Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution* vol. 18 189–197 (2003).
15. Bradburd, G. S. & Ralph, P. L. Spatial Population Genetics: It's about Time. *Annual Review of Ecology, Evolution, and Systematics* vol. 50 427–449 (2019).
16. Ficetola, G. F., Mazel, F. & Thuiller, W. Global determinants of zoogeographical boundaries. *Nat. Ecol. Evol.* **1**, 89 (2017).
17. Antonelli, A. Biogeography: Drivers of bioregionalization. *Nat. Ecol. Evol.* **1**, 0114 (2017).
18. Boutilier, S. T., Taylor, S. A., Morris-Pocock, J. A., Lavoie, R. A. & Friesen, V. L. Evidence for genetic differentiation among Caspian tern (*Hydroprogne caspia*) populations in North America. *Conserv. Genet.* **15**, 275–281 (2014).
19. Zalewski, A., Piertney, S. B., Zalewska, H. & Lambin, X. Landscape barriers reduce gene flow in an invasive carnivore: Geographical and local genetic structure of American mink in Scotland. *Mol. Ecol.* **18**, 1601–1615 (2009).
20. Mcrae, B. H., Beier, P., Dewald, L. E., Huynh, L. Y. & Keim, P. Habitat barriers limit gene flow and illuminate historical events in a wide-ranging carnivore, the American puma. *Mol. Ecol.* **14**, 1965–1977 (2005).

21. Keyghobadi, N., Roland, J. & Strobeck, C. Influence of landscape on the population genetic structure of the alpine butterfly *Parnassius smintheus* (Papilionidae). *Mol. Ecol.* **8**, 1481–1495 (1999).

22. Frantz, A. C. *et al.* Comparative landscape genetic analyses show a Belgian motorway to be a gene flow barrier for red deer (*Cervus elaphus*), but not wild boars (*Sus scrofa*). *Mol. Ecol.* **21**, 3445–3457 (2012).

23. Munshi-South, J. Urban landscape genetics: Canopy cover predicts gene flow between white-footed mouse (*Peromyscus leucopus*) populations in New York City. *Mol. Ecol.* **21**, 1360–1378 (2012).

24. Hartmann, S. A., Steyer, K., Kraus, R. H. S., Segelbacher, G. & Nowak, C. Potential barriers to gene flow in the endangered European wildcat (*Felis silvestris*). *Conserv. Genet.* **14**, 413–426 (2013).

25. Velo-Antón, G., Zamudio, K. R. & Cordero-Rivera, A. Genetic drift and rapid evolution of viviparity in insular fire salamanders (*Salamandra salamandra*). *Heredity (Edinb).* **108**, 410–418 (2012).

26. Hurston, H. *et al.* Effects of fragmentation on genetic diversity in island populations of the Aegean wall lizard *Podarcis erhardii* (Lacertidae, Reptilia). *Mol. Phylogenet. Evol.* **52**, 395–405 (2009).

27. Hayes, F. E. & Sewlal, J.-A. N. The Amazon River as a dispersal barrier to passerine birds: effects of river width, habitat and taxonomy. *J. Biogeogr.* **31**, 1809–1818 (2016).

28. Dobzhansky, T. *Genetics and the Origin of Species*. vol. 11 (Columbia University Press, 1982).

29. Nosil, P. Speciation with gene flow could be common. *Mol. Ecol.* **17**, 2103–2106 (2008).

30. Mayr, E. *Systematics and the origin of species*. (Columbia University Press, 1942).

31. Grant, P. R. *Evolution on Islands*. (Oxford University Press, 1998).

32. Warren, B. H. *et al.* Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecology Letters* vol. 18 200–217 (2015).

33. MacArthur, R. H. & Wilson, E. O. *The theory of island biogeography*. (Princeton University Press, 1967).

34. Losos, J. B. & Ricklefs, R. E. Adaptation and diversification on islands. *Nature* **457**, 830–6 (2009).

35. Mayr, E. Change of genetic environment and evolution. in *Evolution as a Process* (eds. Huxley, J., Hardy, A. C. & Ford, E. B.) 157–180 (Allen & Unwin, 1954).

36. Barton, N. H. Natural selection and random genetic drift as causes of evolution on islands. *Philos. Trans. R. Soc. B Biol. Sci.* **351**, 785–795 (1996).

37. Masel, J. Genetic drift. *Curr. Biol.* **21**, R837–R838 (2011).

38. Frankham, R. Inbreeding and extinction: Island populations. *Conserv. Biol.* **12**, 665–675 (1998).

39. Valente, L. *et al.* A simple dynamic model explains the diversity of island birds worldwide. *Nature* 1–5 (2020) doi:10.1038/s41586-020-2022-5.

40. Whittaker, R. J. & Fernández-Palacios, J. M. *Island Biogeography: Ecology, Evolution and Conservation*. (Oxford University Press, 2007).

41. Valen, L. Van. Body Size and Numbers of Plants and Animals. *Evolution (N. Y).* **27**, 27 (1973).

42. Foster, J. B. Evolution of Mammals on Islands. *Nature* **202**, 234 (1964).

43. Gillespie, R. Community assembly through adaptive radiation in Hawaiian spiders. *Science (80-.).* **303**, 356–359 (2004).

44. Gavrilets, S. & Losos, J. B. Adaptive radiation: Contrasting theory with data. *Science* vol. 323 732–737 (2009).

45. Gillespie, R. G. *et al.* Long-distance dispersal: a framework for hypothesis testing. *Trends Ecol. Evol.* **27**, (2012).

46. Felicísimo, Á. M., Muñoz, J. & González-Solis, J. Ocean surface winds drive dynamics of transoceanic aerial movements. *PLoS One* **3**, 2928 (2008).

47. Cassey, P., Blackburn, T. M., Sol, D., Duncan, R. P. & Lockwood, J. L. Global patterns of introduction effort and establishment success in birds. *Proc. R. Soc. B Biol. Sci.* **271**, S405–S408 (2004).

48. Armstrong, C. *et al.* Genomic associations with bill length and disease reveal drift and selection across island bird populations. *Evol. Lett.* **2**, 22–36 (2018).

49. Spurgin, L. G., Illera, J. C., Jorgensen, T. H., Dawson, D. A. & Richardson, D. S. Genetic and phenotypic divergence in an island bird: Isolation by distance, by colonization or by adaptation? *Mol. Ecol.* **23**, 1028–1039 (2014).

50. Grant, P. & R. The adaptive significance of some size trends in island birds. *Evolution (N. Y).* **19**, 355–367 (1965).

51. Clegg, S. M. & Owens, I. P. F. The 'island rule' in birds: Medium body size and its ecological explanation. *Proc. R. Soc. B Biol. Sci.* **269**, 1359–1365 (2002).

52. Lokatis, S. & Jeschke, J. M. The island rule: An assessment of biases and research trends. *J. Biogeogr.* **45**, 289–303 (2018).

53. Covas, R. Evolution of reproductive life histories in island birds worldwide. *Proc. R. Soc. B Biol. Sci.* **279**, 1531–1537 (2012).

54. Sayol, F., Downing, P. A., Iwaniuk, A. N., Maspons, J. & Sol, D. Predictable evolution towards larger brains in birds colonizing oceanic islands. *Nat. Commun.* **9**, 2820 (2018).

55. Wright, N. A., Steadman, D. W. & Witt, C. C. Predictable evolution toward flightlessness in volant island birds. *Proc. Natl. Acad. Sci.* **113**, 4765–4770 (2016).

56. Doutrelant, C. *et al.* Worldwide patterns of bird colouration on islands. *Ecol. Lett.* **19**, 537–545 (2016).

57. Grant, P. R. Plumage and the Evolution of Birds on Islands. *R. Grant Source Syst. Zool.* **14**, 47–52 (1965).

58. Clegg, S. M. Evolutionary changes following island colonization in birds: empirical insights into the roles of microevolutionary processes. in *The Theory of Island Biogeography Revisited* 293–325 (Oxford University Press, 2010).

59. Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M. & Fleischer, R. C. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Curr. Biol.* **21**, 1838–1844 (2011).

60. Navalón, G., Marugán-Lobón, J., Bright, J. A., Cooney, C. R. & Rayfield, E. J. The consequences of craniofacial integration for the adaptive radiations of Darwin's finches and Hawaiian honeycreepers. *Nat. Ecol. Evol.* **4**, 270–278 (2020).

61. Grant, P. R. *Ecology and Evolution of Darwin's Finches*. (Princeton University Press, 1999).

62. Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).

63. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

64. Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).

65. Dalloul, R. A. *et al.* Multi-platform next-generation sequencing of the Domestic Turkey (*Meleagris gallopavo*): Genome assembly and analysis. *PLoS Biol.* **8**, e1000475 (2010).

66. Warren, W. C. *et al.* The genome of a songbird. *Nature* **464**, 757–762 (2010).

67. Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–20 (2014).

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

69. Ellegren, H. *et al.* The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**, 756–60 (2012).

70. Zhang, G. *et al.* Comparative genomics reveals insights into avian genome evolution and adaptation. *Science (80-.).* **346**, 1311–1320 (2014).

71. Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science (80-.).* **346**, 1320–1331 (2014).

72. Zhang, G. Bird sequencing project takes off. *Nature* **522**, 34 (2015).

73. Feng, S. *et al.* Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).

74. Beichman, A. C., Huerta-Sanchez, E. & Lohmueller, K. E. Using genomic data to infer historic population dynamics of nonmodel organisms. *Annu. Rev. Ecol. Evol. Syst.* **49**, 433–456 (2018).

75. Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust Demographic Inference from Genomic and SNP Data. *PLoS Genet.* **9**, 1003905 (2013).

76. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, 1000695 (2009).

77. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

78. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–25 (2014).

79. Delmore, K. E. & Liedvogel, M. Avian Population Genomics Taking Off: Latest Findings and Future Prospects. in *Statistical Population Genomics* (ed. Dutheil, J. Y.) vol. 2090 413–433 (Humana Press, 2020).

80. Oswald, J. A., Overcast, I., Mauck, W. M., Andersen, M. J. & Smith, B. T. Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds. *Mol. Ecol.* **26**, 1386–1400 (2017).

81. Funk, E. R. *et al.* Phylogenomic data reveal widespread introgression across the range of an Alpine and Arctic specialist. *Syst. Biol.* **70**, 527–541 (2021).

82. Vianna, J. A. *et al.* Genome-wide analyses reveal drivers of penguin diversification. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22303–22310 (2020).

83. Burri, R. *et al.* Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* **25**, 1656–1665 (2015).

84. Poelstra, J. W. *et al.* The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science (80-.).* **344**, 1410–4 (2014).

85. Kardos, M. *et al.* Genomic consequences of intensive inbreeding in an isolated wolf

population. *Nat. Ecol. Evol.* **2**, 124–131 (2018).

86. Wolf, J. B. W. & Ellegren, H. Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* (2016) doi:10.1038/nrg.2016.133.

87. Seehausen, O. *et al.* Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192 (2014).

88. Newton, I. *Speciation & Biogeography of Birds*. (Academic Press, 2003).

89. Uva, V., Päckert, M., Cibois, A., Fumagalli, L. & Roulin, A. Comprehensive molecular phylogeny of barn owls and relatives (Family: Tytonidae), and their six major Pleistocene radiations. *Mol. Phylogenet. Evol.* **125**, 127–137 (2018).

90. Aliabadian, M., Alaei-Kakhki, N., Mirshamsi, O., Nijman, V. & Roulin, A. Phylogeny, biogeography, and diversification of barn owls (Aves: Strigiformes). *Biol. J. Linn. Soc.* **119**, 904–918 (2016).

91. Roulin, A. & Salamin, N. Insularity and the evolution of melanism, sexual dichromatism and body size in the worldwide-distributed barn owl. *J. Evol. Biol.* **23**, 925–934 (2010).

92. Ducrest, A.-L. *et al.* New genome assembly of the barn owl (*Tyto alba alba*). *Ecol. Evol.* **10**, 2284–2298 (2020).

93. Machado, A. P. *et al.* Unexpected post-glacial colonisation route explains the white colour of barn owls (*Tyto alba*) from the British Isles. *bioRxiv* (2021) doi:10.1101/2021.04.23.441058.

94. Savalli, U. M. The Evolution of Bird Coloration and Plumage Elaboration. in *Current Ornithology* (ed. Power, D. M.) 141–190 (Springer US, 1995). doi:10.1007/978-1-4615-1835-8_5.

95. Roulin, A. & Dijkstra, C. Genetic and environmental components of variation in eumelanin and phaeomelanin sex-traits in the barn owl. *Heredity (Edinb).* **90**, 359–364 (2003).

96. Romano, A., Séchaud, R., Hirzel, A. H. & Roulin, A. Climate-driven convergent evolution of plumage colour in a cosmopolitan bird. *Glob. Ecol. Biogeogr.* (2019) doi:10.1111/geb.12870.

97. Roulin, A., Wink, M. & Salamin, N. Selection on a eumelanic ornament is stronger in the tropics than in temperate zones in the worldwide-distributed barn owl. *J. Evol. Biol.* **22**, 345–354 (2009).

98. Roulin, A. Geographic variation in sexual dimorphism in the barn owl *Tyto alba*: a role for direct selection or genetic correlation? *J. Avian Biol.* **34**, 251–258 (2003).

99. Voous, K. H. On the distributional and genetical origin of the intermediate population populations of th<e barn owl (*Tyto alba*) in Europe. in *Syllegomena biologica* (eds. Jordans, A. & Peus, F.) 420–443 (Ziemsen Verlag, Wittenberg, Germany, 1950).

100. Antoniazza, S. *et al.* Natural selection in a postglacial range expansion: The case of the colour cline in the European barn owl. *Mol. Ecol.* **23**, 5508–5523 (2014).

101. Antoniazza, S., Burri, R., Fumagalli, L., Goudet, J. & Roulin, A. Local adaptation maintains clinal variation in melanin-based coloration of European barn owls (*Tyto alba*). *Evolution (N. Y).* **64**, 1944–1954 (2010).

102. Burri, R. *et al.* The genetic basis of color-related local adaptation in a ring-like colonization around the Mediterranean. *Evolution (N. Y).* **70**, 140–153 (2016).

103. Roulin, A. The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. *Biol Rev Camb Philos Soc* **79**, 815–848 (2004).

104. Hegyi, G. *et al.* Reflectance variation in the blue tit crown in relation to feather structure. *J. Exp. Biol.* **221**, (2018).

105. Shawkey, M. D. & D'Alba, L. Interactions between colour-producing mechanisms and their effects on the integumentary colour palette. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).

106. Baião, P. C. & Parker, P. G. Evolution of the melanocortin-1 receptor (MC1R) in boobies and gannets (Aves, Suliformes). *J. Hered.* **103**, 322–329 (2012).

107. Brush, A. H. Metabolism of carotenoid pigments in birds. *FASEB J.* **4**, 2969–2977 (1990).

108. Brown, C. J. & Bruton, A. G. Plumage colour and feather structure of the bearded vulture (*Gypaetus barbatus*). *J. Zool.* **223**, 627–640 (1991).

109. Riley, P. A. Melanin. *Int. J. Biochem. Cell Biol.* **29**, 1235–1239 (1997).

110. Kijas, J. M. H. *et al.* Melanocortin receptor 1 (MC1R) mutations and coat color in pigs. *Genetics* **150**, 1177–1185 (1998).

111. Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. & Crossland, J. P. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science (80-.).* **313**, 101–104 (2006).

112. Valverde, P., Healy, E., Jackson, I., Rees, J. L. & Thody, A. J. Variants of the melanocyte–stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat. Genet.* **11**, 328–330 (1995).

113. Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R. & Mundy, N. The molecular basis of an avian plumage polymorphism in the wild: a point mutation in the melanocortin-1 receptor is perfectly associated with melanism in the bananaquit (*Coereba flaveola*). *Curr. Biol.* **11**, 550–557 (2001).

114. Mundy, N. I. *et al.* Conserved genetic basis of a quantitative plumage trait involved in mate choice. *Science (80-.).* **303**, 1870–1873 (2004).

115. San-Jose, L. M. *et al.* Effect of the MC1R gene on sexual dimorphism in melanin-based colorations. *Mol. Ecol.* **24**, 2794–2808 (2015).

116. San-Jose, L. M. *et al.* MC1R variants affect the expression of melanocortin and melanogenic genes and the association between melanocortin genes and coloration. *Mol. Ecol.* **26**, 259–276 (2017).

117. Johnson, J. A., Ambers, A. D. & Burnham, K. K. Genetics of plumage color in the gyrfalcon (*Falco rusticolus*): Analysis of the melanocortin-1 receptor gene. *J. Hered.* **103**, 315–321 (2012).

118. Yu, W. *et al.* Non-synonymous SNPs in MC1R gene are associated with the extended black variant in domestic ducks (*Anas platyrhynchos*). *Anim. Genet.* **44**, 214–216 (2013).

119. Ducrest, A.-L., Keller, L. & Roulin, A. Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends in Ecology and Evolution* vol. 23 502–510 (2008).

120. Roulin, A. & Ducrest, A.-L. Association between melanism, physiology and behaviour: A role for the melanocortin system. *Eur. J. Pharmacol.* **660**, 226–233 (2011).

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

122. Hewitt, G. M. Mediterranean Peninsulas: The Evolution of Hotspots. in *Biodiversity Hotspots* (eds. Zachos, F. E. & Habel, J. C.) 123–147 (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-20992-5_7.

123. Hewitt, G. M. Postglacial re-colonisation of European biota. *Biol. J. Linn. Soc.* **68**, 87–112 (1999).

124. Taberlet, P., Fumagalli, L., Wust-Saucy, A. G. & Cosson, J. F. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* **7**, 453–464 (1998).

125. Stewart, J. R. & Lister, A. M. Cryptic northern refugia and the origins of the modern biota. *Trends Ecol. Evol.* **16**, 608–613 (2001).

126. Bilton, D. T. *et al.* Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. *Proc. R. Soc. B Biol. Sci.* **265**, 1219–1226 (1998).

127. Deffontaine, V. *et al.* Beyond the Mediterranean peninsulas: Evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Mol. Ecol.* **14**, 1727–1739 (2005).

128. García-Vázquez, A., Pinto Llona, A. C. & Grandal-d'Anglade, A. Post-glacial colonization of Western Europe brown bears from a cryptic Atlantic refugium out of the Iberian Peninsula. *Hist. Biol.* **31**, 618–630 (2019).

129. Herman, J. S. *et al.* Post-glacial colonization of Europe by the wood mouse, *Apodemus sylvaticus*: Evidence of a northern refugium and dispersal with humans. *Biol. J. Linn. Soc.* **120**, 313–332 (2017).

130. Montgomery, W. I., Provan, J., McCabe, A. M. & Yalden, D. W. Origin of British and Irish mammals: Disparate post-glacial colonisation and species introductions. *Quat. Sci. Rev.* **98**, 144–165 (2014).

131. Coles, B. J. Doggerland: a Speculative Survey. *Proc. Prehist. Soc.* **64**, 45–81 (1998).

132. Ward, I., Larcombe, P. & Lillie, M. The dating of Doggerland – post-glacial geochronology of the southern North Sea. *Environ. Archaeol.* **11**, 207–218 (2006).

133. Kelly, A., Charman, D. J. & Newnham, R. M. A last glacial maximum pollen record from bodmin moor showing a possible cryptic Northern refugium in Southwest England. *J. Quat. Sci.* **25**, 296–308 (2010).

134. Teacher, A. G. F., Garner, T. W. J. & Nichols, R. A. European phylogeography of the common frog (*Rana temporaria*): Routes of postglacial colonization into the British Isles, and evidence for an Irish glacial refugium. *Heredity (Edinb).* **102**, 490–496 (2009).

135. Snell, C., Tetteh, J. & Evans, I. H. Phylogeography of the pool frog (*Rana lessonae* Camerano) in Europe: Evidence for native status in Great Britain and for an unusual postglacial colonization route. *Biol. J. Linn. Soc.* **85**, 41–51 (2005).

136. Boston, E. S. M., Ian Montgomery, W., Hynes, R. & Prodöhl, P. A. New insights on postglacial colonization in western Europe: The phylogeography of the Leisler's bat (*Nyctalus leisleri*). *Proc. R. Soc. B Biol. Sci.* **282**, (2015).

137. Lister, A. M. Evolutionary and ecological origins of British deer. *Proc. R. Soc. Edinburgh. Sect. B. Biol. Sci.* **82**, 205–229 (1984).

138. O'Meara, D. B. *et al.* Genetic structure of Eurasian badgers *Meles meles* (Carnivora: Mustelidae) and the colonization history of Ireland. *Biol. J. Linn. Soc.* **106**, 893–909 (2012).

139. Brace, S. *et al.* The colonization history of British water vole (*Arvicola amphibius* (Linnaeus, 1758)): Origins and development of the Celtic fringe. *Proc. R. Soc. B Biol. Sci.* **283**, (2016).

140. Martínková, N., McDonald, R. A. & Searle, J. B. Stoats (*Mustela erminea*) provide evidence of natural overland colonization of Ireland. *Proc. R. Soc. B Biol. Sci.* **274**, 1387–1393 (2007).

141. Martin, J. R. *The Barn Owl: Guardian of the Countryside*. (Whittet Books, 2017).

142. Roulin, A. & Randin, C. F. Barn owls display larger black feather spots in cooler regions of the British Isles. *Biol. J. Linn. Soc.* (2016) doi:10.1111/bij.12814.

143. Roulin, A. Ring recoveries of dead birds confirm that darker pheomelanic Barn Owls disperse longer distances. *J. Ornithol.* **154**, 871–874 (2013).

144. van den Brink, V., Dreiss, A. N. & Roulin, A. Melanin-based coloration predicts natal dispersal in the barn owl, *Tyto alba*. *Anim. Behav.* **84**, 805–812 (2012).

145. Altwegg, R., Roulin, A., Kestenholz, M. & Jenni, L. Demographic effects of extreme winter weather in the barn owl. *Oecologia* **149**, 44–51 (2006).

146. Chin, C. S. *et al.* Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054 (2016).

147. Smit, A. F. & Hubley, R. RepeatModeler Open-1.0. http://www.repeatmasker.org.

148. Smit, A. F., Hubley, R. & Green, P. RepeatMasker Open-4.0. http://www.repeatmasker.org.

149. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016).

150. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database. *bioRxiv* 2020.08.10.245134 (2020) doi:10.1101/2020.08.10.245134.

151. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-Genome Annotation with BRAKER. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 65–95 (Springer New York, 2019). doi:10.1007/978-1-4939-9173-0_5.

152. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

153. Stanke, M., Schöffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**, 62 (2006).

154. van der Auwera, G. A. *et al.* From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **43**, 11.10.1-11.10.33 (2013).

155. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).

156. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

157. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).

158. Pickrell, J. & Pritchard, J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nat. Preced.* (2012) doi:10.1038/npre.2012.6956.1.

159. Weir, B. S. & Goudet, J. A unified characterization of population structure and relatedness. *Genetics* **206**, 2085–2103 (2017).

160. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution (N. Y).* **38**, 1358–1370 (1984).

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

162. Speek, G., Clark, J. A., Rohde, Z., Wassenaar, R. & Van Noordwijk, A. J. *The EURING exchange-code 2000*. (2001).

163. du Feu, C. R., Clark, J. A., Schaub, M., Fiedler, W. & Baillie, S. R. The EURING Data Bank – a critical tool for continental-scale studies of marked birds. *Ringing Migr.* **31**, 1–18 (2016).

164. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

165. Warren, D. L. & Seifert, S. N. Ecological niche modeling in Maxent: The importance of

model complexity and the performance of model selection criteria. *Ecol. Appl.* **21**, 335–342 (2011).

166. Excoffier, L. & Foll, M. fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**, 1332–1334 (2011).

167. Ravinet, M., Harrod, C., Eizaguirre, C. & Prodöhl, P. A. Unique mitochondrial DNA lineages in Irish stickleback populations: Cryptic refugium or rapid recolonization? *Ecol. Evol.* **4**, 2488–2504 (2014).

168. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).

169. Kocher, T. *et al.* Dynamics of mitochondrial DNA evolution in animals : Amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 6196–6200 (1989).

170. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* vol. 7 203–214 (2000).

171. Machado, A. P., Clément, L., Uva, V., Goudet, J. & Roulin, A. The Rocky Mountains as a dispersal barrier between barn owl (*Tyto alba*) populations in North America. *J. Biogeogr.* 1288–1300 (2018) doi:10.1111/jbi.13219.

172. *Barn Owl Report*. (2019).

173. Lambeck, K., Rouby, H., Purcell, A., Sun, Y. & Sambridge, M. Sea level and global ice volumes from the Last Glacial Maximum to the Holocene. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15296–15303 (2014).

174. Leorri, E., Cearreta, A. & Milne, G. Field observations and modelling of Holocene sea-level changes in the southern Bay of Biscay: Implication for understanding current rates of relative sea-level change and vertical land motion along the Atlantic coast of SW Europe. *Quat. Sci. Rev.* **42**, 59–73 (2012).

175. Doutrelant, C. *et al.* Worldwide patterns of bird colouration on islands. *Ecol. Lett.* **19**, 537–545 (2016).

176. Romano, A., Séchaud, R. & Roulin, A. Evolution of wing length and melanin-based coloration in insular populations of a cosmopolitan raptor. *J. Biogeogr.* **48**, 961–973 (2021).

177. McDevitt, A. *et al.* Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox (*Vulpes vulpes*), in Europe. *bioRxiv* (2020) doi:10.1101/2020.02.21.954966.

178. Py, I., Ducrest, A.-L., Duvoisin, N., Fumagalli, L. & Roulin, A. Ultraviolet reflectance in a melanin-based plumage trait is heritable. *Evol. Ecol. Res.* **8**, 483–491 (2006).

179. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

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

181. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 1–13 (2014).

182. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

183. R Development Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2016).

184. Graffelman, J. Exploring Diallelic Genetic Markers: The {HardyWeinberg} Package. *J. Stat.*

Softw. **64**, 1–23 (2015).

185. Graffelman, J. & Morales-Camarena, J. Graphical tests for Hardy-Weinberg Equilibrium based on the ternary plot. *Hum. Hered.* **65**, 77–84 (2008).

186. Exposito-Alonso, M. rbioclim: Improved getData function from the raster R package to interact with past, present and future climate data from worldclim.org. (2017).

187. BirdLife International, . The IUCN Red List of Threatened Species. Version 6.2. *BirdLife International* https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T22688504A155542941.en (2019).

188. Fourcade, Y. Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. *Ecol. Inform.* **36**, 8–14 (2016).

189. Swets, J. A. Measuring the accuracy of diagnostic systems. *Science (80-.).* **240**, 1285–1293 (1988).

190. Li, Y., Li, M., Li, C. & Liu, Z. Optimized maxent model predictions of climate change impacts on the suitable distribution of cunninghamia lanceolata in China. *Forests* **11**, 302 (2020).

191. Elith, J., Kearney, M. & Phillips, S. The art of modelling range-shifting species. *Methods Ecol. Evol.* **1**, 330–342 (2010).

192. Pouyet, F., Aeschbacher, S., Thiéry, A. & Excoffier, L. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. *Elife* **7**, (2018).

193. Malaspinas, A.-S. *et al.* A genomic history of Aboriginal Australia. *Nature* **538**, 207–214 (2016).

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

195. Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. Evol.* **17**, 230–241 (2002).

196. Frankham, R. Do island populations have less genetic variation than mainland populations? *Heredity (Edinb).* **78**, 311–327 (1997).

197. Médail, F. & Quézel, P. Biodiversity Hotspots in the Mediterranean Basin: Setting Global Conservation Priorities. *Conserv. Biol.* **13**, 1510–1513 (1999).

198. Bache, F. *et al.* A two-step process for the reflooding of the Mediterranean after the Messinian Salinity Crisis. *Basin Res.* **24**, 125–153 (2012).

199. Emin, D. *et al.* Home range and habitat selection of long-eared owls (*Asio otus*) in Mediterranean agricultural landscapes (Crete, Greece). *Avian Biol. Res.* **11**, 204–218 (2018).

200. Panter, C. T. *et al.* Kites (*Milvus* spp.) wintering on Crete. *Eur. Zool. J.* **87**, 591–596 (2020).

201. Bonhomme, F. *et al.* Genetic differentiation of the house mouse around the Mediterranean basin: Matrilineal footprints of early and late colonization. *Proc. R. Soc. B Biol. Sci.* **278**, 1034–1043 (2011).

202. Cucchi, T., Vigne, J. D., Auffray, J. C., Croft, P. & Peltenburg, E. Introduction involontaire de la souris domestique (*Mus musculus domesticus*) à Chypre dès le Néolithique précéramique ancien (fin IXe et VIIIe millénaires av. J.-C.). *Comptes Rendus - Palevol* **1**, 235–241 (2002).

203. Dubey, S. *et al.* Mediterranean populations of the lesser white-toothed shrew (*Crocidura suaveolens* group): An unexpected puzzle of Pleistocene survivors and prehistoric introductions. *Mol. Ecol.* **16**, 3438–3452 (2007).

204. Cumer, T. *et al.* Landscape and climatic variations of the Quaternary shaped multiple secondary contacts among barn owls (*Tyto alba*) of the Western Palearctic.

205. Clements, J. F. *et al. The eBird/Clements Checklist of Birds of the World: v2019*. https://www.birds.cornell.edu/clementschecklist/download/ (2019).

206. Zhang, M. *et al.* Preparation of megabase-sized DNA from a variety of organisms using the nuclei method for advanced genomics research. *Nat. Protoc.* **7**, 467–478 (2012).

207. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

208. Iwasaki, W. *et al.* Mitofish and mitoannotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.* **30**, 2531–2540 (2013).

209. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

210. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

211. Paradis, E. Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420 (2010).

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

213. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

214. Goudet, J. HIERFSTAT, a package for R to compute and test hierarchical F -statistics. *Mol. Ecol. Notes* **5**, 184–186 (2005).

215. McQuillan, R. *et al.* Runs of Homozygosity in European Populations. *Am. J. Hum. Genet.* **83**, 359–372 (2008).

216. Simaiakis, S. M. *et al.* Geographic changes in the Aegean Sea since the Last Glacial Maximum: Postulating biogeographic effects of sea-level rise on islands. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **471**, 108–119 (2017).

217. Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G. & Ellegren, H. Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences. *Curr. Biol.* **25**, 1375–1380 (2015).

218. Smeds, L., Qvarnström, A. & Ellegren, H. Direct estimate of the rate of germline mutation in a bird. *Genome Res.* **26**, 1211–1218 (2016).

219. Balloux, F. Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evolution (N. Y).* **58**, 1891–1900 (2004).

220. Zecchetto, S. & De Biasio, F. Sea surface winds over the Mediterranean basin from satellite data (2000-04): Meso- and local-scale features on annual and seasonal time scales. *J. Appl. Meteorol. Climatol.* **46**, 814–827 (2007).

221. Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: Windows into population history and trait architecture. *Nature Reviews Genetics* vol. 19 220–234 (2018).

222. Greig, J. R. A. & Warren, P. M. Early Bronze Age Agriculture in Western Crete. *Antiquity* **48**, 130–132 (1974).

223. Pareschi, M. T., Favalli, M. & Boschi, E. Impact of the Minoan tsunami of Santorini: Simulated scenarios in the eastern Mediterranean. *Geophys. Res. Lett.* **33**, n/a-n/a (2006).

224. Lenormand, T. Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**, 183–189 (2002).

225. MacArthur, R. H. & Wilson, E. O. An Equilibrium Theory of Insular Zoogeography. *Evolution (N. Y).* **17**, 373–387 (1963).

226. Gillespie, R., Croom, H. & Hasty, G. Phylogenetic Relationships and Adaptive Shifts among Major Clades of Tetragnatha Spiders (Araneae: Tetragnathidae) in Hawai'i. *Pacific Sci.* **51**, 380–394 (1997).

227. Losos, J. B., Jackman, T. R., Larson, A., De Queiroz, K. & Rodríguez-Schettino, L. Contingency and determinism in replicated adaptive radiations of island lizards. *Science (80-.).* **279**, 2115–2118 (1998).

228. Malinsky, M. *et al.* Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science (80-.).* **350**, 1493–1498 (2015).

229. Anguita, F. & Hernán, F. The Canary Islands origin: A unifying model. *J. Volcanol. Geotherm. Res.* **103**, 1–26 (2000).

230. Norder, S. J. *et al.* Beyond the Last Glacial Maximum: Island endemism is best explained by long-lasting archipelago configurations. *Glob. Ecol. Biogeogr.* **28**, 184–197 (2019).

231. Steinbauer, M. J. *et al.* Topography-driven isolation, speciation and a global increase of endemism with elevation. *Glob. Ecol. Biogeogr.* **25**, 1097–1107 (2016).

232. Carine, M. A., Humphries, C. J., Guma, I. R., Reyes-Betancort, J. A. & Santos Guerra, A. Areas and algorithms: Evaluating numerical approaches for the delimitation of areas of endemism in the Canary Islands archipelago. *J. Biogeogr.* **36**, 593–611 (2009).

233. Thorpe, R. S. & Baez, M. Geographic variation in scalation of the lizard Gallotia stehlini within the island of Gran Canaria. *Biol. J. Linn. Soc.* **48**, 75–87 (1993).

234. Molina-Borja, M. Sexual dimorphism of *Gallotia atlantica atlantica* and *Gallotia atlantica mahoratae* (Lacertidae) from the Eastern Canary Islands. *J. Herpetol.* **37**, 769–772 (2003).

235. Nogales, M., De León, L. & Gómez, R. On the presence of the endemic skink *Chalcides simonyi* Steind. 1891 in Lanzarote (Canary Islands). *Amphibia-Reptilia* **19**, 427–430 (1998).

236. Pestano, J., Brown, R. P., Suárez, N. M., Benzal, J. & Fajardo, S. Intraspecific evolution of Canary Island Plecotine bats, based on mtDNA sequences. *Heredity (Edinb).* **90**, 302–307 (2003).

237. Firmat, C. *et al.* Mandible morphology, dental microwear, and diet of the extinct giant rats *Canariomys* (Rodentia: Murinae) of the Canary Islands (Spain). *Biol. J. Linn. Soc.* **101**, 28–40 (2010).

238. Hutterer, R., Lopez-Jurado, L. F. & Vogel, P. The shrews of the eastern Canary Islands: A new species (mammalia: Soricidae). *J. Nat. Hist.* **21**, 1347–1357 (1987).

239. Senfeld, T. *et al.* Taxonomic status of the extinct Canary Islands Oystercatcher *Haematopus meadewaldoi*. *Ibis (Lond. 1859).* **162**, 1068–1074 (2020).

240. González Melián, E. Sex determination of the Canary Island Chiffchaff *Phylloscopus canariensis* using morphological traits and molecular sexing. *Ornithol Sci* **17**, 37–44 (2018).

241. Lifjeld, J. T. *et al.* Species-level divergences in multiple functional traits between the two endemic subspecies of Blue Chaffinches *Fringilla teydea* in Canary Islands. *BMC Zool.* **1**, 1–19 (2016).

242. Delgado, G., Carrillo, J. & Trujillo, D. Sobre la presencia y distribución de la lechuza común (*Tyto alba*) (Scopoli, 1769) en las islas orientales del archipiélago canario. *Vieraea Folia*

Sci. Biol. Canar. 145–148 (1992).

243. Goudet, J., Kay, T. & Weir, B. S. How to estimate kinship. *Mol. Ecol.* **27**, 4121–4135 (2018).

244. Leathlobhair, M. N. *et al.* The evolutionary history of dogs in the Americas. *Science (80-.).* **361**, 81–85 (2018).

245. Liu, L. *et al.* Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat. Commun.* **10**, 1–9 (2019).

246. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

247. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).

248. Martin, S. H. & Van Belleghem, S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).

249. Weir, B. S., Cardon, L. R., Anderson, A. D., Nielsen, D. M. & Hill, W. G. Measures of human population structure show heterogeneity among genomic regions. *Genome Res.* **15**, 1468–1476 (2005).

250. Ge, S. X., Jung, D., Jung, D. & Yao, R. ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628–2629 (2020).

251. Dolédec, S., Chessel, D. & Gimaret-Carpentier, C. Niche separation in community analysis: A new method. *Ecology* **81**, 2914–2927 (2000).

252. Dray, S. & Dufour, A. B. The ade4 package: Implementing the duality diagram for ecologists. *J. Stat. Softw.* **22**, 1–20 (2007).

253. Machado, A. P. *et al.* Genomic consequences of colonisation, migration and genetic drift in barn owl insular populations of the eastern Mediterranean. *prep*.

254. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993 (2008).

255. Kichaev, G. *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am. J. Hum. Genet.* **104**, 65–75 (2019).

256. Zhu, Z. *et al.* Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.* **145**, 537–549 (2020).

257. Turcot, V. *et al.* Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.* **50**, 26–35 (2018).

258. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases. *Cell* **182**, 1214-1231.e11 (2020).

259. Pulit, S. L. *et al.* Meta-Analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum. Mol. Genet.* **28**, 166–174 (2019).

260. Richardson, T. G., Sanderson, E., Elsworth, B., Tilling, K. & Smith, G. D. Use of genetic variation to separate the effects of early and later life adiposity on disease risk: Mendelian randomisation study. *BMJ* **369**, (2020).

261. Gu, J., Liang, Q., Liu, C. & Li, S. Genomic Analyses Reveal Adaptation to Hot Arid and Harsh Environments in Native Chickens of China. *Front. Genet.* **11**, 582355 (2020).

262. Wain, L. V. *et al.* Novel Blood Pressure Locus and Gene Discovery Using Genome-Wide Association Study and Expression Data Sets from Blood and the Kidney. *Hypertension* **70**, e4–e19 (2017).

263. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.* **41**, 666–676 (2009).

264. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).

265. Hoffmann, T. J. *et al.* Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* **49**, 54–64 (2017).

266. Ehret, G. B. *et al.* The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* **48**, 1171–1184 (2016).

267. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* **48**, 1162–1170 (2016).

268. Kulminski, A. M. *et al.* Strong impact of natural-selection-free heterogeneity in genetics of age-related phenotypes. *Aging (Albany. NY).* **10**, 492–514 (2018).

269. German, C. A., Sinsheimer, J. S., Klimentidis, Y. C., Zhou, H. & Zhou, J. J. Ordered multinomial regression for genetic association analysis of ordinal phenotypes at Biobank scale. *Genet. Epidemiol.* **44**, 248–260 (2020).

270. Halonen, J. I., Zanobetti, A., Sparrow, D., Vokonas, P. S. & Schwartz, J. Relationship between outdoor temperature and blood pressure. *Occup. Environ. Med.* **68**, 296–301 (2011).

271. Darre, M. J. & Harrison, P. C. Heart rate, blood pressure, cardiac output, and total peripheral resistance of single comb White Leghorn hens during an acute exposure to 35 C ambient temperature. *Poult. Sci.* **66**, 541–547 (1987).

272. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415-1429.e19 (2016).

273. Oskarsson, G. R. *et al.* Predicted loss and gain of function mutations in ACO1 are associated with erythropoiesis. *Commun. Biol.* **3**, 1–10 (2020).

274. Chen, M. H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213.e14 (2020).

275. O'Brien, K. A., Simonson, T. S. & Murray, A. J. Metabolic adaptation to high altitude. *Current Opinion in Endocrine and Metabolic Research* vol. 11 33–41 (2020).

276. Witt, K. E. & Huerta-Sánchez, E. Convergent evolution in human and domesticate adaptation to high-altitude environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 374 (2019).

277. Juan, C., Emerson, B. C., Oromí, P. & Hewitt, G. M. Colonization and diversification: Towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology and Evolution* vol. 15 104–109 (2000).

278. Palacios, C. J. Current status and distribution of birds of prey in the Canary Islands. *Bird Conserv. Int.* **14**, 203–213 (2004).

279. Tachmazidou, I. *et al.* Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. *Am. J. Hum. Genet.* **100**, 865–884 (2017).

280. Hagenaars, S. P. *et al.* Genetic prediction of male pattern baldness. *PLoS Genet.* **13**, e1006594 (2017).

281. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants in the Japanese population. *Nat. Commun.* **10**, 1–11 (2019).

282. Chen, M. H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213.e14 (2020).

283. Ramsuran, V. *et al.* Duffy-null-associated low neutrophil counts influence HIV-1

susceptibility in high-risk South African black women. *Clin. Infect. Dis.* **52**, 1248–1256 (2011).

284. van Setten, J. *et al.* Genome-wide association meta-analysis of 30,000 samples identifies seven novel loci for quantitative ECG traits. *Eur. J. Hum. Genet.* **27**, 952–962 (2019).

285. Giri, A. *et al.* Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. *Nat. Genet.* **51**, 51–62 (2019).

286. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).

287. Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518 (2019).

288. Allen, H. L. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).

289. Gudbjartsson, D. F. *et al.* Many sequence variants affecting diversity of adult human height. *Nat. Genet.* **40**, 609–615 (2008).

290. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014).

291. He, M. *et al.* Meta-analysis of genome-wide association studies of adult height in East Asians identifies 17 novel loci. *Hum. Mol. Genet.* **24**, 1791–1800 (2015).

292. Jeong, C. *et al.* Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. *PLoS Genet.* **14**, e1007650 (2018).

293. Jeong, C. *et al.* Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat. Commun.* **5**, (2014).

294. Yang, J. *et al.* Genetic signatures of high-altitude adaptation in Tibetans. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4189–4194 (2017).

295. Ntalla, I. *et al.* Multi-ancestry GWAS of the electrocardiographic PR interval identifies 202 loci underlying cardiac conduction. *Nat. Commun.* **11**, 1–12 (2020).

296. Ganesh, S. K. *et al.* Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat. Genet.* **41**, 1191–1198 (2009).

297. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat. Genet.* **50**, 390–400 (2018).

298. van Rooij, F. J. A. *et al.* Genome-wide Trans-ethnic Meta-analysis Identifies Seven Genetic Loci Influencing Erythrocyte Traits and a Role for RBPMS in Erythropoiesis. *Am. J. Hum. Genet.* **100**, 51–63 (2017).

299. Kowalski, M. H. *et al.* Use of >100,000 NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium whole genome sequences improves imputation quality and detection of rare variant associations in admixed African and Hispanic/Latino populations. *PLoS Genet.* **15**, e1008500 (2019).

300. Hodonsky, C. J. *et al.* Genome-wide association study of red blood cell traits in Hispanics/Latinos: The Hispanic Community Health Study/Study of Latinos. *PLoS Genet.* **13**, e1006760 (2017).

301. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat. Genet.* **42**, 210–215 (2010).

302. Deng, X. *et al.* Genome wide association study (GWAS) of chagas cardiomyopathy in trypanosoma cruzi seropositive subjects. *PLoS One* **8**, e79629 (2013).

303. Morgan, M. D. *et al.* Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. *Nat. Commun.* **9**, 1–10 (2018).

304. Huntley, B. & Webb, T. Migration: Species' Response to Climatic Variations Caused by Changes in the Earth's Orbit. *J. Biogeogr.* **16**, 5–19 (1989).

305. Irwin, D. E., Irwin, J. H. & Price, T. D. Ring species as bridges between microevolution and speciation BT - Microevolution Rate, Pattern, Process. in (eds. Hendry, A. P. & Kinnison, M. T.) 223–243 (Springer Netherlands, 2001). doi:10.1007/978-94-010-0585-2_14.

306. Devitt, T. J., Baird, S. J. E. & Moritz, C. Asymmetric reproductive isolation between terminal forms of the salamander ring species *Ensatina eschscholtzii* revealed by fine-scale genetic analysis of a hybrid zone. *BMC Evol. Biol.* **11**, 245 (2011).

307. Pereira, R. J., Monahan, W. B. & Wake, D. B. Predictors for reproductive isolation in a ring species complex following genetic and ecological divergence. *BMC Evol. Biol.* **11**, 194 (2011).

308. Cacho, N. I. & Baum, D. A. The Caribbean slipper spurge Euphorbia tithymaloides: the first example of a ring species in plants. *Proc. R. Soc. B Biol. Sci.* **279**, 3377–3383 (2012).

309. Stöck, M. *et al.* Cryptic diversity among Western Palearctic tree frogs: Postglacial range expansion, range limits, and secondary contacts of three European tree frog lineages (*Hyla arborea* group). *Mol. Phylogenet. Evol.* **65**, 1–9 (2012).

310. Willeit, M., Ganopolski, A., Calov, R. & Brovkin, V. Mid-Pleistocene transition in glacial cycles explained by declining CO2 and regolith removal. *Sci. Adv.* **5**, (2019).

311. Clark, P. U. *et al.* The Last Glacial Maximum. *Science (80-.).* **325**, 710–714 (2009).

312. Bertl, J., Ringbauer, H. & Blum, M. G. B. Can secondary contact following range expansion be distinguished from barriers to gene flow? *PeerJ* **2018**, e5325 (2018).

313. Boucher, F. C., Zimmermann, N. E. & Conti, E. Allopatric speciation with little niche divergence is common among alpine *Primulaceae*. *J. Biogeogr.* **43**, 591–602 (2016).

314. Tomasello, S., Karbstein, K., Hodač, L., Paetzold, C. & Hörandl, E. Phylogenomics unravels Quaternary vicariance and allopatric speciation patterns in temperate-montane plant species: A case study on the *Ranunculus auricomus* species complex. *Mol. Ecol.* **29**, 2031–2049 (2020).

315. Capblancq, T., Després, L., Rioux, D. & Mavárez, J. Hybridization promotes speciation in *Coenonympha butterflies*. *Mol. Ecol.* **24**, 6209–6222 (2015).

316. Hey, J. The divergence of chimpanzee species and subspecies as revealed in multipopulation isolation-with-migration analyses. *Mol. Biol. Evol.* **27**, 921–933 (2010).

317. Linck, E., Freeman, B. G. & Dumbacher, J. P. Speciation and gene flow across an elevational gradient in New Guinea kingfishers. *J. Evol. Biol.* **33**, 1643–1652 (2020).

318. Rougemont, Q. *et al.* Demographic history shaped geographical patterns of deleterious mutation load in a broadly distributed Pacific Salmon. *PLoS Genet.* **16**, e1008348 (2020).

319. Duranton, M. *et al.* The origin and remolding of genomic islands of differentiation in the European sea bass. *Nat. Commun.* **9**, 1–11 (2018).

320. Martin, M. *et al.* WhatsHap: fast and accurate read-based phasing. *bioRxiv* 085050 (2016) doi:10.1101/085050.

321. Delaneau, O., Zagury, J. F., Robinson, M. R., Marchini, J. L. & Dermitzakis, E. T. Accurate, scalable and integrative haplotype estimation. *Nat. Commun.* **10**, 1–10 (2019).

322. Choi, Y., Chan, A. P., Kirkness, E., Telenti, A. & Schork, N. J. Comparison of phasing strategies for whole human genomes. *PLoS Genet.* **14**, e1007308 (2018).

323. Goudet, J. & Jombart, T. hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.04-22. (2015).

324. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using

dense haplotype data. *PLoS Genet.* **8**, 1002453 (2012).

325. Elith, J. *et al.* A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**, 43–57 (2011).

326. Aguiar, A., Lopes, A. L., Pimenta, M. & Luís, A. Owls (Strigiformes) in Parque Nacional Peneda-Gerês (PNPG) – Portugal. *Nov. Acta Científica Compostel.* **19**, 83–92 (2010).

327. *Ornitologia Italiana - Vol.3 - Stercorariidae Caprimulgidae - Brichetti Pierandrea; Fracasso Giancarlo | Libro Alberto Perdisa Editore 06/2005 - HOEPLI.it*.

328. Schmid, H., Luder, R., Naef-Daenzer, B., Graf, R. & Zbinden, N. *Atlas des oiseaux nicheurs de Suisse : Distribution des oiseaux nicheurs en Suisse et au Liechtenstein en 1993-1996*. (1998).

329. Tucker, G., Heath, M. F., Tomiałojć, L. & Grimmett, R. F. A. *Birds in Europe: their conservation status*. (BirdLife International, 1994).

330. Göçer, E. & Johnson, D. H. The Barn Owl (*Tyto alba*) in Turkey. *Ela J. For. Wildl.* **7**, (2018).

331. Stojak, J., Borowik, T., Górny, M., McDevitt, A. D. & Wójcik, J. M. Climatic influences on the genetic structure and distribution of the common vole and field vole in Europe. *Mammal Res.* **64**, 19–29 (2019).

332. Dubey, S. *et al.* Molecular evidence of Pleistocene bidirectional faunal exchange between Europe and the Near East: The case of the bicoloured shrew (*Crocidura leucodon*, Soricidae). *J. Evol. Biol.* **20**, 1799–1808 (2007).

333. Dapporto, L. & Bruschini, C. Invading a refugium: Post glacial replacement of the ancestral lineage of a Nymphalid butterfly in the West Mediterranean. *Org. Divers. Evol.* **12**, 39–49 (2012).

334. Husemann, M., Schmitt, T., Zachos, F. E., Ulrich, W. & Habel, J. C. Palaearctic biogeography revisited: Evidence for the existence of a North African refugium for Western Palaearctic biota. *J. Biogeogr.* **41**, 81–94 (2014).

335. Lodolo, E. *et al.* Post-LGM coastline evolution of the NW Sicilian Channel: Comparing high-resolution geophysical data with Glacial Isostatic Adjustment modeling. *PLoS One* **15**, e0228087 (2020).

336. Heckel, G., Burri, R., Fink, S., Desmet, J.-F. & Excoffier, L. Genetic structure and colonization processes in European populations of the common vole, *Microtus arvalis*. *Evolution (N. Y).* **59**, 2231–2242 (2005).

337. Mátics, R. & Hoffmann, G. *Location of the transition zone of the Barn Owl subspecies Tyto alba alba and Tyto alba guttata (Strigiformes: Tytonidae)*. *Acta zoologica cracoviensia* vol. 45 (2002).

338. Huang, A. C. *et al.* Barn owls (</>Tyto alba</i>) in western North America: phylogeographic structure, connectivity, and genetic diversity. *Conserv. Genet.* **17**, 357–367 (2016).

339. Brito, P. H. The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in western Europe. *Mol. Ecol.* **14**, 3077–3094 (2005).

340. Miller, E. F. *et al.* Post-glacial expansion dynamics, not refugial isolation, shaped the genetic structure of a migratory bird, the Yellow Warbler (*Setophaga petechia*). *bioRxiv* 2021.05.10.443405 (2021) doi:10.1101/2021.05.10.443405.

341. Gousy-Leblanc, M., Yannic, G., Therrien, J.-F. & Lecomte, N. Mapping our knowledge on birds of prey population genetics. *Conserv. Genet.* **1**, 3 (2021).

342. Hui, R., D'Atanasio, E., Cassidy, L. M., Scheib, C. L. & Kivisild, T. Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Sci. Rep.* **10**, 18542 (2020).

343. Marciniak, S. & Perry, G. H. Harnessing ancient genomes to study the history of human

adaptation. *Nature Reviews Genetics* vol. 18 659–674 (2017).

344. Verdugo, M. P. *et al.* Ancient cattle genomics, origins, and rapid turnover in the Fertile Crescent. *Science (80-.).* **365**, 173–176 (2019).

345. Chen, N. *et al.* Whole-genome resequencing reveals world-wide ancestry and adaptive introgression events of domesticated cattle in East Asia. *Nat. Commun.* **9**, 1–13 (2018).

346. Jouganous, J., Long, W., Ragsdale, A. P. & Gravel, S. Inferring the joint demographic history of multiple populations: Beyond the diffusion approximation. *Genetics* **206**, 1549–1567 (2017).

347. Rippe, J. P., Dixon, G., Fuller, Z. L., Liao, Y. & Matz, M. Environmental specialization and cryptic genetic divergence in two massive coral species from the Florida Keys Reef Tract. *Mol. Ecol.* (2021) doi:10.1111/mec.15931.

348. Prada, C. & Hellberg, M. E. Speciation-by-depth on coral reefs: Sympatric divergence with gene flow or cryptic transient isolation? in *Journal of Evolutionary Biology* vol. 34 128–137 (2021).

349. Valdez, L. & D'Elía, G. Genetic Diversity and Demographic History of the Shaggy Soft-Haired Mouse *Abrothrix hirta* (Cricetidae; Abrotrichini). *Front. Genet.* **12**, 184 (2021).

350. Xiong, P. *et al.* The comparative genomic landscape of adaptive radiation in crater lake cichlid fishes. *Mol. Ecol.* **30**, 955–972 (2021).

351. Mikles, C. S. *et al.* Genomic differentiation and local adaptation on a microgeographic scale in a resident songbird. *Mol. Ecol.* **29**, 4295–4307 (2020).

352. Ragsdale, A. P. & Gravel, S. Unbiased estimation of linkage disequilibrium from unphased data. *Mol. Biol. Evol.* **37**, 923–932 (2020).

353. Ragsdale, A. P. & Gravel, S. Models of archaic admixture and recent history from two-locus statistics. *PLoS Genet.* **15**, e1008204 (2019).