An integrative study of bird migration.

From the migratory phenotype to its gene regulation mechanisms and back.

Dissertation

in fulfillment of the requirements for the degree

Doctor rerum naturalium

of the Faculty of Mathematics and Natural Sciences

at the Christian Albrechts University of Kiel

Submitted by

Juan Sebastian Lugo Ramos

Max Planck Research Group Behavioural Genomics

Max Planck Institute for Evolutionary Biology

Plön, March 2020

First referee: Dr Miriam Liedvogel.
Second referee: Prof. Dr. Hinrich Schulenburg
Examiner: Prof. Dr. Eva Stukenbrock
Date of oral examination: 15 of May 2020

Dedicated to

Maria Paula Latorre Guzman.

Brilliant mind. Loyal friend. Sublime love…Early migrant.

And to my family.

Wherever I might be you are compass to guide me home.

"*Farewell!*" they cried, "*wherever you fare,*
till your eyries receive you at the journey's end!"
That is the polite thing to say among eagles.
"*May the wind under your wings bear you*
where the sun sails and the moon walks,"
answered Gandalf, who knew the correct reply."
J.R.R Tolkien

"The hobbit or There and Back Again"

ABSTRACT

Life goes out of equilibrium; it is in constant movement. Animals, especially, move as part of their life cycle. An outstanding example is bird migration. Some birds adopt migration as a strategy to survive the harsh conditions of weather seasonality in temperate regions. Different sources of evidence indicate that seasonal migration is innate, and it can be inherited. Mutations in such heritable behaviour create an array of diversity in migratory traits: timing, orientation and distance.

The diversity of migratory traits can affect ecological speciation. Migratory divides, for instance, are geographical areas where birds with different migratory orientations hybridise. If the differences in migratory behaviour are strong enough to create reproductive barriers, this could evolve into population divergence and eventually, speciation. However, to understand the potential processes of divergence caused by migratory behaviours, a crucial element is missing: the identity of the molecular mechanisms involved in migration. Genome-wide studies in bird species with migratory divides find several different genomic regions with species-specific signature. Similarly, gene expression approaches in different organs and species find groups of individual differentially expressed genes. These results suggest an intricate mechanism for the genetics of migration with potential species-specific characteristics.

This thesis analyses the migratory behaviour from different angles spanning the phenotype to gene regulation, to contribute to the identification of mechanisms and evolution of migration.

Most of the chapters of this thesis use the Eurasian blackcap (*Sylvia atricapilla*) a species that comprise an extensive repertoire of orientation and distance traits, including entirely resident populations. With blackcaps, we studied the phenotypic variability of migration tracking individuals

throughout the year (Chapter 2). We used light-level geolocators to obtain migratory routes of individuals from populations in Central Europe and the United Kingdom. We describe for the first time the orientation and timing patterns of individuals from a migratory divide and a recently adapted population in the UK.

In chapter 4, we analyse the genomics and evolution patterns of blackcaps. Using whole-genome resequencing of populations covering all the differences in migratory traits, we describe population structure and demography in this species. We found that blackcaps show very little genomic differentiation. The most divergent populations are residents, while migratory populations comprise a single population at the genetic level.

Chapter 5 is the first study of gene regulatory mechanisms in the context of bird migration. We characterised the chromatin accessibility landscape in three brain areas contrasting individuals during migration with individuals out of the migratory season. One of the findings is a general pattern of gene repression in relevant brain regions like the Cluster N. Moreover; we found cis-regulatory modules with particular evolutionary trajectories that may play a role in migration.

Lastly, we did two comparative approaches to study macroevolutionary patterns related to migration. First, we analysed phylogenetic patterns and structural characteristics of previously proposed candidate genes (chapter 3). We found that the candidate genes do not have structural characteristics correlated with the presence of migration across the avian clade as it does within some species. The second comparative approach (Chapter 6), evaluates the repeatability patterns of genomic divergence in pairs of populations from migratory divides. Our results suggests that the degree of repeatability is mainly driven by how apart in the speciation continuum is the population pair located: if the pair is recently diverging,

few repeatability is detected, while if the populations are further apart, repeatability is more plausible.

Overall, this thesis highlights an essential feature for the study of complex traits like migration: integration of different sources of evidence. Ideally, in these cases, the analysis of phenotype, evolutionary patterns and regulatory mechanisms in the same individuals, should be the standard procedure. We are aware that this is an implausible scenario. However, the integration of different studies, help to guide the search of molecular elements involved in bird migration. This thesis is the first - at least that we are aware of - study compilating research on a variety of topics to understand bird migration.

We are still far from getting a definitive understanding of bird migration. Nevertheless, confirming the heritability of the phenotype, describing macro and microevolutionary patterns of migration and specific regulatory elements, will improve the search for new candidate genes for this behaviour.

KURZFASSUNG

Das Leben gerät aus dem Gleichgewicht; es ist in kontinuierlicher Bewegung. Insbesondere Tiere bewegen sich in bestimmten Abschnitten ihres Lebenszyklus. Ein herausragendes Beispiel dafür ist der Vogelzug. Einige Vogelarten nutzen den Vogelzug als Überlebensstrategie um die rauen Wetterbedingungen, verursacht durch die Jahreszeiten in der gemäßigten Zone der Erde, zu überleben. Verschiedene Beweisquellen zeigen, dass der jahreszeitliche Vogelzug angeboren ist und vererbt werden kann. Mutationen in einem solchen vererbten Verhalten erzeugen ein vielfältiges Spektrum in Zugeigenschaften: zeitliche Koordinierung, Richtung und Distanz.

Die Vielfältigkeit der Zugeigenschaften kann ökologische Artbildung beeinflussen. „Zugscheide" zum Beispiel sind geographische Gebiete in denen Vögel mit unterschiedlichen Zugrichtungen hybridisieren. Wenn die Unterschiede im Zugverhalten groß genug sind um Reproduktionsbarrieren zu bilden, kann dies zu Populationsdivergenz und schließlich Artbildung führen. Um allerdings die durch Zugverhalten verursachten potentiellen Prozesse der Divergenz zu verstehen, fehlt ein entscheidendes Element: die Identifizierung der in den Vogelzug involvierten molekularen Mechanismen. Genomweite Studien in Vogelarten mit Zugscheide finden mehrere unterschiedliche genomische Regionen mit artspezifischen Signaturen. Gleichermaßen finden Methoden, die Genexpression betrachten, in verschiedenen Organen und Arten Gruppen von individuell unterschiedlich exprimierten Genen. Diese Ergebnisse legen einen komplizierten Mechanismus für die genetische Grundlage des Vogelzugs mit eventuell artspezifischen Eigenschaften nahe.

Diese Doktorarbeit analysiert das Zugverhalten aus verschiedenen Winkeln, den Phänotypen bis hin zur Genregulation umfassend, um zu der Identifizierung des Mechanismus und Charakterisierung der Evolution des Vogelzugs beizutragen.

Die meisten Kapitel dieser Thesis betreffen die Mönchsgrasmücke (*Sylvia atricapilla*), eine Art, die ein umfangreiches Repertoire an Zugrichtungen und Zugdistanzen umfasst, einschließlich vollkommen residenter Populationen. In der Mönchsgrasmücke untersuchten wir die phänotypische Variabilität des Vogelzugs indem wir einzelne Vögel über das Jahr verfolgten (Kapitel 2). Wir beschreiben erstmals die Orientierung und zeitliche Zugstrategie von Individuen entlang einer Zugscheide, sowie die Brutgebiete einer erst seit kurzem in Großbritannien angesiedelten Überwinterungspopulation.

In Kapitel 4 analysieren wir Genomik und evolutionäre Muster der Mönchsgrasmücke. Mit genomweiter Resquenzierung von Populationen, die alle unterschiedliche Zugeigenschaften umfassen, beschreiben wir Populationsstruktur und Demografie in diese Vogelart. Wir fanden, dass Mönchsgrasmücken sehr wenig genomische Differenzierung zeigen. Die am meisten divergierenden Populationen sind Ziehende und Nicht-ziehende Populationen, wogegen die ziehenden Populationen auf dem genetischen Level eine einzelne Population darstellen.

Kapitel 5 ist die erste Studie über Genregulationsmechanismen im Zusammenhang mit dem Vogelzug. Wir charakterisierten die Chromatin "accessibility landscape" in drei Gehirnarealen, um Individuen während des Vogelzugs von Individuen außerhalb der Zugzeit zu unterscheiden. Eines der Erkenntnisse ist die Identifizierung eines Zugzeit-spezifischen Genexpressionsmusters in relevanten Hirnregionen wie Cluster N. Zudem fanden wir cis-regulative Module mit insbesondere evolutionären Pfaden, welche eine Rolle im Vogelzug spielen können.

Zuletzt verwendeten wir zwei vergleichende Ansätze um mit dem Vogelzug in Verbindung stehende makroevolutionäre Muster zu untersuchen. Zuerst analysierten wir phylogenetische Muster und strukturelle Charakteristika von im Vorhinein vorgeschlagenen Kandidatengenen (Kapitel 3). Wir fanden, dass keine strukturellen Merkmale der Kandidatengene mit dem Vorhandensein des Vogelzugs innerhalb der vogelartigen Klade korrelieren, wie sie es innerhalb einiger Arten tun. Der zweite vergleichende Ansatz

(Kapitel 6) evaluiert die Wiederholbarkeit der genomisch divergierten Muster in Paaren von Populationen der „migratory divides“. Unsere Ergebnisse legen nahe, dass der Grad der Wiederholbarkeit vor Allem dadurch bestimmt wird wie weit sich das Populationspaar im Artbildungskontinuum befindet: wenn das Paar jüngst divergiert wird eine geringe Wiederholbarkeit bemerkt, während, wenn die Populationen weiter auseinander sind, die Wiederholbarkeit eingängiger ist.

Insgesamt hebt diese Doktorarbeit eine essentielle Eigenschaft für das Studieren von komplexen Eigenschaften wie den Vogelzug hervor: die Integration verschiedener Beweisquellen. Idealerweise sollte die Standardprozedur in diesen Fällen die Analyse von Phänotypen, evolutionären Mustern and regulatorischen Mechanismen im selben Individuum sein. Wir sind uns darüber bewusst, dass dies ein unplausibles Szenario ist. Jedoch hilft die Integration verschiedener Studien die Suche nach molekularen Elementen, die in den Vogelzug involviert sind, zu lenken. Diese Doktorarbeit ist die erste – zu Mindestens uns bekannte - Studie, die Forschung vielfältiger Themen zusammenstellt um Vogelzug zu verstehen.

Wir sind immer noch weit von einem endgültigen Verstehen des Vogelzugs entfernt. Trotzdem wird die Suche nach neuen Kandidatengenen für dieses Verhalten durch die Bestätigung der Vererbbarkeit des Phänotyps und die Beschreibung makro und mikro evolutionären Muster und spezifischer regulatorischer Elemente verbessert werden.

Table of contents

OVERVIEW OF THE DISSERTATION

The topics covered in this thesis range from the behavioural components of bird migration, to speciation with a special emphasis around finding the molecular elements of migration and their evolutionary history. This thesis consists of three parts: first, gives a focus into migration at the phenotypic level, second at the gene sequence level and lastly a study parallel evolution and speciation.

In the introductory chapter 1, I review the basic biology of migration, its definition, behavioural features and ecology. I focus on the research of genetics and molecular elements of migration. Next, I discuss the evolution of migration and its potential impact on speciation. Finally, I present the Eurasian Blackcap (*Sylvia atricapilla*) as an excellent model for migration and focal study system used in my thesis.

Part I.

To understand the underlying genetic architecture of any complex behaviour, precise characterization of the focal phenotype is imperative. Chapter 2, *"Versatile migratory strategies and evolutionary insights revealed by tracks of wild Eurasian blackcaps"* is a colaboration where we addressed previous limitations of indirect approaches (ringing recovery, isotope, and funnel orientation analysis) to characterize the migratory phenotype. To do this we fit geolocators (small archival tags that record light intensity values) on blackcaps in Central Europe and the United Kingdom. In the study we characterized routes and timing patterns of 90 individuals, some of which were tracked for two years. This also allowed us to confirm the repeatability of migratory routes in the wild, and patterns of heritability previously described in funnel-based orientation and crossbreeding experiments. Phenotypic characterization and repeatability are crucial requirements to base evidence for the genetic nature of migration.

Part II

In this part, the focus is to explore the molecular elements of migration with three approaches. The first one evaluates the extent of the current gene candidates of migration in an evolutionary framework. Second, we used a genome-wide based approach to look for genes associated with migration in *S.*

atricapilla. Finally, we applied a gene regulation approach using chromatin accessibility to identify cis-regulatory elements involved in migration.In chapter 3, *"Candidate genes do not distinguish between migratory and nonmigratory species"* I used a phylogenetic and molecular evolution approach to analyze a set of previously suggested candidate genes for migration. Using available genome data for 70 bird species, I found that structural sequence characteristics of candidate genes (i.e. allele lenghts) are not a reliable measure for propensity to migration. This suggests that candidate genes for migration found in one species might not be necessary to other groups in the avian clade.

In chapter 4, *"The evolutionary history and genomics of European blackcap migration"* we analyzed population structure and genomic associations of migratory traits in Eurasian blackcaps (*Sylvia atricapilla*). We used whole-genome resequencing data of individuals across its breeding range in Europe, to compare all possible migratory phenotypes. Our analyses indicate that genomic differentiation among populations is very low, only a few SNPs show strong differences between migratory and resident populations. Most of the SNPs with high differentiation are located in non-coding regions suggesting a potential cis-regulatory role for the onset of migration. We also suggest that selection on standing variation is playing a role in recent adaptations of the phenotype.

As in many complex traits, regulatory elements play a relevant role in the evolution and onset of migration. In chapter 5 I took a functional approach to search for molecular elements of bird migration with a gene regulation perspective. In *"Controlling bird migration behaviour through cis-regulatory elements"* we looked for patterns of differential chromatin accessibility in an experimental setup to asses migratory behaviour. Because the migratory phenotype is only exhibited during the migratory season, we can contrast it with an off-season/control phenotype. Specifically, we characterized chromatin accessibility in three brain regions related to migration. We characterised the genomic regions changing chromatin accessibility and their potential Transcription Factor Binding Sites. These genomic regions harbour potential regulatory elements for the migratory phenotype. Additionally, evolutionary patterns show that these elements have gone through population changes that might have shaped the evolution of migration in Eurasian blackcaps.

Overall, part II of this thesis recognise that changes in coding and non-coding sequences are relevant for migratory behaviour. However, many sources of evidence point to a significant, mostly unexplored contribution from cis-regulatory sequences.

PART III

The last part of this thesis explores the elements of speciation in hybrid zones. In chapter 6 *"Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds"*, we analyzed the repeatability of genomic differentiation and divergence in eight pairs of bird populations forming hybrid zones. We found that repeatability can only be recognized once populations are clearly divergent, compared to population pairs that diverged recently. Repeatability in this context depends on where in the speciation continuum the pair of populations is located and if evolution has had sufficient time to leave recognizable selection signatures in the genome.

CHAPTER 1

THE BIOLOGY OF MIGRATION.

The first observations of animal migration dating around ancient Greek philosophers noticed that some animals vanished and reappeared almost suddenly during specific seasons of the year. A rigorous analysis of this phenomena, found that patterns of bird flocks flying, bison herds running, and butterfly swarms moving correlated with the seasonal absence of these animals. Since then, many questions and hypothesis about origins and characteristic of animal migration have been postulated.

A large body of scientific literature from various disciplines has established the foundations of our understanding of animal migration. Many of the questions why, when or how animals move in specific seasons are now answered by different disciplines of biology. Patterns of timing or *when* an animal population starts and stops to move [1] is mainly studied by chronobiology. The routes that a migratory bird could take or *where* do the animals go, is studied mainly by the field of movement ecology. [2]. The physiological adaptations and sensory mechanisms or *how* animals are able to accomplish the challenge of migration [3] is mainly studied by physiological, neuroanatomical and neurosensory approaches to behaviour. Finally, the answers of *why* animals have the urge to go to other areas [4] and what could be the benefits of exhibiting such an energetic intense behaviour, is studied in the context of evolutionary biology.

All of the biological disciplines involved in migration have laid the ground work for many answers. Even in genetics, it is already established that some traits are heritable [5], [6]. However, despite the collected evidence from all the different fields studying migration, the molecular mechanisms that enable certain individuals/populations or species to migrate, remain a complete mystery.

1.1 Definition, patterns and mechanisms of migration.

Migration can be distinguished from other movement behaviours (e.g dispersion or foraging) by two characteristics: 1) a directed, coordinated back and forth journey between two fixed territories, and 2) predictable seasonality. Animals

predictably move between breeding and non-breeding grounds, either alone or in groups. Usually, breeding grounds are

located in higher latitudes (like temperate or polar regions) and non-breeding grounds are located in lower equatorial/tropical latitudes.

Seasonal migration is ubiquitous across the animal kingdom. Examples of seasonal migrations range from blue whales swimming between Costa Rica and the Alaska [7], to the multigenerational monarch butterfly migratory cycle from North America to Mexico and Central America [8]. Though ubiquitous in animals, birds are arguably the taxon with the longest history of migration studies.

Glossary

Zugunruhe: or migratory restlessness: a characteristic behaviour observed in caged migratory songbirds at night during the migratory season. It consists of an increase nocturnal activity characterized mainly by directed hopping hopping and flying, as well as wing whirring while perched. This behaviour can be monitored and quantified through motion sensors and infrared video cameras.

Migratory divide: During glacial times, the geographical distribution of many bird species was confined to equatorial regions (glacial refugia). Allopatric populations developed specific migratory phenotypes. After glaciation, populations expanded and came close in geographical locations. The secondary contact of the populations creates a migratory divide.

Radical pair mechanism: A spin-chemical reaction initiated by light excitation of a donor molecule followed by electron transfer to an acceptor molecule and formation of a transient radical pair. The spin state of the unpaired electrons can naturally change between opposite (singlet state, S) or parallel (triplet state, T) spin orientations that lead to different end products. This interconversion rate can be altered depending on the orientation of the molecule within an ambient magnetic field, consequently shifting the reaction towards one of the two states and subsequently altering the yield in the end product.

Partial migratory populations: are composed of individuals from the same breeding grounds that constitutively migrate and individuals that are all year residents.

Genomic islands of divergence: Many sympatric populations experience gene flow. The genomic islands of divergence are parts of the genome that do not show evidence of gene flow. These regions stand out in the analysis of divergence using measures of relative differentiation (FsT) or absolute divergence (dxy). The genomic islands of divergence contrast with the rest of the genome that is under gene flow and should homogenize the divergence.

Migratory journeys displayed by birds range from a few kilometers to impressive pole-to-pole journeys performed by Artic terns (*Sterna artica*) .
The predictable seasonality of migration is linked to the life cycle of a migratory animal. In birds, it starts with hatching on their breeding grounds, where the fledglings grow and prepare for their first autumn migration. During the migratory season, they travel to the non-breeding grounds where they stay until they are prepared to travel back to their natal breeding grounds to reproduce and start the cycle all over again.

In the following sections (1.1.1 to 1.1.3) I will show the evidence of how birds use different adaptations and information sources to perform a migratory journey.

1.1.1. Choosing or adapting to migratory routes.

Bird migration routes are diverse and might vary even within the same species. Those routes are optimized to avoid geographical or ecological barriers that could be disadvantageous (e.g. high mountains, deserts or sea). Despite the optimization of migratory routes, population-specific variation in migratory routes may have different survival and fitness consequences [9].
Each population has its own consistent route for migration but this repeatability of route depends on age and genetic factors. Age adds components of learning and memory which makes adults more consistent than naïve juveniles that have never been to the area they are heading [10]. Genetic structure also influences repeatability considering that species with low population structure have less consistent routes [11].

In some species, migratory routes have gone through dramatic shifts in very recent times. Barn swallows (*Hirundo rustica*) usually migrate from North America to the Equator. Nonetheless, recent evidence shows that some barn swallows populations have now also populated southern South America, creating a pattern of migration that mirrors those from North American populations: this new population migrates from Argentina to the equator and back south, without going to North America like their relatives [12]. Similarly, introduced populations of House finches (*Carpodacus mexicanus*) [13] and natural populations of Eurasian Blackcaps (*Sylvia atricapilla*) have changed their migratory patterns and

evolved new strategies in recent times. The reasons for such recent changes may have different origins, environmental or genetic, but without knowledge about molecular mechanisms the reasons remains speculative.

1.1.2. Synchronization and timing to know when to migrate.

To know when to start migration, a combination of environmental cues and internal mechanisms work together to synchronize the life cycle of the migratory animals to the annual cycle. Environmental characteristics of the seasons trigger hormonal changes in birds to prepare for migration[14]. One of the preparatory mechanisms for migration is hyperphagia, a behavioural adaptation where birds start consuming large amounts of food, particularly sugar-rich sources, before migration[14]. This has the purpose to store fat in the body as fuel for the energy demand of the migratory journey. Hyperphagia often implies a shift in the regular diet. For instance, by the end of summer/beginning of autumn insectivorous birds shift to a frugivorous diet specializing in fruits with high sugar content like berries. Hyperphagia must be strictly controlled to balance the amount of fat deposition and the gain of mass to conform with the extent of migration distance [15], otherwise, the bird would carry an unnecessary load.

However, ought to potential fluctuations in weather and temperature from year to year, the environment is not reliable as the unique cue all the time. Internal processes keeping track of time (i.e. biological clocks), can take over to trigger and coordinate necessary changes for migration. The precise departures and arrivals of migratory birds during years of unconventional weather suggests the existence of an internal clock mechanism to keep track of life history events like breeding and moult. In experimental settings, *Zugunruhe,* or migratory restlessness, is a characteristic nocturnal behaviour that caged migratory birds exhibit when kept indoors. Recent studies across the genus *Saxicola*, found that the intensity of *Zugunruhe* of individuals correlates well with the phenotype of the population (e.g long distance migrants have longer intervals of higher intensity of zugunruhe)[6], [16]. The evidence from controlled conditions suggests that timing and intensity of this activity can be used as a proxy for migration in the wild.

The synchronisation mechanisms for migration are not entirely understood. However, these mechanisms must agree with physiological need and environmental cues, allowing a migratory bird an appropriate synchronisation to avoid negative consequences in fitness.

1.1.3. Adaptations to find the way.

1.1.3.1. How to navigate in space.

To move in space, animals must navigate and orientate using different external reference systems to keep their inherited [17] directional information. Some of the information comes from celestial cues like the stars or the sun, sunset and polarized light, as well as the Earth's magnetic field. Additionally, odour, landmarks and signposts are important elements to be integrated into the knowledge for the route they must take[15]. It is also well known that the sensory information of various sources is integrated in specific structures of the brain such as the hippocampus (see box 1).

Animal migration requires the development or enhancement of mechanisms that help animals to navigate. Generally, the mechanisms used by animals to navigate include: non-compass orientation, vector navigation, and compass orientation (For a review of all the mechanisms, see [18] in press). In any of these mechanisms, birds use various sources of cues to guide the path to follow during migration.

1.1.3.2. Celestial cues

The movement and position of the sun functions as a reliable orientation cue for diurnal birds. The movement of the sun always follows an east-west pattern and its position changes through the course of the day and year. Birds possess a time compensated sun compass which accounts for time of day when using the sun as a reference cue[15]. The stars can also be used as a reference cue in a different way than the time compensated sun compass. Birds might use the centre of axial rotation as an orientation cue that points poleward.

Sunlight gets deflected when it enters the atmosphere. This creates different amounts of polarization given the inclination of the light source. Experiments changing orientation and filtering light show that the polarization conditions of light change the direction that migratory birds tend to orientate [19]

1.1.3.3. Magnetic compass

The Earths' magnetic field is not uniformly distributed along all the surface. This magnetic field has a polarity going from South to North of the magnetic poles. Similarly, the intensity of the field changes across the globe having a maximum at the poles and diminishing in half towards the equator. Behavioural experiments have shown that birds can sense changes of inclination in the magnetic field and can use that information as directional cues for orientation [20] .

Despite the evidence that birds use the Earth's magnetic field to orient, we still do not clearly understand how they sense this information. Two main mechanisms have been proposed to explain this: (I) based on iron mineral structures (like magnetite) and (ii) based on a light-dependent biochemical reaction. Magnetite is accumulated in many living organisms as crystals inside some cells and could act as a compass needle and potentially aid sensing magnetic fields[21]. The light dependent biochemical reaction, called *radical pair mechanism,* has been suggested by theoretical physicists based on the fact that the rate of some biochemical reactions involving electron transfers forming transient radical-pairs, can be affected by magnetic fields as low as the Earth's magnetic field[21]. Currently the most promising candidate molecule that fulfills these requirements is cryptochrome 4 (CRY-4). CRY-4 is a member of a multigene family photosensible to blue light, however in differenc to other members of the family, it doesn fluctuate with the circadian rhytms [22].
Many sensory inputs are processed and integrated in dedicated brain regions, and increasing evidence suggests that magnetic compass information (at least in night-migratory songbirds) is processed in a specific forebrain area part of the visual Wulst, called Cluster N[23] (see box 1). This makes Cluster N a promising brain area to find molecular elements that are involved in shaping migratory behaviour.

BOX 1. The brain structure of a migrant bird

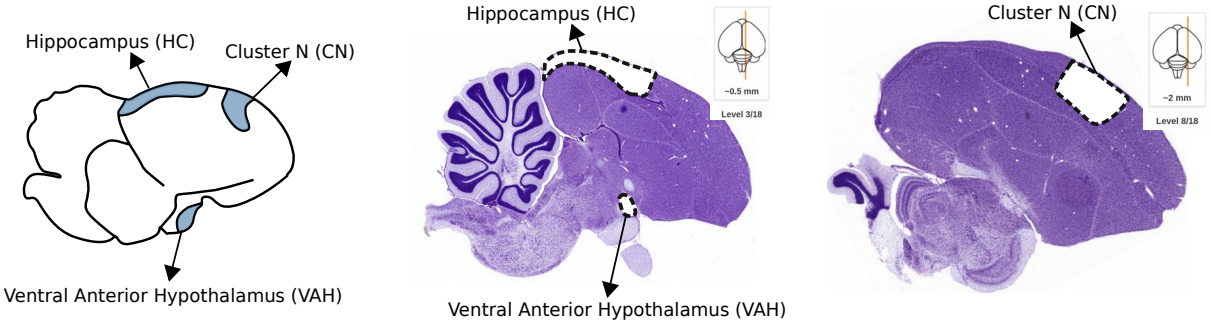
The anatomical structure of the avian brain is different than that of mammals. The avian brain organization maintains big structures homologous to the mammalian brain (i.e. telencephalon, cerebellum, thalamus, midbrain and hindbrain) but the organization inside such structures is different. The avian brain has well compartimentalized regions, called nuclei, characterized by different cell types and functional features (for a review see[26]). Spatial movement, circadian cycles, and sensorial input are key elements of migration related to structures in the brain.

The hippocampus acts as multisensorial integrator of olfaction and the visual cortex [27], additionally, all the functions related to spatial memory and cognition are mainly located in this region[28]. Besides the functional associations with migration, some evidence suggests a hippocampal volume increase during the migratory season [29] and the number of cells in the hippocampus differs between migratory and resident species .

Another essential factor for migration is timing. Observations and seasonality of migratory restlessness, suggest a circannual clock in birds likely linked to circadian cycles[15], [30]. The Suprachiasmatic Nucleus (SCN) is the core regulatory unit of circadian cycles in mammals. Likely, it has the same function in aves. Surgical ablation of the avian SCN disrupts the normal circadian cycles in birds.

The visual regions of birds are mainly located in the palial layer of the brain. When birds are stimulated with light during the day, several different regions of this part of the brain have early gene expression activity. One specific region shows active gene expression during night vision in migratory songbirds[26] . This region, located on forebrain is called Cluster N. No evidence of activation of this area was found during daytime, when nonmigratory birds were tested or the eyes of migratory birds were covered [23]. Chemical lesions in Cluster N disrupt magnetic compass orientation in migratory birds[31], but does not affect orientation capabilities using sunset cues or an artificial star compass.

Fig B1. Brain regions in a migratory bird brain. **Left:** Location of three relevant regions for migration: Hippocampus (HC), Cluster N (CN) and Ventral Anterior Hypothalamus (VAH). **Center** Sagital cut at approx 0.5 mm from the medial line. It indicates the location of HC and VAH. **Right** Saggital cut at 2mm from medial line. Depicts CN area



1.2 The search for genetic and molecular regulators of migration.

The set of physiological and behavioural adaptations like hyperphagia, circadian and circannual timing of key life-history events, and navigation skills are sometimes referred as the "migratory syndrome"[24] . If such "migratory syndrome" exists it might be controlled by major genetic components. Evidence from quantitative genetics of common garden experiments [25]and displacement experiments in the wild[17], suggest a clear heritable component of several migratory traits (See box 2). Despite several approaches to find potential genes underlying migration behaviour, their identity remains elusive.

1.2.1 Candidate Gene Approaches

The objective of a candidate gene is to find associations between genetic markers and a specific phenotype. This approach relies on the orthology and conservation of genes to infer functional homology across species with similar traits. The usual approach to select a candidate gene starts finding a gene of known function in model organisms with a polymorphic genetic marker in the target species (microsatellite lengths repeats or characteristic polymorphisms) that correlates with a certain trait of the phenotype. In migration, the correlations are usually with traits like timing, orientation, distance or migratory restlessness[32]. Some candidate gene approaches work well for simple traits. However in complex traits like many diseases, this approach receives criticisms about low replication, and lack of thoroughness and inclusivity [33].

As migration is a timing related behaviour, naturally the search for candidate gene(s) for migration started with the molecular machinery controlling circadian cycles. One of the first candidates was the *CLOCK* gene, a master regulator of the circadian cycle in mammals. The first associations found with *CLOCK* were the breeding latitude and the length of a poly Glutamine(Q, polyQ) repeats in the exon 12 of this gene in blue tits [34]. In several other species, polyQ length correlates with timing of seasonal traits[35]–[37], and migratory distance [38]. A second candidate gene, *ADCYAP1* showed a positive correlation between lengths polymorphism in the 3′ UTR and migratory restlessness (*zugunruhe*) in blackcaps [39]. However the results are inconclusive. On one hand, correlations of length

polymorphism with breeding latitude and timing were also found in other species [35], [38], [40], [41]. On the other hand, replications of the same approach in other species do not find a correlation between the length polymorphisms of any the previous gene candidates (*CLOCK* and *ADCYAP1*) and traits related to migration [40], [42]–[44].

One of the main drawbacks of using candidate gene approaches to study migration is the lack of genetic structure assessment on correlations with migratory traits that covary with geography[32]. Breeding latitude and migratory distance could have significant correlations with migration due to demography and not strictly with the migratory phenotype. There is still a need for gene candidates of migration. However, the search for new candidates has more benefits using genome-wide approaches.

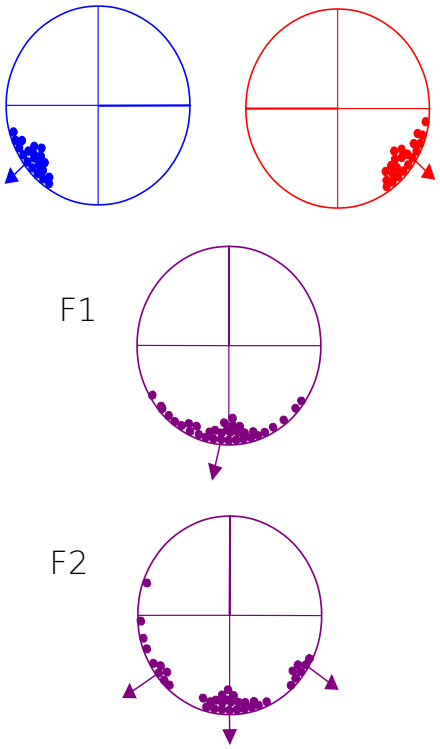
1.2.2. Population Genomics

Taking advantage of the access to new sequencing technologies in any species now allows us to expand the search for genetic factors of migration from candidate genes and marker-based approaches to whole genome examination.

Population genomics approaches on species with divergent patterns of migration have used summary statistics to identify broad genomic regions of divergence (*genomic islands of divergence*) potentially harboring genes related to migratory behaviour. One of the first examples came from the Swainson's thrush (*Catharus ustulatus*), a migratory species with populations showing different orientation patterns. Using a windowed FsT approach, the authors found genomic islands of differentiation between these populations. The genomic islands harbor genes implicated in circadian cycles and Heat shock proteins [45], [46]. Following a similar approach in willow warblers (*Phylloscopus trochilus)* the genomic regions differentiating between sympatric populations with opposite orientation patterns, have long stretches of elevated differentiation and clear boundaries which suggests structural variation located in chromosomal inversions [47]. The genes in such genomic regions are involved in fatty acid metabolic pathways, and transcription factors. Nonetheless, the regions found are not similar to those found in the Swainson's thrush, and do not include any of the early candidate genes (i.e. ADCYAP1 or CLOCK).

BOX 2. The genetics of migration.

To characterize the molecular machinery that controls migration, it is necessary to identify whether the phenotype is influenced by genetic and/or environmental factors. Using quantitative genetic analyses of cross breeding and selective breeding in a common garden environment, [1,2] demonstrated that migratory traits are heritable. In blackcaps, birds breeding on either side of the central European migratory divide show distinctly different migratory orientation directions: in autumn, birds breeding on the west of the divide, migrate South West, and populations breeding east of the divide head to a South East direction(see section 1.4). Orientation preference of selectively bred offspring resembles the same direction as the parents. When crossing individuals from either side of the divide, crossbred offspring follows an intermediate orientation pattern. The orientation in the F2 shows increased variance, but also recovers the pattern of the parental orientation preference, suggesting the genetics of orientation is based on only few genomic elements with big size effects.



The figure depicts the cross breeding experiment done by Helbig et al 1991 [1]. Funnel experiment results are represented as circles locating the cardinal locations. Each point is the average orientation of an individual. Blue and red are pure populations (southwest and southeast, respectively). The F1 result of crossbreeding blue and red individuals shows an intermediate orientation. Furthermore an F2 obtained from inbreeding of F1 individuals, shows that the intermediate and parental phenotypes are recovered.

The study conducted in Chapter 2, confims this classic experiment revealing previously inaccessible insight into the huge variability in orientation direction across a migratory divide. In the contact zone, several individuals will follow an intermediate orientation, confirming what was found with earlier classic experiments.

As an addition to the broad genomic population differentiation patterns found throug population genomics genome-wide-association studies (GWAS) are beginning to narrow down the potential genes involved in migration. To characterize the phenotype, the miniaturization of tracking devices now allows a

more complete picture of the variability in migratory phenotypes of small songbird species like the Blackcap. With such tracking devices, now it is possible to make associations between the routes taken by individual birds and their genome. The relationship between tracks and single-nucleotide polymorphisms (SNPs) in Swainson's thrushes, revealed a genomic region located in chromosome 4 associated with differences of route in migration. In this region there are several genes including CLOCK[48], endopeptidases of the nervous system and cell signaling genes. With a similar approach, a study associating individual migratory routes and genomes in blue/golden winged warblers (*Vermivora spp*) found a small region differentiated between these two subspecies in the chromosome Z [49]. The only gene found in that region is VPS13A a gene associated with movement disorders. However, unlike other species like blackcaps, the Swainson's thrushes and the blue/golden winged warblers populations not only differ in migration. These subspecies also have differences in plumage colouring that could confound the results obtained population genomic approaches[49], [50].

The disagreements between the results with the methodologies used suggest that the mechanisms employed by different species might not be the same. Alternatively, it could also suggest that other signatures not hard coded in the genomes, like DNA methylation or histone modifications might play a role in the regulation of bird migration.

1.2.3. Regulatory and functional genomics (transcriptomics and epigenomics)

The underlying difficulty of approaches using population genomics to study bird migration is that, in many bird species, demography and other evolutionary processes influence the divergence between populations. This divergence does not necessarily translate directly into genetic elements related with migration.

Recent transcriptomic studies have started to unravel the complexity of the migratory phenotype, analyzing differentially expressed genes (hereafter, DEG) of several tissues (blood, muscle, heart, liver, brain and ventral hypothalamus) in different species [51]–[57]. Results from these studies are so far inconclusive about the general mechanisms or signaling pathways involved in migratory

behaviour. Depending on the species, experimental design and focal tissue, the number of DEG can range from just 4 ([51] – blackbirds, tracking devices, blood) to around 188 ([52] - Swainson's thrush, common garden, ventral hypothalamus). Moreover, the latter study found little overlap between the differentially expressed genes with those found in a sequence based GWAS in the same species [48]. So far only one study found DEG evidence for ADCYAP1 one of the traditional candidate genes for migration [57]. However, the little overlap could be due to the heterogeneity of approaches and tissues included in the studies or the complexity of a trait like migration involving many genes in similar pathways.

Approaches related to gene regulation like DNA methylation are starting to become integrated to study migration in non-avian species. The comparison of differentially methylated regions across the whole genome in fins of F2 inbred migratory and non-migratory trouts, have identified regions close to genes of the circadian rhythm pathway and nervous system development [58]. Changes in methylation can affect gene expression [59], therefore the changes found in these fish could potentially affect genes involved in migration.

So far non-coding sequences have not been investigated in the context of migration. These approaches were previously mostly limited to model organisms and difficult to adapt to non-model species, like migratory birds. However, with recent advancements of genomic techniques and the possibility to study chromatin accessibility using approaches, such as ATAC-seq (see box 3) the study of non-coding regions can now be applied to basically any species of interest.

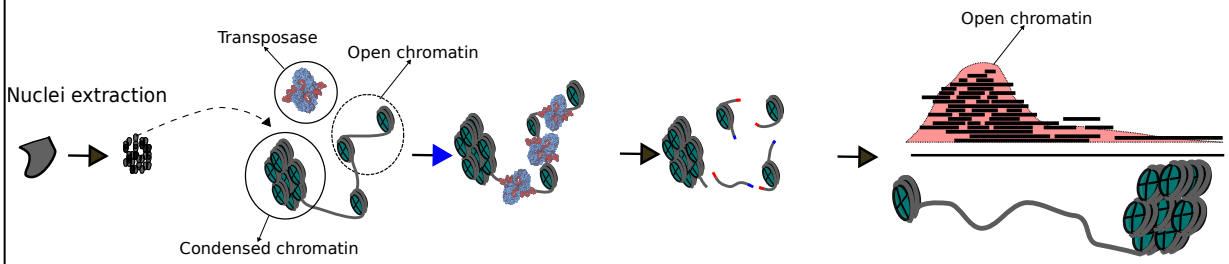
BOX 3. Techniques used to study chromatin accessibility.

In eukaryotes and archaea, the DNA wrapped around an octamer complex of histone proteins is called a nucleosome. Nucleosomes are arranged in a hierarchical organization to get what Is commonly known as chromatin. This hierarchical nature of chromatin allows different mechanisms of gene regulation. One of those mechanisms is chromatin accessibility, consisting on the distribution and occupancy of nucleosomes to enable physical access of the DNA sequence. This accessibility is the translation outcome of the information encoded in the histone and DNA modifications of nucleosomes. The most accessible DNA also called open chromatin regions (OCRs hereafter), is often bound by transcription factors or RNA polymerases. Different degrees of accessibility along the chromatin, create a landscape reflecting the regulatory snapshot of a given cell in a specific condition. Hence, changes in accessibility can relate to specific cis-regulatory sequences controlling the expression of neighbouring genes and long-range interactions.

Several techniques have been developed to analyse chromatin accessibility. FAIRE-seq uses a gradient of phenol-chloroform to separate open chromatin regions from DNA in the nucleosomes. DNAse-seq employs an enzyme that cuts on accessible DNA. A limitation of these techniques is the requirement of large amounts of cells/tissue. This makes it difficult to study small focal areas or parts of organs whenever large amounts of tissue are not available.

More recently, a technique called ATAC-seq has been shown to work reliably even in situations where only small amounts of samples are available. This technique is based on a modified tn5 transposase that targets only regions of the genome that are not bound to nucleosomes. This enzyme cuts open stretches of DNA and pastes adapters that can be used for sequencing with Illumina based technologies. After aligning the reads to a reference genome, the regions with higher frequency of mapped reads indicates where the chromatin is more accessible than the background.

Figure B3. ATAC-seq overview. After a nuclei extraction of any tissue, the transposase enzyme will target exclusively regions of open chromatin. It cuts the DNA and paste adapters. This fragments are sequenced and mapped back to a Genome of reference. Regions of the genome where there are high frequencies of reads are regions of Open chromatin.



1.3 How migration evolves, appears and disappears in birds.

1.3.1 Origins of migration

The origin of migration is not clear, in fact, its presence across the animal kingdom indicates a very ancient origin. In birds, however, it is difficult to pinpoint at what time and how migration arose. It is more accurate to establish when migration appears or disappears, and still, it is a very complex picture. In terms of species, migration can be present in two sister species, but not in their outgroup, or sister species may have one lineage with obligatory migration, and the other completely resident[60]. More strikingly, in the same species there could be a complete spectrum from completely resident passing by partial migrants to obligate migratory populations.

There are three main hypotheses for the origin of migration. The first one supports that migration reduces intra-specific competition during breeding season. Birds start to move from tropical to temperate regions because the latter offers more resources during breeding season, therefore, adaptation to a migratory life style might translate into increased fitness for migratory birds. However, harsh conditions in temperate regions during winter make migratory birds return to tropical areas and come back to their breeding grounds[61]. A second hypothesis holds the opposite view. Birds posses site fidelity to their breeding grounds and migration is one of the strategies to avoid harsh conditions in temperate regions. A phylogenetic approach in a large family of songbirds (*Emberizae)*, supports this idea. Several migratory clades of Emberizae are related to non-migratory lineages breeding in temperate regions [60]. The third hypothesis proposes a relevant role of the species historical contingency, with weather conditions as a switch to activate migration. Niche modelling supports the idea that reduction of potential refugia during glacial maxima forces birds to populate tropical areas, while keeping their breeding grounds [62]. Elaborating on this idea, *Zink et al* [63] proposes that migration appears depending on which strategy improves fitness at a given time. For instance, the conditions on the beginning and end of glacial maxima can act as a migratory switch to activate or disactivate migration. Despite these hypothesis are well supported, none of the three invalidate or hold a stronger support than the other hypotheses. The origins and establishment of migration are still in debate.

Partial migratory populations provide a good opportunity to study the patterns that enable or suppress migration and its adaptive process [5]. Fitness effects can favour migration or not, leading to frequency changes of strategies within one population. However, if both strategies have very small differences in fitness, that could indicate the emergence of a partial migratory population [64]. In some partial migratory populations the differences in fitness can be minimized when considering all year round [65]. Despite the potential dangers that might come with migration, birds facing this threat may have more chances to survive during the winter compared to all year residents. To balance fitness all year round residents must have at least ~60% higher breeding success than migrants, a requirement that these individual met having 2 or 3 broods per year. All these adaptations in residents can even the fitness of a migratory strategy.

Alternative hypotheses describe migratory behaviour as a threshold model, which implies the interaction of environmental and genetic elements to express one or other migratory phenotype. The model assumes that migration is a continuous trait with a normal distribution. The genetic and environmental conditions place an individual inside the distribution of the trait. If the individual is above a threshold, it will express the migratory phenotype; otherwise, it will not express the migratory phenotype[66]. This model confers characteristics of phenotypic plasticity to migration, which is a powerful mechanism for adaptation.

1.3.2. Evolutionary consequences of seasonal migration.

Behavioural isolation can be a channelling factor towards speciation favoring selection of sexual secondary traits like plumage and song [67], [68]. At the same time, behaviour isolation can create or enhance prezygotic reproductive barriers that evolve either in allopatry or sympatry. When populations meet in secondary contact, postzygotic reinforcement of the differences could happen via lower hybrid fitness. In several migratory species this process leads to *migratory divides*.

Several examples of migratory divides are known, for example willow warblers in Scandinavia, Blackcaps in central Europe and Swainson's thrushes in North America. Migratory divides provide a useful case to study how migratory orientation evolves. Changes in the behaviour of individuals in a population can cause divergence and differentiate at the genomic level showing patterns of disruptive selection[68]. Different patterns of migratory divides, can reflect differences in timing and preference of wintering grounds. Secondarily, if migratory divides indirectly start to affect differences in morphology or sexual traits like song or plumage, this can catalyze the effect on speciation[67]. Although most of those mechanisms are happening through prezygotic isolation, postzygotic isolation can occur in the case of the lower fitness of hybrids on the migratory divide. However, the importance of migration in selection against hybrids has not been fully addressed.

1.4 The Eurasian Blackcap *(Sylvia atricapilla)* the ideal model for the genetics of seasonal migration.

The Eurasian blackcap (*S. atricapilla*) is a bird of the genus *Sylvia* (*Passeriformes*), a group of small songbirds similar to warblers. The distinctive black colored feathers on the upper part of his head gives it the common name blackcap. Blackcaps are common breeders across Eurasia with a wide breeding distribution throughout Europe including areas of Norway and Russia. The largest number of blackcaps overwinter in the western and central areas of the Mediterranean Sea and north of the African continent. Some blackcaps migrate across the Sahara Desert to wintering grounds in countries like Senegal and Sudan.

Blackcaps, like most songbirds, are nocturnal and solitary migrants. Fledglings are under parental care, but once they have moulted and accumulated enough fat, they are ready to perform their first migratory journey on their own (REF). This characteristic of individual nocturnal migration suggests the existence of innate mechanisms that equip the bird with information about when to start and where/which direction to migrate. This discards any learning process for

migration, and highlights the predominance of genetic factors to exhibit the migratory phenotype.

Blackcaps have the full spectrum of migratory distances: there are populations with long, medium, and short distances, in addition to year-round residents. Orientation patterns of migration in blackcaps vary from south western to south eastern orientation. Some closely neighbouring populations migrate in distinct different orientations. Specifically, populations breeding east of the migratory divide breeding east of migrate towards a southeast direction via Greece and Turkey towards sub-Saharan countries. Individuals from the west side travel to south Spain and north Africa[69]. In the middle of the migratory divide, there are hybrid individuals that theoretically, should travel across the Alps and the middle of the Mediterranean Sea. A new orientation pattern seems to have emerged recently: in the UK, an increasing population of blackcaps has been observed recently overwintering, suggesting a new north west orientation pattern[70].

It is uncommon to find species exhibiting a whole range of behavioural phenotypes. Many of the differences in behaviour are between species or subspecies. Having the complete range of migratory phenotypes in a single species, plus differences in the propensity, distance and orientation of migration, are ideal prerequisites that make the blackcaps an ideal model organism to study the genetics of this behaviour[71] .

REFERENCES

[1] S. Åkesson, M. Ilieva, J. Karagicheva, E. Rakhimberdiev, B. Tomotani, and B. Helm, "Timing avian long-distance migration: From internal clock mechanisms to global flights," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 372, no. 1734, 2017.

[2] E. A. McKinnon and O. P. Love, "Ten years tracking the migrations of small landbirds: Lessons learned in the golden age of bio-logging," *Auk*, vol. 135, no. 4, pp. 834–856, 2018.

[3] C. G. Guglielmo, "Obese super athletes: Fat-fueled migration in birds and bats," *J. Exp. Biol.*, vol. 121, 2018.

[4] M. Liedvogel, S. Åkesson, and S. Bensch, "The genetics of migration on the move," *Trends Ecol. Evol.*, vol. 26, no. 11, pp. 561–569, 2011.

[5] P. Berthold and A. Helbig, "The genetics of bird migration: stimulus, timing, and direction," *Ibis (Lond. 1859).*, vol. 134, pp. 35–40, 1992.

[6] P. Berthold, A. Helbig, G. Mohr, and U. Querner, "Rapid microevolution of migratory behaviour in a wild bird species," *Nature*, vol. 2, no. 3, pp. 173–179, 1992.

[7] B. Abrahms *et al.*, "Memory and resource tracking drive blue whale migrations," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 12, pp. 5582–5587, 2019.

[8] C. Merlin and M. Liedvogel, "The genetics and epigenetics of animal migration and orientation: Birds, butterflies and beyond," *J. Exp. Biol.*, vol. 222, pp. 1–12, 2019.

[9] C. M. Hewson, K. Thorup, J. W. Pearce-Higgins, and P. W. Atkinson, "Population decline is linked to migration route in the Common Cuckoo," *Nat. Commun.*, vol. 7, pp. 1–8, 2016.

[10] F. Sergio *et al.*, "Individual improvements and selective mortality shape lifelong migratory performance," *Nature*, vol. 515, no. 7527, pp. 410–413, 2014.

[11] M. G. DeSaix *et al.*, "Population assignment reveals low migratory connectivity in a weakly structured songbird," *Mol. Ecol.*, vol. 28, no. 9, pp. 2122–2135, 2019.

[12] D. W. Winkler *et al.*, "Long-Distance Range Expansion and Rapid Adjustment of Migration in a Newly Established Population of Barn Swallows Breeding in Argentina," *Curr. Biol.*, vol. 27, no. 7, pp. 1080–1084, 2017.

[13] K. P. Able and J. R. Belthoff, "Rapid 'evolution' of migratory behaviour in the introduced house finch of eastern North America," *Proc. R. Soc. B Biol. Sci.*, vol. 265, no. 1410, pp. 2063–2071, 1998.

[14] A. Hegemann, A. M. Fudickar, and J. Å. Nilsson, "A physiological perspective on the ecology and evolution of partial migration," *J. Ornithol.*, vol. 160, no. 3, pp. 893–905, 2019.

[15] L. Hansson and S. Åkesson, *Animal Movement Across Scales*. Oxford: Oxford University Press, 2014.

[16] B. M. V. Doren, M. Liedvogel, and B. Helm, "Programmed and flexible: long-term Zugunruhe data highlight the many axes of variation in avian migratory behaviour," *J. Avian Biol.*, vol. 48, no. 1, pp. 155–172, 2017.

[17] A. C. Perdeck, "An experiment of the orientation of juvenile Starlings during spring migration - an addendum," *Ardea*, vol. 71, no. 2, p. 255, 1983.

[18] G. Durieux and M. Liedvogel, "Orientation and navigation in the animal world.," in *Position Navigation & Timing technologies in the 21st Century.*, F. van D. J. Morton and B. Parkinson, Eds. Wiley-IEEE., 2020.

[19] S. Schwarze *et al.*, "Migratory blackcaps can use their magnetic compass at 5 degrees inclination, but are completely random at 0 degrees inclination," *Sci. Rep.*, vol. 6, no. August, pp. 1–10, 2016.

[20] R. Wiltschko and W. Wiltschko, "Magnetoreception in birds," *J. R. Soc. Interface*, vol. 16, no. 158, pp. 0–3, 2019.

[21] P. J. Hore and H. Mouritsen, "The Radical-Pair Mechanism of Magnetoreception," *Annu. Rev. Biophys.*, vol. 45, no. 1, pp. 299–344, 2016.

[22] M. Liedvogel and H. Mouritsen, "Cryptochromes - A potential magnetoreceptor: What do we know and what do we want to know?," *J. R. Soc. Interface*, vol. 7, no. SUPPL. 2, 2010.

[23] H. Mouritsen, G. Feenders, M. Liedvogel, K. Wada, and E. D. Jarvis, "Night-vision brain area in migratory songbirds," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 23, pp. 8339–8344, 2005.

[24] T. PIERSMA, J. PÉREZ-TRIS, H. MOURITSEN, U. BAUCHINGER, and F. BAIRLEIN, "Is There a 'Migratory Syndrome' Common to All Migrant Birds?," *Ann. N. Y. Acad. Sci.*, vol. 1046, no. 1, pp. 282–293, 2005.

[25] J. Helbig, "SE- and SW-migrating populations in Central in the contact zone," vol. 670, pp. 657–670, 1991.

[26] E. D. Jarvis *et al.*, "Global view of the functional molecular organization of the avian cerebrum: Mirror images and functional columns," *J. Comp. Neurol.*, vol. 521, no. 16, pp. 3614–3665, 2013.

[27] M. C. Kahn and V. P. Bingman, "Avian hippocampal role in space and content memory," *Eur. J. Neurosci.*, vol. 30, no. 10, pp. 1900–1908, 2009.

[28] Y. Atoji and J. M. Wild, "Anatomy of the avian hippocampal formation," *Rev. Neurosci.*, vol. 17, no. 1–2, pp. 3–15, 2006.

[29] D. F. Sherry and S. A. MacDougall-Shackleton, "Seasonal change in the avian hippocampus," *Front. Neuroendocrinol.*, vol. 37, pp. 158–167, 2015.

[30] V. M. Cassone, J. K. Paulose, C. E. Harpole, Y. Li, and M. Whitfield-Rucker, "Avian circadian organization," *Biol. Timekeep. Clocks, Rhythm. Behav.*, vol. 35, no. 1, pp. 241–256, 2017.

[31] M. Zapka *et al.*, "Visual but not trigeminal mediation of magnetic compass information in a migratory bird," *Nature*, vol. 461, no. 7268, pp. 1274–1277, 2009.

[32] C. Steinmeyer, J. C. Mueller, and B. Kempenaers, "Search for informative polymorphisms in candidate genes: Clock genes and circadian behaviour in blue tits," *Genetica*, vol. 136, no. 1, pp. 109–117, 2009.

[33] H. K. Tabor, N. J. Risch, and R. M. Myers, "Candidate-gene approaches for studying complex genetic traits: practical considerations.," *Nature reviews. Genetics*, vol. 3, no. 5. England, pp. 391–397, May-2002.

[34] A. Johnsen *et al.*, "Avian Clock gene polymorphism: Evidence for a latitudinal cline in allele frequencies," *Mol. Ecol.*, vol. 16, no. 22, pp. 4867–4880, 2007.

[35] G. Bazzi *et al.*, "Clock gene polymorphism and scheduling of migration: a geolocator study of the barn swallow Hirundo rustica," *Sci. Rep.*, vol. 5, no. January, p. 12443, 2015.

[36] A. Bourret and D. Garant, "Candidate gene-environment interactions and their relationships with timing of breeding in a wild bird population," *Ecol. Evol.*, vol. 5, no. 17, pp. 3628–3641, 2015.

[37] N. Saino *et al.*, "Polymorphism at the *Clock* gene predicts phenology of long-distance migration in birds," *Mol. Ecol.*, vol. 24, no. 8, pp. 1758–1773, 2015.

[38] M. P. Peterson, M. Abolins-Abols, J. W. Atwell, R. J. Rice, B. Milá, and E. D. Ketterson, "Variation in candidate genes CLOCK and ADCYAP1 does not consistently predict differences in migratory behavior in the songbird genus Junco.," *F1000Research*, vol. 2, no. 0, p. 115, 2013.

[39] J. C. Mueller, F. Pulido, and B. Kempenaers, "Identification of a gene associated with avian migratory behaviour," *Proc. R. Soc. B Biol. Sci.*, vol. 278, no. 1719, pp. 2848–2856, 2011.

[40] G. Bazzi *et al.*, "Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds.," *Mol. Ecol.*, vol. 25, no. 24, pp. 6077–6091, Dec. 2016.

[41] J. Ralston *et al.*, "Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (Setophaga striata)," *Ecol. Evol.*, vol. 9, no. 15, pp. 8840–8855, 2019.

[42] A. Contina, E. S. Bridge, J. D. Ross, J. R. Shipley, and J. F. Kelly, "Examination of clock and Adcyap1 gene variation in a neotropical migratory passerine," *PLoS One*, 2018.

[43] G. Bazzi *et al.*, "Candidate genes have sex-specific effects on timing of spring migration and moult speed in a long-distance migratory bird.," *Curr. Zool.*, vol. 63, no. 5, pp. 479–486, Oct. 2017.

[44] R. Dor *et al.*, "Population genetics and morphological comparisons of migratory European (Hirundo rustica rustica) and sedentary East-Mediterranean (Hirundo rustica transitiva) barn swallows.," *J. Hered.*, vol. 103, no. 1, pp. 55–63, 2012.

[45] K. E. Delmore *et al.*, "Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation," *Mol. Ecol.*, vol. 24, no. 8, pp. 1873–1888, 2015.

[46] K. Ruegg, E. C. Anderson, J. Boone, J. Pouls, and T. B. Smith, "A role for migration-linked genes and genomic islands in divergence of a songbird," *Mol. Ecol.*, pp. 4757–4769, 2014.

[47] M. Lundberg *et al.*, "Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks," *Evol. Lett.*, pp. 155–168, 2017.

[48] K. E. Delmore, D. P. L. Toews, R. R. Germain, G. L. Owens, and D. E. Irwin, "The Genetics of Seasonal Migration and Plumage Color," *Current Biology*. 2016.

[49] D. P. L. Toews, S. A. Taylor, H. M. Streby, G. R. Kramer, and I. J. Lovette, "Selection on VPS13A linked to migration in a songbird," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 37, pp. 18272–18274, 2019.

[50] K. E. Delmore, D. P. L. Toews, R. R. Germain, G. L. Owens, and D. E. Irwin, "The Genetics of Seasonal Migration and Plumage Color," *Curr. Biol.*, vol. 26, no. 16, pp. 2167–2173, 2016.

[51] P. Franchini *et al.*, "Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species," *Mol. Ecol.*, 2017.

[52] R. A. Johnston, K. L. Paxton, F. R. Moore, R. K. Wayne, and T. B. Smith, "Seasonal gene expression in a migratory songbird," *Mol. Ecol.*, vol. 25, no. 22, pp. 5680–5691, 2016.

[53] C. M. Bauer *et al.*, "Seasonally sympatric but allochronic: Differential expression of hypothalamic genes in a songbird during gonadal development," *Proc. R. Soc. B Biol. Sci.*, vol. 285, no. 1889, 2018.

[54] A. M. Fudickar, M. P. Peterson, T. J. Greives, J. W. Atwell, E. S. Bridge, and E. D. Ketterson, "Differential gene expression in seasonal sympatry: Mechanisms involved in diverging life histories," *Biol. Lett.*, 2016.

[55] J. Boss *et al.*, "Gene expression in the brain of a migratory songbird during breeding and migration," *Mov. Ecol.*, pp. 1–11, 2016.

[56] A. Sharma and V. Kumar, "Metabolic plasticity mediates differential responses to spring and autumn migrations: Evidence from gene expression patterns in migratory buntings.," *Exp. Physiol.*, vol. 104, no. 12, pp. 1841–1857, Dec. 2019.

[57] W. J. Horton, M. Jensen, A. Sebastian, C. A. Praul, I. Albert, and P. A. Bartell, "Transcriptome Analyses of Heart and Liver Reveal Novel Pathways for Regulating Songbird Migration.," *Sci. Rep.*, vol. 9, no. 1, p. 6058, Apr. 2019.

[58] M. R. Baerwald *et al.*, "Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout," *Mol. Ecol.*, vol. 25, no. 8, pp. 1785–1800, 2016.

[59] M. L. Sciences and F. Antequera, "Review Structure, function and evolution of CpG island promoters," vol. 60, pp. 1647–1658, 2003.

[60] B. M. Winger, F. K. Barker, and R. H. Ree, "Temperate origins of long-distance seasonal migration in New World songbirds," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 33, pp. 12115–12120, 2014.

[61] B. M. Winger, G. G. Auteri, T. M. Pegan, and B. C. Weeks, "A long winter for the Red Queen: rethinking the evolution of seasonal migration," *Biol. Rev.*, vol. 94, no. 3, pp. 737–752, 2019.

[62] A. Louchart, "Emergence of long distance bird migrations: a new model integrating global climate changes.," *Naturwissenschaften*, vol. 95, no. 12, pp. 1109–1119, Dec. 2008.

[63] R. M. Zink and A. S. Gardner, "Glaciation as a migratory switch," *Sci. Adv.*, vol. 3, no. 9, pp. 1–9, 2017.

[64] A. Hegemann, P. P. Marra, and B. I. Tieleman, "Causes and Consequences of Partial Migration in a Passerine Bird," vol. 186, no. 4, 2015.

[65] D. Zúñiga *et al.*, "Migration confers winter survival benefits in a partially migratory songbird," *Elife*, vol. 6, pp. 1–12, 2017.

[66] F. Pulido, "Evolutionary genetics of partial migration – the threshold model of migration revis(it)ed," no. September, pp. 1776–1783, 2011.

[67] J. A. C. Uy, D. E. Irwin, and M. S. Webster, "Behavioral Isolation and Incipient Speciation in Birds," *Annu. Rev. Ecol. Evol. Syst.*, vol. 49, no. 1, pp. 1–24, 2018.

[68] S. P. Turbek, E. S. C. Scordato, and R. J. Safran, "The Role of Seasonal Migration in Population Divergence and Reproductive Isolation," *Trends in Ecology and Evolution*, 2017.

[69] R. Aymí, G. G., and D. A. Christie, "No Title," in *Handbook of the Birds of the World Alive*, J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, Eds. Barcelona: Lynx Edicions, 2020.

[70] S. Bearhop *et al.*, "Evolution: Assortative mating as a mechanism for rapid evolution of a migratory divide," *Science (80-.).*, vol. 310, no. 5747, pp. 502–504, 2005.

[71] a Helbig, "Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palaearctic warblers (Aves: Sylviidae)," *J. Exp. Biol.*, vol. 199, no. Pt 1, pp. 49–55, 1996.

PART I

PREFACE

Currently, tracking data is the most accurate way to describe the migratory phenotype in the wild. For small songbirds, the current development of tracking devices does not allow real-time data acquisition. Light-level geolocators are the most up to date devices for tracking small songbirds. These are archival devices that store information of light intensity and length. This information allows for a rough positional estimation through all the year of a single individual.

Here we used light-level geolocators to track individuals of the Eurasian blackcaps (*S. atricapilla*) known to have a wide array of migratory orientations and distances. We were particularly interested in describing the migratory phenotype of two populations: 1) individuals from a migratory divide and 2) individuals wintering in the United Kingdom. The first individuals are of interest to describe how the migratory divides can act as hybrid zones, as secondary contact of populations with opposite migratory directions. The second individuals are a potential case of recent adaptation. Individuals migrating to the UK have been reported since the 1960s, and their population is increasing. To know from where these individuals are coming from, can give clues on how the migratory behaviours can be flexible to adapt in a world of constant change.

We retrieved 98 individuals from all over Europe to describe how a broad spectrum of orientation patterns is present in the migratory divide of Eurasian blackcaps and the unexpected wide distribution of breeding sites for the populations wintering in the UK.

CHAPTER 2

Versatile migratory strategies and evolutionary insights revealed by tracks of wild Eurasian blackcaps

Kira Delmore*,[1,2],✉, Benjamin M. Van Doren*,[3,1],✉, Greg J. Conway[4], Teja Curk[5,6], Tania Garrido-Garduño[1], Ryan R. Germain[7], Timo Hasselmann[1], Dieter Hiemer[5], Henk P. van der Jeugd[6], Hannah Justen[1,2], Juan Sebastian Lugo Ramos[1], Ivan Maggini[8], Britta S. Meyer[1], Robbie J. Phillips[9], Magdalena Remisiewicz[10], Graham C. M. Roberts[4], Ben C. Sheldon[3], Wolfgang Vogl[8], and Miriam Liedvogel[1],✉
* These authors contributed equally to this work.
1 Max Planck Institute for Evolutionary Biology, MPRG Behavioural Genomics, 24306, Plön, Germany
2 Texas A&M University, 3528 TAMU, College Station, TX, 77843, USA
3 Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK
4 British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU, UK
5 Max Planck Institute of Animal Behaviour, Am Obstberg 1, 78315 Radolfzell, Germany
6 Vogeltrekstation – Dutch Centre for Avian Migration and Demography, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6700 AB Wageningen, The Netherlands
7 Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
8 Konrad-Lorenz Institute of Ethology, University of Veterinary Medicine Vienna, Savoyenstraße 1a, 1160 Wien, Austria
9 University of Exeter, Penryn, Cornwall, TR10 9FE, UK
10 Bird Migration Research Station, Faculty of Biology, University of Gdańsk, Poland

Contributions: I was a core member of the field crew across the the migratory divide from beginning to end. i.e. four consecutive breeding seasons. I provided comments on interpretations of the results and comments on the manuscript draft.

Submitted manuscript: This manuscript was submitted to PloS Biology on March 6 2020.

Versatile migratory strategies and evolutionary insights revealed by tracks of wild Eurasian blackcaps

Kira Delmore[*,1,2,✉], Benjamin M. Van Doren[*,3,1,✉], Greg J. Conway[4], Teja Curk[5,6], Tania Garrido-Garduño[1], Ryan R. Germain[7], Timo Hasselmann[1], Dieter Hiemer[5], Henk P. van der Jeugd[6], Hannah Justen[1,2], Juan Sebastian Lugo Ramos[1], Ivan Maggini[8], Britta S. Meyer[1], Robbie J. Phillips[9], Magdalena Remisiewicz[10], Graham C. M. Roberts[4], Ben C. Sheldon[3], Wolfgang Vogl[8], and Miriam Liedvogel[1,✉]

[*] These authors contributed equally to this work.

[1] Max Planck Institute for Evolutionary Biology, MPRG Behavioural Genomics, 24306, Plön, Germany
[2] Texas A&M University, 3528 TAMU, College Station, TX, 77843, USA
[3] Edward Grey Institute, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK
[4] British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU, UK
[5] Max Planck Institute of Animal Behaviour, Am Obstberg 1, 78315 Radolfzell, Germany
[6] Vogeltrekstation – Dutch Centre for Avian Migration and Demography, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6700 AB Wageningen, The Netherlands
[7] Section for Ecology and Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, 2100 Copenhagen, Denmark
[8] Konrad-Lorenz Institute of Ethology, University of Veterinary Medicine Vienna, Savoyenstraße 1a, 1160 Wien, Austria
[9] University of Exeter, Penryn, Cornwall, TR10 9FE, UK
[10] Bird Migration Research Station, Faculty of Biology, University of Gdańsk, Poland

✉ Correspondence: Kira Delmore <kdelmore@bio.tamu.edu>, Benjamin M. Van Doren <bmvandoren@gmail.com>, Miriam Liedvogel <liedvogel@evolbio.mpg.de>

Revised 15 February 2020.

Summary

Migration is ubiquitous in the animal kingdom and may play a key role in promoting reproductive isolation [1–4] and underpinning responses to environmental change [5,6]. Migratory divides are contact zones between populations with different migratory phenotypes and ideal natural laboratories for studying the evolution of migration [7,8]. The Eurasian blackcap (*Sylvia atricapilla*) exhibits a migratory divide in Central Europe between populations that migrate southwest (SW) and southeast (SE) in autumn [3,9,10] and has recently established a wintering population in Britain [1,5,11,12]. We tracked 106 annual migrations of 98 blackcaps captured across their range to characterize both the migratory divide and novel wintering strategy. Blackcaps to the west and east of the divide used predominantly SW and SE directions, respectively, but close to the contact zone many individuals took intermediate (S) routes. At 14.0ºE, we documented a sharp transition (22 km) in migratory direction from SW to SE, implying a strong selection gradient across the divide. Blackcaps wintering in Britain took northwesterly migration routes from continental European breeding grounds. They originated from a surprisingly extensive area, spanning 2000 km of the breeding range. British winterers bred in sympatry with SW-bound migrants but arrived 10 days earlier on the breeding grounds, suggesting some potential for assortative mating by timing. Overall, our data reveal complex variation in songbird migration and suggest that selection can maintain variation in migration direction across short distances while enabling the spread of a novel strategy across a wide range.

Results and Discussion

Pioneering studies of blackcaps revealed that songbird migration has a genetic basis and can rapidly evolve, and these findings underlie much of our current understanding of bird migration [1,5,9,13–21]. Today, blackcaps may offer important insight into adaptation to environmental change, as recent population increases [22] and new routes [5] illustrate how this species has successfully kept pace with a changing world. A major limitation of past studies on blackcaps has been a reliance on indirect experiments in captivity (see [23,24]) and infrequent recaptures of ringed birds to infer phenotypes. We sought to bridge this gap by intensively tracking blackcaps in the wild across the species' range,

examining the processes shaping migratory divides and contemporary migratory change, and placing our results in an evolutionary context.

Tracking blackcaps across a migratory divide

Ringing and orientation studies suggest that a migratory divide exists in Central Europe between blackcaps that migrate SW and SE, running north-south at 14ºE [3,10]. We tracked 41 annual migrations of 36 adult male blackcaps from breeding territories across the divide in Austria. To contrast behavioral variation inside and outside the divide, we also tracked blackcaps (3 F, 39 M) from breeding sites in the Netherlands (N=21), west Austria (N=6), central Germany (N=4), northern Poland (N=8), and east Austria (N=3). We expected to find a mix of strategies in the divide versus pure SW and SE directions at sites west and east of the divide, respectively.

Our tracks from the divide area clearly demonstrate the existence of a migratory divide (Figures 1 and 2, Figure S5). We estimated each blackcap's autumn migration direction by drawing a rhumb line between breeding and wintering areas. Migration directions varied between 130 and 288º. Intermediate (S) routes were more common (53.7%) than SE (26.8%) and SW (17.1%) strategies (Figure 1A). One individual from within the divide migrated NW to winter in Britain. Multi-year tracks reveal highly repeatable routes (Figure S6). Among-individual variation in migratory direction was considerably greater in the divide (Figure 3), suggesting that the contact between migratory phenotypes gives rise to increased diversity of behaviours.

A cline analysis using migration directions suggests that strong selection is maintaining the divide. Specifically, we examined the change in directions from western Austria (entirely SW), through the divide to eastern Austria (largely SE) (Figure 2; see Methods). We fit a cline through these directions to characterize its center and width. Clines maintained by selection should be narrow relative to dispersal distance, with a rapid transition between phenotypes [25]. Our data showed this pattern: the center of the cline occurred at 14.0ºE [interval within two log-likelihood units: 13.8–14.2º] and its width was only 22 km [2LL: 14–93 km]. This transition from SW to SE directions is very narrow compared to average natal dispersal distance in blackcaps (41.2 km [26]).

Our data do not allow direct identification of the source of selection, but possible processes include prezygotic selection for assortative mating and postzygotic selection reducing the fitness of hybrids. We discuss the potential for assortative mating in the next section. Helbig [9] selectively mated SW and SE blackcaps in captivity and observed intermediate orientations in their offspring. He argued that these hybrids would experience lower fitness through reduced survival, as they would have to cross the Alps, Mediterranean Sea, and Sahara Desert. This is a widely held hypothesis today [4,9,27], but our data do not necessarily support it, as a considerable number of the birds we tracked successfully took intermediate routes, survived, and returned to be recaptured. Most of these birds encountered portions of the Alps, but many did not cross the Mediterranean, in which case they never encountered this barrier or the Sahara Desert. Many of the birds that wintered in Africa navigated around the Mediterranean, and others used Italy as a land bridge (Figure 1 and Figure S5).

There is one important caveat: to maximize recapture success, we exclusively tracked adult birds, which had already completed at least one migration. It is possible that some blackcaps attempt to migrate over the Mediterranean and Sahara but do not survive to adulthood. Indeed, there is a striking deficit of birds wintering in Africa around 5ºE and 15ºE (Figure 1 and Figure S5; note birds from Dutch and Polish populations did winter in these areas). This deficit would not have been present in Helbig's work because he was not tracking free-flying birds. Alvarado et al. [28] argued similarly after failing to recover hybrids in a divide between hermit thrushes (*Catharus guttatus*). At present, tracking of small songbirds is limited to archival tags not capable of transmitted daily location estimates, so we cannot address this idea further.

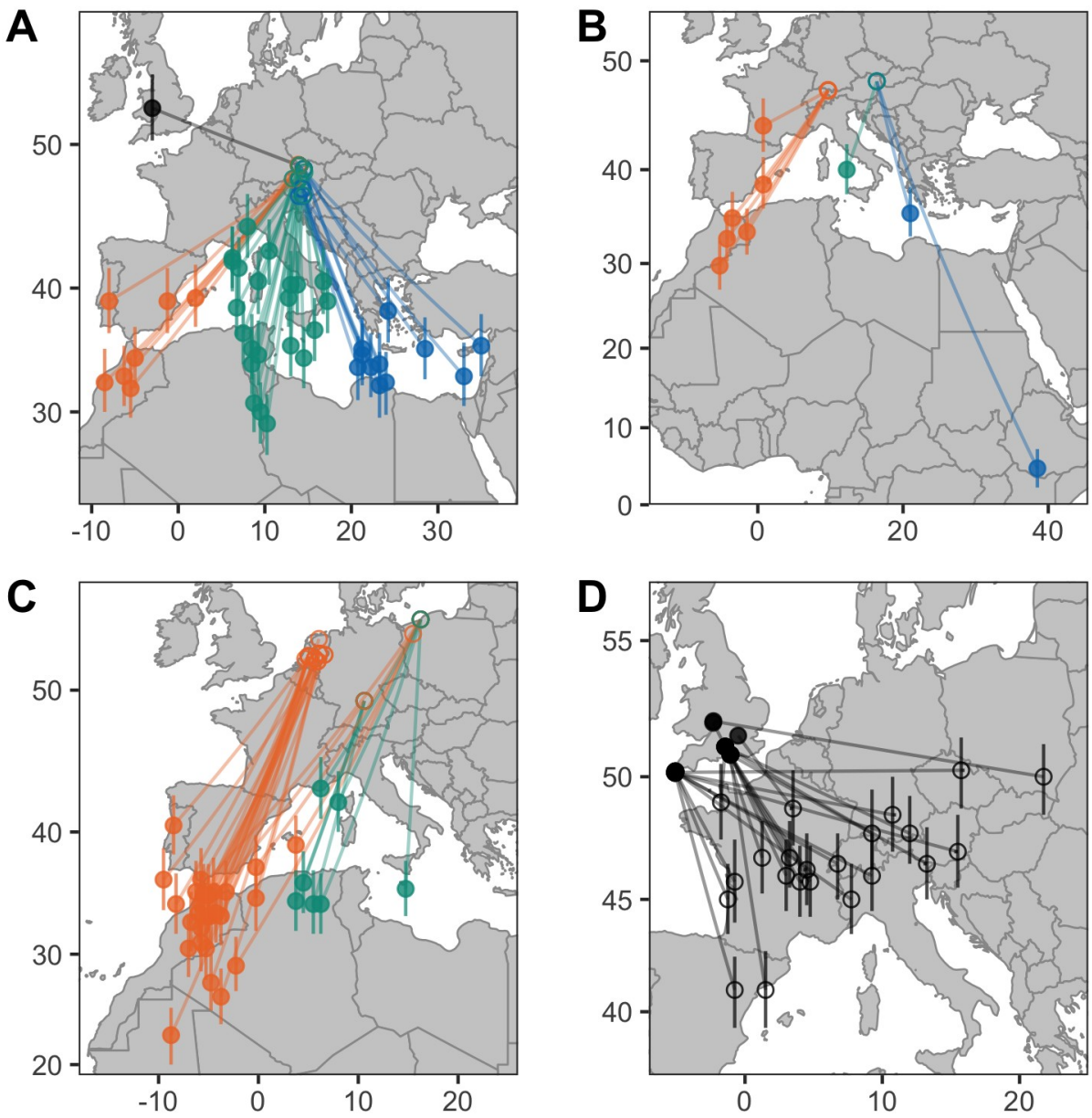


Figure 1: ***Wintering (i.e. non-breeding) and breeding locations of migratory blackcaps.*** *Wintering and breeding location estimates made with GeoLight shown with closed and open circles, respectively. Uncertainty in latitude estimation is indicated with vertical bars, which show estimates for sun angles higher and lower than the calibrated sun angle by 1º (following [29]). Colors indicate SW (orange)/intermediate (green)/SE (blue)/Britain (black) phenotypes, categorized by wintering location.* ***(A)*** *Winter sites of blackcaps breeding within the central European migratory divide transect in Austria.* ***(B)*** *Winter sites of blackcaps breeding in Austria east or west of the migratory divide.* ***(C)*** *Winter sites of*

blackcaps breeding in the Netherlands, southern Germany, and northern Poland. ***(D)*** *Breeding sites of blackcaps wintering in Britain.*

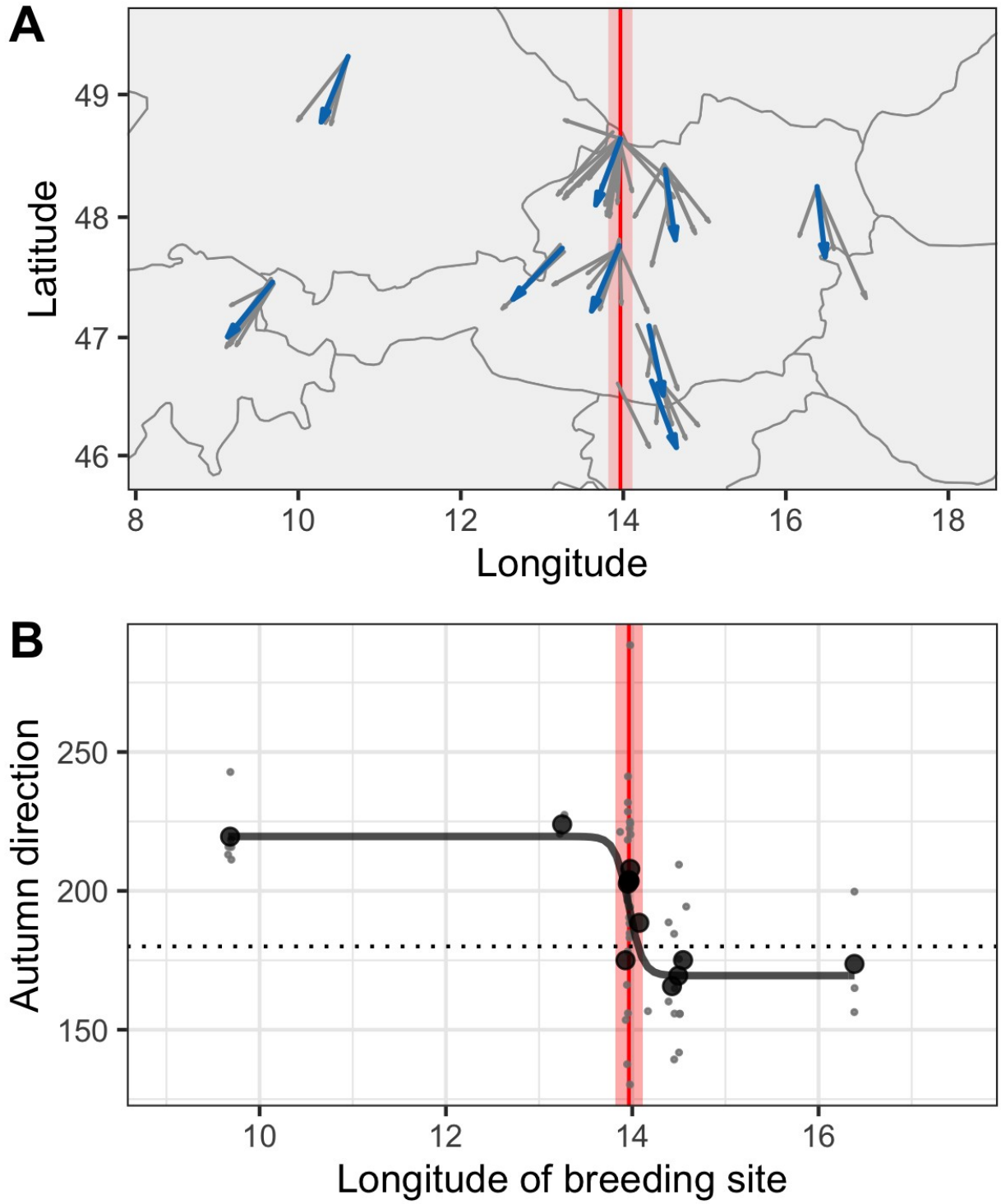


Figure 2: ***Autumn migration directions of blackcaps in Central Europe. (A)*** *Gray lines indicate migration directions of individual blackcaps, and blue lines indicate the mean direction at each capture site. In both panels, the solid vertical red line*

indicates the estimated cline center, and the red shading shows estimated cline width. ***(B)*** *Autumn migration direction by breeding longitude for Austrian blackcaps, with the maximum likelihood cline plotted. Small gray dots show the directions of individual blackcaps, and large black dots represent groupings of birds treated as sites for the analysis with hzar, which requires site-based input data. The dotted horizontal line is 180º (due south).*

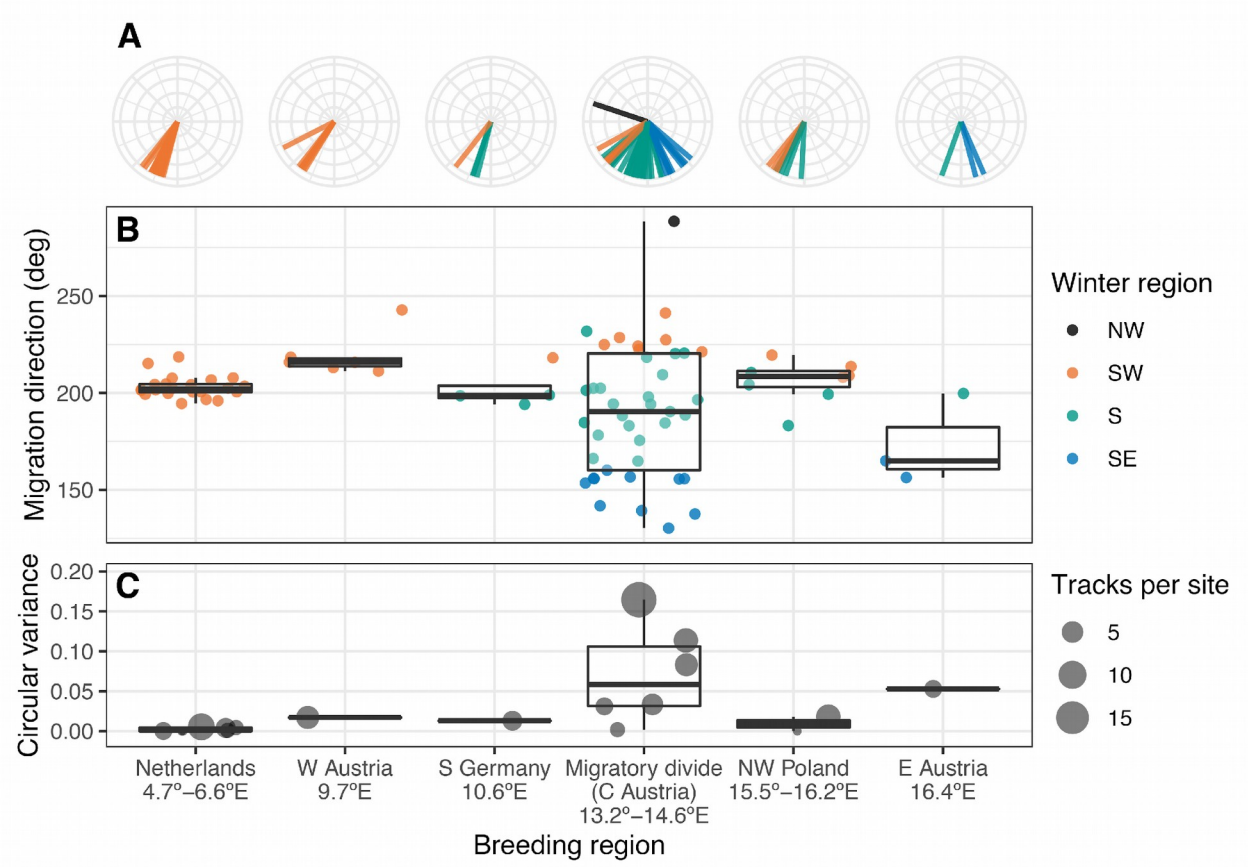


Figure 3: ***Variation in autumn migration direction by breeding area.*** ***(A)*** *Migration direction of tracked blackcaps caught at breeding sites across continental Europe. Each line points in the direction of autumn migration and is colored by winter region (SW=orange, intermediate=green, SE=blue, and NW (Britain)=black). Levene's test among sites with 5 or more tracked birds showed significantly higher variation in the area of the migratory divide: divide vs. Netherlands* $F_{1,61}=29.3$, $P<0.0001$; *divide vs. west Austria* $F_{1,45}=6.36$, $P=0.015$; *divide vs. Poland* $F_{1,47}=7.68$, $P=0.008$ *(excluding the NW migrant does not appreciably change this result),* ***(B)*** *Each dot shows the migration direction of one tracked blackcap (colored as in A).* ***(C)*** *Circular variance of autumn migration directions at each capture site, categorized by breeding region. Dot size shows the sample size at each site.*

Migration timing in the divide

Migration timing is an important component of the annual cycle that affects reproductive success [30,31] and mate selection [1]. Assortative mating based on migration phenotype might occur if migration timing and breeding differ consistently among phenotypes [1]. This could result in divergence between populations with different strategies and explain the rapid transition from SW to SE phenotypes [4]. However, we found no differences in spring arrival timing between birds using SW and SE autumn strategies (effect = -0.3 days, t_{23}=-0.069, P=0.95), nor in any other migration timing trait (Figure 4, Table S1). Data from eight individual blackcaps tracked over two years suggests repeatability in timing was higher on spring migration (spring migration start: R [95% CI]=0.86 [0.56,0.99], end: R [95% CI]=0.77 [0.24,0.96]; autumn migration start: R [95% CI]=0 [0,0.78], end: R [95% CI]=0 [0,0.73]), albeit with considerable uncertainty in all estimates. We therefore find no evidence that the migratory divide is maintained by temporal premating isolation. Variation across the divide in other traits, including body size (approximated by tarsus length or wing length) is also absent from our dataset.

So what is maintaining this migratory divide? One intriguing possibility is revealed by an analysis of timing that includes intermediate (S) migratory strategies. These blackcaps began spring migration on average 15 days earlier than SE and SW migrants (effect = -14.6 days, t_{23}=-2.7, P=0.014) and arrived 9 days earlier on the breeding grounds (effect = -9.4 days, t_{23}=-2.6, P=0.015) (Figure 4A, Table S1). This pattern is apparent even if we do not categorize individuals into discrete groups (Figure 4B). Early spring arrival may relate to the fact that blackcaps following intermediate strategies have the shortest distances to migrate (Figure S7D), so cues on the wintering site may predict conditions on the breeding grounds [32,33]. Importantly, early arrival may lead to assortative mating among intermediates, allowing them to exist relatively independently of pure SW and SE migrating populations within the 22 km cline. Selection against birds deviating from an immediately intermediate route (discussed previously) could limit the area where intermediates are favored to the observed cline width.

We used simulations to test if our measured distribution of arrival times would generate assortative mating among intermediate birds, comparing simulations where mate choice is dependent or independent of arrival time. The proportion of matings between intermediates was substantial and increased when we added mate selection based on timing (from 28% with no timing to 41% with timing), suggesting early arrival on the breeding grounds may facilitate assortative mating among intermediates, especially given their high relative abundance. Hybrid zones maintained by increased hybrid fitness are referred to as zones of bounded superiority[34]. Additional work is needed to support this idea, including direct observations of mated pairs and their offspring in the divide. We also note that genetic differentiation across this divide is low [35]. However, all of the genetic work on this system has focused on allopatric populations distant from the divide [21,36–38].

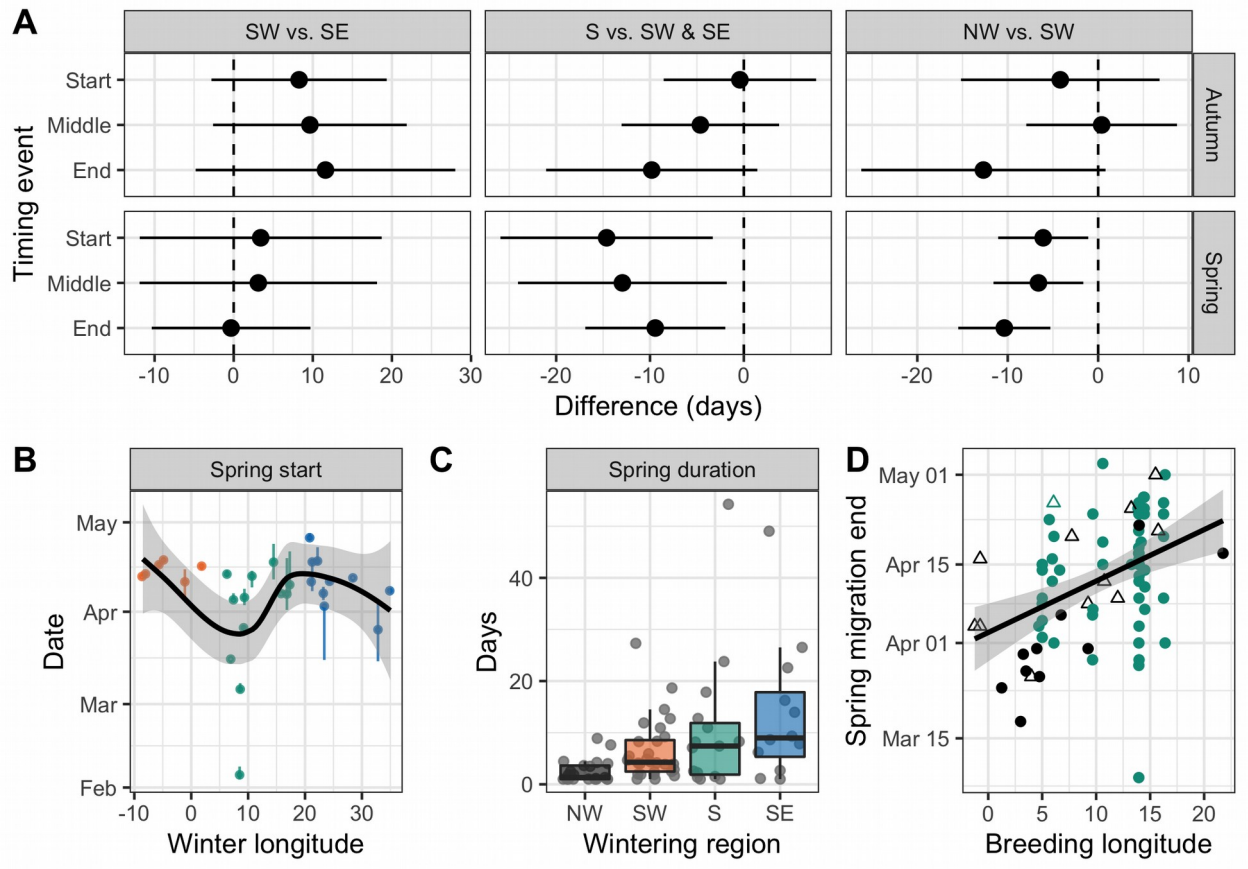


*Figure 4: **Blackcap migration timing**. **(A)** Timing within the migratory divide, showing model results for two timing comparisons: SW vs. SE (left) and intermediate (S) vs. SW/SE (right). Dots give model estimate and bars 95% confidence interval. Negative values indicate that SW or intermediate (S) groups,*

respectively, had earlier timing or shorter migrations. (B) Timing of the start of spring migration for birds tracked within the migratory divide. Points colored by wintering area, and vertical lines indicate the interquartile range of timing estimates made with FLightR. Curve is a loess smooth. (C) Boxplots showing spring migration duration by wintering area. Gray points correspond to individual tracks. (D) Breeding longitude vs. spring migration timing, with NW migrants in black and other birds in green. Triangles show females and circles show males.

Origins of blackcaps wintering in Britain

Blackcaps wintering in the UK in increasing numbers represent a rapid and recent change in migratory behavior, illustrating the speed at which movement strategies can evolve [11,12]. Early experiments supported a genetic basis for this migratory phenotype [5], but its nature is still poorly understood. Foremost is a lack of knowledge of the breeding grounds of birds wintering in Britain. No studies have tracked the direct migrations of free-living blackcaps to understand how many adopt this novel phenotype and determine whether those breeding in Britain are also changing their behavior by adopting residency. We fitted geolocators to blackcaps wintering in the UK and obtained 22 tracks from 20 blackcaps (11 F, 9 M), in addition to the one NW migrant tracked from our central Austrian cohort.

Blackcaps wintering in Britain originated from breeding areas in an unexpectedly broad expanse covering much of western and central Europe, remarkably extending south to latitudes occupied by the species in winter (Figure 1D). Their autumn migrations ranged from northerly (e.g. from Spain) to westerly (e.g. from Poland). This strategy enabled them to use short migration routes, on average 939±374 km; in contrast, birds tracked from central Europe flew on average 1865±717 km when they chose a southerly direction (Figure S7A). Although British winterers had the shortest routes in our sample, most also bred relatively close to suitable southerly wintering areas. To determine how far a blackcap would need to fly if it selected an alternative southerly migration route instead of a northerly route to the UK, we calculated the distance from the breeding site of each British winterer to the 10 closest wintering locations of tracked continental breeders. In 17 out of 23 cases (including two repeat tracks), the tracked route to the UK was longer than the average of the 10 possible southerly routes, often by

400-600 km (Figure S7C). This suggests that migration distance is of limited importance in explaining the British overwintering strategy. The availability of reliable supplemental food in British gardens may be a key driver [6] by positively influencing body condition and survival.

Only one of 41 individuals tracked from within the central European divide spent the winter in Britain (2.4%, 95% CI [0.13, 14]), and neither did any of the remaining 43 individuals tracked elsewhere in continental Europe. Previous studies estimated that northwest migrants comprise 6.8–25% of individuals breeding in Central Europe, based on ringing data, cage experiments, and stable isotopes [3,10,39]. One cage-orientation study suggested that as many as 50% of birds breeding in the vicinity of Linz, Austria migrate northwest [3]. Our results from free flying birds suggest these may be overestimates. Blackcaps wintering in Britain appear to breed across most of Europe at low densities, instead of occurring locally at higher densities.

Timing of northwest migrants

We tested for timing differences between NW migrants (British winterers) and SW migrants that might lead to reproductive isolation. Such timing differences have long been anticipated: Terrill and Berthold [40] predicted that differences in photoperiod should lead British winterers to depart and arrive c. 5 and 16 days earlier, respectively, and Bearhop et al. [1] reported evidence of assortative mating by wintering latitude based on stable isotopes from claw samples. Given that the NW phenotype appears to occur at low densities across Europe, assortative mating could be key to explaining how it is maintained in the population.

Other important factors may influence migration timing in blackcaps. For example, protandry is common among migratory songbirds and documented in blackcaps [41]. In our study, females were primarily sampled from among blackcaps wintering in Britain, where females showed later spring timing than their male counterparts (Table S2). In addition, different parts of continental Europe experience different spring phenology. In our dataset, blackcaps breeding further west in Europe underwent earlier spring migrations (Table S2, Figure 4D).

After including breeding latitude, longitude, sex, and year as predictors to account for their effects on timing, we found that NW migrants spending the winter in Britain reached their breeding grounds earlier than SW migrants that wintered in Iberia and northwest Africa (effect = -10.4 days, t_{44}=-4.1, P=0.00017; Table S2, Figure 4). They accomplished this by leaving the wintering grounds earlier (effect = -6.1 days, t_{43}=-2.5, P=0.018; compare [40]) and having shorter migration durations (ratio = 0.4x, t_{44}=-3.3, P=0.0019). In autumn, there were no timing differences between NW and SW migrants (Figure 4, Table S2).

Our data support the hypothesis that differences in arrival timing may contribute to reproductive isolation among blackcaps wintering in Britain, likely due to a combination of differing photoperiodic cues and shorter migrations [40]. Early-arriving individuals from Britain may experience fewer hazards during faster journeys, they may be in better condition due to supplemental food in British gardens [1,6], and they may be able to use local weather cues to judge the suitability of their continental breeding areas. In turn, these individuals may be able to secure higher quality territories. However, it is unclear whether the magnitude of the timing difference (10 days) could result in effective reproductive isolation. Rolshausen et al. [39] modeled assortative mating based on a timing difference of 10 days and a relative abundance of NW migrants of 1 out of 13 breeding individuals, concluding that NW migrants had a 28% chance of mating assortatively. Although we only tracked one NW migrant from within the migratory divide and therefore cannot capture the distribution of arrival dates in this particular breeding population, our similar average timing difference and lower relative abundance of NW migrants corroborate their conclusion of weak evidence for effective isolation solely based on timing. However, differences in microhabitat selection by migration phenotype [39] or body condition could still contribute to reproductive isolation.

Conclusion

We find considerable variation in blackcap migratory behavior across the central European migratory divide and diverse breeding origins for blackcaps exhibiting the novel British overwintering strategy. A narrow cline in migration direction across the divide suggests that selection on migratory strategy is strong. Assortative mating among birds orienting immediately south and selection

against those deviating from this direction may help maintain this narrow cline (but see [42]). British winterers arrived on continental breeding grounds earlier than migrants from Mediterranean wintering areas, but the difference in timing may be insufficient to drive assortative mating. Accurately characterizing the migrations of individual blackcaps reveals fascinating variability in the migratory behavior of this species, paving the way for targeted studies of the genetic basis of migration and adaptation to global change.

Methods

Geolocator application and retrieval

From 2016-2019, we deployed 806 archival light-level geolocators on breeding blackcaps in Austria (N=376, May–June), Germany (N=57, *MONTHS?*), the Netherlands (N=189, *MONTHS?*), and Poland (N=53, April–May and August), and on wintering Blackcaps in the United Kingdom (N=131, January–March) (Table S3). In Austria, we focused our sampling on the anticipated location of the migratory divide, where blackcaps with eastern and western migratory routes meet, and including populations that prior studies suggested contained NW migrants [3,10].

Birds were captured using mist nets and tape luring with audio recordings of the male blackcap territorial song. In the UK, we captured birds attending feeding stations in suburban gardens from January to March with mist nets and potter traps. We used leg-loop harnesses [43] made from elastic, viton, or nylon to attach geolocators. Tags were various models manufactured by Migrate Technologies, Inc. (see Table S3). Overall, we retrieved 115 devices, of which 106 contained data from at least one complete migration. We concurrently marked control cohorts of in the United Kingdom and the Netherlands (see Table S3). Return rates did not significantly differ between control and tagged birds (Fisher's exact test, UK: P=0.28; Netherlands: P=1).

Analysis of light data

We first used the *preprocessLight* function in the TwGeos [44] R package to define twilight events. We used a light threshold of 1.5 lux because blackcaps often occupy darker understory and mid-story habitats [45]. To maximize

repeatability, we minimized manual processing. We manually removed only obviously erroneous twilights, focusing on calibration periods. After manual processing, we used the *twilightEdit* function in TwGeos to perform additional automated editing and deletion of erroneous twilights. We used the following settings in *twilightEdit*: window = 4, outlier.mins = 30, and stationary.mins = 15. In the case of zero device with substantial shading of the light sensor, *twilightEdit* removed too many twilights to use in downstream analysis; in this case, we used only manually processed twilight times.

We used FLightR [46,47] to determine migration timing. FLightR uses the slope of the light curve around twilight to estimate locations and is therefore sensitive to data quality. In our dataset, several devices experienced substantial shading due to mantle feathers covering the light sensor, especially after the summer molt of body feathers. Geolocators with shorter "light pipes" ("-7" models, see Table S3) or with the light sensor on the body of the device itself (deployed in Poland, see Table S3) were prone to this issue, whereas devices with a light sensor at the end of a 11-mm "light stalk" ("-11" models) never experienced shading. We therefore performed an automated step to remove highly shaded light curves. For each twilight event, we took the mean of all "log.light" values returned by FLightR and removed twilights with values less than 1. We removed no more than 10% of twilights with this method; if more than 10% of twilights were heavily shaded, we removed the worst 10%. This approach improved performance for most individuals with light to moderate shading of the light sensor, but we were unable to obtain FLightR tracks for 6 heavily shaded devices. These were excluded from the FLightR timing analysis.

To identify birds' migration destinations (i.e. breeding or wintering sites, depending on the season of deployment), we used the R package GeoLight [48]. GeoLight contains a function *siteEstimate* for estimating a bird's location during a given time period, specifically designed for blackcaps and other birds for which shading of the light sensor can be a problem [29]. We succeeded in using *siteEstimate* to obtain location estimates for all birds, including those for which FLightR had failed. For devices deployed in summer, we used twilights from 15 December to 15 January to estimate wintering locations. For devices deployed in winter, we used twilights from 1 June to 1 August to estimate summer breeding

locations. In both cases, we set these time periods in mid-winter and mid-summer, when they are least likely to overlap with spring and autumn movements. We used the same time window for all birds to obtain comparable locations across individuals.

Both GeoLight and FLightR require that users define calibration periods during which the bird was stationary in a known location. We set calibration periods by visually inspecting plots of the log of observed versus expected light slopes for the deployment site over time (*plot_slopes_by_location* function in FLightR). When a bird moves away from the deployment site, the observed and expected slopes visually diverge [49]. For some individuals, visual resighting data were available after deployment and before recapture to aid calibration. After running FLightR, we refined calibration periods if the analysis suggested that movement had occurred during calibration periods. Some devices had insufficient calibration periods, if, for example, the bird departed shortly after tagging and the device stopped recording before the return migration. In these cases, and cases where the resulting track showed clear signatures of poor calibration (e.g. latitudinal drift during stationary periods or widely varying location estimates), we used a global calibration made from the combined data of all devices. For this global calibration, we used a linear model to estimate the overall mean calibration slope, accounting for the magnitude of shading to the light sensor. We did not include devices that lacked light pipes or light stalks, which made the light data qualitatively different from those collected by the other devices.

In GeoLight, we used the same calibration periods as for FlightR, with one additional refining step: we used *siteEstimate* to estimate the location of deployment and compared the result to the actual deployment location; if a lower or higher sun angle (±0.5º increments) resulted in a more accurate estimate of the deployment site, we used the adjusted sun angle instead.

We defined the FLightR model search grid between 10ºS and 65ºN latitude and 20ºW and 52ºE longitude. We chose these settings after visually inspecting light data with the *thresholdPath* function in the R package SGAT [48,50] to confirm that no tracks were likely to occur outside of this area. FLightR contains a prior for the decision to move, which has a default of 0.05. We adjusted this setting

outside of the migration season (i.e. from Dec 15–Mar 1 and May 15–Aug 15) to a value of 0.001. For the final run of each individual, we ran the particle filter with the recommended 1 million particles.

Migratory phenotypes

For comparative analyses of migratory phenotypes, we used both (1) winter longitude and (2) autumn migration direction. We estimated the bird's direction on autumn migration as the rhumb line connecting breeding and wintering sites (*bearingRhumb* in R package geosphere, [51]). We used this simplified representation of the route for calculating migration direction because geolocator tracks over short distances are sensitive to bias caused by imperfect calibration, especially close to an equinox.

In geolocation analyses of bird migration, longitude can generally be estimated with greater precision than latitude [52–54]. Latitude estimates are derived from daylengths, which are affected by shading and unreliable around the spring and autumn equinoxes. We compared destination longitudes estimated with GeoLight (*siteEstimate*) to estimates derived from FLightR. The two methods were highly correlated (ρ=0.99), affirming destination longitude as a reliable measure of migratory phenotype. that is insensitive to the choice of analysis method. Destination latitude showed a lower correlation between the two methods (ρ=0.82).

On 8 occasions, we were able to track the same individual for two subsequent years (5 from the migratory divide, 1 from the Netherlands, and 2 from Britain). From these data, we estimated individual repeatability using R package rptR [55] as the proportion of total variation explained by bird identity, where the total includes both variation from bird identity and among-year variation among birds.

We assigned individuals to four categories based on wintering location. For birds wintering north of 37.5ºN, we considered those west of 5ºE to be southwest (SW) migrants, those east of 20ºE to be southeast (SE) migrants, and those between 5-20ºE to have intermediate southerly (S) routes. For birds wintering south of 37.5ºN, we used a cutoff of 0º to distinguish SW from S because these longer routes require less of a westerly component to reach the same longitude.

We used Levene's test to compare variances (*leveneTest* R function in the car package) to determine whether the distribution of autumn migration directions differed among breeding sites. We controlled for multiple testing by applying a false discovery rate correction using the *p.adjust* R function.

Timing

We calculated migration timing using the *find.times.distribution* function in FLightR. To use this function, the user defines a spatial area, and the function reports the time at which the bird was likely to have crossed into and out of that area. For each individual, we used the shortest-distance route (i.e. a great circle route) between summer and winter areas to aid in defining migration progress. Specifically, we calculated paths perpendicular to the shortest-distance route at 30%, 50%, and 70% of the way between summer and winter locations, and we used *find.times.distribution* to determine when on migration the bird crossed these thresholds. We chose values of 30 and 70% because we found using values closer to the endpoints of the journey (e.g. 15%/85%) caused a higher proportion of calculations to fail, which typically occurs when the bird does not transit cleanly across the threshold. Close to summer and winter sites, local movements and geolocation uncertainty over time may lead to the modeled bird's path approaching the threshold more than twice per year. We treated these thresholds (30%, 50%, 70%) as representing early, middle, and late stages of the migratory journey, and we considered a bird to have reached each point at the 0.50 quantile time returned by *find.times.distribution.* As a measure of migration duration, we found the number of days it took each bird to travel from early (30%) to late (70%) migration stages, setting the value to one if it was estimated as less than one day. We calculated the speed of migration by dividing migration distance by duration. Because timing estimates of north-south movements can be inaccurate near the equinox, we did not retain timing estimates of movements taking place within 7 days of an equinox along a route within 15º of due north or south.

We validated FLightR timing estimates using simple longitude coordinate output from GeoLight (*crds* function), which we used to derive alternative measures of migration timing across an east-west axis. With this method, we considered a bird to be halfway through its migration when its estimated longitude was closer

to the longitude of its destination than its origin. We defined the start of migration as the time when a bird crossed a threshold from its starting longitude and did not return. Our threshold was defined as 10% of the difference between origin longitude and destination longitude. We defined the end of migration as the point when a bird crossed to within 10% of its destination longitude. We expected migration timing estimated from longitude data to be most comparable to FLightR estimates for birds that primarily used east-west routes. For birds that primarily moved along a north-south axis, the component of movement across longitudes is small relative to the component across latitudes. Therefore, we excluded birds with strongly southerly migration directions (150º–210º) from this validation. The timing of spring migration was consistent across methods (all Spearman $\rho>0.77$). In autumn, ρ ranged from 0.60 to 0.77).

We constructed linear models to compare the timing of migration for three different comparisons. For individuals tracked within the Austrian migratory divide, we tested for differences (1) between SW and SE parental phenotypes, and (2) between intermediate (S) and parental (SW/SE) phenotypes. Finally, we (3) tested for differences between NW (i.e. UK) and SW phenotypes. In all cases, we tested fixed effects of wintering area (NW/SW/S/SE) and year. We attempted to fit a random effect of bird identity, but our sample size of repeat tracks (N=8) was insufficient to estimate a variance component of bird identity, resulting in singular fits. Therefore, for birds with repeat tracks we randomly chose one track to include in the timing analysis, so that only one data point per individual was included for each timing measure. For comparison 3 (NW vs. SW), we also included effects of sex and breeding latitude and longitude. These effects were not relevant for comparisons 1 and 2 because all birds were tracked from a single breeding area (the contact zone), and all tracked birds were males. We used the R package emmeans [56] to construct the proper contrasts for comparisons 1 and 2. To maximize the precision of our estimates given a limited sample size, we removed terms with P-values greater than 0.10. For migration speed and duration, which had right-skewed distributions, we log-transformed the response variable before fitting the model.

We used simulations to test if our measured arrival timing differences in the migratory divide among SW, SE, and S (intermediate) could lead to substantial

assortative mating. In each simulation, we used the observed relative abundances of S, SW, and S phenotypes in the divide to draw a random sample of birds of equal number, following a multinomial distribution. Then, we used density curves fit to the original data to draw a sample of arrival dates for each phenotype group. Finally, for each individual, we selected a random mate based on the proportions of individuals present five days after its simulated arrival date. We used this delay because pair formation occurs within days after arrival [57] and females tend to arrive later than males. We repeated this simulation 1000 times and extracted the proportion of pairings that occurred between two intermediate individuals.

Routes

We used route output from FLightR. For tags that stopped in late winter or close to the spring equinox, track estimates could be unreliable. In these cases (n=16), we ignored location estimates for dates after 1 January if the tag stopped operation within three weeks of the spring equinox.

Cline analysis

We used the R package hzar [58] to estimate the location and width of the cline marking the transition from westerly to easterly migratory directions in the migratory divide. We used code from the supplementary materials of [58] as the basis for the analysis. Because hzar assumes that data come from a one-dimensional transects (in our case, an east-west transect), we limited the sites we included to the narrow range of latitudes within Austria. The analysis requires input data in the form of sites (not individuals), so we grouped individuals in the following way: we treated individuals as belonging to the same site group if their breeding territories were within 0.2 degrees of longitude, setting a maximum group size of 5 unless doing so would create an individual without a group. In this way, we assigned individuals to similarly-sized groups based on the longitude of their breeding site in Austria.

Author contributions

Conceptualization: KD, BMVD, BCS, ML; Methodology: KD, BMVD; Formal Analysis: KD, BMVD; Fieldwork: KD, BMVD, TC, TGG, RRG, TH, DH, HJ, JSLR, IM, BSM, RJP,

MR, GCMR, HPJ, WV, ML; Writing –Original Draft: BMVD with input from KD and ML; Writing –Review & Editing: KD, BMVD, GJC, TC, TGG, RRG, JSLR, IM, BSM, MR, BCS, HPJ, WV, ML; Visualization: BMVD; Supervision: GJC, MR, HPJ, BCS, ML; Project Administration: WV, IM, HPJ, GJC, MR, ML; Funding Acquisition: BMVD, MR, HPJ, ML.

Acknowledgements

For fieldwork assistance and logistical support, we thank Mayra Zamora, Gillian Durieux, Karen Bascon Cardozo, Andrea Bours, Shraddha Lall, Lisa Kettemer, Vasiliki Tsapalou, Josef Hemetsberger, Hans Winkler, Hemma Gressel, Alwin Schönenberger, Gerd Spreitzer, OSR Dir. Reinhold Petz, Mikkel Willemoes, Anne Hloch, Clara Leutgeb, Lisa Rosenich, Marius Adrion, Simon Kofler, Wolfgang Fiedler, Sally Amos, Jon Avon, Jake Bailey, Penny Barret, Stuart Bearhop, Rob and Liz Boon, Stuart Brown, Malcolm Burgess, Emily Cuff, Kate Dalziel, Ian Duncan, Rachel Durham, Phil Evans, Sheila Evans, Kate Fox, Roger Francis, Lyn Gammage, Gill Garrett, Sheila Gowers and Paul Ensom, Mark Grantham, Jodie Mae Henderson, John and Jane Holmes, Emma Inzani, Brian Isles, Michael and Helen Johnson, Ben Porter, Mel Mason, Irene McGregor, Keith McMahon, Nicole Milligan, Dee and Jonnie Reeves, Fiona Roberts, Dr ET Roberts, Gary Samways, Ash Sendell-Price, Ana Shapiro, Anna Smith, Dave Stoddard, Esmé Tackley, John Webber, Kester Wilson, Penny Witcombe, and other contributors, ringers, and homeowners. Special thanks to Glynne Evans for generous support and guidance. We thank Krzysztof Stępniewski, Katarzyna Stępniewska, Michał Redlisiak, and the Operation Baltic team of citizen scientists at Bukowo, Poland; Tijs van den Berg, Henri Bouwmeester, Ruud Foppen, Arend Timmerman, Morrison Pot and Hans Vlottes; James Fox and Migrate Technology Ltd for reliable geolocators and excellent technical support; and Eldar Rakhimberdiev and Simeon Lisovski for invaluable technical insights.

Funding:

This work was supported through funding from the Max Planck Society (MPRG grant to ML), the Marshall Aid Commemoration Commission (to BMVD), the American Ornithological Society (Mewaldt-King Research Award, to BMVD), the

Society for the Study of Evolution (Rosemary Grant Award, to BMVD), the Frank M. Chapman Memorial Fund (to BMVD), the British Trust for Ornithology (to BMVD). The purchase of 266 geolocators for use in the Netherlands and Germany was supported by grants from the 3V-Fonds from the Royal Netherlands Academy of Sciences and from NIOO-KNAW [both to HJ], Collecting data in Poland was supported by Special Research Facility grants (SPUB) of the Polish Ministry of Science and Higher Education to the Bird Migration Research Station, University of Gdańsk.

Permits:

AUSTRIA

Fieldwork in Austria was approved by the institutional ethics and animal welfare committee and the national authority (GZ 68.205/0048-WF/V/3b/2016) according to §§ 26ff. of Animal Experiments Act, Tierversuchsgesetz 2012 – TVG 2012.

GZ BMWFW-68.205/0048-WF/V/3b/2016 and BMWFW-68.205/0139-WF/V/3b/2016 (AT), UID: ATU36801500, MA22-24411/2016 (wien): BHBR-I-7100.00-69/2016-13 (VA), ABT13-53V-10/1998-42 (steiermark), 205-05RI/549/58/7-2016 (salzburg), N-2016-197947/8-Pin (OÖ), VL3-NS-3068/2016 /005/2016) (KÄ Villach), SV19-ALL-938/2016 (004/2016) (KÄ St Veit), SP3-NS-2823/2016 (007/2016) (KÄ Spittal), HE3-NS-1280/2016 (005/2016) (KÄ, Hermagor), FE3-NS-2127/2016 (006/2016) (KÄ Feldkirchen), 5/N.AB-10120-8-2016, (Burgenland), RU5-BE-286/011-2016 (NÖ)

UK

In the UK, geolocator deployments were approved by the University of Oxford Animal Welfare Ethical Review Body. Work was conducted under licenses from the British Trust for Ornithology, approved by the Special Methods Technical Panel.

PL

Approved by the General Directorate for Environmental Protection within the permit to capture and ring wild birds (DZP-WG.6401.03.36.2015.km, DZP-

WG.6401.03.98.2016.km, DZP-WG.6401.03.97.2017.jro, DZP-WG.6401.03.2.2018.jro)

GER

Permit issued by the Regierung von Mittelfranken, Bavaria. Permit number: 54-2532.1-13/14

NL

permit issued by the Centrale Commissie Dierproeven. Permit number: AVD801002016519 valid 27-6-2016 through 31-5-2021

References

1. Bearhop, S., Fiedler, W., Furness, R.W., Votier, S.C., Waldron, S., Newton, J., Bowen, G.J., Berthold, P., and Farnsworth, K. (2005). Assortative Mating as a Mechanism for Rapid Evolution of a Migratory Divide. Science *310*, 502–504.

2. Bensch, S., Andersson, T., and S. (1999). Morphological and Molecular Variation Across a Migratory Divide in Willow Warblers, Phylloscopus Trochilus. Evolution *53*, 1925–1935.

3. Helbig, A.J. (1991). SE- and SW-migrating Blackcap (Sylvia atricapilla) populations in Central Europe: Orientation of birds in the contact zone. J. Evolution. Biol. *4*, 657–670.

4. Irwin, D.E., and Irwin, J.H. (2005). Siberian migratory divides: The role of seasonal migration in speciation. In Birds of Two Worlds: The Ecology and Evolution of Migration, R. Greenberg and P. P. Marra, eds. (Baltimore: Johns Hopkins University Press), pp. 27–40.

5. Berthold, P., Helbig, A.J., Mohr, G., and Querner, U. (1992). Rapid microevolution of migratory behaviour in a wild bird species. Cah. Rev. The. *360*, 668–670.

6. Plummer, K.E., Siriwardena, G.M., Conway, G.J., Risely, K., and Toms, M.P. (2015). Is supplementary feeding in gardens a driver of evolutionary change in a migratory bird species? Glob. Change Biol. *21*, 4353–4363.

7. Delmore, K.E., Toews, D.P.L., Germain, R.R., Owens, G.L., and Irwin, D.E. (2016). The genetics of seasonal migration and plumage color. Curr. Biol. *26*, 2167–2173.

8. Delmore, K.E., and Irwin, D.E. (2014). Hybrid songbirds employ intermediate routes in a migratory divide. Ecol. Lett. *17*, 1211–1218.

9. Helbig, A.J. (1991). Inheritance of migratory direction in a bird species: A cross-breeding experiment with SE- and SW-migrating blackcaps (Sylvia atricapilla). Behav. Ecol. Sociobiol. *28*, 9–12.

10. Helbig, A.J. (1992). Population differentiation of migratory directions in birds: Comparison between ringing results and orientation behaviour of hand-raised migrants. Oecologia *90*, 483–488.

11. Berthold, P., and Terrill, S.B. (1988). Migratory behaviour and population growth of Blackcaps wintering in Britain and Ireland: Some hypotheses. Ringing & Migration *9*, 153–159.

12. Leach, I.H. (1981). Wintering Blackcaps in Britain and Ireland. Bird Study *28*, 5–14.

13. Berthold, P. (1988). Evolutionary aspects of migratory behavior in European warblers. J. Evolution. Biol. *1*, 195–209.

14. Berthold, P., and Pulido, F. (1994). Heritability of migratory activity in a natural bird population. Proc. R. Soc. Lond. B *257*, 311–315.

15. Helbig, A. (1996). Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palaearctic warblers (Aves: Sylviidae). J. Exp. Biol. *199*, 49–55.

16. Mueller Jakob C., Pulido Francisco, and Kempenaers Bart (2011). Identification of a gene associated with avian migratory behaviour. Proceedings of the Royal Society B: Biological Sciences *278*, 2848–2856.

17. Pulido, F., and Berthold, P. (2010). Current selection for lower migratory activity will drive the evolution of residency in a migratory bird population. Proceedings of the National Academy of Sciences *107*, 7341–7346.

18. Pulido, F., Berthold, P., and van Noordwijk, A.J. (1996). Frequency of migrants and migratory activity are genetically correlated in a bird population: Evolutionary implications. Proceedings of the National Academy of Sciences *93*, 14642–14647.

19. Pulido, F. (2007). The genetics and evolution of avian migration. BioScience *57*, 165–174.

20. Pulido, F., Berthold, P., Mohr, G., and Querner, U. (2001). Heritability of the timing of autumn migration in a natural bird population. Proceedings of the Royal Society of London B: Biological Sciences *268*, 953–959.

21. Rolshausen, G., Segelbacher, G., Hobson, K.A., and Schaefer, H.M. (2009). Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. Curr. Biol. *19*, 2097–2101.

22. EBCC/BirdLife/RSPB/CSO (2018). Trends of common birds in Europe, 2018 update.

23. Van Doren, B.M., Liedvogel, M., and Helm, B. (2017). Programmed and flexible: Long-term Zugunruhe data highlight the many axes of variation in avian migratory behaviour. J. Avian Biol. *48*, 155–172.

24. Z'uñiga, D., Falconer, J., Fudickar, A.M., Jensen, W., Schmidt, A., Wikelski, M., and Partecke, J. (2016). Abrupt switch to migratory night flight in a wild migratory songbird. Scientific Reports *6*, srep34207.

25. Barton, N.H., and Gale, K.S. (1993). Genetic Analysis of Hybrid Zones. In Hybrid Zones and the Evolutionary Process (Oxford University Press), pp. 13–45.

26. Paradis, E., Baillie, S.R., Sutherland, W.J., and Gregory, R.D. (1998). Patterns of natal and breeding dispersal in birds. J. Anim. Ecol. *67*, 518–536.

27. Bensch, S., Grahn, M., Müller, N., Gay, L., and S. (2009). Genetic, morphological, and feather isotope variation of migratory willow warblers show gradual divergence in a ring. Mol. Ecol. *18*, 3087–3096.

28. Alvarado, A.H., Fuller, T.L., and Smith, T.B. (2014). Integrative tracking methods elucidate the evolutionary dynamics of a migratory divide. Ecology and Evolution *4*, 3456–3469.

29. Hiemer, D., Salewski, V., Fiedler, W., Hahn, S., and Lisovski, S. (2017). First tracks of individual Blackcaps suggest a complex migration pattern. J. Ornithol., 1–6.

30. Taylor, R.S., and Friesen, V.L. (2017). The role of allochrony in speciation. Mol. Ecol. *26*, 3330–3342.

31. Winker, K. (2010). On the Origin of Species Through Heteropatric Differentiation: A Review and a Model of Speciation in Migratory Animals - Sobre el Origen de las Especies Mediante Diferenciación Heteropátrica: Una Revisión y un Modelo de Especiación en Animales Migratorios. Ornithological Monographs *69*, 1–30.

32. Both, C., Turnhout, C.A.M.V., Bijlsma, R.G., Siepel, H., Strien, A.J.V., and Foppen, R.P.B. (2009). Avian population consequences of climate change are most severe for long-distance migrants in seasonal habitats. Proceedings of the Royal Society of London B: Biological Sciences, rspb20091525.

33. Butler, C.J. (2003). The disproportionate effect of global warming on the arrival dates of short-distance migratory birds in North America. Ibis *145*, 484–495.

34. Moore, W.S. (1977). An Evaluation of Narrow Hybrid Zones in Vertebrates. The Quarterly Review of Biology *52*, 263–277.

35. Delmore, K.E., Illera, J.C., P'erez-Tris, J., Segelbacher, G., Lugo Ramos, J.S., Durieux, G., Ishigohoka, J., and Liedvogel, M. The evolutionary history and genomics of European blackcap migration. eLife.

36. Mueller, J.C., Pulido, F., and Kempenaers, B. (2011). Identification of a gene associated with avian migratory behaviour. Proceedings of the Royal Society B: Biological Sciences *278*, 2848–2856.

37. P'erez-Tris, J., Bensch, S., Carbonell, R., Helbig, A., and Teller'ıa, J.L. (2004). Historical Diversification of Migration Patterns in a Passerine Bird. Evolution *58*, 1819–1832.

38. Rolshausen, G., Segelbacher, G., Hermes, C., Hobson, K.A., and Schaefer, H.M. (2013). Individual differences in migratory behavior shape population genetic structure and microhabitat choice in sympatric blackcaps (Sylvia atricapilla). Ecology and Evolution *3*, 4278–4289.

39. Rolshausen, G., Hobson, K.A., and Schaefer, H.M. (2009). Spring arrival along a migratory divide of sympatric blackcaps (Sylvia atricapilla). Oecologia *162*, 175.

40. Terrill, S.B., and Berthold, P. (1990). Ecophysiological aspects of rapid population growth in a novel migratory blackcap (Sylvia atricapilla) population: An experimental approach. Oecologia *85*, 266–270.

41. Rainio, K., Tøttrup, A.P., Lehikoinen, E., and Coppack, T. (2007). Effects of climate change on the degree of protandry in migratory songbirds. Clim. Res. *35*, 107–114.

42. Irwin, D. (2020). Assortative mating in hybrid zones is remarkably ineffective in promoting speciation. The American Naturalist.

43. Rappole, J.H., and Tipton, A.R. (1991). New harness design for attachment of radio transmitters to small passerines (nuevo diseño de arnés para atar transmisores a passeriformes pequeños). J. Field Ornithol. *62*, 335–337.

44. Wotherspoon, S., Sumner, M., and Lisovski, S. (2016). TwGeos: Basic data processing for light-level geolocation archival tags.

45. Rakhimberdiev, E., Senner, N.R., Verhoeven, M.A., Winkler, D.W., Bouten, W., and Piersma, T. (2016). Comparing inferences of solar geolocation data against high-precision GPS data: Annual movements of a double-tagged black-tailed godwit. J. Avian Biol. *47*, 589–596.

46. Rakhimberdiev, E., Saveliev, A., Piersma, T., and Karagicheva, J. (2017). FLightR: An R package for reconstructing animal paths from solar geolocation loggers. Methods Ecol. Evol. *NA*, NA–NA.

47. Rakhimberdiev, E., Winkler, D.W., Bridge, E., Seavy, N.E., Sheldon, D., Piersma, T., and Saveliev, A. (2015). A hidden Markov model for reconstructing animal paths from solar geolocation loggers using templates for light intensity. Movement Ecology *3*.

48. Lisovski, S., and Hahn, S. (2012). GeoLight - processing and analysing light-based geolocation in R. Methods Ecol. Evol.

49. Lisovski, S., Bauer, S., Briedis, M., Davidson, S.C., Dhanjal-Adams, K.L., Hallworth, M.T., Karagicheva, J., Meier, C.M., Merkel, B., and Ouwehand, J. *et al.* (2020). Light-level geolocator analyses: A user's guide. J. Anim. Ecol. *89*, 221–236.

50. Sumner, M.D., Wotherspoon, S.J., and Hindell, M.A. (2009). Bayesian Estimation of Animal Movement from Archival and Satellite Tags. PLOS ONE *4*, e7324.

51. Hijmans, R.J. (2017). Geosphere: Spherical trigonometry.

52. Lisovski, S., Hewson, C.M., Klaassen, R.H.G., Korner-Nievergelt, F., Kristensen, M.W., and Hahn, S. (2012). Geolocation by light: Accuracy and precision affected by environmental factors. Methods Ecol. Evol. *3*, 603–612.

53. Ekstrom, P.A. (2004). An advance in geolocation by light. *58*, 210–226.

54. Fudickar, A.M., Wikelski, M., and Partecke, J. (2012). Tracking migratory songbirds: Accuracy of light-level loggers (geolocators) in forest habitats. Methods Ecol. Evol. *3*, 47–52.

55. Stoffel, M.A., Nakagawa, S., and Schielzeth, H. (2017). rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol. Evol. *8*, 1639???1644.

56. Lenth, R. (2019). Emmeans: Estimated marginal means, aka least-squares means.

57. Bairlein, F. (1978). Über die Biologie einer südwestdeutschen Population der Mönchsgrasmücke (Sylvia atricapilla). Journal für Ornithologie *119*, 14–51.

58. Derryberry, E.P., Derryberry, G.E., Maley, J.M., and Brumfield, R.T. (2014). Hzar: Hybrid zone analysis using an R software package. Mol. Ecol. Resour. *14*, 652–663.

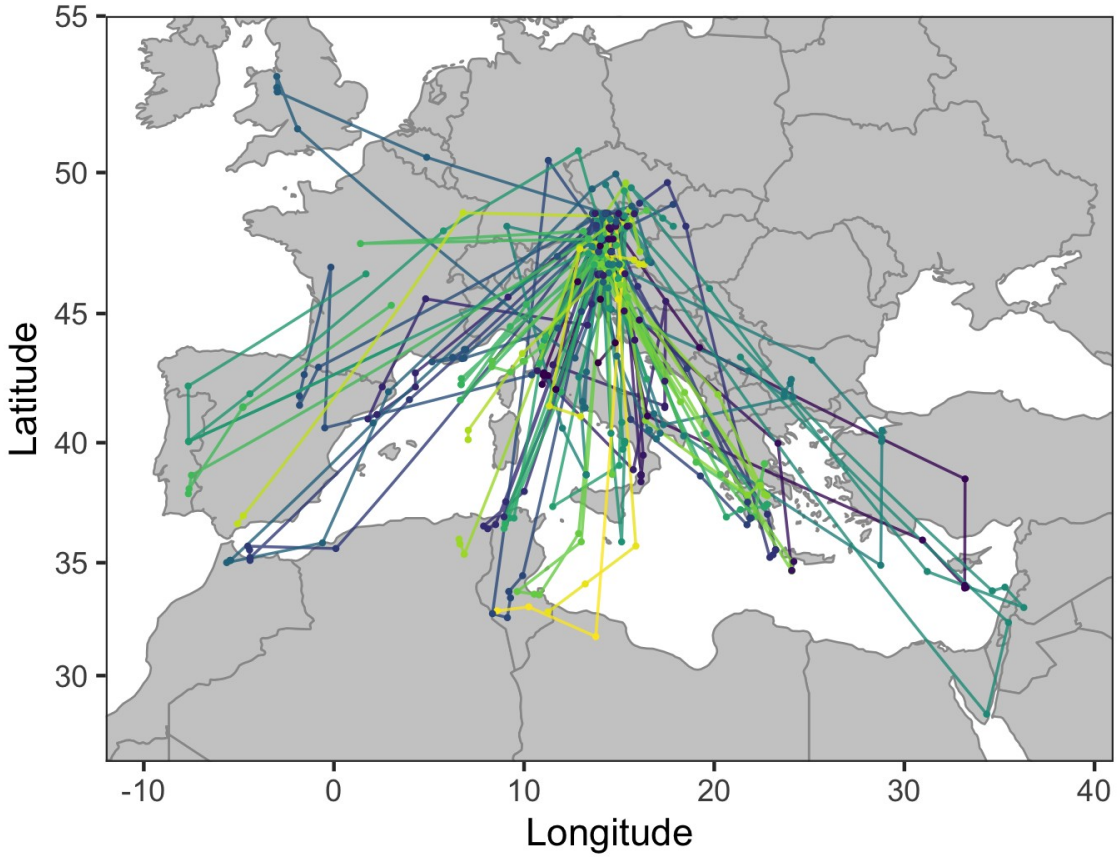


*Figure 5: **Full tracks of blackcaps from the migratory divide**. Tracks estimated with FLightR, with each track in a different color. To reduce clutter, one point is shown for each month and error bars are omitted. FLightR estimated some wintering locations at slightly higher latitudes than the siteEstimate function in GeoLight; for example, some FLightR tracks that end in the southern Balkan Peninsula have GeoLight estimates on the northeast coast of Libya (Figure 1A). Note that headings over short distances are sensitive to the calibration used and may not be fully trustworthy.*

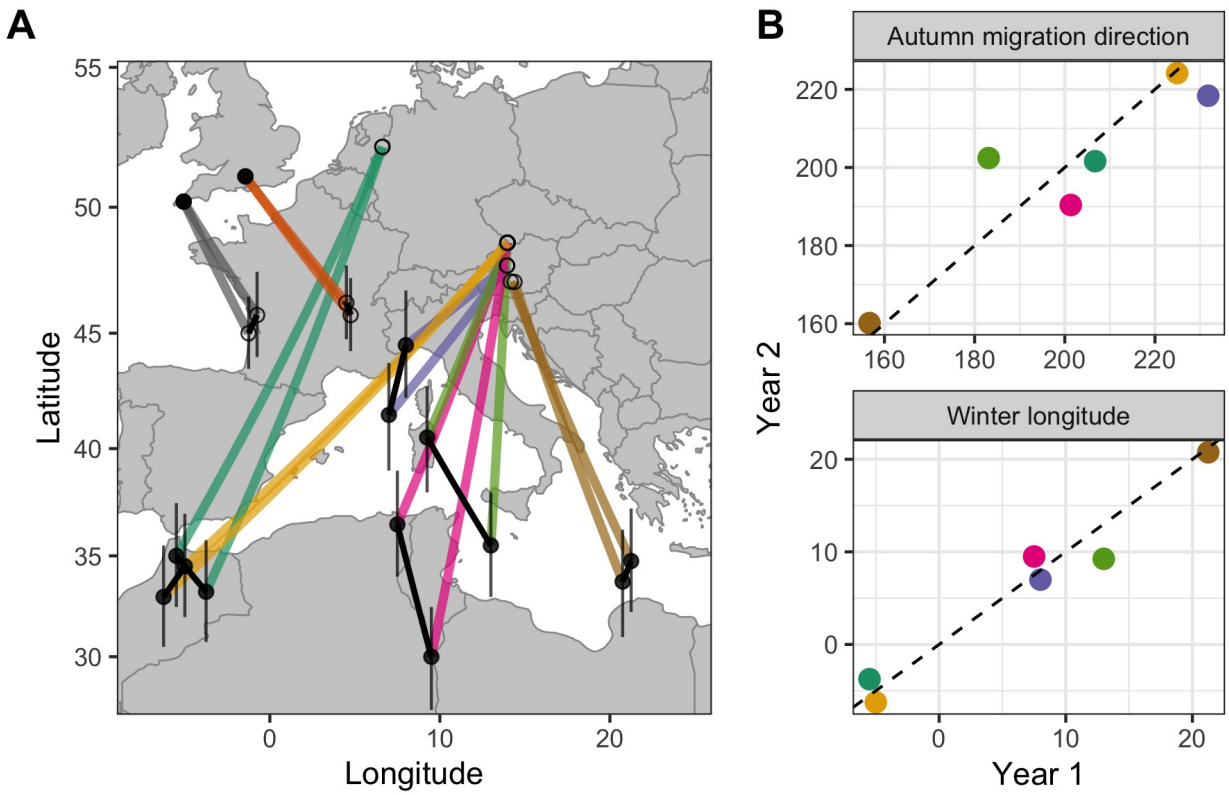


Figure 6: ***Repeatability of migratory phenotypes within individuals.*** ***(A)*** *Each color represents one individual tracked over two subsequent years, with solid black lines connecting location estimates for the same individual. Breeding and non-breeding sites and error bars as in Figure 1. For the two British winterers, our repeated location estimates were very similar (59 and 92 km apart, respectively), strongly suggesting that they bred in the same area.* ***(B)*** *Migratory phenotype estimates for individuals tracked from continental Europe for two years (excluding those tagged in Britain). The dashed line is the identity line. We estimated repeatability in winter longitude as R [95% CI]=0.99 [0.96,1] and repeatability in migration direction as R [95% CI]=0.91 [0.78,1]. The winter location estimates for these individuals averaged 385±253 km apart in consecutive winters.*

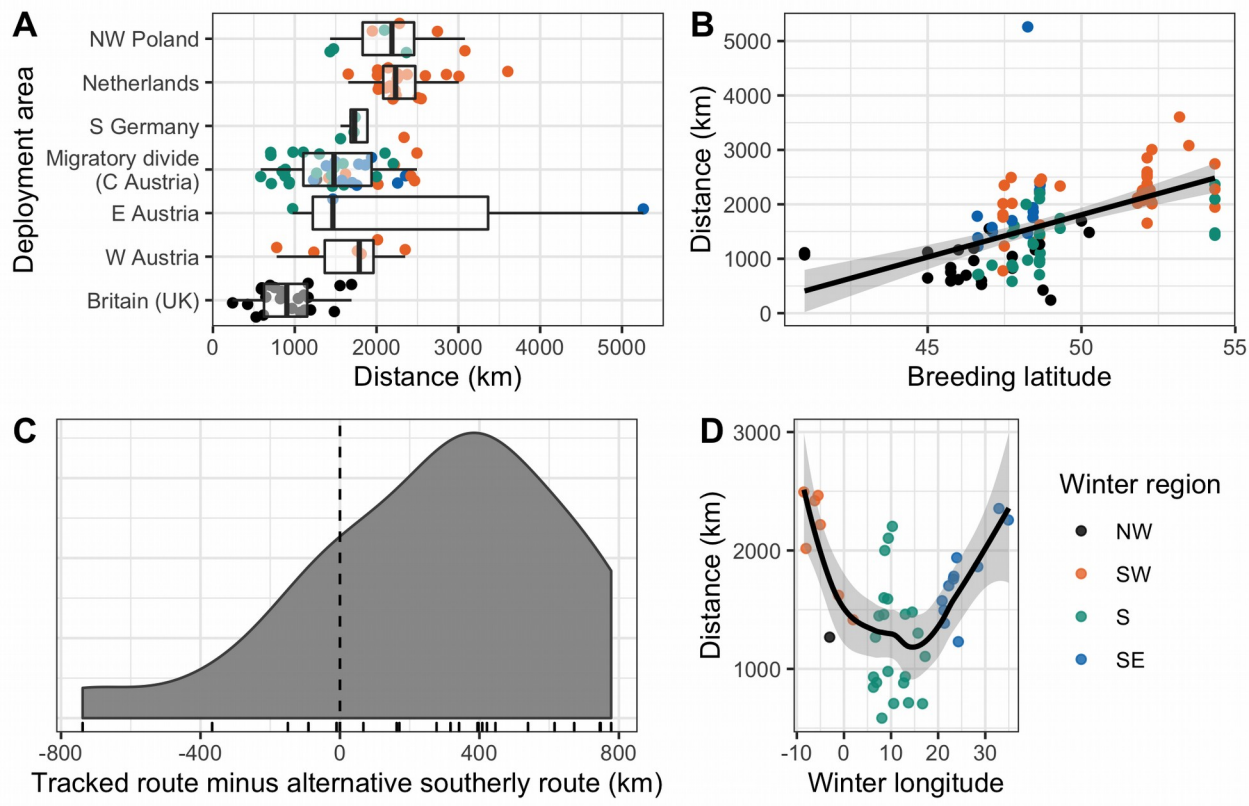


Figure 7: ***Migration distances.*** *Colors indicate SW (orange)/intermediate (green)/SE (blue)/Britain (black) phenotypes, categorized by wintering location.* ***(A)*** *Boxplots showing the distance between breeding and wintering sites for all blackcaps tracked, by deployment area.* ***(B)*** *Migration distance by breeding latitude, for all blackcaps tracked.* ***(C)*** *British winterers fly farther than necessary. Values shown are the difference between the observed migration distance, and the average of the distances to the 10 closest tracked individuals that wintered in traditional southerly areas, instead of in the UK.* ***(D)*** *Migration distance by wintering longitude for blackcaps tracked within the migratory divide only. Intermediate individuals had the shortest migration distances.*

Table 1: Model results comparing migration timing in the migratory divide between SW and SE phenotypes and between intermediate (S) and SW/SE phenotypes. Log-transformed variables indicated by "log" in parentheses.

Contrast	Season (response)	Estimate	SE	df	t-ratio	P-value
SW vs. SE	Spring start	3.42	7.39	23	0.46	0.648
SW vs. SE	Spring middle	3.11	7.25	23	0.43	0.672

SW vs. SE	Spring end	-0.33	4.85	23	-0.07	0.946
SW vs. SE	Autumn start	8.27	5.39	26	1.53	0.137
SW vs. SE	Autumn middle	9.63	5.99	30	1.61	0.118
SW vs. SE	Autumn end	11.60	8.04	30	1.44	0.159
SW vs. SE	Autumn duration (log)	0.19	0.70	25	0.27	0.792
SW vs. SE	Spring duration (log)	-0.80	0.55	23	-1.44	0.163
SW vs. SE	Autumn speed (log)	-0.05	0.68	25	-0.08	0.938
SW vs. SE	Spring speed (log)	0.94	0.55	23	1.72	0.099
S vs. SW & SE	Spring start	-14.62	5.47	23	-2.67	0.014
S vs. SW & SE	Spring middle	-12.94	5.38	23	-2.41	0.025
S vs. SW & SE	Spring end	-9.44	3.61	23	-2.62	0.015
S vs. SW & SE	Autumn start	-0.42	3.95	26	-0.11	0.917
S vs. SW & SE	Autumn middle	-4.63	4.11	30	-1.13	0.269
S vs. SW & SE	Autumn end	-9.82	5.51	30	-1.78	0.085
S vs. SW & SE	Autumn duration (log)	-0.99	0.53	25	-1.89	0.070
S vs. SW & SE	Spring duration (log)	0.17	0.42	23	0.41	0.686
S vs. SW & SE	Autumn speed (log)	0.39	0.51	25	0.76	0.454
S vs. SW & SE	Spring speed (log)	-0.65	0.42	23	-1.56	0.133

Table 2: Model results comparing migration timing of British winterers (NW migrants) to SW migrants. All models tested for timing differences between NW and SW phenotypes; other predictor variables were removed if P>0.1 and are therefore omitted from the table. Log-transformed variables indicated by "log" in parentheses. NW and SW phenotypes differed significantly in all spring timing measures and no autumn timing measures. Likewise, protandry was evident in all spring timing measures and none in autumn. Breeding longitude was most strongly associated with the timing of migration spring. Breeding latitude was not significantly associated with any timing trait. Year effects were evident only in autumn.

Predictor	Season (response)	Estimate	SE	df	t-ratio	F-value	P-value
NW vs. SW	Spring start	-6.08	2.47	43	-2.46	-	0.018
NW vs. SW	Spring middle	-6.61	2.46	44	-2.68	-	0.01
NW vs. SW	Spring end	-10.38	2.53	44	-4.10	-	<0.001
NW vs. SW	Autumn start	-4.19	5.47	50	-0.77	-	0.447
NW vs. SW	Autumn middle	0.39	4.15	51	0.09	-	0.926
NW vs. SW	Autumn end	-12.69	6.72	49	-1.89	-	0.065
NW vs. SW	Autumn duration (log)	-1.10	0.38	48	-2.89	-	0.006
NW vs. SW	Spring duration (log)	-0.81	0.24	44	-3.30	-	0.002
NW vs. SW	Autumn speed (log)	0.43	0.44	48	0.98	-	0.331
NW vs. SW	Spring speed (log)	-0.07	0.23	44	-0.33	-	0.745
Male vs. Female	Spring start	-9.34	2.86	43	-3.27	-	0.002
Male vs. Female	Spring middle	-9.10	2.86	44	-3.18	-	0.003
Male vs. Female	Spring end	-11.37	2.94	44	-3.87	-	<0.001
Male vs. Female	Autumn duration (log)	-0.99	0.47	48	-2.09	-	0.041
Breeding longitude	Spring start	1.21	0.20	43	6.01	-	<0.001

Breeding longitude	Spring middle	1.18	0.20	44	5.86	-	<0.001
Breeding longitude	Spring end	1.14	0.21	44	5.53	-	<0.001
Breeding longitude	Autumn end	0.87	0.40	49	2.21	-	0.032
Breeding longitude	Autumn duration (log)	0.10	0.03	48	3.33	-	0.002
Breeding longitude	Autumn speed (log)	-0.07	0.03	48	-2.80	-	0.007
Breeding latitude	Autumn end	-1.60	0.91	49	-1.76	-	0.084
Breeding latitude	Autumn speed (log)	0.13	0.07	48	1.93	-	0.059
Year (F-test)	Autumn start	-	-	-	-	6.44	0.001
Year (F-test)	Autumn middle	-	-	-	-	7.20	<0.001
Year (F-test)	Autumn end	-	-	-	-	2.23	0.097
Year (F-test)	Autumn duration (log)	-	-	-	-	2.85	0.047
Year (F-test)	Autumn speed (log)	-	-	-	-	3.13	0.034

Table 3: Geolocator deployment summary. All devices manufactured by Migrate Technology Ltd.

Region	Year	Devices deployed	Devices returned	Devices retrieved	Harness material	Device
Austria	2016	202	24 (19 nylon; 5 viton)	24	nylon braid 1 mm; viton cord 0.6 mm	P65Z1top2 end-11
Austria	2017	159	28	27	nylon braid 1 mm	P50Z11-11
Austria	2018	15	4	3	nylon braid 1 mm	P65Z1top1 -11
Netherlands	2016	61	5	5	nylon braid 1 mm	P50B1-11
Netherlands	2017	61	8	7	nylon braid 1 mm	P50B1-11

Netherlands	2018	67	14	13	stretch elastic 0.7 mm	P30Z11-7-DIP-NOT; P65B1-7-NOT
Poland	2015	12	1	1	nylon braid 1 mm	W30Z11-DIP-NOT ; W65B1-DIP NOT
Poland	2016	9	1	1	nylon braid 1 mm	W30Z11-DIP-NOT ; W65B1-DIP NOT
Poland	2017	12	4	4	nylon braid 1 mm	W65B1-DIP NOT ; W30Z11-DIP-NOT
Poland	2018	20	4	3	nylon braid 1 mm	W65B1-DIP NOT ; W30Z11-DIP-NOT
S Germany	2018	57	7	5	stretch elastic 0.7 mm	P30Z11-7-DIP-NOT; P65B1-7-NOT
UK	2016-17	36	8	6	stretch elastic 0.8 mm	P50Z11-11-NOT
UK	2017-18	48	10	7	stretch elastic 0.8 mm	P50Z11-7-DIP-NOT

PART II

PREFACE

It has been a long-standing question to know what are those molecular elements that make birds migrate, that give birds the adaptations for such an enduring journey in their life cycle. The first approaches were adopting candidate genes from phenotype that potentially are involved in migration, like elements of the circadian cycle in mammals. These approaches started the studies of population genetics to find correlations between molecular markers (e.g. microsatellites) with population characteristics related to the migratory behaviour, (i.e. migratory restlessness and geographic breeding location).

Nowadays, with the extensive use of next-generation sequencing the focus has changed to not look into a single marker in a single gene, but to examine population differences along the genome. This approach has been used in some migratory bird species with particular findings that point to different evolutionary trajectories creating the difference between migratory populations in each species. The different studies find evolutionary mechanisms like genomic islands of divergence and structural rearrangements as sources of genetic variability in different migratory species.

More recently, the studies of the molecular biology of migration have started to explore beyond the hard coded DNA sequence elements. Approaches to find differences in gene expression comparing different migratory states (e.g. during and outside the migratory season) has started to propose new candidates that may play a role in migration.

Here we first propose a molecular evolution approach to evaluate any potential candidate gene in a macroevolutionary framework, making use of publicly available genome sequences. We found that none of the candidate genes (by the time we conducted the analysis) can be reliably suggested as migration genes

along the avian clade. We argue for genome-wide approaches that can suggest more candidates to evaluate if all the migratory birds have a common gene or set of genes that make them able to migrate.

Second, we used a genome-wide approach to study the evolution and genomic associations with the migration of the Eurasian blackcap. We found that despite having a wide array of variability in migratory traits, the population structure in this species is low. Most of the differences are with year-round resident populations in the southern part of the European continent and islands of the Atlantic Sea. We find as well that the recently established UK wintering population has a signature of standing variation in some regions the genome, that potentially let them adapt rapidly to this new orientation pattern of migration.

Lastly, we conclude this second part with (to our knowledge) the first approach of gene regulation in migration. We analyse chromatin accessibility in three focal brain areas related to migration: hippocampus, Cluster N and the ventral anterior hypothalamus. We compared the chromatin accessibility landscape in individuals outside the migratory season (when the migratory behaviour is not exhibited) with individuals during the migratory season (when the migratory behaviour is expressed). We find a pattern of genome-wide repression of the chromatin in individuals that are migrating, probaly due to strict control of any unnecessary energetically demanding task in the cells of these brain areas. We also identify cis-regulatory modules that harbour potential elements to modulate the migratory behaviour.

CHAPTER 3

Candidate genes for migration do not distinguish migratory and non-migratory birds

Juan S. Lugo Ramos*, Kira E. Delmore*, Miriam Liedvogel*

*Max Planck Institute for Evolutionary Biology, AG Behavioural Genomics, August-Thienemann-Str. 2,24306 Plön, Germany.

Contributions: The study was designed by me and M. Liedvogel. I conducted all analysis and visualizations. I wrote the first draft and all authors contributed with revisions and result interpretations.

Published Article: "Candidate genes for migration do not distinguish migratory and non-migratory birds". Juan S. Lugo Ramos, Kira E. Delmore,· Miriam Liedvogel. *J Comp Physiol A* 203, 383–397 (2017).

J Comp Physiol A (2017) 203:383–397
DOI 10.1007/s00359-017-1184-6

ORIGINAL PAPER

Candidate genes for migration do not distinguish migratory and non-migratory birds

Juan S. Lugo Ramos[1] · **Kira E. Delmore**[1] · **Miriam Liedvogel**[1]

Received: 22 December 2016 / Revised: 12 May 2017 / Accepted: 15 May 2017 / Published online: 5 June 2017

Abstract Migratory traits in birds have been shown to have a strong heritable component and several candidate genes have been suggested to control these migratory traits. To investigate if the genetic makeup of one or a set of these candidate genes can be used to identify a general pattern between migratory and non-migratory birds, we extracted genomic sequence data for 25 hypothesised candidate genes for migration from 70 available genomes across all orders of Aves and characterised sequence divergence between migratory and non-migratory phenotypes. When examining each gene separately across all species, we did not identify any genetic variants in candidate genes that distinguished migrants from non-migrants; any resulting pattern was driven by the phylogenetic signal. This was true for each gene analysed independently, but also for concatenated sequence alignments of all candidate genes combined. We also attempted to distinguish between migrant and non-migrants using structural features at four candidate genes that have previously been reported to show associated with migratory behaviour but did not pick up a signal for migratory phenotype here either. Finally, a screen for d*N*/d*S* ratio across all focal candidate genes to probe for putative features of selection did not uncover a pattern, though this might not be expected given the broad phylogenetic scale used here. Our study demonstrates the potential of public genomic data to test for general patterns of migratory gene candidates in a cross-species comparative context, and raise questions on the applicability of candidate gene approaches in a macro-evolutionary context to understand the genetic architecture of migratory behaviour.

Keywords Candidate genes · Migratory traits · Bird migration · Next-generation sequencing · Genomic data

Introduction

Bird migration is one of the most fascinating and well-studied behaviours among birds, including work on the physiological and morphological adaptations required for successful migration and ecological correlates of this behaviour. Considerable interest has also focused on understanding how variation in the migratory phenotype is generated, with several studies demonstrating high variability for various migratory traits using selective breeding studies (e.g. Berthold et al. 1992), displacement experiments (e.g. Perdeck 1958; Chernetsov et al. 2008) and quantitative genetics analyses (e.g. Pulido and Berthold 2010). Nevertheless, the underlying genetic architecture shaping this behaviour remains poorly understood (Liedvogel et al. 2011; Delmore and Liedvogel 2016). One popular approach to enhance our understanding of the molecular basis of migratory traits has been the use of candidate genes for behavioural traits suggested to contribute to variation in migratory phenotype. These candidates are often selected by their molecularly characterised specific function in other (often model) organisms. A candidate gene approach attempts to identify an association between genetic variation in that gene and the focal phenotype, here our focal behaviour is migration. The underlying rationale of this approach is to focus

Electronic supplementary material The online version of this article (doi:10.1007/s00359-017-1184-6) contains supplementary material, which is available to authorized users.

✉ Miriam Liedvogel
liedvogel@evolbio.mpg.de

[1] Max Planck Institute for Evolutionary Biology, AG Behavioural Genomics, August-Thienemann-Str. 2, 24306 Plön, Germany

on genetic variation in specific candidate regions of the genome that have been suggested to directly impact the function of the candidate gene and ultimately the target phenotype in other species as well. In the context of migration, the traits receiving most attention are circadian behaviour and personality traits (e.g. exploratory or anxiety-related behaviour); focal candidate genes include *CLOCK, ADCYAP1, CREB1* and *NPAS2* for circadian rhythm, and *DRD4* and *SERT* for personality traits (see Müller et al. 2011 as one of the pioneer studies in the context of migration).

CLOCK and *ADCYAP1* are among the most studied candidate genes in the field of bird migration. The overall pattern for *CLOCK* variability indicates a latitudinal cline in repeat lengths at the variable region, possibly reflecting local adaptation to seasonal variation at different latitudes. The circadian *CLOCK* gene is highly conserved among birds throughout most of its sequence with the exception of one C-terminal region that contains a variable poly-glutamine(Q) (poly-Q) motif with variability in the number of glutamine repeats both among and within species (e.g. Johnsen et al. 2007; Liedvogel et al. 2009; Bazzi et al. 2016; but also see Liedvogel and Sheldon 2010; Dor et al. 2011). This region has been suggested to influence the transcription activating potential of the protein, potentially altering rhythms in both physiology and behaviour. The aforementioned pattern of a latitudinal cline in lengths polymorphism has been recovered in several species and in a migration context has been suggested to reflect adaptive features related to migration. Similarly for *ADCYAP1,* a neuropeptide-coding gene, one 3′ UTR microsatellite length variation has shown associations with migratory behaviour in blackcaps (Müller et al. 2011; Mettler et al. 2015) with longer alleles associated with higher migratory activity. However, the pattern so far lacks consistency across other avian species (e.g. Peterson et al. 2013) in order to confirm its suggested role as regulatory unit of migratory behaviour.

More recently, Delmore et al. (2015a) used next-generation sequencing data to estimate genomic differentiation between two subspecies groups of *Catharus ustulatus* that exhibit differences in their migratory behaviour (both timing and orientation). They characterised the entire genome to probe for enrichment of candidate genes for migration in areas of elevated differentiation across the genome. The resulting list of candidates included 25 genes that had been identified using a literature search (Ruegg et al. 2014). As predicted, genes from the list of candidates were enriched in areas of elevated differentiation, suggesting selection around these candidates could not only contribute to variation documented in their migratory behaviour, but also help maintain differences between the groups. Delmore et al. (2016) expanded on this work using hybrids and a genome-wide association study to identify one region in particular that is associated with variation in migratory orientation and harbours the *CLOCK* gene.

Despite several clearly species-specific differences in behavioural traits that make up the migratory phenotype, all migratory birds have to meet a general consensus schedule of key adaptations in order to cope with the challenge of migration. Specifically, migratory birds need a set of navigation and orientation mechanisms to migrate successfully, and must complete several key physiological traits such as hyperphagia and feather moulting in time prior to their migratory journey. Migratory individuals also share common morphological and anatomical adaptations, such as elongated and more pointed wings, lighter bone structure, maybe also brain volume (e.g. Lockwood et al. 1998; Fuchs et al. 2014). Given this common set of features among migratory birds and demonstrations of both their heritability and potentially shared genetic basis (e.g. the aforementioned involvement of *CLOCK* and *ADCYAP1* across species), it has been postulated that there may be a shared set of genes for migration among birds (Berthold 1999).

Here, we test the hypothesis that sequence variation at one or a common set of suggested candidate genes for migration has been exploited to adapt to the requirements of migratory behaviour in different avian lineages using a cross-species comparative approach. Specifically, we focused on each of the 25 candidate genes for migration proposed by Ruegg et al. (2014) and compared patterns of genetic variability separately for each candidate gene between migratory and non-migratory species of birds. In addition to analyses per candidate gene, we also analysed concatenated sequence data from all candidate genes. This work benefited from the wealth of genomic data that have been accumulated in the last half-decade, with the assembly of several draft reference genomes for birds and large-scale initiative of the Avian phylogenomics project (http://avian.genomics.cn/en/index.html; more recently expanding to the B10K project). Our full dataset included 17 obligate migratory species, 32 sedentary or non-migratory species and 21 species with an additional intermediate movement phenotype (e.g. dispersive, partial migrant) (Table S1). We compare patterns of evolutionary divergence of each candidate gene using three different approaches: (1) comparing observed topologies for candidate genes to trees built using phylogenetic relationships with and without distinguishing migratory species from non-migratory species; (2) performing a gene-wide and branch-specific d*N*/d*S* analysis to identify if selective pressures on these candidate genes play a role related to migration; and (3) focussing on structural features that previous studies have shown to correlate with migratory behaviour (e.g. microsatellites at *ADCYAP1* and *CREB1* 3′ UTR regions and poly-Q regions in *NPAS2*

and *CLOCK*), running linear models between these variants and two predictors of migratory behaviour. If genetic variation at candidate genes included in our study code for differences in migration, we predicted that gene trees based on candidate genes that play a role in shaping migratory behaviour would group migratory species together. Further, selective pressures on those candidate genes with a clear role in shaping the focal phenotype should be picked up in lineages with migratory species and linear models would show strong associations between genetic variation at structural features and predictors of migratory phenotype.

Materials and methods

Migratory phenotype characterisation

We classified each of the 70 species included in our study, according to their migratory phenotype. Our classification was based on a careful literature review of bird guides (Svensson et al. 1999), as well as BirdLife (http://birdlife.org) and Handbook of birds of the World (HBW) (http://www.hbw.com). We defined the following categories: clearly non-migratory (resident, sedentary) (0/R; $n = 32$), obligate migrant (2/M; $n = 17$). However, sometimes it is not easy to clearly define a species as either clearly non-migrant or obligate migrant, this is especially true when a migratory trait segregates within a population such as in partial migrants where only some individuals of the population migrate, consequently we added a third category (1; $n = 21$). This category includes partial migratory species (i.e. not all individuals of the population migrate), and species that exhibit other kind of migration-independent movement behaviour (e.g. dispersal, homing, foraging flights). For partial migrant species we used additional information of the individual used for generating the reference, such as date and geographical origin of sample collection, in order to clearly define migratory phenotype and grouped that individual/species accordingly whenever possible.

Genome sequences, extraction and alignment

We downloaded genome sequences and annotations for most of the species used in our study from the NCBI database (Supplementary Table S1). We further included genome sequences of five additional migratory species here: Siberian stonechat *Saxicola maurus* (Van Doren et al. 2017), Swainson's thrush *Catharus ustulatus* (Delmore et al. 2015a), European blackcap *Sylvia atricapilla* (Delmore et al., in preparation), Willow warbler *Phylloscopus trochilus* (Lundberg et al. 2017, accepted), and Greenish warbler *Phylloscopus trochiloides* (Irwin et al. 2016).

Once we had sequences data and annotation for all of the species, we used the Bedtools (Quinlan and Hall 2010) getfasta module to extract genomic sequences for each of the 25 candidate genes for every species. Sequences for unpublished genomes or genomes without annotations (for details see Table S1) were generated using Blastn and chicken cDNAs from the Ensembl database as a reference. All genomic sequences were aligned with MAFFT (Katoh and Standley 2013) and manually edited in AliView. Coding (CDS) sequences were also obtained from a multiple alignment of the genomic sequences and Ensemble cDNA sequences (including untranslated regions, UTRs) for the flycatcher, chicken, and Zebra Finch. Only sequences covering 50% or more of the chicken genes were considered for further analysis.

Statistical analyses

Topological comparisons

Evolutionary trees were constructed for each candidate gene using whole genomic sequences and cDNA as reference, using a Neighbour Joining approach in MEGA v5.2 (Tamura et al. 2011). The reliability of the trees was evaluated performing a bootstrap analysis of 1000 replicates with the Kimura 2 Parameters model. To visualise the trees we used Figtree (http://tree.bio.ed.ac.uk/).

We compared the pattern of evolutionary divergence of these gene trees with three different hypothetical scenarios: divergence driven by phylogeny ('phylogenetic topology'); divergence constrained by migration, i.e. different migratory phenotypes clustering in separate braches, while keeping the evolutionary relationship of the phylogenetic topology within each branch, ('migratory phenotype topology'); and random divergence ('random topology'). These comparisons were carried out for (a) the full dataset including all three migratory phenotypes; (b) a restricted dataset only contrasting exclusively obligatory migratory and completely non-migratory (resident) species, and (c) a clade-specific analysis exclusively focusing on the genus of Passeriformes, as this is the only monophyletic clade in our dataset with a sufficiently high number of species for both obligate migrants and non-migratory species, thus allowing for a more fine-tuned assessment on a narrower phylogenetic scale. The clade-specific subset allows us to test if the migratory phenotype might be controlled by a different clade-specific subset of genes. This comparative approach allows us to identify the presence or absence of general patterns, using genetic variation at candidate genes to distinguish between patterns related to phylogenetic relationships and migratory behaviour. The divergence driven by phylogeny (i.e. the gene trees matching the species tree, 'phylogenetic topology') was constructed using

the total evidence nucleotide species tree (TENT) phylogeny, published by Jarvis et al. (2014). For species not included in the TENT phylogeny we used timetree (Hedges et al. 2015) divergence times to position these species in our phylogeny. The divergence constrained by migration scenario ('migratory phenotype topology') was constructed by clustering each phenotype (once exclusively focusing on migratory versus resident species for the restricted dataset; and also for the full dataset including other movement as a third phenotype category) in one separate branch while keeping the evolutionary relationships of the phylogenetic topology within each branch. Random divergence ('random tree') was generated shuffling branches randomly from the gene trees obtained, in order to avoid bias regarding the method of random trees generation by TOPD/fmts that only randomises taxa, but not branches for the statistical comparison. Restricting these analyses to exclusively Passerine species allowed us to analyse the effects of the evolutionary patterns of each candidate genes on a smaller scale. An example of the topologies is illustrated for the candidate gene *PER3* in Fig. S1.

Comparisons of these three focal topologies were carried out in TOPD/fmtS (Puigbò et al. 2007) using three different approaches: nodal, splits and disagree from the program. In brief, the 'nodal approach' counts the number of nodes that separate two taxa in a given topology and calculates the root mean squared deviation (RSMD) between each pair of trees. For identical topologies RMSD results in a value of zero. To calculate the significance of the RMSD obtained, TOPD/fmts calculates the distance between two contrasted tree pairs and 100 random trees obtaining one standard deviation (SD) confidence interval (CI). Compared topologies are characterised as statistically similar, within noise or different, depending on their distance with respect to CI. Specifically distances below CI denote similar topologies; distances above CI indicate statistical difference (distances around CI are within noise). The 'disagree method' characterises how many taxa need to be removed from the compared topology in order to end up with the exact same topologies for both trees (assessed as count of taxa/total taxa). Consequently, a value of 0/total indicates identical topologies. The 'splits method' evaluates if there are common branches between both trees, the lower the distance the more branches the tree pair shares.

dN/dS analysis

Accounting for the fact that different species might have found different ways to alter similar phenotypes in the same gene (i.e. different changes in sequence), we also analysed synonymous and non-synonymous mutations of all candidate genes across species. In order to pick up on putative selective pressures on candidate genes for migration, a gene-wide d*N*/d*S* analysis was carried out on the Datamonkey server (Pond and Frost 2005). We used three different datasets for each candidate gene: one including all the species, one restricted to migratory, and another restricted to non-migratory species.

Gene-wide d*N*/d*S* ratios (w) were estimated by maximum likelihood (ML) methods using a different model for each gene. Each model was obtained from the CMS module of the server. Neighbour Joining (NJ) phylogenies obtained for each candidate gene were used as input to assess likelihood of the tree comparing the neutral null model M1 ($w < 1$) and the model M2 that allows $w > 1$. Positive selection was assessed if the likelihood shows a $p < 0.05$.

To evaluate if lineages with migratory species show a signature of selection, a branch-specific analysis of dN/dS was also carried on Datamonkey with the Branch-Site REL program (Pond et al. 2011). The dataset for each candidate included migratory and non-migratory species. Branches under episodic diversifying selection were identified with a Holm–Bonferroni corrected $p \leq 0.05$.

Structural features and predictors of migration

We used a linear regression analysis to test for a correlation between the genotype of migratory species at each focal candidate gene and both breeding latitude and migratory distance. Models for both predictors were run separately. The genotype used for each gene was the microsatellite length (as number of bases) of the 3′UTR of *ADCYAP1* and *CREB1*, or poly-Q (as number of predicted glutamine amino acids) on exon 20 of *CLOCK* and *NPAS*. For the *CLOCK* gene we included two separate polymorphic regions with variable poly-Q repeats in our analysis (both variable regions are located in the same exon). The significance of the fit was assessed with a simple linear regression, using a significance threshold $p \leq 0.05$.

Within and across population variability in candidate gene sequence

Our comparative analyses focus exclusively on the sequence of one reference genome; inter-individual variation is not taken into account, mostly due to the limitation of available data to examine this level of variation. In order to make an attempt to see if variance within on candidate gene could be higher/lower in a specific migratory phenotype, we focused on *CLOCK* gene polymorphism, the only candidate gene with a sufficiently high number of individual sequence data available for several species ($n = 10$), including both migratory ($n = 8$) and resident ($n = 2$) species. Here we compare datasets of individually genotyped migratory species: flycatcher *Ficedula hypoleuca* (Saino et al. 2015; $n = 226$), willow warbler *Phylloscopus*

trochilus ssp (unpublished data, $n = 384$), chiffchaff *Phylloscopus collybita* ssp (unpublished data, $n = 56$), nightingale *Luscinia megarhynchos* (Saino et al. 2015; $n = 151$), tree pipit *Anthus trivialis* (Saino et al. 2015; $n = 144$), barn swallow *Hirundo rustica* (Dor et al. 2011; $n = 830$), whinchat *Saxicola rubetra* (Saino et al. 2015, $n = 374$); and two non-migratory species: blue tit *Cyanistes caerulea* (Liedvogel et al. 2009; $n = 950$), great tit *Parus major* (Liedvogel and Sheldon 2010; $n = 804$). We compared averages and variances among different pairs of groups or species, employing a Welch t test and F test, respectively. We assume as statistically similar, distributions with a $p > 0.001$.

Results

Comparing phylogenetic, migratory phenotype and random topologies

We show constructed gene trees for a select number of candidate genes in Fig. 1, right column. In general, candidate gene trees do not separate migratory from non-migratory birds. Instead, they resemble the phylogenetic topologies expected based on Jarvis et al. (2014). Nonetheless, some groups of birds that comprised exclusively of migrants group together in most of the candidate gene trees. For example, *Falconifomes*, *Accipithridae* and some *Passeriformes* show a clustering pattern. Nevertheless, this is most likely due to their higher levels of similarity and common ancestry rather than relationships based on migratory phenotype (also see Fig. S2). Note that in addition to separate analyses for each of the 25 candidate genes, we also concatenated sequence data from all genes by species and constructed a combined candidate genes tree (Fig. 2). In this concatenated tree the clustering patterns persist in the aforementioned lineages.

Results from nodal, splits and disagree methods for statistically distinguishing between phylogenetic, migratory phenotype and random trees can be found in Table 1 (full dataset). Our results clearly show that the phylogenetic topology tree provides the best fit to most gene trees. The lack of support for the migratory phenotype topology shows no evidence for a monophyletic origin for a migratory phenotype across avian taxa. This pattern did not change when restricting our analysis to only the extreme phenotypes, i.e. exclusively obligate migrants and clear non-migrants. We further zoomed into just one clade (specifically Passerines). Although overall, we see general patterns either more similar to the speciation phylogeny or not showing differences from a comparison with random topologies (see nodal approach in Table 1), the monophyletic clade analysis on Passeriformes indicates that respective gene topologies for some candidates, specifically *HRSP12* and *HSPA5,* are consistently more similar to the migration topology than the speciation topology (Fig. 3). This clustering pattern is also evident in the tree topologies (see Fig. S2) where most of the migrant species within Passerine tend to cluster in one branch. The other approaches did not show evidence for a general trend towards one or a set of candidate genes being recurrently involved in distinguishing migrants from non-migrants.

Recall that we also concatenated sequence data from all genes by species and constructed a complete gene tree. This concatenated phylogeny shows that most of the taxonomic groups clustering migratory species together are statistically well supported and show larger branch lengths. This could suggest an evolutionary process of acceleration in these lineages for some of the candidate genes.

Selection in candidate genes is not related to migration

To identify selective forces in lineages with migratory species, we performed a gene-wide and branch-specific dN/dS analysis. The gene-wide analysis for selection in candidate genes across all bird species indicates that *CLOCK, DRD4, NEK2, HSPA5* and *CSNKE1*, have been under positive selection at $p < 0.05$ (Table 2). Nonetheless, and important within the focus of our analysis, this signature does not seem to be linked to the migratory phenotype in any of our comparisons, as independent datasets of separately analysed migratory, partial migratory and non-migratory species generally do not show any signature of selection. The only exception here was the *CLOCK* gene that showed significant values for selection in partial migrants. Thus, despite the general overall pattern of selection across all avian clade, none of these candidate genes showed lineages under selective pressure.

Association between structural features and predictors of migration

Alignments of the genomic sequences (Fig. S3) of any of the 25 gene candidates did not separate migratory and non-migratory species. *ADCYAP1* has variable sequence lengths of a microsatellite with an AG repeat at the 3′ UTR region ranging from 26 to 56 bp for migratory species, and 10–54 bp for non-migratory species. *CREB1* also has a microsatellite with a TG/CG repeat motive at the 3′ UTR region, ranging from 12 to 26 bp in migratory and 14–42 bp in non-migratory species, respectively (the longest length variant was exclusively found in tits). *NPAS* has one variable region of poly-Q repeats, in migratory species; this has length variation between 6 and 10 amino acids (aa), in non-migratory species the repeat length varied between 5 and 12 aa repeats. The *CLOCK* gene has 2

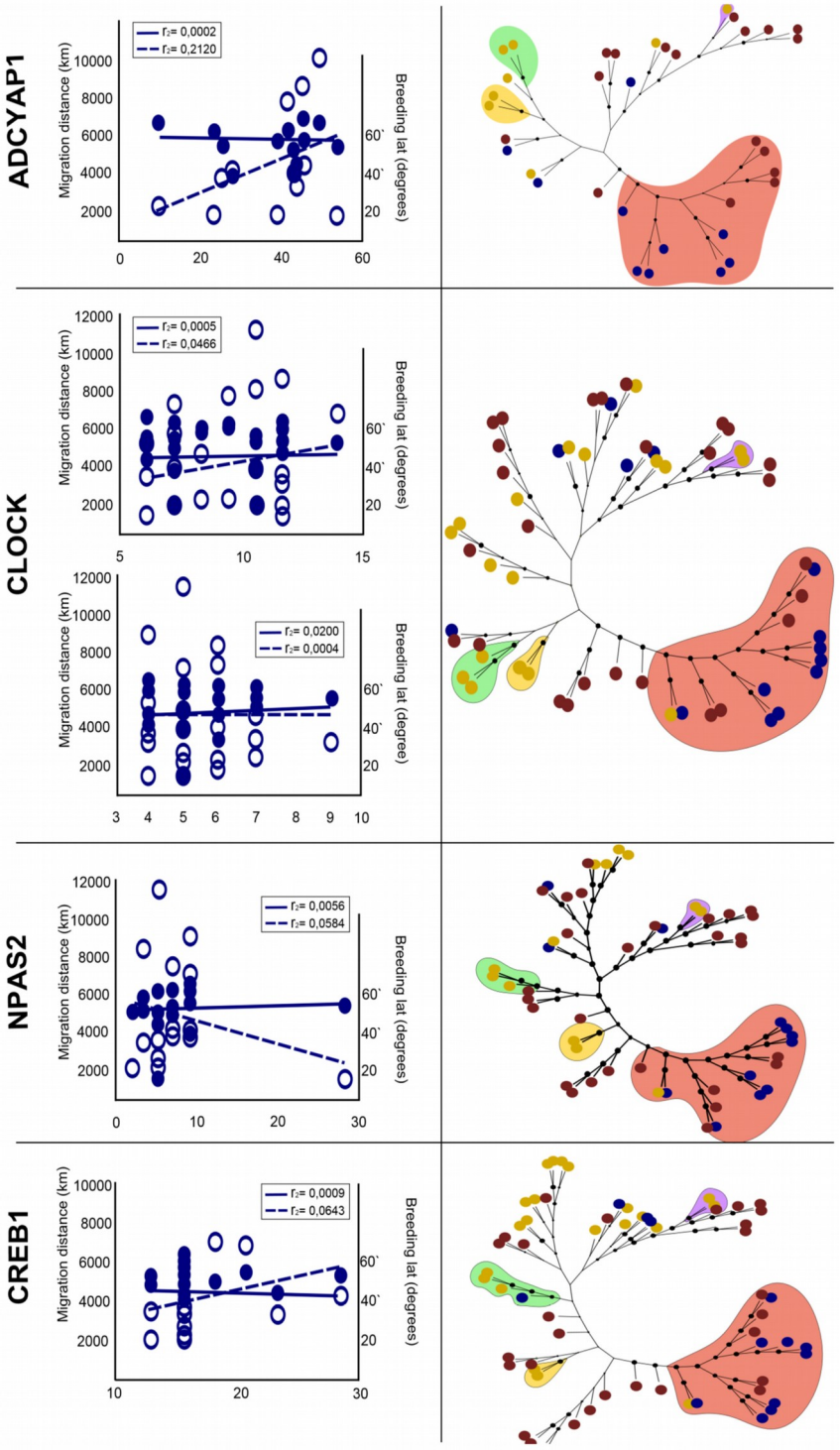


Fig. 1 Clustering patterns of migratory candidate genes. *Right panel* shows gene phylogeny for each candidate gene. Neighbour Joining (NJ) analyses shown for the four most widely used candidate genes for migration. *Coloured dots* indicate migratory (*blue*), non-migratory (*red*) and partial migrant/dispersive (*yellow*) taxa: node support is indicated by the size of nodes. Coloured clouds highlighted in *red*, *yellow*, *green* and *purple* highlight represent *Passeriformes* (*red*), *Falconiformes* (*yellow*), *Accipitridae* (*green*) and *Anseriformes* (*purple*), respectively. *Left panel* shows repeat lengths of focal genetic variants of each candidate gene, exemplarily illustrated for the most widely used candidate genes in the context of migration: *ADCYAP1, CLOCK, NPAS* and *CREB1*. Genotype is plotted in relation to migratory distance (*open circles*, *dashed lines*) and breeding latitude (*filled circles*, *continuous lines*). *Dashed* and *continuous lines* indicate the trend for linear regression of repeat variation at the candidate locus versus migratory distance and breeding, respectively. Variation at the *CLOCK* genes is shown for both variable length repeat regions (also see Fig. S3). *R* squared values are state for fitted linear models

poly-Q variable regions, both located on exon 20. The first region (R1) shows length variation between 6 and 13 poly-Q repeats in migrants and 4 to 14 repeats in non-migratory species. Length polymorphism in variable region two (R2) varied between 4 and 9 poly-Q repeats in both phenotypes. Lengths variation in neither of the focal candidates showed significant differences between migratory and non-migratory birds.

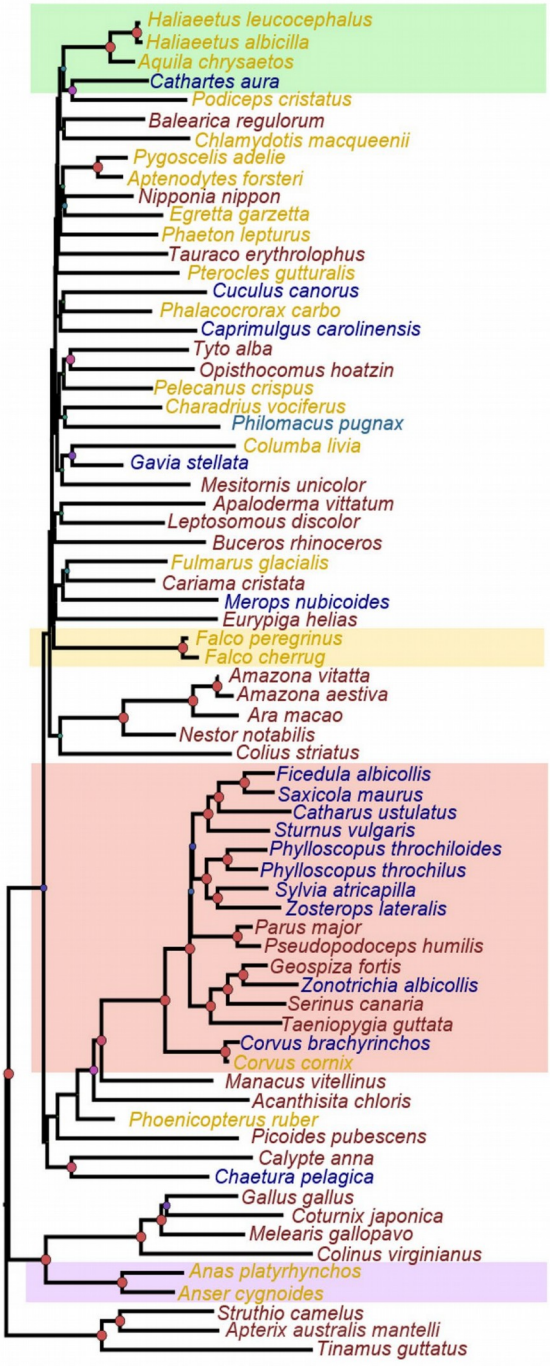


Fig. 2 Phylogenetic tree of concatenated candidate gene sequences. Neighbour Joining topology of genomic sequences for all 25 candidate genes. *Colour scheme* and node support as in Fig. 1

The intra-specific analysis focusing on individual variability of *CLOCK* gene polymorphism shows a tendency for higher variability in migratory species (Fig. S4); however, given that only two non-migratory species were included (namely great tits and blue tits—great tits being mostly monomorphic, blue tits highly variable), this does not allow us to draw any statistically supported conclusions (Table S3). Nonetheless, this individual based analysis of variability suggests that most common allele lengths (i.e. number of poly-Q repeats) varied considerably between species, as did the overall degree of variability within species (Fig. S4, Table S3). For example, migratory species like *H. rustica* and *F. hypoleuca*, show constrained levels of variability, being mostly monomorphic, while *C. caeruleus*, a non-migrant species, shows a degree of variance comparable to those migrant species. Statistical differences were found between the group of migrant and non-migrant species. Nonetheless when the comparisons were performed between individual species, migrant species and non-migrant species showed statistical differences, but also comparisons between a migrant and a non-migrant species did not show statistical differences (see Table S3).

We used linear models to quantify the correlation of structural repeats (poly-Q in *CLOCK* and *NPAS2*; and allele lengths in *ADCYAP1* and *CREB1*) and both breeding latitude and migratory distance (Fig. 1, left column). The fit of the data to a linear regression model did not show significant associations between genotype and migratory phenotype (Table 3). The best fit, albeit not significant, that was following a similar pattern as earlier work showing a correlation between candidate gene variation and migratory phenotype results from our comparison between *ADCYAP1* genotype variation in relation to migratory distance ($r^2 = 0.2120$, $p = 0.01132$). Although not significant, here the regression trend is in line with earlier studies, and might be taken as support for earlier findings reporting genotype variation at this locus with relevance to the migratory phenotype on the within-species level.

Discussion

We used publicly available archived genomic data from non-migratory and migratory bird species to test for the presence or absence of a general clustering pattern in candidate genes for migration. In a cross-species comparative approach, we characterised sequence features in 25 candidate genes in a dataset including birds across all orders of Aves. Our study thus illustrates the potential of public genomic data to test for general patterns of migratory gene candidates in a cross-species comparative context. Despite this powerful dataset we were not able to identify genetic variation that allowed us to distinguish migratory from non-migratory birds based on sequence difference in any of the candidate genes included here.

Most patterns we recover based on candidate gene sequence variation were driven by species phylogeny, and

Table 1 Topology comparison among candidate gene trees and target trees

	RANDOM			SPECIATION			MIGRATION		
	Dis	Nod	Split	Dis	Nod	Split	Dis	Nod	Split
PASSERINES restricted dataset									
ADCYAP1	12/12	2.741378	1	1/12	**0.696311**	**0.111111**	2/12	0.912871	0.222222
CLOCK	12/12	2.383656	1	1/13	**1.450022**	**0.4**	1/13	1.450022	0.4
CRY1	8/15	2.146980	0.916667	1/15	**1.751190**	**0.416667**	1/15	1.751190	0.416667
CRY2	17/17	3.226727	1	4/17	2.711631	**0.5**	4/17	**2.637401**	0.5
DRD4	17/17	3.391165	0	0/17	**0**	**0**	5/17	2.730546	0.5
HRSP12	15/15	2.700970	1	3/15	1.825742	0.5	2/15	**1.621287**	**0.416667**
HSP90B1	14/14	2.896522	1	1/14	**1.657386**	**0.364656**	3/14	1.675089	0.374074
HSPA5	16/16	3.345395	1	6/16	2.534758	**0.4**	3/16	**1.897367**	0.4
HSPA8	14/14	3.275938	1	1/14	**2.305879**	**0.435897**	3/14	2.399986	0.564103
NEK2	14/15	2.640600	0.638889	1/15	**1.067071**	**0.272222**	3/15	1.977253	0.438889
NFIL3	15/16	3.221151	1	2/16	**2.416335**	**0.5**	4/16	2.507021	0.5
PARL	17/17	3.429972	1	5/17	2.667892	0.642857	4/17	**2.546624**	**0.5**
PER2	13/13	2.480695	1	1/14	**1.500915**	**0.363636**	1/14	1.500915	0.363636
PER3	17/17	3.564531	1	5/17	2.825826	0.642857	4/17	**2.662374**	**0.5**
SLC2A1	12/12	2.424005	1	2/12	**1.299174**	**0.309259**	5/12	1.371361	0.546296
AANAT	16/16	2.966479	1	3/16	**2.081666**	**0.307692**	4/16	2.677063	0.692308
CSNK1E	13/13	3.125577	1	4/13	2.124340	**0.5**	4/13	**1.860521**	0.5
SLC1A3	12/14	3.119223	1	1/14	**2.401755**	**0.465397**	1/14	2.568437	0.472222
HSP90AA1	13/15	3.072418	0.823900	3/15	**1.345224**	**0.5**	5/15	2.042943	0.35
TTR	14/15	3.034009	1	2/15	**2.113851**	**0.407280**	2/15	2.193276	0.437451
YPEL1	13/13	2.916833	0.990741	3/13	**1.874818**	**0.5**	3/13	1.894358	0.5
NPAS2	10/10	2.653709	1	2/10	**1.341210**	**0.322046**	4/10	1.674568	0.386420
ARNTL	12/12	3.286592	1	2/12	**2.412694**	**0.4**	2/12	2.664591	0.4
CPNE4	13/14	3.221151	1	2/14	**2.416335**	**0.527778**	2/14	2.707021	0.472222
CREB1	17/17	3.735757	1	4/17	**2.711631**	**0.5**	4/17	2.902374	0.5
MIGRANTS versus NON-MIGRANTS restricted dataset									
ADCYAP1	26/26	3.478727	0.9565	5/27	**2.951652**	**0.375**	9/27	3.148735	0.5
CLOCK	40/40	5.640058	1	11/41	**5.180639**	**0.578947**	17/41	5.264237	0.631579
CRY1	39/39	5.528725	0.9722	7/40	**5.137968**	**0.621622**	17/40	5.142956	0.648649
CRY2	41/41	5.931387	0.9766	10/41	5.340413	0.683058	17/41	**5.126161**	**0.618569**
DRD4	41/41	5.086807	1	3/17	**0.469668**	**0.071429**	12/42	4.301095	0.512821
HRSP12	42/42	4.994887	0.9744	11/43	5.041138	**0.675**	23/43	**4.723354**	0.75
HSP90B1	46/47	5.639855	0.9852	16/47	**5.230941**	**0.640355**	21/47	5.391719	0.666943
HSPA5	47/47	5.483471	1	12/48	**5.055016**	**0.622222**	27/48	5.831560	0.733333
HSPA8	44/44	5.468022	1	13/44	**4.792124**	**0.670420**	18/44	4.877529	0.713964
NEK2	44/44	4.559393	0.9783	16/44	**4.066146**	**0.476974**	20/44	4.206486	0.565789
NFIL3	46/47	5.878573	0.9852	15/47	**5.185939**	**0.661762**	21/47	5.276429	0.686346
PARL	42/42	4.882503	0.9762	17/42	**4.423420**	**0.525190**	17/42	4.518643	0.593409
PER2	42/42	5.092866	1	13/43	**4.831376**	**0.65**	22/43	5.369616	0.675
PER3	46/46	4.579834	1	23/46	4.538616	0.697674	15/46	**3.413103**	**0.558140**
SLC2A1	28/28	5.259547	0.9853	13/28	**5.003874**	**0.646401**	16/28	5.037441	0.693649
AANAT	38/38	4.836443	1	10/39	**4.577714**	**0.611111**	20/39	5.174160	0.833333
CSNK1E	40/40	6.099601	1	20/40	5.006534	0.729730	13/40	**4.580897**	**0.594595**
SLC1A3	46/46	5.205493	0.9822	13/46	**5.003494**	**0.648874**	26/46	5.078642	0.691216
HSP90AA1	39/39	4.910991	0.9556	19/39	**4.893480**	**0.573299**	26/39	4.915253	0.631431
TTR	21/22	5.185258	0.9915	12/22	**4.888213**	**0.665140**	19/22	4.990166	0.712988
YPEL1	47/48	4.874233	0.9928	13/48	**4.553259**	**0.395**	22/48	4.683716	0.619087

Table 1 continued

	RANDOM			SPECIATION			MIGRATION		
	Dis	Nod	Split	Dis	Nod	Split	Dis	Nod	Split
NPAS2	40/40	5.479820	1	19/40	**4.936480**	**0.660727**	21/40	5.174525	0.678437
ARNTL	36/36	4.673910	0.9848	17/36	**4.342727**	**0.566613**	28/36	4.646077	0.572446
CPNE4	38/39	5.878573	0.9852	18/39	**5.185939**	**0.661762**	25/39	5.276429	0.676346
CREB1	40/48	6.699674	1	24/48	**5.973568**	0.644444	15/48	6.074231	**0.511111**
Full dataset									
ADCYAP1	30/34	3.794169	0.935484	8/35	3.906039	**0.468750**	15/35	**3.467496**	0.593750
CLOCK	59/59	5.931323	1	20/60	**5.510288**	**0.684211**	27/60	5.647208	0.701754
CRY1	56/60	5.894122	0.982456	23/62	**5.500493**	**0.677966**	28/62	5.589141	0.677966
CRY2	61/61	5.690368	0.991729	19/61	5.419998	0.676801	24/61	**5.307191**	**0.662440**
DRD4	60/60	5.837537	1	3/17	**0.469668**	**0.071429**	20/61	4.578520	0.568966
HRSP12	62/62	5.237258	1	27/63	5.349055	**0.7**	30/63	**5.082625**	0.733333
HSP90B1	67/67	4.944896	0.965674	25/67	**4.556093**	**0.604968**	30/67	4.976348	0.672276
HSPA5	67/67	6.240832	1	18/68	**5.691938**	**0.630769**	35/68	5.965220	0.692308
HSPA8	62/63	5.620202	0.985844	22/63	**5.125876**	**0.667334**	27/63	5.353811	0.682397
NEK2	65/66	6.254055	0.983173	25/66	**5.649675**	**0.651162**	27/66	5.737866	0.607372
NFIL3	67/67	5.970355	0.991207	21/67	**5.528167**	**0.637798**	32/67	5.724862	0.624456
PARL	66/66	5.523274	0.989792	20/66	**5.203310**	**0.632417**	29/66	5.403632	0.628094
PER2	59/59	5.391185	1	22/60	**5.182543**	**0.666667**	28/60	5.781717	0.684211
PER3	66/66	5.478672	1	30/67	5.203242	0.671875	20/67	**4.601044**	**0.562500**
SLC2A1	41/41	5.860697	0.987179	15/41	**5.275366**	**0.663248**	20/41	5.524032	0.670378
AANAT	54/54	5.090447	1	14/55	**4.810622**	**0.653846**	30/55	5.136389	0.788462
CSNK1E	51/55	5.950073	0.961538	25/55	4.951618	0.692308	19/55	**4.825159**	**0.634615**
SLC1A3	65/65	4.862746	0.967742	16/65	**4.708164**	**0.576480**	25/65	4.557352	0.647752
HSP90AA1	52/53	5.206538	0.972647	17/53	**4.972273**	**0.610309**	23/53	4.991282	0.657823
TTR	22/22	6.378516	0.965144	11/22	**5.565637**	**0.666466**	14/22	5.380218	0.582933
YPEL1	48/48	5.687568	0.994152	16/48	**5.453279**	**0.687392**	30/48	5.639658	0.704351
NPAS2	56/56	5.182652	0.980130	20/56	**4.961226**	**0.619555**	31/56	4.986559	0.622692
ARNTL	55/56	5.949312	0.974225	17/56	**5.253574**	**0.667268**	24/56	5.519978	0.648638
CPNE4	54/54	5.503541	0.991195	19/54	**5.202493**	**0.676264**	27/54	5.373337	0.696865
CREB1	61/67	6.806959	0.968750	29/67	6.179656	0.640625	18/67	**5.935276**	**0.531250**

To contrast topology patterns we compared single gene trees derived for each candidate gene with three target trees: (1) randomly generated trees based on randomly permutating braches of single gene tree topology 'random topology'; (2) genes based on overall phylogeny (Jarvis et al. 2014), thus representing evolutionary divergence among all the species included in the study 'phylogenetic topology', and (3) topology classified by migratory behaviour, i.e. an artificial tree clustering migratory species and non-migratory species in two different branches 'topology based on migratory phenotype'

Bold characters highlight the most similar topology for each candidate gene tree for the three methods. Tables shown for three datasets: Monophyletic analysis on Passerines; exclusively contrasting obligate migrants versus clear residents; and the full dataset including the third phenotype category with partial migrants, dispersive species

Topologies were compared and analysed by three methods included in the TOPD software: *Dis* disagree, *Nod* nodal, *split* split (see "Materials and methods" for details)

* and ** Denote significance below and above one SD of 100 randomly generated trees, respectively

did not show indications separating migratory and non-migratory birds. This could suggest that candidate genes do not play a general role in controlling migration across the full avian clade, or that patterns cannot be evident in a wide phylogenetic analysis. In addition to analysing each candidate gene individually, we also analysed concatenated sequence data from all candidate genes, which also did not allow clear separation of migratory and non-migratory birds. The latter finding might not be too surprising, as this scenario would assume divergence of candidate genes among lineages in parallel to species divergence, a rather unlikely scenario.

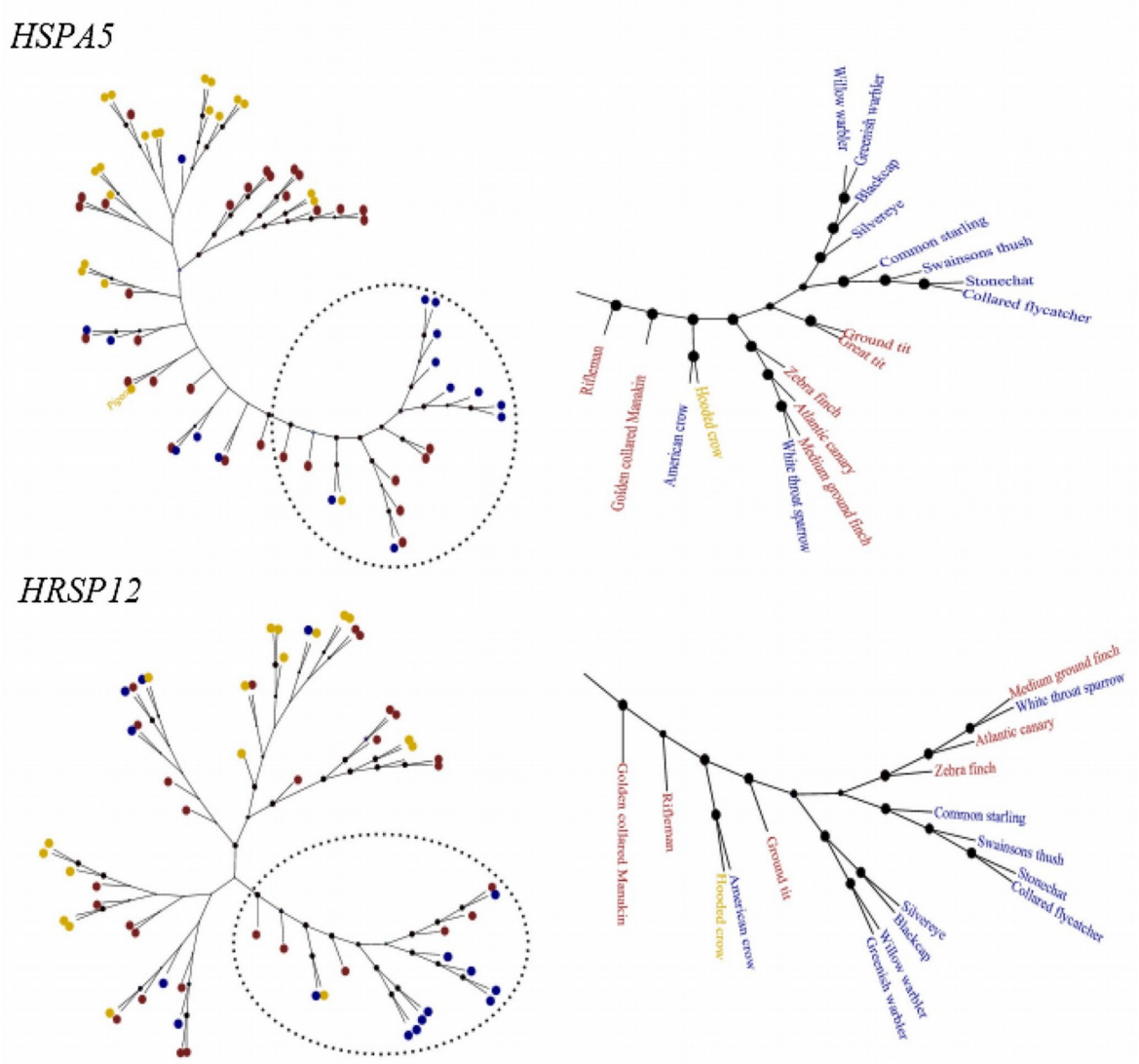


Fig. 3 Topologies on a constrained phylogenetic scale. Clade-specific analysis exclusively focusing on all species within the genus Passerines. The candidate gene trees obtained from the Neighbour Joining analysis for HRSP12 and HSPA5, two candidate genes that follow clustering patterns consistently more similar to the migration topology than the speciation topology. *Colouring scheme* and node support as in Fig. 1

In order to address the fact that we cannot exclude that the relevance and contribution of certain traits may differ between avian lineages in a way that different traits may be controlled by different mutations in the same genes, and/or varying patterns of epistatic interactions with other genes, we also screened for putative features of selection within these candidate genes. However, our dN/dS approach did not allow identifying patterns of selection at either the gene-wide or branch-sites level for the migratory lineages. This result does not provide support for lineage-specific selection on the gene candidates.

Indicative patterns might get lost if not approached at the appropriate phylogenetic scale. Our more focused analyses on a restricted data assembly exclusive to the Passeriformes allow aims to address this issue at least partly allowing us to potentially recover general trends towards some or all candidate genes being recurrently involved in the control of migration. Analyses limited to this family show improved resolution and come close to the margin of randomness in nodal comparisons. However, even on this more focal scale only a few of these comparisons show better values than the phylogeny species tree, suggesting that even on a more constrained phylogenetic scale, the genes keep following

Table 2 Gene-wide dN/dS analysis of migratory gene candidates

Gene	Phenotype	lnL M1	lnL M2	LRT	*P* value
ARNTL	ALL	−12014.300	−12014.300	1.30E−04	1.000
	Migratory	−2443.540	−2443.540	8.43E−05	1.000
	Partial migrants	−5751.000	−5883.250	3.42E−03	1.000
	Residents	−9022.375	−9022.055	1.17E+00	0.623
CLOCK	**ALL**	**−17203.100**	**−17199.300**	**7.49E+00**	**0.0236713****
	Migratory	−2784.523	−2784.444	2.89E−04	0.899
	Partial migrants	**−6084.900**	**−6801.840**	**6.13E+00**	**0.0467****
	Residents	−12321.600	−12321.600	1.27E−04	1.000
PER3	ALL	−41720.300	−41718.500	3.49E+00	0.175
	Migratory	−2124.250	−2124.250	2.34E−06	1.000
	Partial migrants	−5883.250	−5883.250	6.48E−05	1.000
	Residents	−13206.766	−13206.727	7.11E−02	0.969
CRY1	ALL	−11297.000	−11297.000	6.54E−05	1.000
	Migratory	−2618.571	−2618.509	5.14E−01	0.949
	Partial migrants	−6543.055	−6543.055	2.94E−05	0.999
	Residents	−7465.354	−7465.354	4.34E−05	1.000
CRY2	ALL	−12140.600	−12140.600	−5.26E−04	1.000
	Migratory	−2718.098	−2718.030	2.57E−01	0.869
	Partial migrants	**−5137.220**	**−5134.740**	**4.96E+00**	**0.084**
	Residents	−11345.504	−11345.444	1.11E−01	0.952
NEK2	**ALL**	**−15013.300**	**−15008.900**	**8.86E+00**	**0.0119051****
	Migratory	−3123.490	−3123.400	1.03E+00	0.798
	Partial migrants	−6882.490	−6882.400	2.48E−06	1.000
	Residents	−13729.370	−13729.286	1.55E−01	0.932
NPAS2	ALL	−9773.415	−9773.415	3.58E−07	1.000
	Migratory	−2788.055	−2787.988	6.78E−04	1.000
	Partial migrants	−6711.240	−6711.240	5.43E−06	1.000
	Residents	−8384.493	−8384.493	4.83E−05	0.998
PER2	ALL	−34024.900	−34024.900	2.68E−04	1.000
	Migratory	−2703.313	−2703.248	3.57E−04	0.975
	Partial migrants	−5601.859	−6548.185	−1.29E−04	1.000
	Residents	−14439.024	−14438.968	1.02E−01	0.955
DRD4	**ALL**	**−9382.370**	**−9377.070**	**1.06E+01**	**0.00499358****
	Migratory	**−3260.480**	**−3257.910**	**5.14E+00**	**0.077**
	Partial migrants	−5392.430	−6882.400	5.45E−06	1.000
	Residents	−6979.373	−6979.033	1.25E+00	0.599
AANAT	ALL	−5698.000	−5698.000	2.84E−05	1.000
	Migratory	−2699.671	−2699.604	6.46E−02	0.914
	Partial migrants	−6545.730	−6545.730	2.02E−05	0.998
	Residents	−4721.560	−4721.560	−7.37E−05	1.000
CPNE4	ALL	−11228.606	−11228.588	3.60E−02	1.000
	Migratory	−2452.620	−2452.575	5.76E−04	1.000
	Partial migrants	−5633.000	−6539.900	3.20E−04	0.999
	Residents	−9385.890	−9385.890	3.73E−04	0.999
HSPA5	**ALL**	**−16331.700**	**−16328.200**	**6.92E+00**	**0.0314173****
	Migratory	−2707.037	−2706.969	3.26E−02	0.796
	Partial migrants	−5595.625	−6566.695	1.26E−03	0.993
	Residents	−12404.581	−12404.509	1.33E−01	0.942

Table 2 continued

Gene	Phenotype	lnL M1	lnL M2	LRT	*P* value
NFIL3	ALL	−9638.910	−9638.910	1.55E−04	1.000
	Migratory	−2788.055	−2788.055	5.14E−01	0.949
	Partial migrants	−5565.000	−6711.240	−2.83E−05	1.000
	Residents	−8583.364	−8583.304	1.11E−01	0.952

Significance of positive selection was assessed with $p < 0.1$ indicated with bold font and $p < 0.05$ indicated with **

the trends of phylogeny or speciation rather than clustering according to migratory phenotype.

Most studies of candidate genes look for correlates of length polymorphism in a limited number of species. Our approach aims to extend this framework to assessing sequence variability in candidate genes to a broader evolutionary scale. However, our data do not uncover statistical differences between species with different phenotypes, neither do they recover correlations with specific migratory features, specifically distance and breeding latitude. *ADCYAP1* did show a tendency for positive, though non-significant, correlation with migratory distance in our dataset, supporting earlier findings in species-specific analyses. As the length polymorphism in this gene is located in a 3′ UTR this could suggest that a specific pattern of gene expression is playing a role for this gene, and this feature might be relevant in long-distance migrants.

Our focus is put on general sequence differences in candidate genes for migration across species, but we also note that there is a lot of variation in the level of individual and inter-population diversity across avian species in many of the candidate genes included here, such as *CLOCK* poly-Q. Within-species variability at polymorphic loci often is high in some, but certainly not all species, irrespective of migratory phenotype. Previous work has suggested that variability in lengths polymorphism of *CLOCK* related to the timing of migration is enhanced in migratory birds (Saino et al. 2015). Our analyses testing for a role of inter-individual variation in *CLOCK* lengths polymorphism between migratory and non-migratory species falls short in strengthening this suggestion. We detect similarly high levels of polymorphisms as well as scenarios with almost monomorphic genotypes for various species, irrespective of migratory phenotype. This could mean that the observed variability in *CLOCK* is more tightly linked to other parameters that shape the variability of the migratory phenotype, such as breeding latitudes, distances and timing regimes, rather than specifically controlling timing traits in a migration context.

Table 3 Predictors of linear models

	df	Res err	*F*	*p*
ADCYAP1 vs distance	11	2228	2.962	0.1132
ADCYAP1 vs breeding	11	7.299	0.05144	0.8252
CREB1 vs distance	13	2093	0.8929	0.3619
CREB1 vs breeding	13	6.126	0.2411	0.6316
CLOCK R1 vs distance	23	2675	1.125	0.3
CLOCK R1 vs breeding	23	12.8	0.01369	0.9079
CLOCK R2 vs distance	21	2726	0.3082	0.5846
CLOCK R2 vs breeding	21	17.73	0.1876	0.6693
NPAS2 vs distance	17	2701	0.647	0.432
NPAS2 vs breeding	17	11.29	0.0788	0.7825

For every comparison *df* degree of freedom, *res* residual error, *F* statistic and significance levels are shown

Significance at $p < 0.05$

But even when leaving inter-individual variability aside and focusing on just one reference sequence, we expect general patterns that allow separating migrants from non-migratory species of suggested candidate genes for migration to show up if they were in fact under parallel selection. Our results do not support this hypothesis and thus highlight limitations and question the usefulness of a candidate gene approach in the context of understanding migratory behaviour. Consequently, we call for caution and highlight the limitation of candidate gene approaches in macro-evolutionary contexts, as in most cases functionality cannot be easily inferred. Also, most candidate gene studies do not allow for a clear distinction from genetic drift, come with uncertainty about the amount of loci involved in the trait and do not allow controlling for possible effect of linkage disequilibrium. Our study further points out how this approach can erroneously simplify a highly complex phenotype that most probably is a multilocus trait.

One further point we feel is especially important to keep in mind when using a candidate gene approach in the context of migration: many candidate genes have been identified in genetic model organisms—none of that migrate or show any correlate to a migratory phenotype, and functional validation or relevance is mostly lacking. In many of these studies, the subset of candidate genes for migration were selected based on their assumed effect on anticipated candidate traits that feed into the complex phenomenon of migration. Importantly, one possibility explaining the lack of pattern might be the limitation of the starter set of ad

hoc hypothesised candidate genes we investigate and may not be completely unexpected. Once more genomic data for migratory and non-migratory species will become available, the genomic toolbox allowing to investigate the genetic basis of migratory traits will grow and allow for an increasingly more informed list of de novo identified candidates to be tested in migratory birds.

Acknowledgements Open access funding provided by Max Planck Society. Funding was provided by the Max Planck Society and the International Max Planck Research School (IMPRS) for Evolutionary Biology.

Compliance with ethical standards

Conflict of interest The authors confirm that they do not have any conflict of interest.

References

Bazzi G, Cecere JG, Caprioli M, Gatti E, Gianfranceschi L, Podofillini S, Possenti CD, Ambrosini R, Saino N, Spina F, Rubolini D (2016) Clock gene polymorphism, migratory behaviour and geographic distribution: a comparative study of trans-Saharan migratory birds. Mol Ecol. doi:10.1111/mec.13913

Berthold P (1999) A comprehensive theory for the evolution, control and adaptability of avian migration. Ostrich 70:1–11

Berthold P, Helbig AJ, Mohr G, Querner U (1992) Rapid microevolution of migratory behaviour in a wild bird species. Nature 360:668–670. doi:10.1038/360668a0

Chernetsov N, Kishkinev D, Mouritsen H (2008) A long-distance avian migrant compensates for longitudinal displacement during spring migration. Curr Biol 18:188–190. doi:10.1016/j.cub.2008.01.018

Delmore KE, Liedvogel M (2016) Investigating factors that generate and maintain variation in migratory orientation: a primer for recent and future work. Front Behav Neurosci 10:3. doi:10.3389/fnbeh.2016.00003

Delmore KE, Hübner S, Kane Schuster R, Andrew RL, Câmara F, Guigó R, Irwin DE (2015a) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol Ecol 24:1873–1888

Delmore KE, Kenyon HL, Germain RR, Irwin DE (2015b) Phenotypic divergence during speciation is inversely associated with differences in seasonal migration. Proc R Soc B Biol Sci 282:20151921

Delmore KE, Toews DPL, Germain RR, Owens GL, Irwin DE (2016) The genetics of seasonal migration and plumage color. Curr Biol 26:2167–2173

Dor R, Lovette IJ, Safran RJ, Billerman SM, Huber GH, Vortman Y et al (2011) Low variation in the polymorphic *Clock* gene poly-Q region despite population genetic structure across barn swallow (*Hirundo rustica*) populations. PLoS One 6(12):e28843

Fuchs R, Winkler H, Bingman VP, Ross JD, Bernroider G (2014) Brain geometry and its relation to migratory behavior in birds. J Adv Neurosci Res 1:1–9

Hedges SB, Marin J, Suleski M, Paymer M, Kumar S (2015) Tree of life reveals clock-like speciation and diversification. Mol Biol Evol 32:835–845

Irwin DE, Alcaide M, Delmore K, Irwin JH, Owens GL (2016) Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Mol Ecol 25:4488–4507

Jarvis ED, Aberer AJ, Li B, Houde P, Li Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1126–1138

Johnsen A, Fidler AE, Kuhn S, Carter KL, Hoffmann A, Barr IR, Biard C, Charmantier A, Eens M, Korsten P, Siitari H, Tomiuk J, Kempenaers B (2007) Avian *Clock* gene polymorphism: evidence for a latitudinal cline in allele frequencies. Mol Ecol 16:4867–4880. doi:10.1111/j.1365-294X.2007.03552.x

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Liedvogel M, Sheldon BC (2010) Low variability and absence of phenotypic correlates of *Clock* gene variation in a great tit *Parus major* population. J Avian Biol 41:543–550

Liedvogel M, Szulkin M, Knowles SCL, Wood MJ, Sheldon BC (2009) Phenotypic correlates of *Clock* gene variation in a wild Blue Tit population: evidence for a role in seasonal timing of reproduction. Mol Ecol 18:2444–2456

Liedvogel M, Åkesson S, Bensch S (2011) The genetics of migration on the move. TREE 26:561–569. doi:10.1016/j.tree.2011.07.009

Lockwood R, Swaddle JP, Rayner JMV (1998) Avian wingtip shape reconsidered: wingtip shape indices and morphological adaptations to migration. J Avian Biol 29:273–292. doi:10.2307/3677110

Lundberg M, Liedvogel M, Larson K, Sigeman H, Grahn M, Wright A, Akesson S, Bensch S (2017) Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. Evol Lett (**accepted manuscript**)

Mettler R, Segelbacher G, Schaefer HM (2015) Interactions between a candidate gene for migration (*ADCYAP1*), Morphology and sex predict Spring arrival in blackcap populations. PLoS ONE 10(12):e0144587. doi:10.1371/journal.pone.0144587

Müller JC, Pulido F, Kempenaers B (2011) Identification of a gene associated with avian migratory behaviour. Proc R Soc B Biol Sci 278:2848–2856. doi:10.1098/rspb.2010.2567

Perdeck AC (1958) Two types of orientation in migrating starlings, *Sturnus vulgaris* L., and chaffinches, *Fringilla coelebs* L., as revealed by displacement experiments. Ardea 46:1–37

Peterson MP, Abolins-Abols M, Atwell JW et al (2013) Variation in candidate genes *CLOCK* and *ADCYAP1* does not consistently predict differences in migratory behavior in the songbird genus *Junco*. F1000 Res 2:115. doi:10.12688/f1000research.2-115.v1

Pond SLK, Frost SDW (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–2533

Pond SLK, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K (2011) A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol. doi:10.1093/molbev/msr125

Puigbò P, Garcia-Vallvé S, McInerney JO (2007) TOPD/FMTS: a new software to compare phylogenetic trees. Bioinf 23:1556–1558

Pulido F, Berthold P (2010) Current selection for lower migratory activity will drive the evolution of residency in a migratory bird population. PNAS 107:7341–7346. doi:10.1073/pnas.0910361107

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinf 26:841–842

Ruegg KC, Anderson EC, Paxton KL, Apkenas V, Lao S, Siegel RB, DeSante DF, Moore F, Smith TB (2014) Mapping migration in a songbird using high-resolution genetic markers. Mol Ecol 23:5726–5739. doi:10.1111/mec.12977

Saino N, Bazzi G, Gatti E, Caprioli M, Cecere JG, Possenti CD, Gallimberti A, Orioli V, Bani L, Rubolini D, Gianfranceschi L, Spina F (2015) Polymorphism at the *Clock* gene predicts phenology of long-distance migration in birds. Mol Ecol 24:1758–1773. doi:10.1111/mec.13159

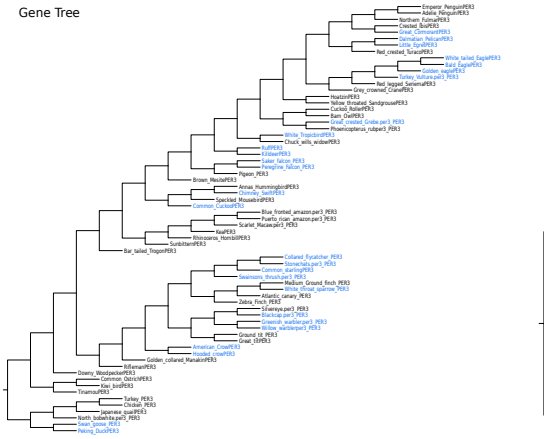
Svensson L, Grant PJ, Mullarney K, Zetterstroem D (1999) Der neue Kosmos-Vogelführer—Alle Arten Europas, Nordafrikas und Vorderasiens. Franckh-Kosmos Verlags-GmbH & Co., Stuttgart. ISBN 3-440-07720-9

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetic analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

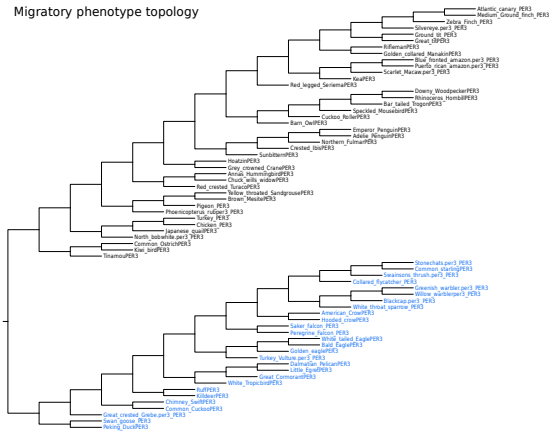
Van Doren BM, Campagna L, Helm B, Illera JC, Lovette IJ, Liedvogel M (2017) Correlated patterns of genetic diversity and differentiation across an avian family. Mol Ecol. doi:10.1111/mec.14083

Supplementary Figures and tables

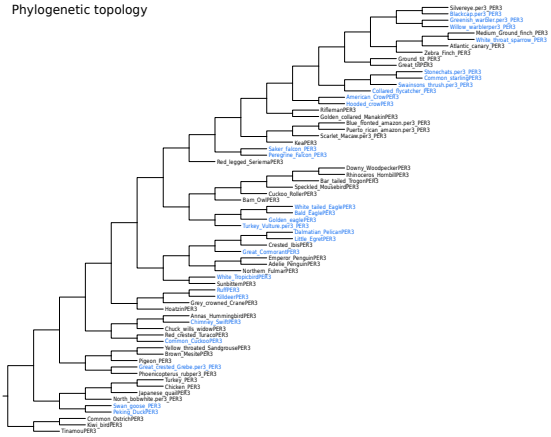
Gene Tree



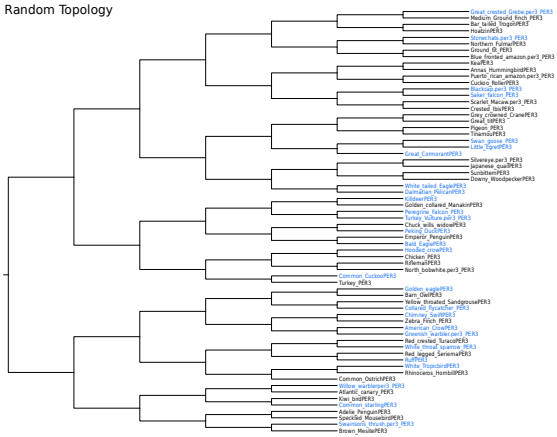
Migratory phenotype topology



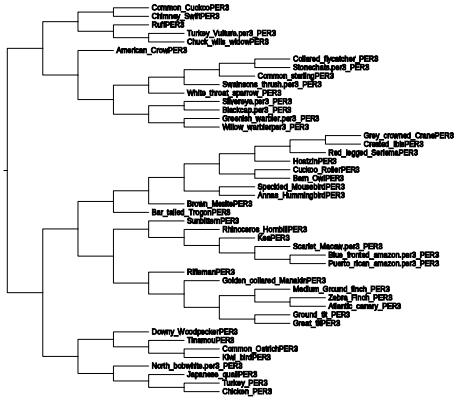
Phylogenetic topology

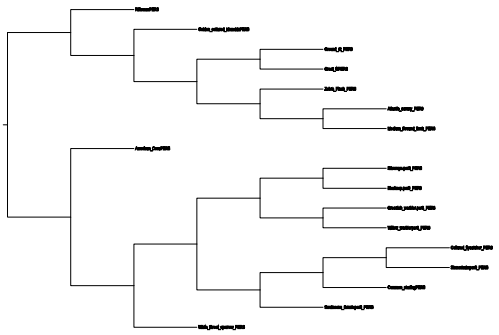


Random Topology



Extremes (0 and 2) topology





Spectrum (0, 1 and 2) topology

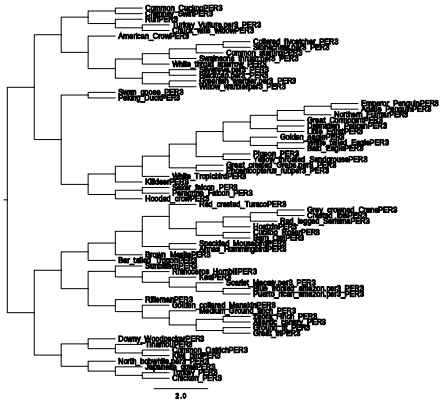


Figure S1 Trees employed for topology comparison. Seven trees are generated for each candidate gene. Migratory taxa are highlighted in blue. Here the gene tree for *PER3* obtained from the Neighbour Joining Analysis (A) is compared to a migratory phenotype topology (B), phylogenetic topology (C), random (D). (E) Shows Topology comparisons exclusively based on the reduced dataset exclusively containing obligate migrants and non-migratory species (E), as well as the pattern for the full dataset including all three phenotype classes (F)

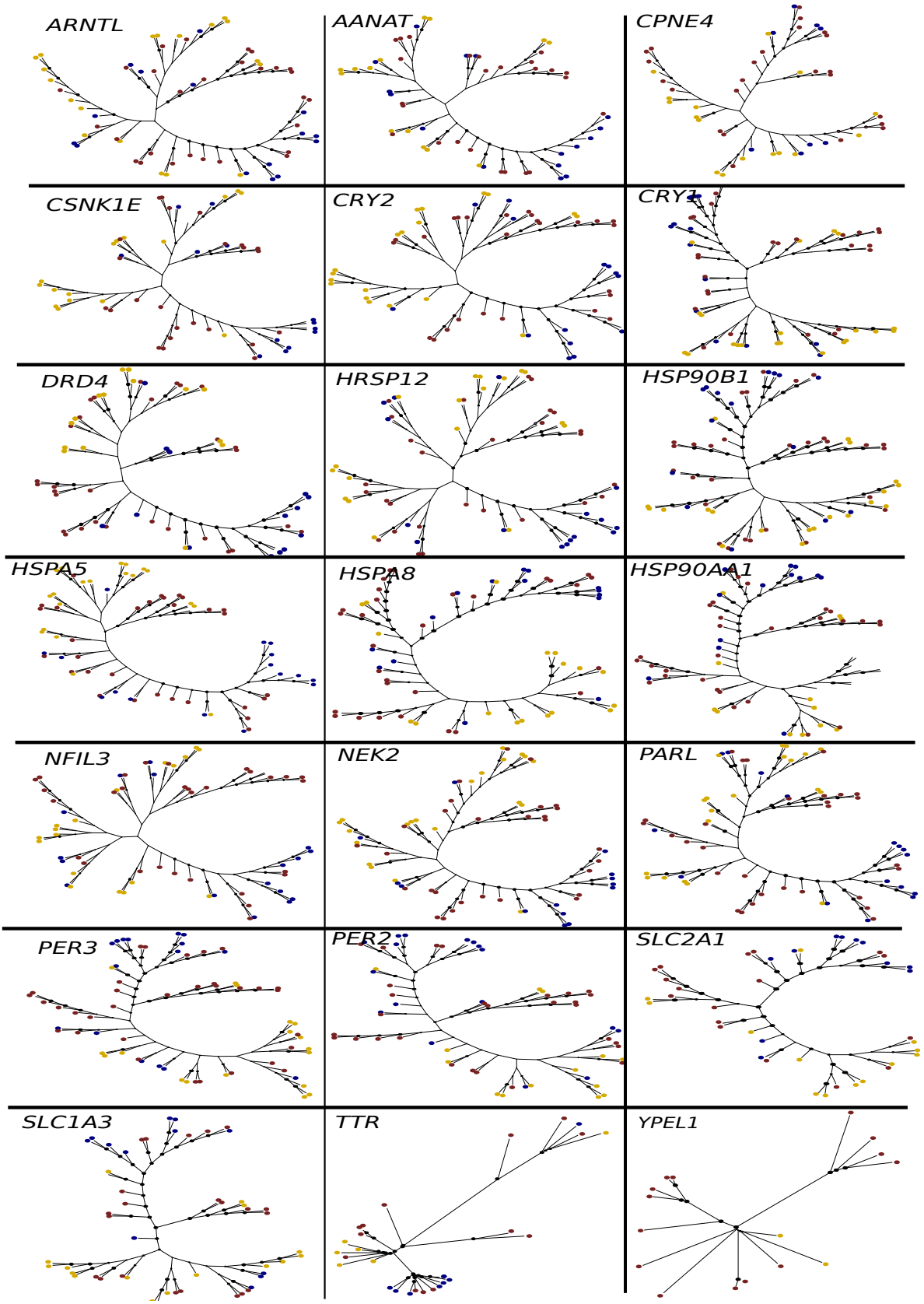


Figure S2 Simplified topologies of all gene candidates. The gene trees obtained from the Neighbour Joining analysis for each candidate gene. Colouring scheme and node support as Figure 1

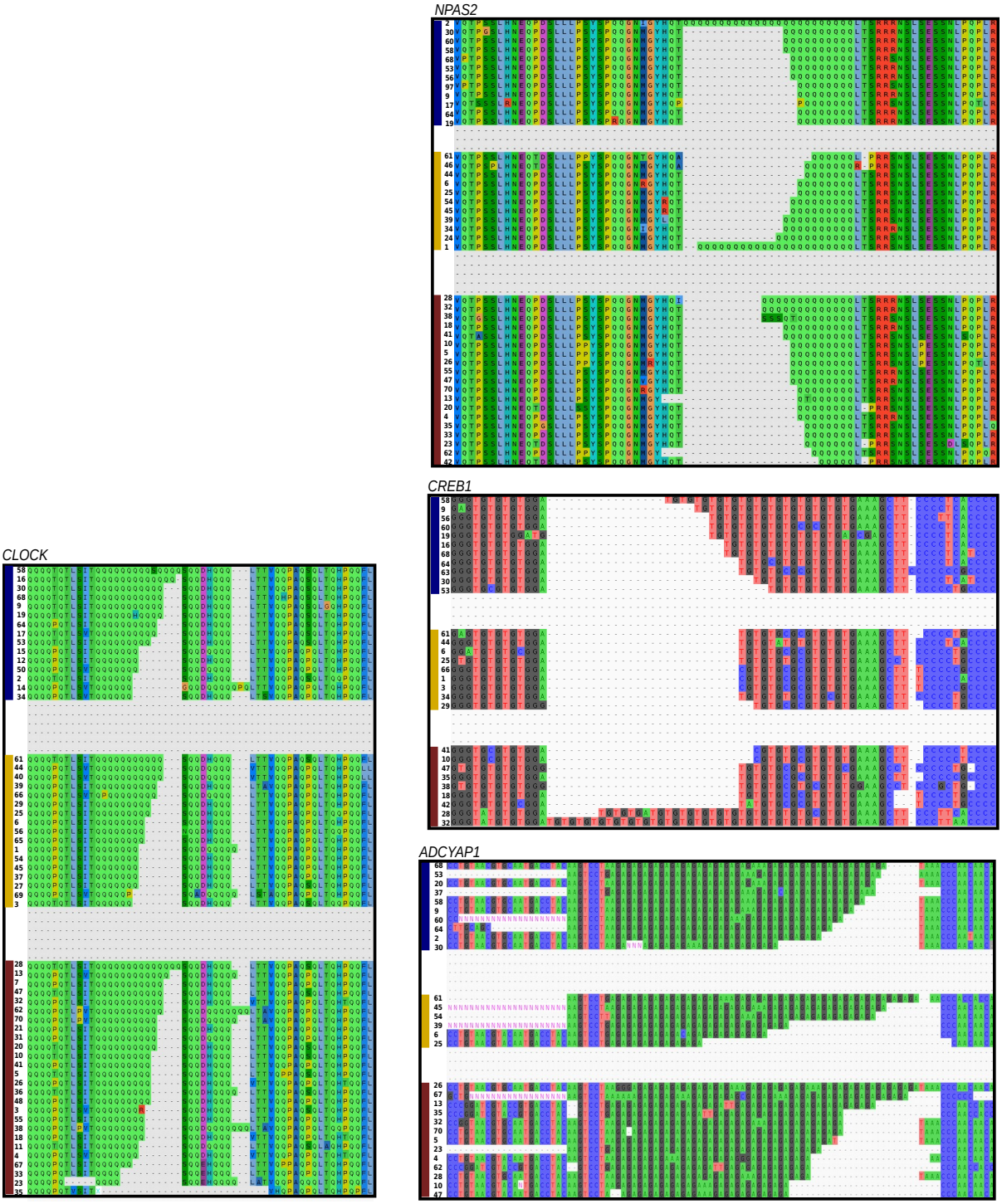


Figure S3 Cross-species comparative pattern of length polymorphisms at candidate genes for migration exemplarily illustrated for the four most widely used candidates (*ADCYAP1*, *CLOCK*, *NPAS* and *CREB1*). Upper panel (highlighted by a blue bar to the left) shows migratory species aligned by decreasing number of repeats at the variable locus; lower panel comprises non-migratory species arranged by increasing length polymorphisms. Only one sequence per species was used as reference

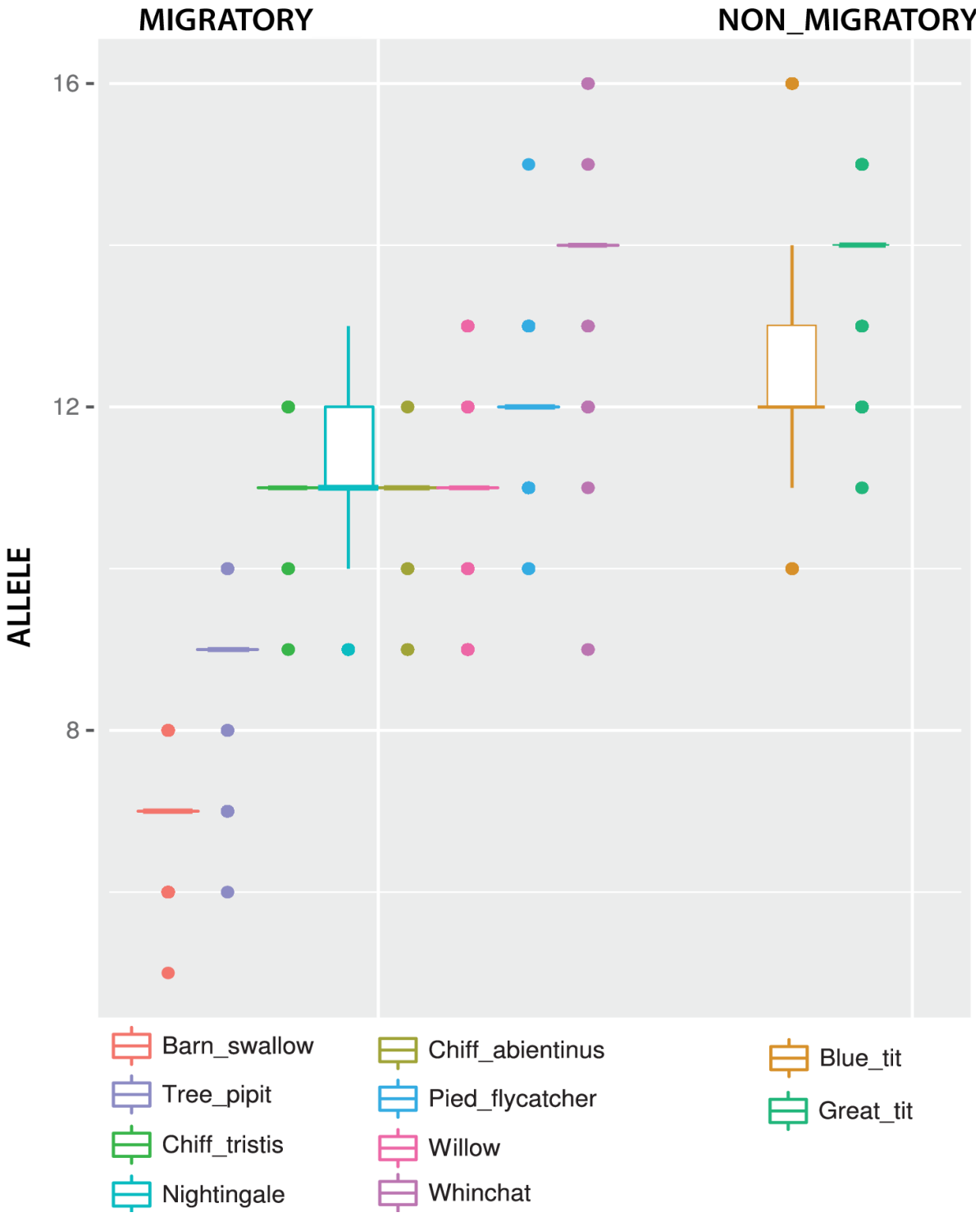


Figure S4. Intra-specific variation of the *CLOCK* gene in migratory and non-migratory species. Boxplot of allele lengths at the poly-Q region in migratory (Barn swallow, Chiffchaff, Nightingale, Pied flycatcher, Tree pipit, Whinchats and Willow warblers) and non-migratory species (Blue tit, Great tit). Lines indicate the most common allele of each species

Table S1 Reference information for all avian species included in the cross-species comparison. NCBI ID reference is included if available. Breeding latitude is stated as degrees from equator. Migratory distance calculated as kilometres between breeding and wintering grounds (only for migratory species). Category is a classification into 0: completely non-migratory (resident), 1: partial migratory/dispersive, 2: obligate migratory species. Distance and breeding latitude obtained following an approach explained in Delmore et al. 2015b

Annotation	Common name	Scientific name	Order	phenotype classes	mig/ resident	breeding latitude	Migratory distance (km)
Y	Adelie Penguin	*Pygoscelis adeliae*	*Sphenisciformes*	1			
Y	American Crow	*Corvus brachyrhynchos*	*Passeriformes*	2	M	52.728282	1449.675551
Y	American Flamingo	*Phoenicopterus ruber*	*Phoenicopteriformes*	1			
Y	Anna's Hummingbird	*Calypte anna*	*Trochiliformes*	0	R		
N	Atlantic canary	*Serinus canaria*	*Passeriformes*	0	R		
Y	Bald Eagle	*Haliaeetus leucocephalus*	*Accipitriformes*	1		57.253298	2196.110047
Y	Bar-tailed Trogon	*Apaloderma vittatum*	*Trogoniformes*	0	R		
Y	Barn Owl	*Tyto alba*	*Strigiformes*	0	R		
N	Blackcap	*Sylvia atricapilla*	*Passeriformes*	2	M	44.402317	3365.931805
N	Blue fronted amazon	*Amazona aestiva*	*Psittaciformes*	0	R		
Y	Brown Mesite	*Mesitornis unicolor*	*Mesithorniformes*	0	R		
Y	Carmine Bee-eater	*Merops nubicoides*	*Coraciiformes*	2	M	-15.601757	1843.256656
Y	Chicken	*Gallus gallus*	*Galliformes*	0	R		
Y	Chimney Swift	*Chaetura pelagica*	*Apodiformes*	2	M	40.90111	5212.796889
Y	Chuck-will's-widow	*Antrostomus carolinensis*	*Caprimulgiformes*	2	M	34.275962	2244.239074
Y	Collared flycatcher	*Ficedulla albicolis*	*Passeriformes*	2	M	50.009516	7141.161536
Y	Common Cuckoo	*Cuculus canorus*	*Cuculiformes*	2	M	50.70067	8342.20431
Y	Common Ostrich	*Struthio camelus*	*Struthioniformes*	0	R		
Y	Common Starling	*Sturnus vulgaris*	*Passeriformes*	2	M		
Y	Crested Ibis	*Nipponia nippon*	*Ciconiformes*	0	R		
Y	Cuckoo Roller	*Leptosomus discolor*	*Leptosiformes (Coraciiformes)*	0	R		
Y	Dalmatian Pelican	*Pelecanus crispus*	*Pelecaniformes*	1		49.397822	2430.699453
Y	Downy Woodpecker	*Picoides pubescens*	*Piciformes*	0	R		
Y	Emperor Penguin	*Aptenodytes forsteri*	*Sphenisciformes*	1			
N	Golden eagle	*Aquila chrysaetos*	*Accipitriformes*	1		60.715493	2506.946417
Y	Golden-collared Manakin	*Manacus vitellinus*	*Passeriformes*	0	R		
Y	Great Cormorant	*Phalacrocorax carbo*	*Suliformes*	1		48.993437	5823.882464
Y	Great tit	*Parus major*	*Passeriformes*	0	R		
Y	Great-crested Grebe	*Podiceps cristatus*	*Podicipediformes*	1		49.092926	4430.38658
N	Greeenish Warbler	*Phylloscopus trochiloides*	*Passeriformes*	2	M	44.402317	3365.931805
Y	Grey-crowned Crane	*Balearica regulorum*	*Gruiformes*	0	R		
Y	Ground tit	*Pseudopodoceps humilis*	*Passeriformes*	0	R		
Y	Hoatzin	*Opisthocomus hoazin*	*Opisthocomiformes*	0	R		
Y	Hooded crow	*Corvus cornix*	*Passeriformes*	1		63.443722	3623.510217
N	Japanese quail	*Coturnix japonica*	*Galliformes*	0	R		
Y	Kea	*Nestor notabilis*	*Psittaciformes*	0	R		
Y	Killdeer	*Charadrius vociferus*	*Charadriifores*	1		51.734141	4741.388576
Y	Kiwi bird	*Apterix australis mantelli*	*Apterigiformes*	0	R		
Y	Little Egret	*Egretta garzetta*	*Pelecaniformes*	1			
Y	MacQueen's Bustard	*Chlamydotis macqueenii*	*Otidiformes*	1		43.625066	1728.089995
Y	Medium Ground-finch	*Geospiza fortis*	*Passeriformes*	0	R		
N	Northern bobwhite	*Colinus virginianus*	*Galliformes*	0	R		
Y	Northern Fulmar	*Fulmarus glacialis*	*Procellariformes*	1			
Y	Peking Duck	*Anas platyrhynchos*	*Anseriformes*	1		55.396228	3447.874244
Y	Peregrine Falcon	*Falco peregrinus*	*Falconiformes*	1		60.833053	7458.828129
Y	Pigeon	*Columba livia*	*Columbiformes*	1			
N	Puerto-Rico-Amazone	*Amazona vittata*	*Psittaciformes*	0	R		
Y	Red-crested Turaco	*Tauraco erythrolophus*	*Musophagiformes*	0	R		
Y	Red-legged Seriema	*Cariama cristata*	*Cariaformes*	0	R		
Y	Red-throated Loon	*Gavia stellata*	*Gaviiformes*	2	M	62.164326	2334.642332
Y	Rhinoceros Hornbill	*Buceros rhinoceros*	*Bucerotiformes*	0	R		
Y	Rifleman	*Acanthisitta chloris*	*Passeriformes*	0	R		
N	Ruff	*Calidris pugnax*	*Charadriifores*	2	M	60.324554	7905
Y	Saker falcon	*Falco cherrug*	*Falconiformes*	1		48.885053	3974.464717
Y	Scarlet Macaw	*Ara macao*	*Psittaciformes*	0	R		
Y	Silvereye	*Zosterops lateralis*	*Passeriformes*	2	M		
Y	Speckled Mousebird	*Colius striatus*	*Coliiformes*	0	R		
N	European Stonechat	*Saxicola rubicola*	*Passeriformes*	2	M	44.402317	3365.931805
Y	Sunbittern	*Eurypyga helias*	*Eurypygiformes*	0	R		
N	Swainsons thrush	*Cathartus ustulatus*	*Passeriformes*	2	M	44.402317	3365.931805
Y	Swan goose	*Anser cygnoides*	*Anseriformes*	1		49.80434	2067.017276
Y	Turkey	*Meleagris gallopavo*	*Galliformes*	0	R		
N	Turkey Vulture	*Cathartes aura*	*Cathartiformes*	2	M	41.715309	6405.52089
Y	White-throated Sparrow	*Zonotrichia albicollis*	*Passeriformes*	2	M	53.993132	2099.292521
Y	White-tailed Eagle	*Haliaeetus albicilla*	*Accipitriformes*	1		55.516736	4917.947226
Y	White-tailed Tropicbird	*Phaeton lepturus*	*Pelecaniformes*	1		-15.786057	11530.04737
Y	White-throated Tinamou	*Tinamus guttatus*	*Tinamiformes*	0	R		
N	Willow warblers	*Phylloscopus trochilus*	*Passeriformes*	2	M	44.402317	3365.931805
Y	Yellow-throated Sandgrouse	*Pterocles gutturalis*	*Pterocidiformes*	1			
Y	Zebra Finch	*Taeniopygia guttata*	*Passeriformes*	0	R		

Table S2. Number of bird species per candidate gene for which we were able to obtain full sequence information to be included in respective analyses.

Gene	Number of spp.
AANAT	55
ADCYAP1	37
ARNTL	63
CLOCK	61
CPNE4	60
CREB1	67
CRY1	64
CRY2	61
CSNK1E	55
DRD4	61
HRSP12	63
HSP90B1	68
HSPA5	68
HSPA8	63
HSPA90AA1	63
NEK2	66
NFIL3	68
NPAS	58
PARL	66
PER2	60
PER3	67
SLC1A3	53
SLC2A1	41
TTR	22
YPEL1	48

Table S3. Welch t test and F-test for intra- and inter specific comparison on the polymorphic *CLOCK* locus. Comparisons of *CLOCK* gene variability among and between migratory and non-migratory species.

	t-test			F-test			
Comparison	**t value**	**df**	**p**	**F value**	**df num**	**df den**	**p**
Migratory *vs* non migratory	-69,779	4733,3	<0.0001	6,1509	3579	3503	<0.0001
Great tit *vs* Blue tit	-86,2	2646,3	<0.0001	5,5388	1895	1607	<0.0001
WW *vs* Nightingale	7,5993	475,1	<0.0001	1,4549	301	753	<0.0001
Flycatcher *vs* Bluetit	-8,0875	834,3	<0.0001	0,62122	451	1895	<0.0001
Fly *vs* Great	-65,692	526,71	<0.0001	3,4408	451	1607	<0.0001
WW *vs* Chiff abi	-2,7165	69,38	0,008323	1,2333	61	753	0,2303
WW *vs* Chiff tri	0,84535	54,38	0,4016	1,2378	49	753	0.2646
Chiff tri *vs* Night	-4,2575	69,497	<0.0001	0,85082	49	301	0,5007
Chiff abi *vs* Night	-6,3066	93,018	<0.0001	0,8477	61	301	0,4404
Night *vs* Blue	-20,158	402,62	<0.0001	1,0026	301	1865	0,9597
Great tit *vs* Whinchat	1,813	476,4	0,07046	0,27898	1607	415	<0.0001
Chiff abi *vs* Chiff tri	-1,2265	104,95	0,2227	0,99633	61	49	0,9813

CHAPTER 4

The evolutionary history and genomics of European blackcap migration

Kira E Delmore[1,2]*, Juan Carlos Illera[3], Javier Pérez-Tris[4], Gernot Segelbacher[5], Juan S. Lugo Ramos[1], Gillian Durieux[1], Jun Ishigohoka[1], Miriam Liedvogel[1]*

* Corresponding authors (delmore@evolbio.mpg.de and liedvogel@evolbio.mpg.de)

[1]Behavioural Genomics, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany

[2]Current address: Department of Biology, Texas A&M University, 77843 College Station, Texas, US

[3]Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, 33600 Mieres, Spain

[4]Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, 28040 Spain

[5]Wildlife Ecology and Management, University Freiburg, 79106 Freiburg, Germany

Contributions: I performed the gene prediction and functional annotation of the Eurasian blackcap (*Sylvia atricapilla*) genome. I also contributed with analysis of SFS for the standing variation selection. I contributed review and editing of the final draft.

Publication: K. Delmore, JC Illera, J Perez-Tris, G. Sechelbacher, JS Lugo Ramos, J Ishigohoka, M. Liedvogel, "The evolutionary history and genomics of European blackcap migration" Elife, vol. 9, p. e54462, Apr. 2020.

RESEARCH ARTICLE

The evolutionary history and genomics of European blackcap migration

Kira Delmore[1†]*, Juan Carlos Illera[2], Javier Pérez-Tris[3], Gernot Segelbacher[4], Juan S Lugo Ramos[1], Gillian Durieux[1], Jun Ishigohoka[1], Miriam Liedvogel[1]*

[1]Behavioural Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany; [2]Research Unit of Biodiversity (UO-CSIC-PA), Oviedo University, Mieres, Spain; [3]Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, Madrid, Spain; [4]Wildlife Ecology and Management, University Freiburg, Freiburg, Germany

Abstract Seasonal migration is a taxonomically widespread behaviour that integrates across many traits. The European blackcap exhibits enormous variation in migration and is renowned for research on its evolution and genetic basis. We assembled a reference genome for blackcaps and obtained whole genome resequencing data from individuals across its breeding range. Analyses of population structure and demography suggested divergence began ~30,000 ya, with evidence for one admixture event between migrant and resident continent birds ~5000 ya. The propensity to migrate, orientation and distance of migration all map to a small number of genomic regions that do not overlap with results from other species, suggesting that there are multiple ways to generate variation in migration. Strongly associated single nucleotide polymorphisms (SNPs) were located in regulatory regions of candidate genes that may serve as major regulators of the migratory syndrome. Evidence for selection on shared variation was documented, providing a mechanism by which rapid changes may evolve.

***For correspondence:** delmore@evolbio.mpg.de (KD); liedvogel@evolbio.mpg.de (ML)

Present address: †Department of Biology, Texas A&M University, College Station, United States

Competing interests: The authors declare that no competing interests exist.

Funding: See page 19

Received: 16 December 2019
Accepted: 13 March 2020
Published: 21 April 2020

Reviewing editor: Elizabeth Scordato, California State Polytechnic University, United States

Introduction

Bird migration is a fascinating and highly variable behaviour that integrates many traits – morphological, physiological and behavioural. Research on a wide range of species has provided important insight into this behaviour, from the incredible distances that birds cover during their journeys (***Alerstam et al., 2003***; ***Egevang et al., 2010***), to the fine-tuned and precisely controlled timing of migration (***Gwinner and Helm, 2003***) and the fascinating sensory modalities that allow birds to navigate with amazing precision (***Mouritsen, 2018***; ***Wiltschko and Wiltschko, 1972***). The European blackcap *Sylvia atricapilla* is an iconic migratory species that is well suited to work on the genetics of migration. Blackcaps exhibit dramatic differences in migratory behaviour (***Figure 1a***), spanning the entire spectrum from exclusively migratory populations in the northern portion of their range to short distance and partially migratory populations in the Mediterranean; non-migratory, or resident, populations occur on both the European continent (Iberian Peninsula) and the Atlantic islands. In addition to variation in the propensity to migrate and the distance covered, blackcaps vary in migratory orientation, with a migratory divide (contact zone between populations that breed adjacent to one another but take different migratory routes) occurring between populations that migrate southwest (SW) and southeast (SE) from their breeding grounds in Central Europe in the autumn. A novel migratory route also evolved very recently in this species, with an increasing number of birds migrating northwest (NW) from the Central European breeding grounds in the autumn (***Cramp, 1992***).

Variation in the migratory behaviour of European blackcaps was harnessed in a series of influential papers published in the 1980s and 1990s that detailed the genetic basis of migration. Common garden experiments showed that selectively bred individuals that were reared in isolation from their

eLife digest Every year as the seasons change, thousands of animals migrate huge distances in search of food or better climates. As far as migrations go, there might be none so impressive as the trans-oceanic flights made by small migrating songbirds. These birds can weigh as little as three grams and travel up to 15,000 kilometres. Most migrate alone and at night and yet still manage to return to the same location each year. Several strands of research suggest there could be a genetic basis to their migratory behaviour, but exactly which genes control this phenomenon remains poorly understood.

One small songbird that has been studied for decades is the European blackcap. This species exhibits a real variety of migration patterns. Some blackcaps travel rather short distances, others much further, and some populations do not migrate at all. Populations that share the same breeding grounds in the summer may migrate in different directions in the autumn. These features make it a good species to study the genetic variation between populations that migrate in different directions and over different distances. However, only in recent years has advancing technology made it possible to comprehensively study an animal's entire genome, leaving no gene unturned.

Now, Delmore et al. have used high-throughput sequencing technologies to trace the evolutionary history of migration in European blackcap and started by assembling a reference genome for the species. Then, the genomes of 110 blackcaps from several populations that take different annual migrations were compared to the reference. This revealed that the populations began to diverge some 30,000 years ago and that there was some apparent gene mixing between groups of migrating and resident blackcaps around 5,000 years ago. The analysis showed only a small set of genes code for their differences in migration. Additionally, while the candidate genes were shown to be common among blackcaps, the genes identified did not match those reported from studies of other migrating songbirds. Finally, Delmore et al. also noted that the differences between the populations tend to be in the parts of the genome that control whether a given gene is switched on or off, which could explain how new migratory behaviours can rapidly evolve.

This study is one of the most comprehensive genomic analysis of migration to date. It is important work as songbirds, like other animals, are responding to increasing pressures of environmental and climate change. In time, the findings could be used to support conservation efforts whereby genetic analyses could determine if certain populations possess enough variation to respond to coming changes in their habitats.

parents maintain population-specific behaviour, suggesting that there is a genetic component to migration (***Helbig, 1994***; ***Helbig et al., 1989***; ***Helbig, 1991a***; ***Berthold and Querner, 1981***; ***Pulido and Berthold, 2010***). F_1 hybrids crossbred between populations that differ in migratory traits exhibited intermediate phenotypes (orientation, distance and propensity to migrate), suggesting that these traits are additively inherited (***Berthold and Querner, 1981***; ***Pulido and Berthold, 2010***; ***Helbig, 1991b***). Further work with F_2 hybrids showed a wider distribution of phenotypes and the recovery of parental phenotypes, indicative of traits that are controlled by only a few major genes (***Helbig, 1996***), and selection experiments mating birds according to migratory status showed that the transition between resident and migratory behaviour can occur in just a few generations (***Pulido et al., 1996***; ***Berthold et al., 1990***). The rapidity with which migratory behaviour can evolve has been supported in natural populations; the NW route taken by some birds was only established in the past 70 years and probably in response to increased food availability during the winter in the United Kingdom (***Berthold et al., 1992***).

The blackcap has also been the subject of extensive phylogeographic study. ***Pérez-Tris et al. (2004)*** used mitochondrial (control) data from 241 birds and 12 populations across the entire breeding range to show that migratory variation in this species arose recently (4,000–13,000 years ago [ya]) and has not yet resulted in significant population differentiation. These results could suggest that the genes that control migratory variation have small effect sizes or are restricted to a small portion of the genome. The only populations showing substantial genetic differentiation occurred in the Central European migratory divide (i.e., between SW and SE migrants), indicating that differences in orientation may help to maintain population differentiation. Resident populations showed evidence

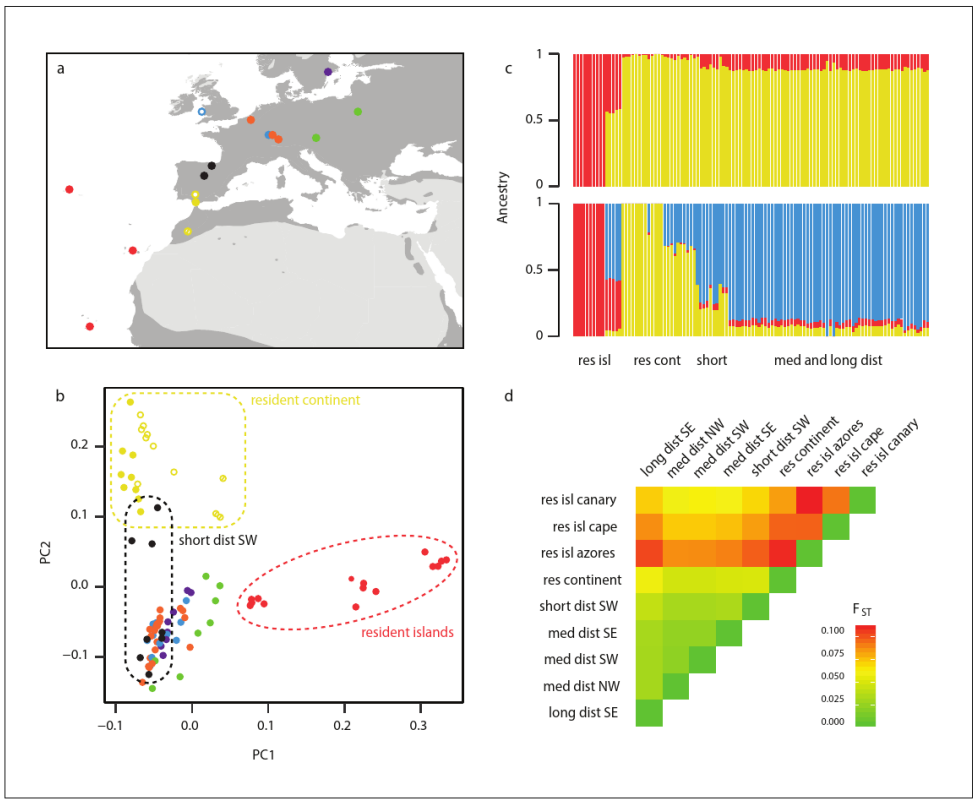


Figure 1. Sampling design and population structure. (**a**) Sampling sites and migratory phenotypes. Samples were collected from the breeding grounds, except for a subset of NW migrants that were sampled during winter in the UK (open blue circle) (details in ***Supplementary file 4***). (**b–d**) Population structure represented by a Principle Component Analysis (PCA) (b), NGSadmix (K = 2 and 3 shown) (c) and pairwise estimates of F_{ST} (d), showing differentiation between migrants and residents (as well as among residents themselves). Long dist SE = long distance migrants that orient SE in autumn (purple), med dist = medium distance migrants that orient in the corresponding heading during autumn migration (SE = green, SW = orange and NW = blue), res continent = residents found on the continent (yellow), short dist SW = short distance migrants that orient SW (black), res isl = resident birds on islands (cape = Cape Verde, canary = Canary Islands). Among continental residents, open circles indicate Cazalla de la Sierra, open circles with dash Asni, and filled circles Gibraltar. A PCA excluding islands can be found in ***Figure 1—figure supplement 1***; results from NGSadmix at larger values of K can be found in ***Figure 1—figure supplement 2***.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Principal component analysis matching that in ***Figure 1*** but excluding island populations.

Figure supplement 2. Complimentary figure to ***Figure 1c***, showing ancestry proportions estimated by ADMIXTURE at larger cluster values (k = 4 through 7).

of historical bottlenecks followed by sudden expansions, suggesting that blackcaps lost their ability to migrate after secondary colonization of mild areas in southern Europe and on the Atlantic islands. This finding was supported by ***Voelker and Light (2011)*** who used mitochondrial (ND2 and cytb) data to reconstruct ancestral states within the genus *Sylvia*. Limited genetic differentiation between blackcaps was also documented by ***Dietzen et al. (2008)*** using mitochondrial (cytb) data; these authors also estimated dates for the colonization of Atlantic islands and for an earlier colonisation of the Canary Islands (the latter occurring 300,000–3,000,000 ya vs. 4,000–40,000 ya for colonisation of other islands including the Azores and Cape Verde).

The phylogeographic studies described above provided important insight into the evolution of migration in blackcaps and other temperate species more generally. When these studies and experimental work are considered together, the thorough set of studies conducted on blackcap migration is arguably unequalled in other species. Surprisingly, these classic experiments and molecular marker-based approaches have been followed by a dearth of genetic work on migration in

blackcaps. Here, we leverage our knowledge of this excellent study system by using high-throughput sequencing techniques to provide the first genome-wide characterization of the blackcap. The major objectives of this study were to assemble a high-quality reference genome de novo, and to use whole genome resequencing data from 110 blackcaps (including birds from each migratory phenotype and encompassing the entire breeding range, ***Figure 1a***) to (1) examine population structure and demography in this system, and (2) study the genetic basis of three migratory traits in unison: the propensity to migrate, migratory distance and orientation.

Analyses of population structure and demography revealed novel insights that are important for understanding both the evolutionary history of migration in blackcaps and the underlying genetics of this behaviour. A small number of studies on the genomics of migration have been conducted in songbirds (***Delmore et al., 2016***; ***Lundberg et al., 2017***). We compare our results to theirs, evaluating a long held hypothesis of a common genetic basis to migratory behaviour (***Liedvogel et al., 2011***). Our results are not only relevant to understanding the genetics of migration in the blackcap. Data on the genetics of complex behaviours is at a premium in the evolutionary literature, which has focused primarily on morphological traits, and migration probably plays an important role in the early stages of speciation in many systems. Our results will speak to the genetic basis of this process.

Results and discussion

Assembly of a high-quality draft reference genome

We used whole genome sequencing (WGS) data and an optical map (Illumina and Bionano Irys technology, ***Supplementary files 1–3***) to de novo assemble a hybrid reference genome for the blackcap (BioProject number PRJNA545868; Guojie Zhang, personal communication). The final genome is 1.02 Gb in length, comprised of only 96 scaffolds and has a large N_{50} scaffold length of 22 Mb. Ninety scaffolds mapped to the collared flycatcher *Ficedula albicollis* genome (average three scaffolds/chromosome; ***Supplementary file 3***) and our annotation strategy, which used both in silico and evidence-based approaches, identified 17,982 protein-coding genes. Results from BUSCO and an analysis of UCEs (ultra-conserved elements) suggest that our reference is nearly complete, with 92% of single-copy orthologues unique to birds and 97% of UCEs identified in amniotes (***Faircloth et al., 2012***; ***Supplementary file 2***).

Population structure and demography

We aligned WGS data from 110 blackcaps (including the two birds used in our assembly) to this reference (average coverage 17.5x, ***Figure 1a***, ***Supplementary file 4***) and estimated genotype likelihoods at genome-wide single nucleotide polymorphisms (SNPs). Genomic differentiation was low between migratory populations of different distances and orientations, but unlike earlier work using mitochondrial data (***Pérez-Tris et al., 2004***; ***Dietzen et al., 2008***), we documented considerable differentiation between migrant and resident populations on both the continent and islands. A PCA separated resident island birds from continental populations on PC1, and resident continental birds from migrants on PC2 (***Figure 1b***). Migrants were not clearly distinguished on either PC (we obtained the same result when we re-ran the PCA excluding islands; ***Figure 1—figure supplement 1***). Results from an ADMIXTURE analysis and estimates of F_{ST} confirm this pattern. At a cluster value of two, ADMIXTURE distinguished between island and continental birds (similar to PC1). At a cluster value of three, populations on the continent were further divided into resident and migratory groups (similar to PC2), and resident island and continent birds showed some admixture with the migratory group (***Figure 1c***). No further structure was observed beyond these three clusters (***Figure 1—figure supplement 2***). Estimates of F_{ST} ranged from 0.018 to 0.11, with the highest estimates occurring between migrants and both resident groups (0.06–0.11 for islands, 0.042–0.05 for continent residents; ***Figure 1d***). Evidence for limited population differentiation combined with dramatic differences in the migratory behaviour of blackcaps is ideal for identifying genomic regions that are associated with this focal trait. Specifically, genomic regions associated with migration should standout against this backdrop of limited differentiation, although analyses involving residents will need to account for elevated values of differentiation.

A phylogeographic analysis using mitochondrial data suggested that variation in migratory behaviour evolved recently, 4000–13,000 ya (**Pérez-Tris et al., 2004**). Our results move this date further back in time, to 30,000 ya and the start of the last glacial maximum (**Clark et al., 2009**). Specifically, we used multiple sequentially Markovian coalescent (MSMC, implemented in MSMC2) (**Schiffels and Durbin, 2014**; **Malaspinas et al., 2016**) to characterize the demographic history of blackcaps. The demographic trajectories of migratory, resident continent and resident island birds began to diverge ~30,000 ya. The effective population size of migrant and resident island populations expanded and contracted, respectively, while continental residents exhibited a relatively constant effective population size (**Figure 2a**; **Figure 2—figure supplement 1**). Relative cross-coalescence rates (CCR) between all three groups exhibited a concomitant drop ~30,000 ya (**Figure 2b**; **Figure 2—figure supplement 2**). The drop of relative CCR between migratory and resident island populations was steeper than that between migratory and resident continent populations (**Figure 2b**), suggesting that genetic separation following the colonization of islands resulted in greater separation than that between continental populations of migrants and residents. Increased differentiation between migrants and resident island birds (vs. resident continent birds) was also documented in our PCA (**Figure 1**). Results for medium-distance migrants (NW, SW and SE) and long-distance migrants are indistinguishable (**Figure 2—figure supplement 4**; **Figure 2—figure supplement 5**; **Figure 2—figure supplement 6**).

One interesting finding from our demographic analyses is that of apparent gene flow between migrant and resident continent birds ~5000 years ago. Specifically, the relative CCR between migrant and resident continent populations started to increase at ~5000 ya (~25,000 years after initial divergence; **Figure 2—figure supplement 3**). This admixture event may reflect secondary contact between migrant and resident continent populations and is line with our results from ADMIXUTRE, with admixture documented between these groups at a cluster value of three. The last glacial maximum ended ~19,000–11,500 ya (**Clark et al., 2009**). After this time, populations would have expanded out of their glacial refugia, and perhaps migrant and resident continent populations came into secondary contact ~5000 years after these expansions began. Similar to our results on population differentiation, island populations exhibit their own evolutionary trajectories following divergence. This result is in line with results from **Dietzen et al. (2008)**, who suggested that at least one separate colonization to the Atlantic islands occurred (earlier than that to the Canaries).

Genetic basis of migratory traits

Here, we transition to study local patterns of genomic differentiation, identifying specific genomic regions that have signatures of selection related to three phenotypes: the propensity to migrate, orientation of migration and distance of migration (resident continent, short distance SW, medium distance NW, SW, SE and long distance SE populations). We excluded resident island birds from these analyses (because of limited sample size [n = 5] and potential effects from founder events) and focused on a single resident continent population (Gibraltar, we obtained similar results using Cazalla de la Sierra, total number of birds included in these analyses = 82, **Supplementary file 4**).

Positive selection was more common in residents and limited to a few, small genomic regions (**Table 1a**). For example, hapFLK is a tree-based method that controls for hierarchical population structure. Global and local NJ trees are constructed using haplotype frequencies and regions under selection show longer branch lengths. Only nine regions were found to be under selection (permutation test, see 'Materials and methods') according to this method, and six of these appeared in residents. **Figure 3a** shows estimates of hapFLK for the entire genome, and **Figure 4a** exemplifies one region on Super-Scaffold 99 (syntenic with flycatcher chromosome 3). The average size of these regions was 16.7 kb and only six genes occurred within them. We used CAVIAR (**Rochus et al., 2018**) to identify variants showing strong associations with selection in these regions. Each region included one to four variants, all of which occurred in non-coding regions (**Supplementary file 5**). Previous phylogeographic work suggested that migration is the ancestral state in blackcaps (**Pérez-Tris et al., 2004**; **Voelker and Light, 2011**). Accordingly, the selection in genomic regions that we identified here is probably involved in the transition from migrant to resident phenotypes.

We complemented results from hapFLK with a modified version of the Population Branch Statistics (PBS) (**Yi et al., 2010**) and the number of segregating sites by length (nSL) (**Ferrer-Admetlla et al., 2014**). PBS is an F_{ST}-based statistic that estimates allele frequency differences between three or more populations. This parameter can be elevated by linked purifying selection (or

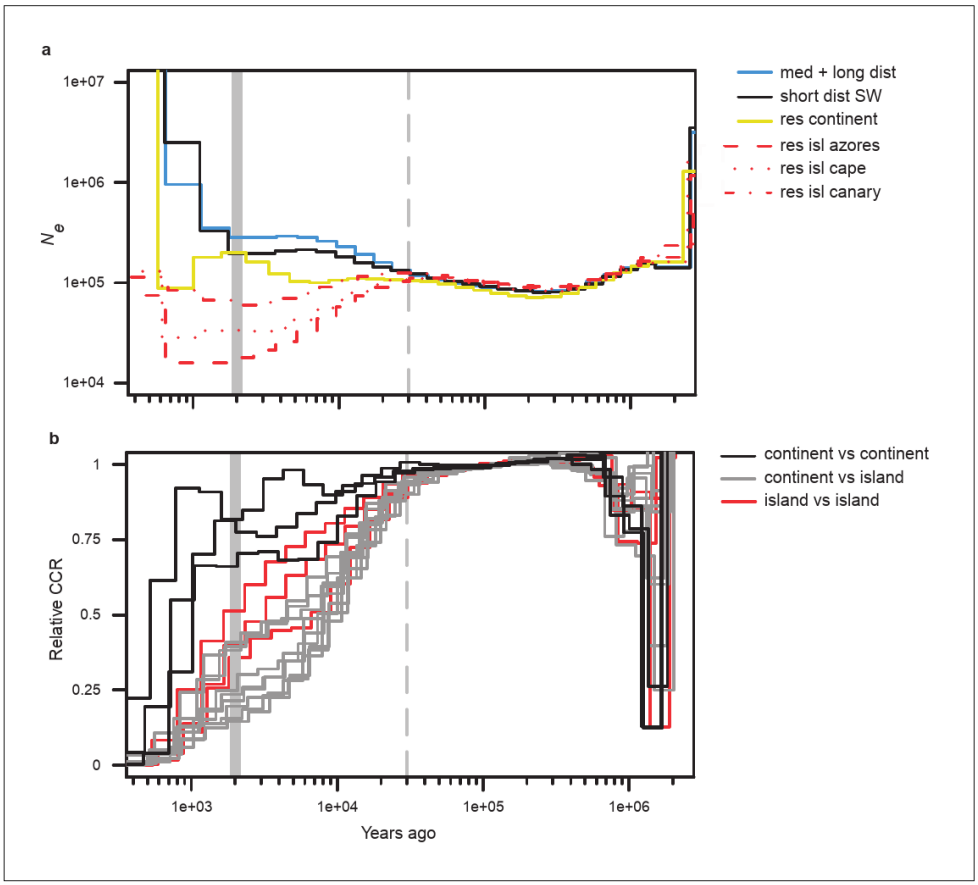


Figure 2. Demographic history. (**a**) Effective population size by time estimated by MSMC2 using five individuals per blackcap phenotype. Note that the most recent time segment is regarded as being unreliable in MSMC2 results. (**b**) Relative cross-coalescence rate estimated by MSMC2. 15 lines with three colours indicate relative cross-coalescence rate for all pairwise combinations of the six populations (three for comparisons between populations on the continent [continent vs. continent], three for comparisons between populations on the islands [island vs. island], and nine for comparisons between continent and island populations [continent vs. island]). The dotted vertical line indicates the inferred time of population separation. Results from down-sampling can be found in ***Figure 2—figure supplements 1***, ***2*** and ***3***; results for medium- and long-distance migrants run separately can be found in ***Figure 2—figure supplements 4***, ***5*** and ***6***.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Down-sampling for demography analysis of effective population size.
Figure supplement 2. Down-sampling for demography analysis of relative cross-coalescence rate.
Figure supplement 3. Demography analysis of relative cross-coalescence rate.
Figure supplement 4. Medium distance NW, SW and SE migrants and long distance migrants show similar demographic histories.
Figure supplement 5. Medium distance NW, SW and SE migrants show similar demographic histories.
Figure supplement 6. Medium distance NW, SW and SE migrants show similar demographic histories.

background selection) within populations that is unrelated to positive selection (in our case selection related to migration). We removed these confounding effects by scaling PBS and subtracting the maximum value of PBS in orthologous windows from that in the non-focal population (hereafter 'ΔPBS', following ***Vijay et al., 2017***). nSL is a haplotype-based statistic that focuses on patterns within populations, using segregating sites to measure the length of haplotypes. Linked selection should increase haplotype lengths at genomic regions that are under positive selection. Eight of the nine regions identified by hapFLK also exhibited extreme values of ΔPBS and nSL (in the top 1% of the distribution) in the same population as that identified by hapFLK (ΔPBS ***Table 1a***, ***Figure 3b*** for resident birds [estimates for short distance SW, medium distance SE, SW, NW, and long distance

Table 1. Genetic variants underlying variation in migration.

(a) Results from analyses including all continental birds and (b) results from analyses limited to medium-distance migrants. Results from hapFLK include the size, the population where the signal was found and genes within the region. Estimates of ΔPBS and (PBS) in the same regions are shown; they are bolded if in the top 1% of the focal population's distribution and new sizes are estimated using neighbouring windows above this threshold (if larger than the limits from hapFLK, additional genes are specified). Estimates of PBS were re-estimated using island populations (vs.continent resident populations). Regions in the top 1% of an island population's distribution are indicated in section (a) (recorded as 'NA' if the initial population under selection was not resident). 'Scaf' refers to the scaffold within the blackcap genome where the region is found and 'chr' refers to the flycatcher chromosome that these scaffolds map to. For the number of strongly associated SNPs identified by CAVIAR and estimates of nSL, see ***Supplementary file 5***.

(a)

		hapFLK				Δpbs			
Scaf	Chr	Size (bp)	Log p-value	Population	Genes	Size (Mb)	ΔPBS (PBS)	Island replacement	Genes
12	4A	14,059	9.4	Resident	LOC100859173	52	**18.7 (0.40)**	Azores	EDA2R
13	11	29,195	8.3	Resident	CHST4, TERF2IP, KARS	303	**41.0 (0.87)**	Cape Verde	DHX38, DHODH, IST1, C2H2, ATXN1, AP1G1, PHLPP2, TAT, GABARAPL2, TMEM231, CHST6
17	3	7610	9.5	Short SW		316.5	**0 (0.02)**	NA	NKAIN1
22	9	53,890	8.8	Med SE	CLSTN2	1,005.5	**21.9 (0.19)**	NA	DUF4637, PIK3CB, FOXL2, MRPS22, COPB2, RBP2, NMNAT3
30	2	13,756	11.5	Resident		42.5	**8.14 (0.19)**	Cape Verde	
30	2	7902	8.8	Resident		1,029.5	**19.1 (0.42)**	Canaries, Cape Verde	
41	8	10,341	8.3	Resident		11.5	**15.0 (0.33)**		
46	1A	412	7.9	Med SE		9.5	**9.0 (0.03)**	NA	
99	3	13,140	7.8	Resident	TTBK1	192	**28.6 (0.61)**	Azores, Canaries, Cape Verde	LOC101820716, ACSS1, NEIL1, SLC22A7, TTL

(b)

		hapFLK				Δpbs	
Scaf	Chr	Size (bp)	Log p-value	Population	Genes	Size (Mb)	ΔPBS (PBS)
17	3	3258	9.04	Med NW	SDC1	5	14.49 (0.20)
30	2	311	8.85	Med NW		7	11.31 (0.16)
46	1A	461	8.71	Med NW		3	8.14 (0.15)
63	1A	1088	9.55	Med SE			1.14 (0.05)
67	6	995	9.46	Med SW		5	9.03 (0.18)
73	5	3611	11.81	Med NW	ATG2B, BDKRB2	3	30.41 (0.35)

SE, ***Figure 2—figure supplement 4***; ***Figure 2—figure supplement 5***; ***Figure 2—figure supplement 6***]; nSL ***Supplementary file 5***).

As noted already, population structure and linked selection can elevate differentiation between populations. We controlled for these effects using hapFLK and ΔPBS, respectively, and emphasise that genomic differentiation between populations of blackcaps is low to begin with (***Figure 1d***). In addition, linked purifying selection would be expected to increase PBS in all populations (i.e., not just the focal population), but this is not the case. This is exemplified in ***Figure 5a*** where estimates of ΔPBS for all populations are shown but these are only elevated in the resident continent population. As a final test of population structure, we re-estimated ΔPBS using resident island birds (instead of resident continent birds). We conservatively excluded these populations from our initial analyses because their sample sizes are small and because genetic drift can affect estimates of differentiation in island populations. Nevertheless, the island populations are also resident and thus these estimates could help to validate the genomic regions that were identified as being under selection in resident populations on the continent. ***Table 1a*** summarizes these results, noting which genomic regions

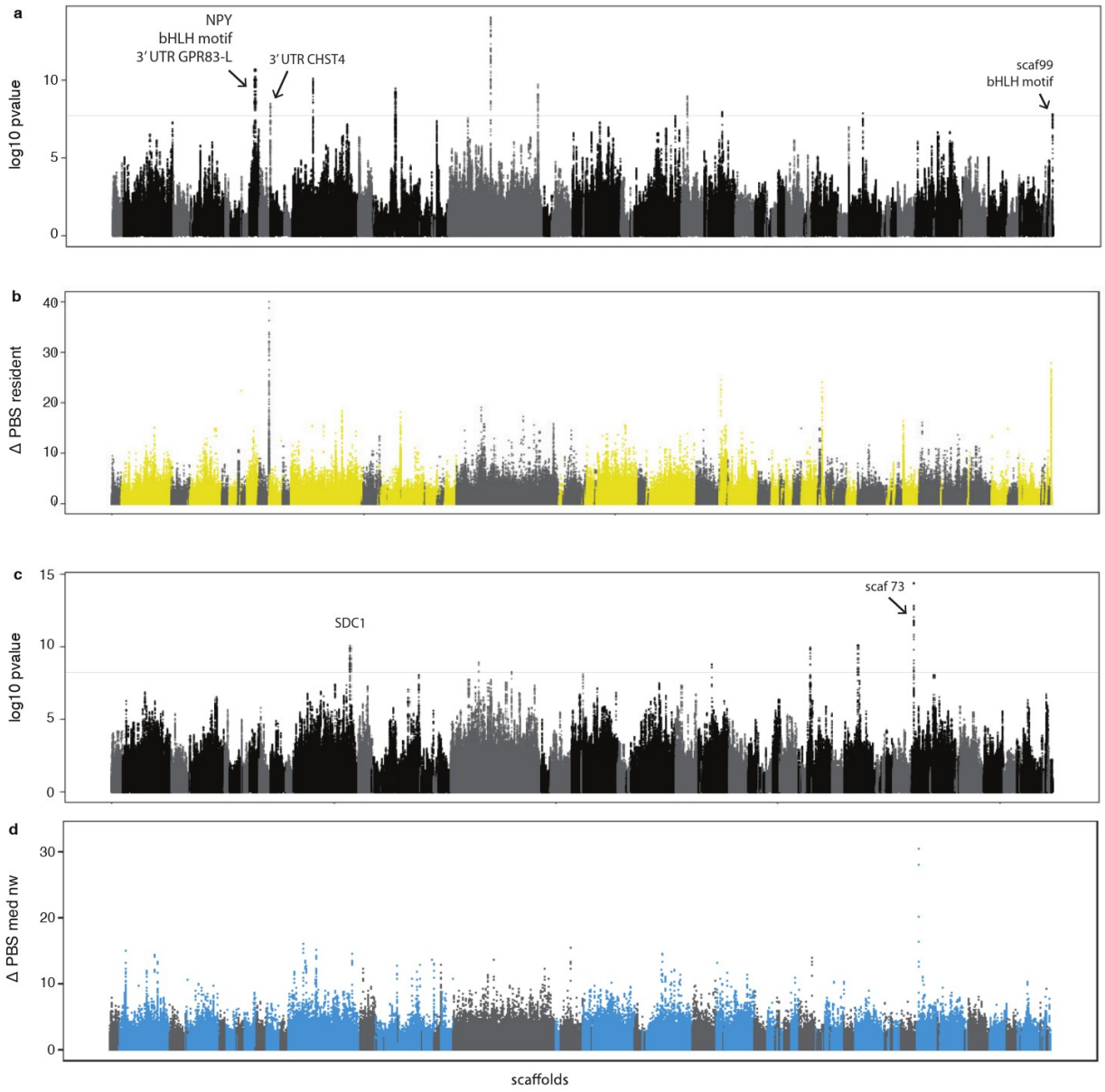


Figure 3. Genome-wide local estimates of population differentiation. Results from hapFLK using haplotype frequencies (**a**,**c**) and ΔPBS using SNP frequencies (**b**,**d**; 2,500 bp windows). Estimates of ΔPBS for resident continent (**b**) and medium-distance NW migrants (**d**) are shown; results for the remaining populations can be found in ***Figure 3—figure supplement 1***. Genetic elements, scaffolds and genes discussed in the text are highlighted. The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Genome-wide local estimates PBS for the remaining populations.

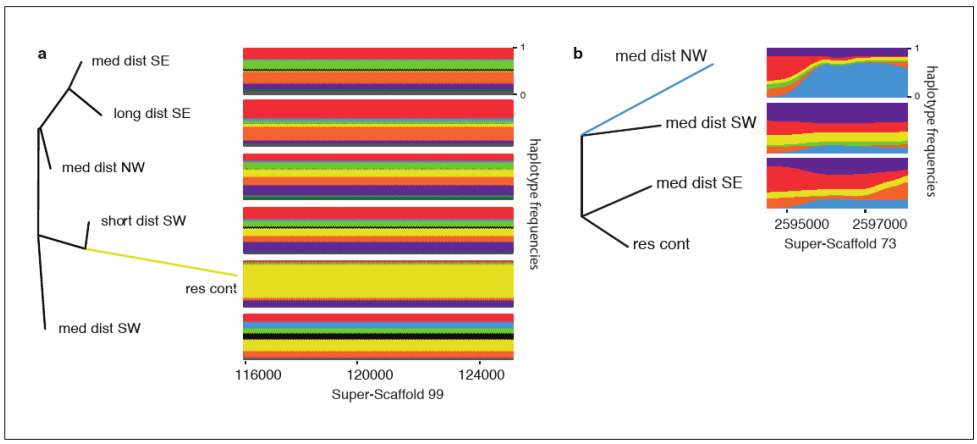


Figure 4. Exemplifying genomic regions under positive selection. Local neighbour joining trees for regions under selection in (**a**) the resident continent population on Super-Scaffold 99, and (**b**) medium-distance NW population on Super-Scaffold 73. Selection is indicated by longer branch lengths in each population than is the case in global trees built using data from all genomic regions (***Figure 4—figure supplement 1***). Panels to the right of the trees show the corresponding frequency of haplotypes in each population of the tree. Haplotype clusters are colour coded (colours of haplotype clusters do not correspond to the population colour coding used in other figures), and frequencies are plotted along the Y axis. Haplotype frequency plots show the near fixation of a single dominating haplotype in (**a**) resident continent (yellow) and (**b**) medium-distance NW populations (blue). The location (in bp) of these regions on each Super-Scaffold is shown below these panels and the resident continent group is only included to root the tree in panel (b), and thus has no haplotype frequencies.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Global neighbor joining trees built using hapFLK data and data from all genomic regions, for comparison with local trees showing positive selection in ***Figure 4***.

exhibited elevated values of PBS on islands. Of particular interest, PBS was elevated in all three island populations at the genomic region on Super-Scaffold 99 (***Figure 5b***). Combined with findings from hapFLK (controlling for population structure and relying on haplotypes), ΔPBS (controlling for linked selection and relying on SNP data) and nLS (estimated within populations and relying on haplotypes), these results provide strong evidence that this specific region contains important variation for the transition to residency, not only on the continent but also on the islands.

Note that it is possible that the signatures of positive selection that we document here reflect selection based on different ecological variables involved with the colonization of areas further south on the continent, but at least in the case of Super-Scaffold 99, we believe that this is rather unlikely as most ecological variables (biotic and abiotic) are quite distinct between islands and the continent (and between the islands themselves) (***Cropper, 2013***; ***Valente et al., 2017***). The transition to residency is shared, probably representing one of the only shared selection pressures experienced by all of these populations. Note that the lack of consistent results for other regions under selection in the resident continent population does not preclude the potential importance of these regions as, for example, genetic drift on islands would affect which genetic variants were present on islands for selection to act on.

Our finding that only a few genomic regions under selection contain genes and that the strongly associated SNPs identified by CAVIAR are in non-coding regions could suggest that cis-regulatory changes are important for the transition from migration to residency. In support of this suggestion, an alignment of predicted mRNAs from several bird species and transcripts from a testis transcriptome of the blackcap placed two of the SNPs from CAVIAR in the 3′ untranslated region (3′ UTRs) of two genes (GPR83-L on Super-Scaffold 12 and CHST4 on Super-Scaffold 13, syntenic with flycatcher chromosomes 11 and 4a, respectively). Three prime3′ UTRs can act as posttranscriptional regulators; they contain binding sides for microRNAs, which can inhibit translation or target mRNA for degradation (***Mayr, 2017***; ***Barrett et al., 2012***). In fact, previous work with monarch butterflies identified 55 conserved microRNAs that are differentially expressed between summer and migratory butterflies (***Zhan et al., 2011***). Future analyses to validate this suggestion could include the use of qPCR to

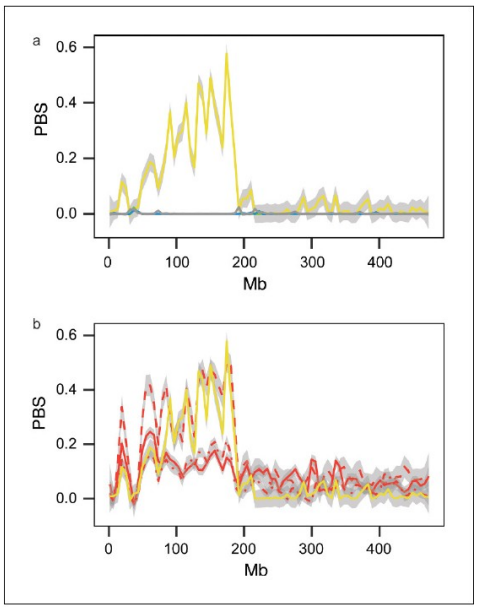


Figure 5. Estimates of ΔPBS on Super-Scaffold 99 corresponding with the region shown in ***Figure 4a*** (smoothed using the geom_smooth function in ggplot to summarize data in 2500-bp windows). (**a**) Estimates for resident continent, medium-distance NW, SW and SE migrants, and short- and long-distance birds. These estimates are only elevated in the resident continent phenotype, ruling out a role for linked selection in generating this signature in residents. (**b**) Estimates for the resident continent and island birds (Azores, Canaries and Cape Verde), which are all elevated, implying that parallel selection is probably involved in the transition from migration to residency in this region. Colours correspond to ***Figure 1a*** with yellow showing data for resident continent birds.

determine whether GPR83-L and/or CHST4 are in fact differentially regulated between the migratory phenotypes.

Future work using techniques aimed at identifying binding sites for transcription factors (e.g., ChIP-seq) could also be useful. We conducted a preliminary analysis here, using HOMER (***Heinz et al., 2010***) to detect known transcription factor motifs in the genomic regions that are under selection in residents. Specifically, ***Ruegg et al. (2014)*** used a literature search to identify 25 candidate genes for migration. Four of these genes are transcription factors whose motifs are in the libraries searched by HOMER: three basic helix-loop-helix transcription factors (bHLH) (*Clock*, *Npas2*, and *Bmal1*) and one basic leucine zipper domain (*Nfil3*). We found a bHLH motif (GHCACGTG) on Super-Scaffolds 12 and 99 (***Figures 3a,b***, ***4a*** and ***5***). The motif on Super-Scaffold 99 is particularly interesting as there is a SNP (G/T) at the beginning of the motif that is nearly fixed in continental residents (the allele frequency for G in Asni, Gibraltar and Cazalla de la Sierra is 1, 0.85 and 0.9, respectively; F_{ST} between Gibraltar and medium-distance NW, SW and SE migratory populations is 0.15, 0.25 and 0.44, respectively). This motif could disrupt or weaken transcription factor binding (***Kasowski et al., 2010***). This is also the genomic region that showed elevated PBS in both resident continent and island populations (***Figure 4a***, ***Figure 5***). *Clock*, *Npas2* and *Bmal1* are involved in maintaining circadian rhythms. Circadian rhythms synchronize circannual clocks, which are important cues controlling seasonal migratory behaviour (***Gwinner, 1996***; ***Visser et al., 2010***).

Concerning the actual identity of genes within regions that are under selection, several have functions that could be related to the transition from migration to residency. For example, LOC100859173 (located on in the genomic region under selection on Super-Scaffold 12, the region with a bHLH motif mentioned above; ***Table 1a***) has been annotated as a probable G-protein coupled receptor that mediates the function of neuropeptide Y (NPY). NPY is localized in the brain of birds and works with Agouti-related peptide (AGRP) and proopiomelanocortin (POMC) to control energy balance. Specifically, NPY/AGRP neurons stimulate appetite, food intake and fat deposition, while POMC inhibits these processes (***Boswell and Dunn, 2017***). It has been hypothesized that the effects of NPY may extend to seasonal changes in energy balance that are important for migration, including hyperphagia and fat deposition (***Boswell and Dunn, 2017***). Beyond its role in energy balance, NPY also facilitates learning and memory via the modulation of hippocampal activity and has an effect on circadian rhythms, reproduction, and the contraction of vascular smooth muscles. It has been suggested that a common genetic mechanism or major regulator may control migratory traits (***Liedvogel et al., 2011***; ***Liedvogel and Lundberg, 2014***). A protein such as NPY, or the transcription factors that bind the bHLH motif identified in the prior analysis, could fill this role.

Analysis focused on migratory orientation and distance

So far, we have considered all three migratory traits exhibited by blackcaps together (propensity, orientation and distance) and our results relate mostly to residents. The elevated population differentiation that we noted between resident and migratory birds could reduce our power to identify selection that is specific to migrants (**Fariello et al., 2013**). Accordingly, we ran a second set of analyses excluding resident birds and examining migratory orientation and distance independently. Starting with orientation and limiting our analysis to medium-distance migrants with varying orientations (medium-distance NW, SW and SE migrants, total number of birds included in these analyses = 54, **Supplementary file 4**), hapFLK identified only six regions that are under positive selection (**Table 1b**). Most of these regions showed selection in the NW phenotype and exhibited extreme values of ΔPBS limited to the population identified by hapFLK (**Figure 3c,d**). **Figure 4b** exemplifies results for hapFLK at one region under selection in the NW migrants (~4 kb on Super-Scaffold 73, syntenic with flycatcher chromosome 5). Results for nSL can be found in **Supplementary file 5**.

The list of genes in genomic regions that are under selection in this analysis focusing on orientation is small, but it also includes genes with functions that are strongly related to the phenotype they are associated with. For example, SDC1 is a region on Super-Scaffold_17 that is under selection in NW migrants. This gene codes for a transmembrane protein that helps to regulate the Wnt signalling pathway. This pathway plays a role in embryonic development and has been shown to influence feather and beak morphogenesis, along with feather molt (**Yu et al., 2004**; **Mallarino et al., 2011**; **Bhullar et al., 2015**; **Widelitz, 2008**). NW migrants have rounder wings and more narrow beaks than southern migrants (**Rolshausen et al., 2009**). Differences in the timing of migration probably mean that birds also molt at different times. This has not been evaluated directly in comparisons between migrants, but variation in molt patterns have been documented between NW migrants and birds that are resident on the continent (**de la Hera et al., 2009**).

Two previous studies attempted to identify de novo genomic regions under selection related to differences in orientation: **Delmore and Liedvogel (2016)** with Swainson's thrushes (*Catharus ustulatus*) and **Lundberg et al. (2017)** with willow warblers (*Phylloscopus trochilus*). **Delmore and Liedvogel (2016)** identified a region on chromosome 4 and **Lundberg et al. (2017)** regions on chromosome 1 and 5 that are associated with orientation. None of these regions overlap with those under selection in our study on blackcaps. It is tempting to suggest that migration may be controlled by similar genes across broad taxonomic scales, with early results from candidate genes (e.g., the poly-glutamine repeat in *Clock*) showing consistent results across groups as divergent as insects, fishes and birds (**Delmore and Liedvogel, 2016**). Nevertheless, several studies have failed to document an association with *Clock*, and a comparison of our results with those of **Delmore and Liedvogel (2016)** and **Lundberg et al. (2017)** adds further caution to this idea of a common basis (at least at the sequence level). This is an important finding as it has long been hypothesized that there may be a shared genetic mechanism for migration, not only in birds but also in other taxonomic groups (**Liedvogel et al., 2011**; **Liedvogel and Lundberg, 2014**; **Liedvogel and Delmore, 2018**).

None of the regions identified by hapFLK and PBS were fixed for alternate haplotypes or alleles. This fact is evident in **Figure 4**, in which the regions under selection still include haplotypes from a different cluster, and it could suggest that selection is acting on shared genetic variation (i.e., variation that is already present in the population rather than newly derived mutations). The idea that transitions between migratory phenotypes have been facilitated by shared genetic variation has been around for quite some time in the blackcap literature, particularly as rapid transitions have been observed and include the evolution of a new NW migratory route in the past 70 years. Shared variation can facilitate these rapid changes as these variants are already present in the population and have been tested by selection (**Barrett and Schluter, 2008**). The fact that regions under selection are quite narrow (**Table 1**) also supports a role for shared genetic variation (**Barrett and Schluter, 2008**) and we provide further evidence below.

First, we estimated the genetic distance between one haplotype in each cluster and an ancestral sequence that we derived using WGS from the two most closely related sister taxa, hill babbler (*Pseudoalcippe abyssinica*, an African resident) and garden warbler (*Sylvia borin*, a long distance migrant) (**Voelker and Light, 2011**). Using the region on Super-Scaffold 73 that shows selection in NW migrants (**Figure 4b**), we predicted that if haplotypes in the light blue cluster were present in the population already, they should exhibit similar levels of divergence from the ancestral sequence

as haplotypes from all other clusters. This is precisely what we found; genetic distance from the ancestral sequence was similar for haplotypes from all clusters (181 differences for the NW haplotype vs. 178, 179 and 181 [x3] and 182 differences in the rest). We reran this analysis limiting our data to synonymous substitutions in predicted coding regions (i.e., those that are likely to be evolving neutrally and located in ATG2B and BDKDB) and obtained similar results. Specifically, we identified six synonymous substitutions between all three medium-distance migrant populations and both garden warblers and hill babblers, suggesting that there is no difference in the age of these haplotypes.

To follow up on the former analysis, we constructed a maximum likelihood (ML) tree using sequence data from the region under selection on Super-Scaffold 73. We built this tree using data from all continental blackcaps, garden warblers, and hill babblers, using the willow warbler as an outgroup, and compared this tree to a consensus tree summarizing ML trees constructed for each scaffold in the blackcap reference genome (i.e., a tree built using genome-wide data; ***Figure 6a***). Supporting previous phylogenetic work in the system, garden warblers and hill babblers formed a sister clade to blackcaps in the consensus tree, and relationships among blackcaps were largely unresolved. By contrast, garden warblers were more closely related to blackcaps than were hill babblers in the tree built using data from the region on Super-Scaffold 73 (***Figure 6b***). In addition, the medium-distance NW population (in which positive selection is acting in this particular region) occurs at the base of the blackcap clade. Recall that garden warblers are obligate migrants whereas hill babblers are residents, supporting the suggestion that haplotypes favoured in the NW phenotype were already present in the population before divergence, perhaps even in ancestral populations. Unfortunately we do not have data from any closely related species to determine how old this haplotype is (i.e., if it is older than the split between garden warblers, hill babblers and blackcaps sensu ***Colosimo et al., 2005***).

In a final analysis, we compared the site frequency spectrum (SFS) for the region on Super-Scaffold 73 to SFSs estimated for 1000 random sequences of the same length from throughout the genome. SFSs for the random sequences are similar to expectations under neutrality, with a preponderance of alleles at low frequencies. By contrast, the SFS of Super-Scaffold 73 shows an excess of mid-frequency alleles (***Figure 6c***). Greater variance in SFSs are expected when selection makes use of standing variation because alleles have been recombining onto different backgrounds in ancestral populations (***Przeworski et al., 2005***; ***Pennings and Hermisson, 2006***).

We conclude our study by examining the genetic architecture of migratory distance. We included all migrants in this analysis, quantified migratory distance as an ordinal variable from short- (1), to medium- (2), to long-distance (3) migrants, and used a Bayesian Sparse Linear Mixed Model (BSLMM, 87) to identify SNPs that are associated with migratory distance (total number of birds included in these analyses = 72, ***Supplementary file 4***). BSLMMs are a form of genome-wide association analysis that includes a term for other factors that influence the phenotype and are correlated with genotype (e.g., population structure and ancestry; a kinship matrix based on genome-wide SNP data) and can be used to estimate both the combined effects of multiple SNPs and the effects of SNPs on their own.

Our results suggest that a large percentage of variance in migratory distance can be explained by our SNP set (PVE = 0.90 ± 0.28), but only one SNP showed a strong association with this focal trait (posterior inclusion probability >0.01). This SNP is located on Super-Scaffold 79, occurs in an area of elevated F_{ST} between long- and short-distance migrants (F_{ST} = 0.31, in 0.018 percentile F_{ST} values) and is 627-bp downstream from the gene KCNIP1, which encodes a potassium channel interacting protein (major determinants of neuronal cell excitability). Combined with the haplotype identified in the hapFLK analysis, which provides a signature of positive selection in short-distance migrants on Super-Scaffold 17 (***Table 1a***), these loci represent good candidates for controlling migratory distance, but future analyses with a larger sample size are needed to confirm the robustness of this finding. Direct information on migratory distance could also inform this analysis by allowing us to code the phenotype as continuous.

Conclusions

Early research on blackcaps was pivotal for demonstrating the existence of a genetic basis of migration and studying its evolution. This is due in large part to the tractability of this species and its variability in migratory behaviour. Here, we have expanded this study system beyond phenotypic and marker-based approaches, launching it into the genomic era and conducting one of the most

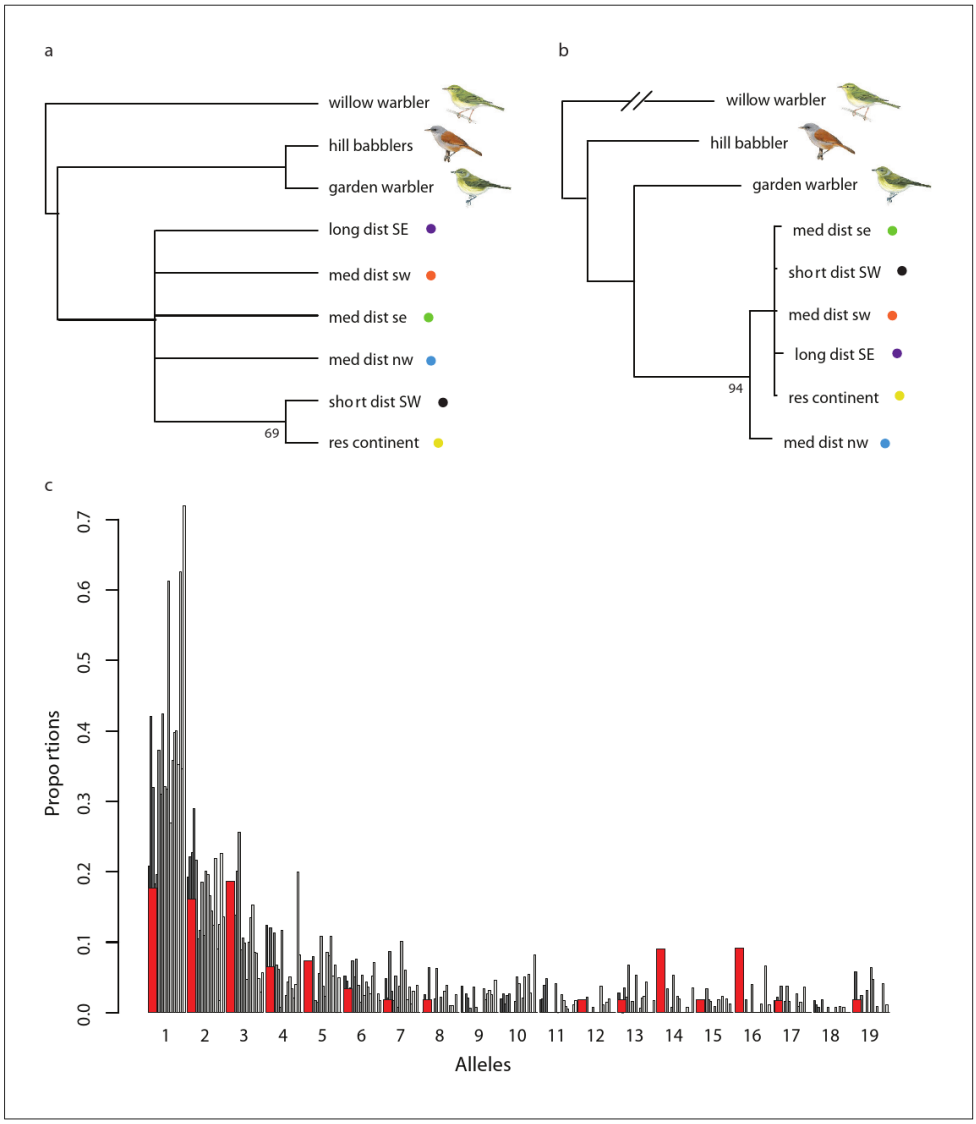


Figure 6. Evidence for the use of shared variation on Super-Scaffold 73. (**a**) A rooted extended majority rule consensus tree summarizing maximum likelihood (ML) trees constructed for all scaffolds in the blackcap reference genome (96 scaffolds). Node numbers indicate the number of scaffolds in which populations were partitioned into two sets. (**b**) A ML tree constructed for the region on Super-Scaffold 73 with migratory garden warbler more closely related to blackcaps and medium-distance NW birds occurring at the base of this clade in ***Figure 4b***. Nodes with bootstrap values <80 are collapsed; nodes without numbers have support values of 100. (**c**) The site frequency spectrum (SFS) for the region on Super-Scaffold 73 (red) compared to SFSs for 1000 random sequences from the genome (varying shades of gray).

comprehensive genome-wide analyses of migration to date. Populations of blackcaps began to diverge ~30,000 years ago, but differentiation remains low between migratory populations. There is evidence for past gene flow between migratory and resident populations on the European continent but comparison of the contemporary structure of these populations suggests that gene flow may be limited. This is certainly the case for resident island birds. It has been suggested that one single genetic mechanism controls migratory traits and may be shared across broad

taxonomic groups. We do not find evidence for one common genetic mechanism across species here, and no protein-coding change is shared across the three focal traits (propensity, distance and orientation) that we examined in unison. Future work on gene expression may identify major regulators that control multiple migratory traits, and both NPY and bHLH transcription factors are good candidates. Combined with the additional results that we presented here (such as the importance of standing genetic variation), this information is vital for understanding how predictable the evolution of migration and other complex behavioural traits may be.

Blackcaps have not only been relevant to work on the evolution and genetics of migration. Early work in this system suggested that differences in migration might serve as reproductive isolating barriers early in speciation. For example, hybrids were shown to exhibit intermediate orientation behaviour that was predicted to be inferior because it would bring hybrids over large ecological barriers that pure forms avoid (***Helbig, 1991b***). More recently, it was shown that NW migrants arrive on the breeding grounds earlier than SW migrants, and that these birds mate assortatively on the basis of arrival time, helping to reduce gene flow between phenotypically distinct groups (***Bearhop et al., 2005***). The role of migration in speciation has gained considerable traction in recent years (***Rolshausen et al., 2009***; ***Bearhop et al., 2005***; ***Irwin and Irwin, 2005***; ***Rohwer and Irwin, 2011***; ***Turbek et al., 2018***; ***Delmore and Irwin, 2014***; ***Bensch et al., 2009***) and results from our study suggest that selection at a very small number of loci may be sufficient to initiate reductions in gene flow very early in the process of population differentiation and speciation.

Materials and methods

Genome assembly

Blood samples from two male blackcaps from the Mooswald breeding population at Freiburg im Breisgau, Germany, classified as medium-distance SW migrants (on the basis of morphometrics and isotope signatures) were used to assemble the reference genome. Full details on all steps in our genome assembly can be found in ***Supplementary file 6*** (BioProject number PRJNA545868; Guojie Zhang, personal communication). Briefly, genomic DNA from one individual was used to sequence Illumina sequencing libraries (fragment and mate pair libraries with insert sizes of 2, 5 and 10 kb). 275.9 Gb of raw high throughput sequence (HTS) data were generated and assembled using ALLPATHS-LG. This assembly was improved several ways (e.g., by removing duplicates and closing gaps). DNA from the second individual was used to generate two BioNano optical maps (one using BspQI and the other BssSI). These maps were used to super-scaffold HTS scaffolds. Statistics for the final assembly and each stage can be found in ***Supplementary file 2*** .

We used SatsumaSynteny (***Grabherr et al., 2010***) to determine which avian chromosome each scaffold was found on (aligning scaffolds to the flycatcher genome, ***Supplementary file 3***). We validated our initial ALLPATHS assembly, the improved ALLPATHS assembly and our final assembly (including BioNano optical maps) using BUSCO (version 3.0.2, AUGUSTUS species chicken and aves_odb9 dataset) and by blasting ultra-conserved elements (UCEs) identified by ***Faircloth (2016)*** using whole-genome alignments for the chicken and zebra finch (***Supplementary file 2***).

Genome annotation

We annotated genes with putative functions and protein domains using MAKER. Gene prediction was performed using a de novo testis transcriptome of blackcaps and cDNAs from three avian species (zebra finch, chicken and flycatchers) from the ensembl database. Following MAKER, we obtained the predicted protein sequences to annotate genes functionally using blastp and Interproscan. For the final annotation, we only included gene predictions that either had an Annotation edit Distance (AED) <0.5 and/or a blastp hit (with the thresholds described above) and/or a predicted protein domain.

Resequencing analysis

We obtained whole genome resequencing (WGS) data from 110 male blackcaps (including WGS data from the two individuals used to generate the reference genomes). High molecular weight DNA was extracted from blood withdrawn from the brachial vein, following a standard salt extraction protocol. Individual samples were collected across the European breeding range including three

island populations (Canary Islands, Cape Verde, and Azores) and covering the entire range of migratory phenotypes. Population phenotype was scored on the basis of morphometry, stable isotope signature and/or ringing-recovery data from selected individuals (see ***Supplementary file 4*** for a description of how each population was phenotyped). Birds were sampled during the breeding season unless indicated otherwise. Specifically, exceptions are a subset of UK overwintering birds (n = 6) sampled during the winter in the British Isles, and a subset of long-distance SE migrants (n = 5) caught during autumn migration and selected on the basis of wing length (see ***Supplementary file 4*** for details). We also obtained WGS data for five garden warblers and three hill babblers, the closest sister taxa to blackcaps, sampled during breeding (***Voelker and Light, 2011***). We prepared small insert libraries using DNA from each individual and sequenced five samples per lane on NextSeq 500 with paired-end 150 bp reads. We trimmed reads with trimmomatic (TRAILING:3 SLIDINGWINDOW:4:10 MINLEN:30) (***Bolger et al., 2014***).

All analyses made use of data from resequencing reads that were aligned to the reference genome using bwa *mem* (***Li and Durbin, 2009***) or stampy in the case of the garden warblers (divergence time of 0.026 based on alignments of UCEs (***Faircloth, 2016***; https://github.com/faircloth-lab/phyluce/). GATK (***McKenna et al., 2010***) and picardtools (http://broadinstitute.github.io/picard) were used to identify and realign reads around indels (*RealignerTargetCreator*, *IndelRealigner*) as well as remove duplicates (*MarkDuplicates*, all default settings).

We recalibrated the resulting bam files using GATK's base quality score recalibration (BQSR). Specifically, we called SNPs for each population separately using three different programs and default settings: UnifiedGenotyper from GATK, samtools (***Li et al., 2009***) and FreeBayes (***Garrison and Marth, 2012***). BQSR requires a set of known variants. We used SNPs identified in all three programs and populations as the set of known variants for the first round of BQSR. We conducted a second round using common SNPs from the three programs that were also of high quality (QUAL >995,~10% of the common SNPs).

Most of our analyses made use of the BQSR recalibrated bams, calling genotype likelihoods (GL) with ANGSD (version 0.910–24-gf84f594, ***Korneliussen et al., 2014***) and filtering reads that did not map to a unique location, did not have a mapping pair, or had mapping qualities below 20 and flags $\geq$256. When it was not possible to use GL as input, we used a vcf that had been run through GATK's variant quality score recalibration (VQSR). VQSR also requires a set of known SNPs. We used the second set of known SNPs (common and high-quality) from BQSR for this analysis and combined variants from all populations into a single vcf file for subsequent analyses. All repetitive regions were excluded from our analyses and those focused on demography did not include the Z chromosome.

Principal components analysis (PCA) and ADMIXTURE analyses

We conducted a PCA using smartpca (EIGENSOFT version 5.0) and the vcf produced from VQSR. Default parameters were used in smartpca except for the addition of a correction for LD across SNPs (nsnpldregress = 2). We conducted an admixture analysis using GLs from ANGSD and running them through ngsADMIX (***Skotte et al., 2013***) with 8 values of K (1–9).

F_{ST}

We estimated F_{ST} between all populations using GLs from ANGSD, starting by estimating unfolded site frequency spectrums (SFS) for each population (doSaf 1, gl 1) and using them to obtain joint frequency spectrums (2DSFS, realSFS) for each pair of populations. These 2DSFSs were used as priors for allele frequencies at each site to estimate F_{ST} (realSFS fst index). In order to estimate unfolded SFS, we needed an ancestral sequence, or the ancestral state of variants segregating in blackcaps. This sequence was generated using WGS from garden warblers and hill babblers. Specifically, we used samtools to generate fasta files for each garden warbler and hill babbler (n=5 and n=3, respectively) and used rules outlined in ***Poelstra et al. (2014)*** to call ancestral states, with alleles that were homozygous in both outgroup species being considered ancestral and excluding remaining sites (those that were triallelic or heterozygous in the outgroup species).

Consensus tree

We obtained consensus fasta sequences for each population using ANGSD (-doFasta 2 –doCounts 1 –minQ 20 –setMinDepth 5) and used IQTREE (***Nguyen et al., 2015***) to construct maximum

likelihood trees for each scaffold in the blackcap genome (there was no difference in the topology obtained for scaffolds mapping to the Z chromosome so they were included in the consensus, data not shown). We summarized the resulting trees using phylip 'consense' and constructing an extended majority-rule consensus tree (in which nodes that were supported by fewer than 50% of the input trees are collapsed).

MSMC2

We used MSMC2 to infer the demographic history of blackcaps in our dataset. MSMC2 implements the multiple sequentially Markovian coalescent (MSMC) model, estimating effective population size by time and relative cross-coalescence rates between any two populations. It allows inference of the expansions and contractions of a population and of the extent and timing of population divergence (***Malaspinas et al., 2016***). Specifically, by running a hidden Markov model (HMM) along all possible pairs of haplotypes, MSMC2 estimates the free parameters for a demography model (a series of effective population sizes as a function of segmented time) and relative cross-coalescence rates between sequences using a maximum-likelihood approach.

After phasing our data using fastphase (***Scheet and Stephens, 2006***), we combined individuals into six groups (medium and long migrants ['med + long'], short-distance SW migrants ('short'), resident continent birds, and resident island birds from the Azores, Cape Verde, and Canary Islands). We grouped medium (NW, SW and SE) and long-distance migrants because they exhibited very little population structure (***Figure 1***) and indistinguishable demographic histories (***Figure 2—figure supplement 4***; ***Figure 2—figure supplement 5***; ***Figure 2—figure supplement 6***). We excluded any birds with less than 15x coverage. This filter left us with all island individuals (five individuals for each island), five short migrants, 19 continental residents, and 44 med + long migrants. To avoid bias associated with the use of unequal numbers of individuals from each group, we randomly down-sampled five individuals from med + long migrants and continental residents to create 10 sample groups. We used the bamCaller.py script provided in the msmc-tools package (https://github.com/stschiff/msmc-tools; ***Khvorykh, 2018***) to create sample-specific callability mask files. We generated a global mappability mask file for the reference genome using GEM (***Derrien et al., 2012***). We inferred effective population size by running MSMC2 separately for each group (***Schiffels and Wang, 2020***). We determined the number of clusters for fastPHASE using a cross-validation procedure (https://github.com/inzilico/kselection/ ***Khvorykh, 2018***). Statistical phasing (i.e., phasing without a reference panel) can be error prone, but fastPHASE is commonly employed in non-model organisms and is well-suited to datasets like ours that include high density SNPs on a physical map (***Scheet and Stephens, 2006***; ***Burri et al., 2015***; ***Kawakami et al., 2017***).

The analysis of cross-coalescence rates requires comparisons between groups and we considered all possible combinations of groups for our analysis (***Schiffels and Wang, 2020***). In other words, we ran analyses for all 15 possible combinations (three between groups on the continent, three between populations on the islands, and nine for comparisons between the three continent groups and three island populations). For each pairwise combination, we ran the combineCrossCoal.py script from msmc-tools (https://github.com/stschiff/msmc-tools) and computed the relative cross-coalescence rate by dividing the between-populations coalescence rate by the average within-population coalescence rate. We scaled results using a mutation rate of 3×10^{-9}/gen/site and a generation time of 2 years (***Nadachowska-Brzyska et al., 2016***; ***Nadachowska-Brzyska et al., 2015***).

hapFLK

hapFLK is a tree-based method that is used to identify genomic regions that are under selection. This program permits the inclusion of two or more populations and accounts for both drift within populations (different N_e) and covariance across them (hierarchical structuring). We used the vcf from VQSR as input for this analysis, applying two additional filters for the inclusion of variants: minimum number of individuals/phenotype = 5 and minor allele frequency of 0.05. hapFLK also requires an estimate of the number of clusters into which haplotypes can be grouped. We ran this analysis for the complete dataset including all populations, and for a restricted dataset including only medium-distance migrants. We determined the number of clusters for each dataset separately using fastPHASE (***Scheet and Stephens, 2006***) and the cross-validation procedure mentioned earlier.

Once hapFLK is estimated, it is normalized using rlm in R, and p-values are computed from the chi-squared distribution. We used a permutation analysis to establish a threshold, beyond which genomic regions would be considered to be experiencing positive selection. Specifically, we randomly shuffled population labels 100 times, re-estimated hapFLK and p-values, recorded the lowest p-value for each randomization and set the threshold to the fifth percentile across randomizations. Once these regions were identified, we determined which population was experiencing selection by comparing branch lengths for a tree built using data from the entire genome and one built using data from the region under selection. Note that results from analyses using medium-distance migrants are plotted using the resident phenotype for illustrative purposes, but the analysis was not run using these birds.

We include birds from three resident continent populations – Cazalla de la Sierra and Gibraltar in the Iberian Peninsula along with Asni in Morocco (only three birds were sampled from this African population, precluding its use in the present analysis; ***Supplementary file 4***). The Iberian Peninsula where the other two populations are found is highly heterogeneous as a result of the effects of mountains and plateaus that create variation in seasonality and, consequently, in the intensity of blackcap migratory behaviour (***Pérez-Tris and Tellería, 2002***; ***Tellería et al., 2001***). There is also some evidence in our PCA to show that this heterogeneity has led to some differentiation between populations, as birds from Cazalla de la Sierra exhibit values more similar to migrants on PC2 (***Figure 1c***). Accordingly, to avoid any confounding effects from population structure, we limited our analysis to birds from Gibraltar. Results using Cazalla de la Sierra instead were very similar. For example, all of the genomic regions identified in ***Table 1b*** were also in the top 1% of the ΔPBS distribution when Cazalla de la Sierra was used as the continental reference population instead of Gibraltar.

CAVIAR

The principle described above for hapFLK focusing on haplotype clusters can also be applied to SNPs (FLK). We used results from an analysis with FLK and limited to genomic regions, which showed evidence of positive selection from hapFLK, to identify independent strongly associated SNPs with CAVIAR (CAusal Variants Identification in Associated Regions [***Hormozdiari et al., 2014***]). CAVIAR was originally designed to identify independent causal SNPs in GWAS studies. We followed methods described in ***Rochus et al. (2018)*** to modify this method for FLK, identifying SNPs with p-values <0.0001 in hapFLK outlier regions and using a correlation matrix generated by FLK by decomposing signals into loading on orthogonal components (vs. p-values from a GWAS and LD as is traditionally done with CAVIAR).

Δ PBS

We used a modified version of PBS (Population Branch Statistic) to complement results from hapFLK. PBS is similar to F_{ST}, but can include more than two populations and identifies regions within each population that exhibit differences in allele frequencies. This statistic was originally designed for three populations, but can be expanded to include more populations (***Zhan et al., 2014***). We used GL from ANGSD to obtain estimates of F_{ST} following the procedure described above (summarized into windows of 2500 kb) and used the equation below to estimate PBS from these values. This equation is an example that was applied to resident populations (R), where T is log transformed F_{ST} between the populations indicated in exponents:

$$\frac{T^{R-NW} + T^{R-SW} + T^{R-SE} - T^{NW-SW} - T^{SW-SE}}{3}$$

Recent papers have noted that F_{ST} can be elevated by reductions in within-population variation alone and that there are many factors that can reduce variation within populations, including linked selection in areas of reduced recombination that may result from purifying selection (background selection, [***Cruickshank and Hahn, 2014***; ***Noor and Bennett, 2009***]). It is unlikely that this process affects our results because recombination rate should elevate estimates of PBS in all populations, but this is not the case (***Figure 5a***). Regardless, we followed methods from ***Vijay et al. (2017)*** to reduce any effects that linked selection may have on our results. ***Vijay et al. (2017)*** used estimates of F_{ST} between allopatric populations of crows that did not differ in their trait of interest to control

for the effects of linked selection, estimating the difference in estimates of F_{ST} in focal populations and maximum F_{ST} in non-focal allopatric populations (ΔFST). F_{ST} in focal populations would have to extend beyond that in non-focal populations to be considered important in generating the trait of interest. We used the same approach for PBS. For example, ΔPBS for resident continent populations was estimated by finding the difference between PBS in residents and maximum PBS in medium-, short- and long-distance migrants.

nSL

The former analyses (hapFLK and PBS) rely on comparisons between phenotypes. In this last analysis, we focus on the affects that selection can have within a population instead. Specifically, selective sweeps can reduce variation at both the locus under selection and its neighbours (***Smith and Haigh, 1974***). Local reductions in variation result in the presence of extended regions of haplotype homozygosity within phenotypes (long haplotypes at high frequency). nSL (number of segregating sites by length) (***Ferrer-Admetlla et al., 2014***) is similar to the more common iHS, but instead of measuring the decay of haplotype identity as a function of recombination distance, it quantifies this decay of how many mutations remain in other haplotypes present in the dataset. In this way, nSL does not require a genetic map and is more robust to variation in not only recombination rate but also mutation rate.

For this analysis, we used selscan (v.1.20a https://github.com/szpiech/selscan) and the same vcf used in hapFLK, but split by phenotype (and scaffold). We ran the data through fastPHASE first to phase haplotypes (using 50 iterations of the EM algorithm, sampling 100 haplotypes from the posterior distribution and using same number of clusters identified for hapFLK). We normalized estimates of nSL into the same 2500-kb windows used for PBS.

Regulatory variants

Two sets of preliminary analyses were used to identify regulatory SNPs in the regions identified by hapFLK and PBS as being under selection. First, we focused on 3′ UTRs, downloading predicted mRNAs from Ensembl and NCBI for several bird species, including the Atlantic canary, White-throated sparrow, American crow, Great tit, Collared flycatcher, Zebra finch, Wild turkey, White-rumped munia, Hooded crow, Blue tit and Ground tit. We aligned these sequences with our annotation for the blackcap, and with transcripts assembled from RNAseq data obtained from the testes of a single male blackcap, to determine whether any of the strongly associated SNPs identified by CAVIAR were within 3′ UTRs. Alignment files are available upon request.

In a second set of analyses, we used HOMER (***Heinz et al., 2010***) to identify known transcription factor binding sites (TFBS) in genomic regions under selection. Specifically, we used findMotifsGenome.pl with default settings to identify known motifs in each region and scanMotifGenomeWide.pl to identify the specific location in each region where the motif could be found (permitting no mismatches). HOMER includes known motifs for thousands of transcription factors (mostly for model organisms); we chose to focus on candidate transcription factors identified by previous studies as having an association with migration (***Ruegg et al., 2014***).

GWAS

In our final analysis on migratory distance, we limited our dataset to short-, medium- and long-distance migrants. We coded distance phenotype as an ordinal variable from 1 to 3 and conducted a GWAS analysis using a Bayesian sparse linear mixed model (BSLMM) (***Zhou et al., 2013***). We chose BSLMM models here (instead of hapFLK) because they allow the inclusion of ordinal variable (vs. categorical with hapFLK). BSLMM models include the phenotype as the response variable and allele frequencies at a set of SNPs as the predictor variable. They also include a term for factors that influence the phenotype and are correlated with genotype (e.g., population structure). BSLMMs are adaptive models that include linear mixed models (LMM) and Bayesian variable selection regression (BVSR) as special cases and that learn the genetic architecture from the data. We ran four independent chains for each BSLMM, with a burnin of 5 million steps and a subsequent 20 million MCMC steps (sampling every 1000 steps). We report one hyperparameter from this model (PVE: the proportion of variance in phenotypes explained by all SNPs, also called chip heritability) and consider SNPs with inclusion probabilities >0.01 following ***Gompert et al. (2013)***. Note, we chose to run this

analysis with GEMMA instead of hapFLK as we did with our other migratory traits (orientation and propensity). This is because our focal variable here (distance) is ordinal in nature and this fact would have been lost in hapFLK. We could not code this variable as continuous because the average distance individuals in each population travel on migration is not exactly known.

Acknowledgements

We acknowledge funding from the Max Planck Society (MPRG grant to ML), NSERC (PDF to KD) and Regional Government of Asturias (GRUPIN to JCI, Ref.: IDI/2018/000151). We thank: Staffan Bensch, Andreas Helbig, Stuart Bearhop, and Thord Fransson for providing samples; Conny Burghardt, Heinke Buhtz and Sven Künzel for lab work; Diethard Tautz, Tobias Kaiser and Sandra Bouwhuis for comments on early drafts; Julien Dutheil, Reto Burri, Kristian Ullrich, Christine Merlin and Aldrin Lugena for discussions of analyses; and Saki Chan at BioNano for the hybrid assembly of our genome. We would also like to thank Elizabeth Scordato, Patricia Wittcopp and three anonymous reviewers for constructive feedback on an earlier version of this manuscript. Permits were provided to JCI for samples collected in Morocco (Haut Commissariat aux Eaux et Forets et a la Lutte Contre la Desertification, 206/2011, 13 Jan 2011), Cape Verde (Ministerio do Ambiente - Habitacao e Ordenamento do Territorio, 18/CITES/DNA, 17 Dec 2015) and the Azores (Instituto da Conservacao da Natureza e da Biodiversidade, 171/2008, 31 Mar 2009). JP-T received permits for samples collected in Gibraltar and Cazalla de la Sierra (Consejeria de Medio Ambiente, 50.725.548-Z, 12 May 2011), Alava (Arabako Foru Aldundia, 50.725.548-Z, 12 Apr 2011) and Guadaramma (Consejeria de Medio Ambiente - Vivenda y Ordenacion del Territorio, 10/160876.9/10, 12 Apr 2010). Thord Fransson received permits for samples collected in Stockholm (Stockholms djurförsöksetiska nämnd Dnr N 16/16 2016-02-25). Permits were provided to GS for samples collected in the remaining locations (Regierungspräsidium Freiburg, 55–8853.17/0).

Additional information

Funding

Funder	Grant reference number	Author
Natural Sciences and Engineering Research Council of Canada	PDF	Kira Delmore
Max-Planck-Gesellschaft	MPRG	Miriam Liedvogel
Regional Government of Asturias	GRUPIN: IDI/2018/000151	Juan Carlos Illera

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Kira Delmore, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing; Juan Carlos Illera, Javier Pérez-Tris, Resources, Data curation, Funding acquisition, Methodology, Writing - review and editing; Gernot Segelbacher, Resources, Data curation, Funding acquisition, Writing - review and editing; Juan S Lugo Ramos, Gillian Durieux, Jun Ishigohoka, Formal analysis, Writing - review and editing; Miriam Liedvogel, Conceptualization, Resources, Data curation, Supervision, Funding acquisition, Methodology, Project administration, Writing - review and editing

Author ORCIDs

Kira Delmore https://orcid.org/0000-0003-4108-9729
Juan Carlos Illera http://orcid.org/0000-0002-4389-0264
Jun Ishigohoka https://orcid.org/0000-0002-5713-9391
Miriam Liedvogel https://orcid.org/0000-0002-8372-8560

Ethics

Animal experimentation: Permits to JCI for samples collected in Morocco (Haut Commissariat aux Eaux et Forets et a la Lutte Contre la Desertification, 206/2011, 13 Jan 2011), Cape Verde (Ministerio do Ambiente - Habitacao e Ordenamento do Territorio, 18/CITES/DNA, 17 Dec 2015) and the Azores (Instituto da Conservacao da Natureza e da Biodiversidade, 171/2008, 31 Mar 2009); Permits to JP-T for samples collected in Gibraltar and Cazalla de la Sierra (Consejeria de Medio Ambiente, 50.725.548-Z, 12 May 2011), Alava (Arabako Foru Aldundia, 50.725.548-Z, 12 Apr 2011) and Guadaramma (Consejeria de Medio Ambiente - Vivenda y Ordenacion del Territorio, 10/160876.9/10, 12 Apr 2010); Permit to Thord Fransson for samples collected in Stockholm (Stockholms djurförsöksetiska nämnd Dnr N 16/16 2016-02-25); Permits to GS for samples collected in the remaining locations (Regierungspräsidium Freiburg, 55-8853.17/0).

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.54462.sa1
Author response https://doi.org/10.7554/eLife.54462.sa2

Additional files

Supplementary files

- Supplementary file 1. Summary of sequencing data used for ALLPATHS-LG assembly. Libraries designated a and b are from the same library preparation but sequenced on two separate lanes.
- Supplementary file 2. Assembly statistics at each stage. The second ALLPATHS assembly follows the removal of duplicates and contaminants along with gap filling.
- Supplementary file 3. Results from satsuma showing which flycatcher chromosome each scaffold in the blackcap reference genome hit. Mean position and orientation refer to the location and orientation of scaffolds on the flycatcher genome. The last six scaffolds did not hit any of the flycatcher chromosomes. Comparing the annotation of the blackcap and zebra finch genomes suggests they match the indicated chromosomes.
- Supplementary file 4. Samples used in the present study, including their locations and details on how phenotypes were determined.
- Supplementary file 5. Extension of ***Table 1*** showing regions identified by hapFLK as being under selection but including the number of causal SNPs identified by CAVIAR and their location within genes. Estimates of nSL are shown (bolded if in top 1% of values, bolded and italicised if in the top 5%).
- Supplementary file 6. Additional details on genome assembly and annotation.
- Transparent reporting form

Data availability

Sequencing data has been deposited under NCBI BioProject PRJNA616371. All other data are included in the manuscript and supporting files.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Delmore K, Illera JC, Pérez-Tris J, Segelbacher G, Lugo Ramos JS, Durieux G, Ishigohoka J, Liedvogel M	2020	European blackcap resequencing	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA616371/	NCBI BioProject, PRJNA616371
B10K Consortium	2019	Bird 10,000 Genomes (B10K) Project - Family phase	https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA545868	NCBI BioProject, PRJNA545868

References

Alerstam T, Hedenstrom A, Akesson S. 2003. Long-distance migration: evolution and determinants. *Oikos* **103**: 247–260. DOI: https://doi.org/10.1034/j.1600-0706.2003.12559.x

Barrett LW, Fletcher S, Wilton SD. 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69**:3613–3634. DOI: https://doi.org/10.1007/s00018-012-0990-9, PMID: 22538991

Barrett RD, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**:38–44. DOI: https://doi.org/10.1016/j.tree.2007.09.008, PMID: 18006185

Bearhop S, Fiedler W, Furness RW, Votier SC, Waldron S, Newton J, Bowen GJ, Berthold P, Farnsworth K. 2005. Assortative mating as a mechanism for rapid evolution of a migratory divide. *Science* **310**:502–504. DOI: https://doi.org/10.1126/science.1115661, PMID: 16239479

Bensch S, Grahn M, Müller N, Gay L, Akesson S. 2009. Genetic, morphological, and feather isotope variation of migratory willow warblers show gradual divergence in a ring. *Molecular Ecology* **18**:3087–3096. DOI: https://doi.org/10.1111/j.1365-294X.2009.04210.x, PMID: 19457197

Berthold P, Mohr G, Querner U. 1990. Steuerung und potentielle evolutionsgeschwindigkeit des obligaten teilzieherverhaltens: ergebnisse eines Zweiweg-Selektionsexperiments mit der mönchsgrasmücke (Sylvia atricapilla). *Journal of Ornithology* **131**:33–45. DOI: https://doi.org/10.1007/BF01644896

Berthold P, Helbig AJ, Mohr G, Querner U. 1992. Rapid microevolution of migratory behaviour in a wild bird species. *Nature* **360**:668–670. DOI: https://doi.org/10.1038/360668a0

Berthold P, Querner U. 1981. Genetic basis of migratory behavior in european warblers. *Science* **212**:77–79. DOI: https://doi.org/10.1126/science.212.4490.77, PMID: 17747634

Bhullar B-AS, Morris ZS, Sefton EM, Tok A, Tokita M, Namkoong B, Camacho J, Burnham DA, Abzhanov A. 2015. A molecular mechanism for the origin of a key evolutionary innovation, the bird beak and palate, revealed by an integrative approach to major transitions in vertebrate history. *Evolution* **69**:1665–1677. DOI: https://doi.org/10.1111/evo.12684

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**:2114–2120. DOI: https://doi.org/10.1093/bioinformatics/btu170, PMID: 24695404

Boswell T, Dunn IC. 2017. Regulation of Agouti-Related protein and Pro-Opiomelanocortin gene expression in the avian arcuate nucleus. *Frontiers in Endocrinology* **8**:75. DOI: https://doi.org/10.3389/fendo.2017.00075, PMID: 28450851

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S, Garamszegi LZ, Hogner S, Moreno J, Qvarnström A, Ružić M, Sæther SA, Sætre GP, Török J, Ellegren H. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research* **25**:1656–1665. DOI: https://doi.org/10.1101/gr.196485.115, PMID: 26355005

Clark PU, Dyke AS, Shakun JD, Carlson AE, Clark J, Wohlfarth B, Mitrovica JX, Hostetler SW, McCabe AM. 2009. The last glacial maximum. *Science* **325**:710–714. DOI: https://doi.org/10.1126/science.1172873, PMID: 19661421

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**:1928–1933. DOI: https://doi.org/10.1126/science.1107239, PMID: 15790847

Cramp S. 1992. *The Birds of the Western Palearctic Vol VI*. Oxford, UK: Oxford University Press.

Cropper T. 2013. The weather and climate of macaronesia: past, present and future. *Weather* **68**:300–307. DOI: https://doi.org/10.1002/wea.2155

Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**:3133–3157. DOI: https://doi.org/10.1111/mec.12796, PMID: 24845075

de la Hera I, Pérez-Tris J, Tellería JL. 2009. Migratory behaviour affects the trade-off between feather growth rate and feather quality in a passerine bird. *Biological Journal of the Linnean Society* **97**:98–105. DOI: https://doi.org/10.1111/j.1095-8312.2008.01189.x

Delmore KE, Toews DPL, Germain RR, Owens GL, Irwin DE. 2016. The genetics of seasonal migration and plumage color. *Current Biology* **26**:2167–2173. DOI: https://doi.org/10.1016/j.cub.2016.06.015

Delmore KE, Irwin DE. 2014. Hybrid songbirds employ intermediate routes in a migratory divide. *Ecology Letters* **17**:1211–1218. DOI: https://doi.org/10.1111/ele.12326, PMID: 25040456

Delmore KE, Liedvogel M. 2016. Investigating factors that generate and maintain variation in migratory orientation: a primer for recent and future work. *Frontiers in Behavioral Neuroscience* **10**:3. DOI: https://doi.org/10.3389/fnbeh.2016.00003, PMID: 26834592

Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, Guigó R, Ribeca P. 2012. Fast computation and applications of genome mappability. *PLOS ONE* **7**:e30377. DOI: https://doi.org/10.1371/journal.pone.0030377, PMID: 22276185

Dietzen C, Garcia-Del-Rey E, Castro GD, Wink M. 2008. Phylogenetic differentiation of Sylvia species (Aves: passeriformes) of the Atlantic islands (Macaronesia) based on mitochondrial DNA sequence data and morphometrics. *Biological Journal of the Linnean Society* **95**:157–174. DOI: https://doi.org/10.1111/j.1095-8312.2008.01005.x

Egevang C, Stenhouse IJ, Phillips RA, Petersen A, Fox JW, Silk JR. 2010. Tracking of arctic terns *Sterna paradisaea* reveals longest animal migration. *PNAS* **107**:2078–2081. DOI: https://doi.org/10.1073/pnas.0909493107, PMID: 20080662

Faircloth BC, McCormack JE, Crawford NG, Harvey MG, Brumfield RT, Glenn TC. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* **61**:717–726. DOI: https://doi.org/10.1093/sysbio/sys004

Faircloth BC. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* **32**:786–788. DOI: https://doi.org/10.1093/bioinformatics/btv646, PMID: 26530724

Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. 2013. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* **193**:929–941. DOI: https://doi.org/10.1534/genetics.112.147231, PMID: 23307896

Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R. 2014. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution* **31**:1275–1291. DOI: https://doi.org/10.1093/molbev/msu077, PMID: 24554778

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv*. https://arxiv.org/abs/1207.3907.

Gompert Z, Lucas LK, Nice CC, Buerkle CA. 2013. Genome divergence and the genetic architecture of barriers to gene flow between *LYCAEIDES IDAS* AND *L. MELISSA*. *Evolution* **67**:2498–2514. DOI: https://doi.org/10.1111/evo.12021

Grabherr MG, Russell P, Meyer M, Mauceli E, Alfoldi J, Di Palma F, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: satsuma. *Bioinformatics* **26**:1145–1151. DOI: https://doi.org/10.1093/bioinformatics/btq102

Gwinner E. 1996. Circadian and circannual programmes in avian migration. *The Journal of Experimental Biology* **199**:39–48. PMID: 9317295

Gwinner E, Helm B. 2003. Circannual and circadian contributions to the timing of avian migration. In: *Avian Migration*. Springer. p. 81–95. DOI: https://doi.org/10.1007/978-3-662-05957-9

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**:576–589. DOI: https://doi.org/10.1016/j.molcel.2010.05.004, PMID: 20513432

Helbig AJ, Berthold P, Wiltschko W. 1989. Migratory Orientation of Blackcaps (Sylvia atricapilla): Population-specific Shifts of Direction during the Autumn. *Ethology* **82**:307–315. DOI: https://doi.org/10.1111/j.1439-0310.1989.tb00510.x

Helbig AJ. 1991a. Experimental and analytical techniques used in bird orientation research. In: Berthold P (Ed). *Orientation in Birds Experientia Supplementum*. Basel: Birkhäuser Basel. p. 270–306. DOI: https://doi.org/10.1007/978-3-0348-7208-9

Helbig A. 1991b. Inheritance of migratory direction in a bird species: a cross-breeding experiment with SE- and SW-migrating blackcaps (Sylvia atricapilla). *Behavioral Ecology and Sociobiology* **28**:9–12. DOI: https://doi.org/10.1007/BF00172133

Helbig AJ. 1994. Genetic base and evolutionary change of migratory directions in a european passerine migrant *sylvia atricapilla* . *Ostrich* **65**:151–159. DOI: https://doi.org/10.1080/00306525.1994.9639677

Helbig A. 1996. Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palaearctic warblers (Aves: sylviidae). *The Journal of Experimental Biology* **199**:49–55. PMID: 9317319

Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. 2014. Identifying causal variants at loci with multiple signals of association. *Genetics* **198**:497–508. DOI: https://doi.org/10.1534/genetics.114.167908

Irwin DE, Irwin JH. 2005. Siberian migratory divides: the role of seasonal migration in speciation. In: Greenberg R, Marra PP (Eds). *Birds of Two Worlds: The Ecology and Evolution of Migration*. Baltimore: Johns Hopkins University Press. p. 27–40.

Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM, Habegger L, Rozowsky J, Shi M, Urban AE, Hong MY, Karczewski KJ, Huber W, Weissman SM, Gerstein MB, Korbel JO, Snyder M. 2010. Variation in transcription factor binding among humans. *Science* **328**:232–235. DOI: https://doi.org/10.1126/science.1183621, PMID: 20299548

Kawakami T, Mugal CF, Suh A, Nater A, Burri R, Smeds L, Ellegren H. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Molecular Ecology* **26**:4158–4172. DOI: https://doi.org/10.1111/mec.14197, PMID: 28597534

Khvorykh G. 2018. Inzilico/kselection. v1.0. Zenodo. https://doi.org/10.5281/zenodo.1173276

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**:1. DOI: https://doi.org/10.1186/s12859-014-0356-4, PMID: 25420514

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079. DOI: https://doi.org/10.1093/bioinformatics/btp352, PMID: 19505943

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760. DOI: https://doi.org/10.1093/bioinformatics/btp324, PMID: 19451168

Liedvogel M, Akesson S, Bensch S. 2011. The genetics of migration on the move. *Trends in Ecology & Evolution* **26**:561–569. DOI: https://doi.org/10.1016/j.tree.2011.07.009, PMID: 21862171

Liedvogel M, Delmore K. 2018. (Micro) Evolutionary Changes and the Evolutionary Potential of Bird Migration. In: *Bird Species*. Springer. p. 109–127. DOI: https://doi.org/10.1007/978-3-319-91689-7

Liedvogel M, Lundberg M. 2014. The genetics of animal movement and migration syndromes. In: Hansson L-A, Åkesson S (Eds). *Animal Movement Across Scales*. New York: Oxford University Press. p. 219–231. DOI: https://doi.org/10.1111/aec.12369

Lundberg M, Liedvogel M, Larson K, Sigeman H, Grahn M, Wright A, Åkesson S, Bensch S. 2017. Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters* **1**:155–168. DOI: https://doi.org/10.1002/evl3.15, PMID: 30283646

Malaspinas AS, Westaway MC, Muller C, Sousa VC, Lao O, Alves I, Bergström A, Athanasiadis G, Cheng JY, Crawford JE, Heupink TH, Macholdt E, Peischl S, Rasmussen S, Schiffels S, Subramanian S, Wright JL, Albrechtsen A, Barbieri C, Dupanloup I, et al. 2016. A genomic history of aboriginal Australia. *Nature* **538**:207–214. DOI: https://doi.org/10.1038/nature18299, PMID: 27654914

Mallarino R, Grant PR, Grant BR, Herrel A, Kuo WP, Abzhanov A. 2011. Two developmental modules establish 3D beak-shape variation in Darwin's finches. *PNAS* **108**:4057–4062. DOI: https://doi.org/10.1073/pnas.1011480108, PMID: 21368127

Mayr C. 2017. Regulation by 3'-Untranslated regions. *Annual Review of Genetics* **51**:171–194. DOI: https://doi.org/10.1146/annurev-genet-120116-024704, PMID: 28853924

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**:1297–1303. DOI: https://doi.org/10.1101/gr.107524.110, PMID: 20644199

Mouritsen H. 2018. Long-distance navigation and magnetoreception in migratory animals. *Nature* **558**:50–59. DOI: https://doi.org/10.1038/s41586-018-0176-1, PMID: 29875486

Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, Ellegren H. 2015. Temporal dynamics of avian populations during pleistocene revealed by Whole-Genome sequences. *Current Biology* **25**:1375–1380. DOI: https://doi.org/10.1016/j.cub.2015.03.047, PMID: 25891404

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Molecular Ecology* **25**:1058–1072. DOI: https://doi.org/10.1111/mec.13540, PMID: 26797914

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268–274. DOI: https://doi.org/10.1093/molbev/msu300, PMID: 25371430

Noor MA, Bennett SM. 2009. Islands of speciation or mirages in the desert? examining the role of restricted recombination in maintaining species. *Heredity* **103**:439–444. DOI: https://doi.org/10.1038/hdy.2009.151, PMID: 19920849

Pennings PS, Hermisson J. 2006. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLOS Genetics* **2**:e186. DOI: https://doi.org/10.1371/journal.pgen.0020186, PMID: 17173482

Pérez-Tris J, Bensch S, Carbonell R, Helbig AJ, Tellería JL. 2004. Historical diversification of migration patterns in a passerine bird. *Evolution* **58**:1819–1832. DOI: https://doi.org/10.1554/03-731

Pérez-Tris J, Tellería JL. 2002. Regional variation in seasonality affects migratory behaviour and life-history traits of two mediterranean passerines. *Acta Oecologica* **23**:13–21. DOI: https://doi.org/10.1016/S1146-609X(01)01129-8

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, Wolf JB. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**:1410–1414. DOI: https://doi.org/10.1126/science.1253226, PMID: 24948738

Przeworski M, Coop G, Wall JD. 2005. The signature of positive selection on standing genetic variation. *Evolution* **59**:2312–2323. DOI: https://doi.org/10.1554/05-273.1

Pulido F, Berthold P, van Noordwijk AJ. 1996. Frequency of migrants and migratory activity are genetically correlated in a bird population: evolutionary implications. *PNAS* **93**:14642–14647. DOI: https://doi.org/10.1073/pnas.93.25.14642, PMID: 8962107

Pulido F, Berthold P. 2010. Current selection for lower migratory activity will drive the evolution of residency in a migratory bird population. *PNAS* **107**:7341–7346. DOI: https://doi.org/10.1073/pnas.0910361107, PMID: 20368446

Rochus CM, Tortereau F, Plisson-Petit F, Restoux G, Moreno-Romieux C, Tosser-Klopp G, Servin B. 2018. Revealing the selection history of adaptive loci using genome-wide scans for selection: an example from domestic sheep. *BMC Genomics* **19**:71. DOI: https://doi.org/10.1186/s12864-018-4447-x, PMID: 29357834

Rohwer S, Irwin DE. 2011. Molt, orientation, and avian speciation. *The Auk* **128**:419–425. DOI: https://doi.org/10.1525/auk.2011.10176

Rolshausen G, Segelbacher G, Hobson KA, Schaefer HM. 2009. Contemporary evolution of reproductive isolation and phenotypic divergence in sympatry along a migratory divide. *Current Biology* **19**:2097–2101. DOI: https://doi.org/10.1016/j.cub.2009.10.061, PMID: 19962309

Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB. 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**:4757–4769. DOI: https://doi.org/10.1111/mec.12842, PMID: 24954641

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *The American Journal of Human Genetics* **78**:629–644. DOI: https://doi.org/10.1086/502802, PMID: 16532393

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics* **46**:919–925. DOI: https://doi.org/10.1038/ng.3015, PMID: 24952747

Schiffels S, Wang K. 2020. MSMC and MSMC2: the multiple sequentially markovian coalescent. In: *Statistical Population Genomics. Methods in Molecular Biology*. New York, NY: Humana. DOI: https://doi.org/10.1007/978-1-0716-0199-0

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* **195**:693–702. DOI: https://doi.org/10.1534/genetics.113.154138, PMID: 24026093

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genetical Research* **23**:23–35. DOI: https://doi.org/10.1017/S0016672300014634, PMID: 4407212

Tellería JL, Pérez-Tris J, Carbonell R. 2001. Seasonal changes in abundance and flight-related morphology reveal different migration patterns in iberian forest passerines. *Ardeola* **48**:27–46.

Turbek SP, Scordato ESC, Safran RJ. 2018. The role of seasonal migration in population divergence and reproductive isolation. *Trends in Ecology & Evolution* **33**:164–175. DOI: https://doi.org/10.1016/j.tree.2017.11.008, PMID: 29289354

Valente L, Illera JC, Havenstein K, Pallien T, Etienne RS, Tiedemann R. 2017. Macroevolutionary dynamics in Atlantic Island avifaunas support MacArthur and Wilson's equilibrium prediction. *Current Biology : CB* **27**:1660–1666. DOI: https://doi.org/10.1016/j.cub.2017.04.053

Vijay N, Weissensteiner M, Burri R, Kawakami T, Ellegren H, Wolf JBW. 2017. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular Ecology* **26**:4284–4295. DOI: https://doi.org/10.1111/mec.14195, PMID: 28570015

Visser ME, Caro SP, van Oers K, Schaper SV, Helm B. 2010. Phenology, seasonal timing and circannual rhythms: towards a unified framework. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**:3113–3127. DOI: https://doi.org/10.1098/rstb.2010.0111

Voelker G, Light JE. 2011. Palaeoclimatic events, dispersal and migratory losses along the Afro-European Axis as drivers of biogeographic distribution in Sylvia warblers. *BMC Evolutionary Biology* **11**:163. DOI: https://doi.org/10.1186/1471-2148-11-163, PMID: 21672229

Widelitz RB. 2008. Wnt signaling in skin organogenesis. *Organogenesis* **4**:123–133. DOI: https://doi.org/10.4161/org.4.2.5859, PMID: 19279724

Wiltschko W, Wiltschko R. 1972. Magnetic compass of european robins. *Science* **176**:62–64. DOI: https://doi.org/10.1126/science.176.4030.62, PMID: 17784420

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, Zheng H, Liu T, He W, Li K, Luo R, Nie X, Wu H, Zhao M, Cao H, Zou J, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**:75–78. DOI: https://doi.org/10.1126/science.1190371, PMID: 20595611

Yu M, Yue Z, Wu P, Wu D-Y, Mayer J-A, Medina M, Widelitz RB, Jiang T-X, Chuong C-M. 2004. The developmental biology of feather follicles. *The International Journal of Developmental Biology* **48**:181–191. DOI: https://doi.org/10.1387/ijdb.031776my

Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The monarch butterfly genome yields insights into Long-Distance migration. *Cell* **147**:1171–1185. DOI: https://doi.org/10.1016/j.cell.2011.09.052

Zhan S, Zhang W, Niitepõld K, Hsu J, Haeger JF, Zalucki MP, Altizer S, de Roode JC, Reppert SM, Kronforst MR. 2014. The genetics of monarch butterfly migration and warning colouration. *Nature* **514**:317–321. DOI: https://doi.org/10.1038/nature13812

Zhou X, Carbonetto P, Stephens M. 2013. Polygenic modeling with bayesian sparse linear mixed models. *PLOS Genetics* **9**:e1003264. DOI: https://doi.org/10.1371/journal.pgen.1003264, PMID: 23408905

Supplementary materials

Additional details on genome assembly and annotation

Library construction Next-generation sequencing libraries were sequenced on a HiSeq 4000 with paired-end, 100 bp reads. DNA for Bionano was suspended in CSB and embedded in agarose-CSB mold. It was labelled following IrysPrep Reagent Kit protocol using two nicking enzymes (BspQI and BssSI). The sample was then loaded onto IrysChips and run on the Irys imaging instrument.

ALLPATHS-LG assembly and improvement Next-generation sequencing libraries were subsampled to 50x coverage and default parameters in ALLPATHS-LG were used in the assembly algorithm, with a ploidy of 2 and choosing the option haploidify=true. This is a relatively new option for diploid genomes that are polymorphic. It removes polymorphisms from reads following error correction, creating a mixed haploid dataset that allows longer contigs and scaffolds to be constructed (because polymorphisms generally have the effect of fragmenting assemblies). Polymorphisms are added back to the consensus sequence near the end of the process. You must provide ALLPATHS-LG with information on insert-size and standard deviation. We obtained these values by generating an assembly using estimated values, mapping a subset of the reads to this reference using bwa 0.7.6 for fragment libraries(79) and stampy 1.0.23(80) for mate pair libraries. We used picardtools 1.97 (http://broadinstitute.github.io/picard) CollectInsertSizeMetrics to estimate parameters for the fragment libraries and stampy automatically outputs these data for mate pairs.

Additional steps were taken to improve the initial ALLPATHS-LG assembly. First, we identified and softmasked repeats using RepeatMasker open-4.0 (-ggcalc, -species aves(81)). 6.60% of the reference was masked (compared to 7.76%, 9.08% and 7.93% for hooded crow, chicken and zebra finch). Similar to previous assembles of avian genomes, most of the repeat elements were retroelements (5.04%). Next, we identified duplicate scaffolds and contained sequences using bbmap v 35.51 (https://jgi.doe.gov/data-and-tools/bbtools/). We did not find any duplicated scaffolds, but 7 sequences were contained in larger scaffolds (total of 15,625 bp) and removed. Finally, we filled gaps in the assembly using GapCloser v 1.12(82). 45,493 gaps were identified in the assembly, totalling 32,284,017 Ns. We finished 12,079 gaps (10,599 filled with sequences, 47 with zero length and 1,433 with negative length), reducing number of Ns to 26,162,538. We removed contigs that did not blastn to bird targets using the NCBI nucleotide database.

Super-scaffolding with optical maps ALLPATHS contigs less than 70kb in size were excluded from the hybrid assembly constructed using Bionano optical maps. On its own the BspQI map had 1,463 scaffolds, with an N50 of 0.83 Mb and total length of 1,013.85 Mb. When combined with the NGS assembly the number of contigs was reduced to 110 with an N50 of 21.85 Mb and total length of 1,038.58 bp. The BssSI map had 1,181 scaffolds, with an N50 of 0.82 and total length of 824.26 Mb. After re-scaffolding with this map we had our final assembly.

Annotation The *de novo* testis transcriptome used in our annotation was obtained by sequencing mRNAs of 1 young and 4 adult male individuals. Library reads were obtained with mid input of 75 bp paired-end sequencing. From 8 to 35 M reads mapped for each individual with an average of 22 M reads. The assembly was obtained for each individual

separately, using TRINITY with default k-mer parameters (k=25) and a minimum contig size of 300 bp. The different individual transcriptomes were merged eliminating transcripts with 95% similarity among individuals to obtain a final transcriptome.

Four cycles of the MAKER pipeline were run as follows: The first cycle included gene prediction exclusively with EXONERATE and all transcripts and transcriptome as evidence. For the second round we obtained a HMM model to train the SNAP gene predictor and an additional run of EXONERATE with all transcript evidence. The third and fourth rounds included an HMM model obtained from the immediately previous cycle to use it into SNAP and the “chicken” HMM model included in Augustus with default settings. In every cycle, only models with initial and stop codons, and > 50AA were included and accepted.

Genes were functionally annotated using blastp and Interproscan. We used blastp against a database of predicted proteins from ensemble (same species as above) with a threshold of 70% similarity and 80% query coverage. Hits under this thresholds, but higher than 50% in both cases, were flagged with a warning annotation. Second, we run Interproscan with default parameters adding GO terms and Pathways and annotations from the following databases: CDD Gene3D Hamap PANTHER Pfam PIRSF PRINTS ProDom ProSitePatterns ProSiteProfiles SMART SUPERFAMILY TIGRFAM and IPR.

Table S1. Summary of sequencing data used for ALLPATHS-LG assembly. Libraries with a and b are from the same library preparation but sequenced on two separate lanes.

Library	Type	Insert (bp)	Raw (Gb)	Used (%)	Used (Gb)	Sequence coverage	Physical coverage
1	Fragment	174 ± 20	30.3	83.6	25.3	15.3	13.3
2a	Fragment	178 ± 20	24.8	84.6	21.0	12.6	11.3
2b	Fragment	178 ± 20	32.3	85.3	27.6	16.5	14.8
Total fragment			87.4		73.9	44.4	39.4
3	Mate	1,468 ± 223	51.31	64.8	33.2	9.3	61.4
4a	Mate	4,617 ± 406	27.24	54.0	14.7	4.1	65.9
4b	Mate	4,617 ± 406	28.05	52.3	14.7	4.1	71.4
5	Mate	9,230 ± 978	24.46	22.4	5.5	1.5	40.5
6	Mate	9,474 ± 805	57.44	40.1	23.0	2.1	50.7
Total mate			188.5		91.1	21.1	289.9
Final assembly			275.9		165.0		

Table S2. Assembly statistics at each stage. The second ALLPATHS assembly follows the removal of duplicates and contaminants along with gap filling.

	Length	#_seq	N50	Ns	Busco[a] (%)	UCE[b] (%)	UCE[c] (%)
ALLPATHS	1,032,745,135	4,071	9,614,374	3,2284,017	92.1	98.0	98.8
ALLPATHS	1,031,299,298	2,896	16,854,067	27,685,636	93.6	98.2	98.9
+Bionano	1,017,107,033	96	21,997,114	33,034,381	92.8	95.8	96.6

[a] percentage of complete BUSCOs (aves_odb9). Full results:
C:93.2%[S:92.1%,D:1.1%],F:3.9%,M:2.9%,n:4915
C:93.6%[S:92.5%,D:1.1%],F:3.7%,M:2.7%,n:4915
C:92.8%[S:91.7%,D:1.1%],F:3.8%,M:3.4%,n:4915
C = complete, S = complete and single-copy, D = duplicated, F = fragmented, M = missing.
[b] percentage of UCEs identified using whole-genome alignments for three amniotes (chicken, anole and zebra finch, total 5472)
[c] percentage of amniote UCEs with greater coverage (of 2560)

Table S3. Results from satsuma showing which flycatcher chromosome each scaffold in the blackcap reference genome hit. Mean position and orientation refer to the location and orientation of scaffolds on the flycatcher genome. The last 6 scaffolds did not hit any of the flycatcher chromosomes. Comparing the annotation of the blackcap and zebra finch genomes suggests they match the indicated chromosomes.

Blackcap scaffold	Flycatcher chr	Mean location	Orientation
Super-Scaffold_33	chr_1	36485	1
Super-Scaffold_8	chr_1	1339683	-1
Super-Scaffold_37	chr_1	4636214	-1
Super-Scaffold_76	chr_1	8096257	1
Super-Scaffold_4	chr_1	10433711	-1
Super-Scaffold_80	chr_1	11098626	-1
Super-Scaffold_32	chr_10	998131	-1
Super-Scaffold_1819	chr_10	2065253	1
Super-Scaffold_13	chr_11	575478	-1
Super-Scaffold_38	chr_11	1568314	1
Super-Scaffold_57	chr_11	2074346	1
Super-Scaffold_61	chr_12	122961	1
Super-Scaffold_89	chr_12	1181156	1
Super-Scaffold_100	chr_13	811989	-1
Super-Scaffold_79	chr_13	1746253	1
Super-Scaffold_72	chr_14	869082	-1
Super-Scaffold_14	chr_15	747687	-1
Super-Scaffold_26	chr_17	618869	-1
Super-Scaffold_60	chr_18	30886	-1
Super-Scaffold_82	chr_18	689775	1
Super-Scaffold_66	chr_19	599067	1
Super-Scaffold_56	chr_1A	916431	-1
Super-Scaffold_31	chr_1A	2191019	1
Super-Scaffold_63	chr_1A	2340900	1
Super-Scaffold_46	chr_1A	4867478	-1
Super-Scaffold_34	chr_1A	5221726	1
Super-Scaffold_94	chr_1A	5486379	1
Super-Scaffold_78	chr_1A	5863296	-1
Super-Scaffold_48	chr_1A	6119647	-1
Super-Scaffold_58	chr_2	76694	-1
Super-Scaffold_40	chr_2	2567093	1
Super-Scaffold_30	chr_2	10337890	1
Super-Scaffold_23	chr_20	25286	1
Super-Scaffold_9	chr_20	90779	-1
Super-Scaffold_71	chr_20	522682	-1
Super-Scaffold_20	chr_20	1226651	1
Super-Scaffold_110	chr_21	304262	1
Super-Scaffold_44	chr_21	743754	1
Super-Scaffold_90	chr_22	85418	1
Super-Scaffold_27	chr_22	356802	-1

Super-Scaffold_64	chr_22	379551	1
Super-Scaffold_107	chr_23	91984	1
Super-Scaffold_98	chr_23	363554	-1
Super-Scaffold_74	chr_23	669009	-1
Super-Scaffold_54	chr_24	400871	1
Super-Scaffold_25	chr_25	203656	1
Super-Scaffold_7	chr_26	149251	1
Super-Scaffold_6	chr_26	525390	-1
Super-Scaffold_5	chr_27	126850	-1
Super-Scaffold_92	chr_27	336482	-1
Super-Scaffold_19	chr_27	506654	1
Super-Scaffold_51	chr_28	101721	1
Super-Scaffold_68	chr_28	407272	1
Super-Scaffold_99	chr_3	752368	-1
Super-Scaffold_24	chr_3	1202731	1
Super-Scaffold_18	chr_3	3257995	-1
Super-Scaffold_17	chr_3	7826899	1
Super-Scaffold_70	chr_4	229429	1
Super-Scaffold_36	chr_4	695200	-1
Super-Scaffold_88	chr_4	788400	-1
Super-Scaffold_10	chr_4	3064426	1
Super-Scaffold_65	chr_4	4511879	1
Super-Scaffold_35	chr_4	5697803	-1
Super-Scaffold_12	chr_4A	485424	1
Super-Scaffold_105	chr_4A	1544995	-1
Super-Scaffold_104	chr_5	466620	1
Super-Scaffold_52	chr_5	627439	-1
Super-Scaffold_1	chr_5	1502460	-1
Super-Scaffold_55	chr_5	2409007	1
Super-Scaffold_47	chr_5	4131534	-1
Super-Scaffold_73	chr_5	4955029	1
Super-Scaffold_50	chr_6	68382	1
Super-Scaffold_29	chr_6	322622	-1
Super-Scaffold_67	chr_6	2085795	1
Super-Scaffold_101	chr_6	3689565	1
Super-Scaffold_39	chr_7	97585	1
Super-Scaffold_83	chr_7	342163	-1
Super-Scaffold_103	chr_7	2165078	-1
Super-Scaffold_109	chr_7	3884732	-1
Super-Scaffold_41	chr_8	1389136	-1
Super-Scaffold_16	chr_8	2256504	-1
Super-Scaffold_22	chr_9	709437	-1
Super-Scaffold_3	chr_9	2046570	1
Super-Scaffold_11	chr_Z	550092	1
Super-Scaffold_75	chr_Z	1636321	-1
Super-Scaffold_43	chr_Z	3625804	-1
Super-Scaffold_93	chr_Z	5025976	1

Super-Scaffold_69	chr_Z	5275119	-1
Super-Scaffold_49	chr_Z	5618693	-1
Super-Scaffold_28[a]	chr_Z_random		
Super-Scaffold_2[a]	chr_25		
Super-Scaffold_102[a]	chr_Un		
Super-Scaffold_2172[a]	chr_Un		
Super-Scaffold_42[a]	chr_Z_random		
Super-Scaffold_59[a]	chr_3_random		

[a] based on the zebra finch annotation

Fig S1. Principal component analysis matching that in Fig. 1 but excluding island populations.

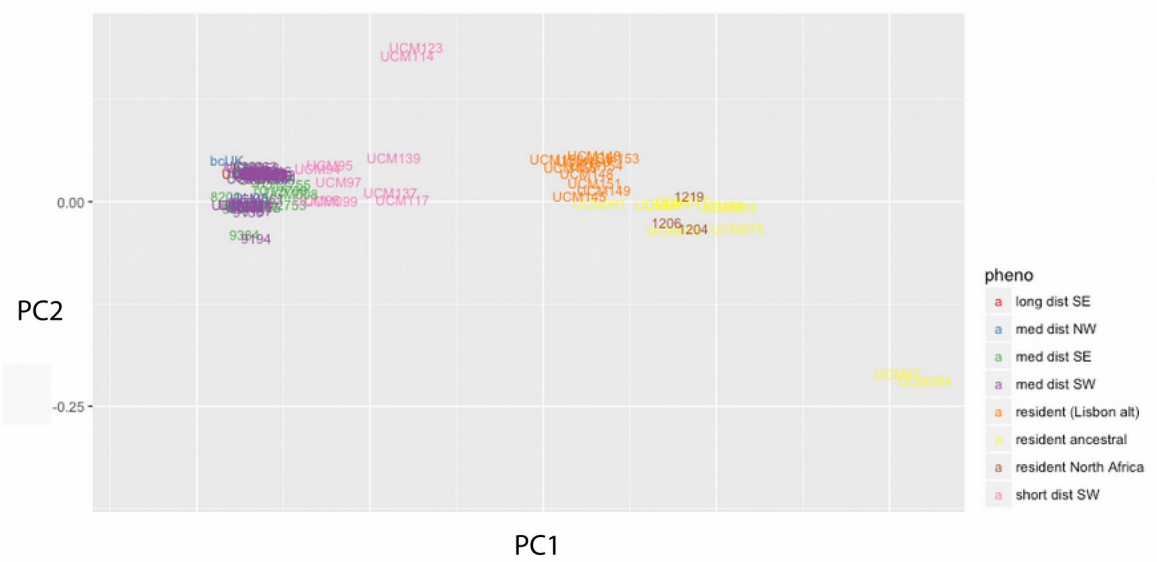


Fig. S2. Complimentary figure to Fig. 1c, showing ancestry proportions estimated by ADMIXTURE at larger cluster values (k=4 through 7).

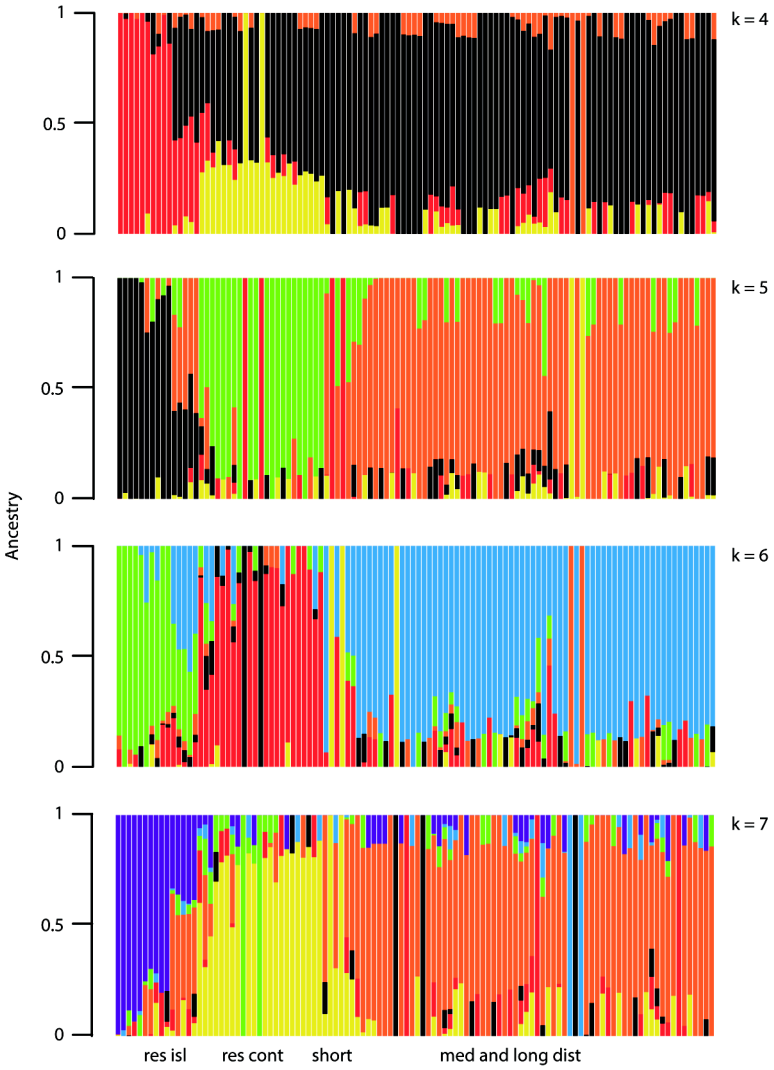


Fig. S3. Down-sampling for demography analysis of effective population size.
Five individuals were randomly sampled from 44 med+long migrants and 19 continental residents for 10 times (down-sampling 1 to 10), which were used for 10 runs of demography analysis with MSMC2. Because there were only five individuals for each of the other four group (short, Azores, Cape Verde, and Azores), the same sample sets were used for all 10 runs of demography analysis. The results of the 10 runs of demography analysis are shown separately. Note that demography estimates of three island populations (red) and short migrants (black) are same across the 10 panels.

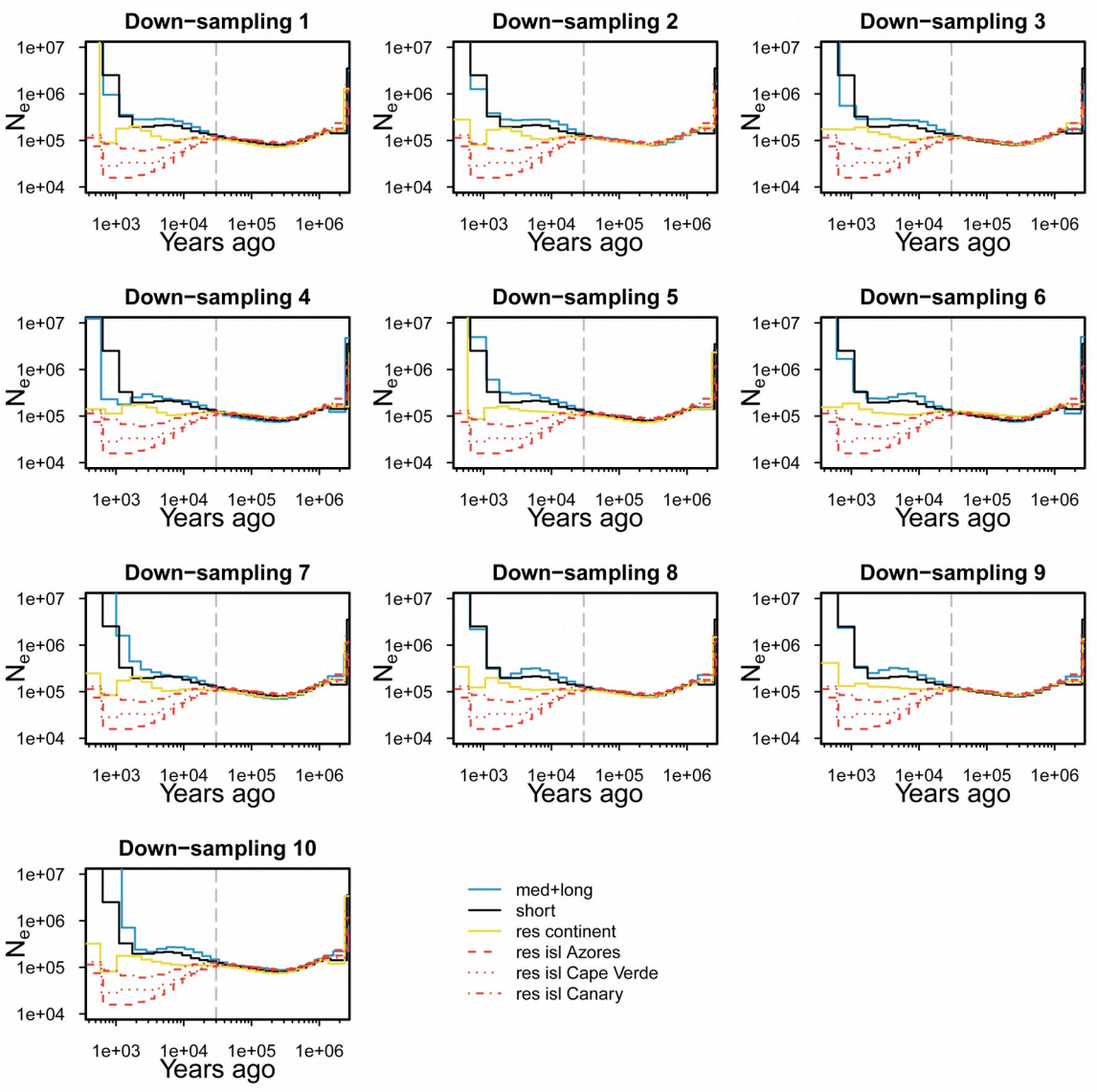


Fig. S4. Down-sampling for demography analysis of relative cross-coalescence rate. The same down-sampled individuals taken for effective population size analysis (Fig. S3) were also used for down-sampling of relative cross-coalescence rate analysis. Although the exact inferences of relative cross-coalescence rate especially between two continental groups (continent vs continent, black) are variable across down-samplings, the general pattern of steeper decline of relative cross-coalescence rate between continental and island groups (continent vs island, gray) than that of continent vs continent is consistent across all 10 down-samplings. Note that some inferences (three of continent vs island and three (all) of island vs island) are same across the 10 down-sampling because the both two phenotypes had only five individuals (see Fig. S5).

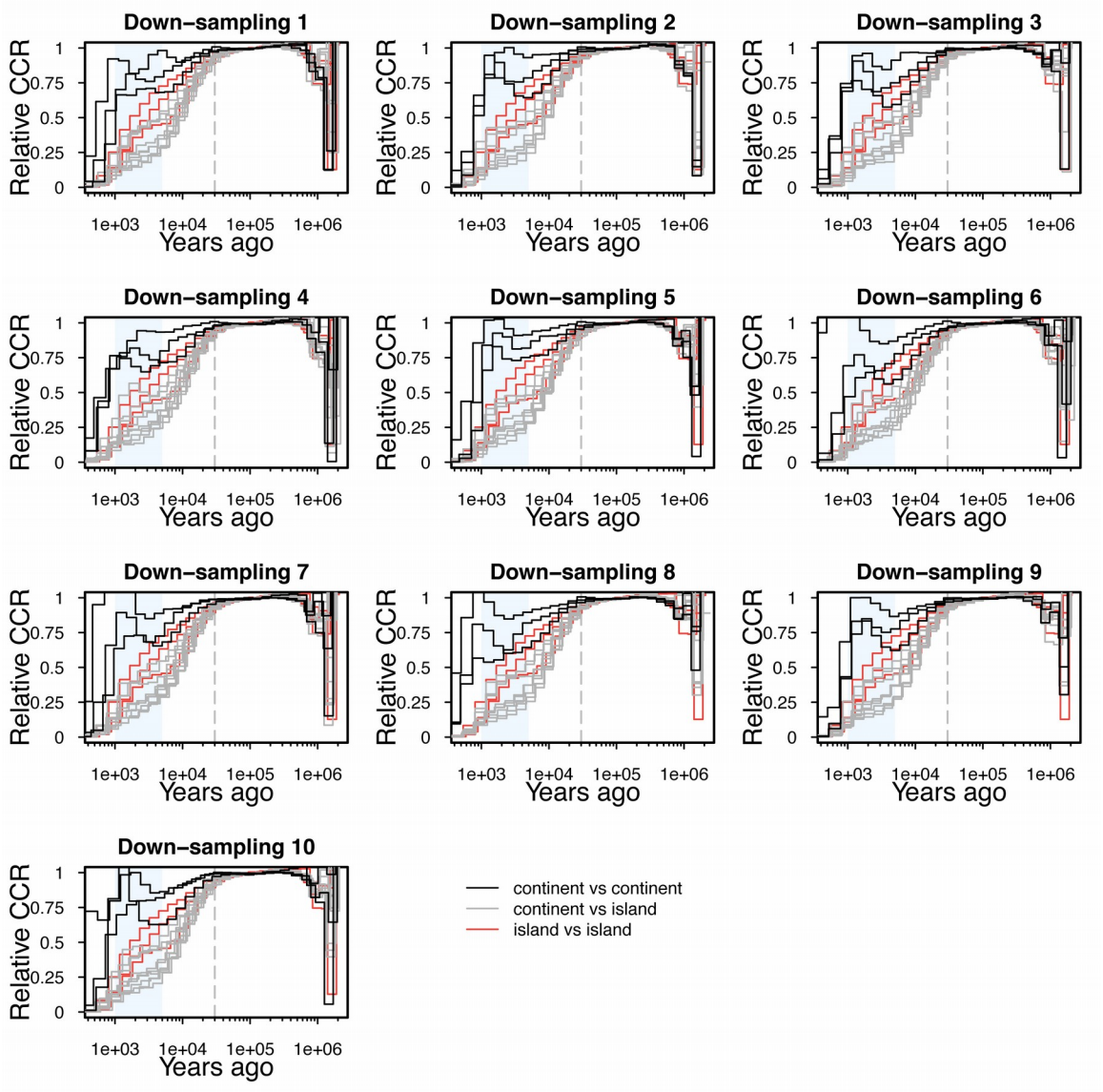


Fig. S5. Demography analysis of relative cross-coalescence rate.
Relative cross-coalescence rate of all 15 possible combinations of six groups are shown. One line represents relative cross-coalescence rate inference of one down-sampling (five individuals per group). The three colours of lines correspond with Fig. 2b and Fig. S4. Relative cross-coalescence rate started to increase at ~5,000 years ago between med+long and continental resident populations (shaded with light blue). Note that there is only one inference for short vs islands (Azores, Cape Verde, Canary) and island vs island (Azores vs Cape Verde, Cape Verde vs Canary, Canary vs Azores) because there are only five individuals for these phenotypes. Also note that the top and bottom diagonals are identical.

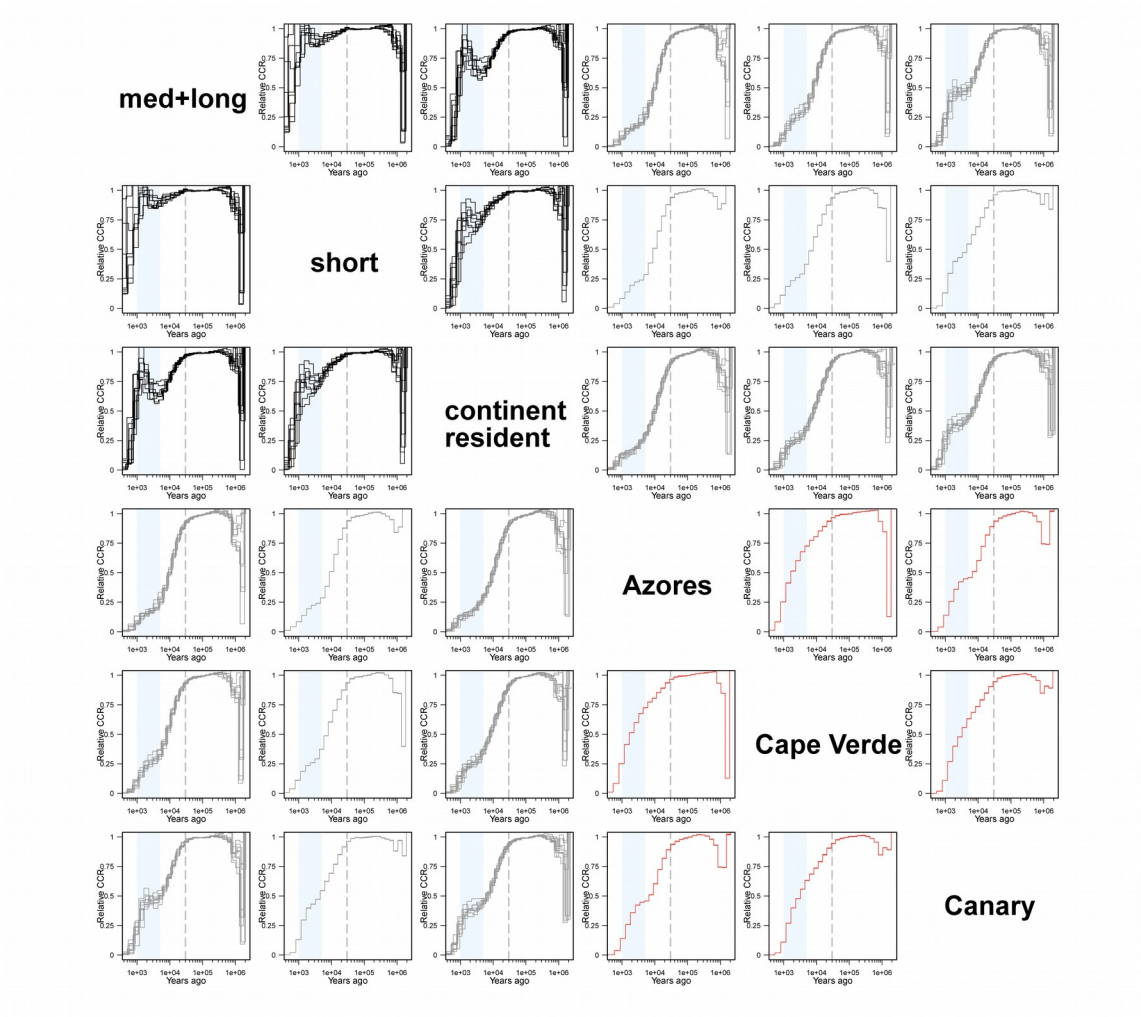
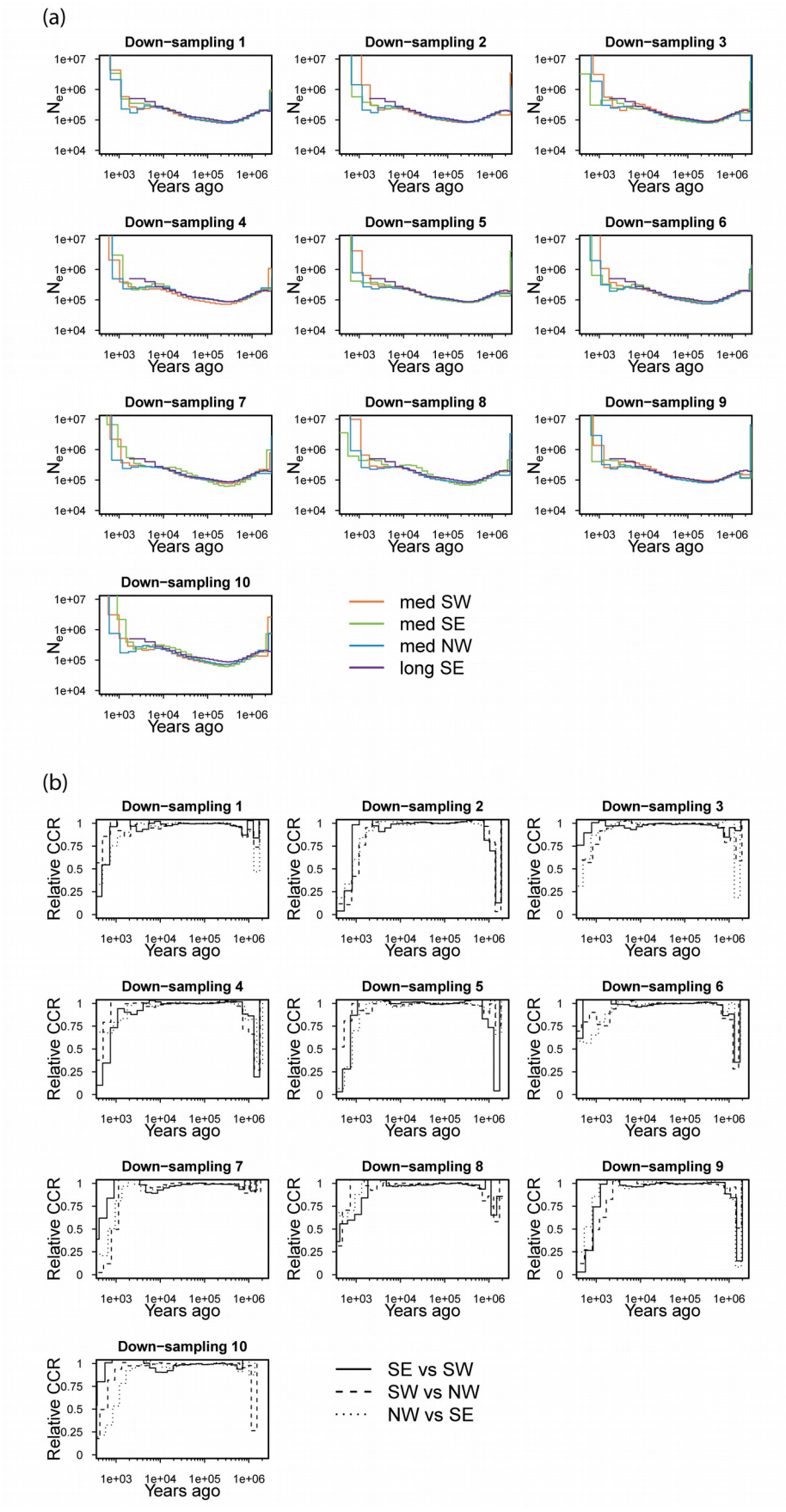


Fig. S6. Long distance migrants and medium distance migrants show similar demographic histories. a) Effective population sizes show the same demographic trajectories. Five individuals were randomly sampled from each medium distance phenotype 10 times (down-sampling 1 to 10), and used for 10 runs of demography analysis with MSMC2. The results of the 10 runs are shown separately. This was not done for long distance migrants as only two individuals met the coverage cutoff to be included in the analysis. b,c) Relative cross-coalescence rates stay high in all three pairwise comparisons between medium distance migrants. Note that the top and bottom diagonals in (c) are identical.


(a)
Down−sampling 1
Down−sampling 2
Down−sampling 3
Down−sampling 4
Down−sampling 5
Down−sampling 6
Down−sampling 7
Down−sampling 8
Down−sampling 9
Down−sampling 10
Ne
1e+07
1e+06
1e+05
1e+04
1e+03
1e+04
1e+05
1e+06
Years ago
med SW
med SE
med NW
long SE
(b)
Relative CCR
1
0.75
0.5
0.25
0
SE vs SW
SW vs NW
NW vs SE

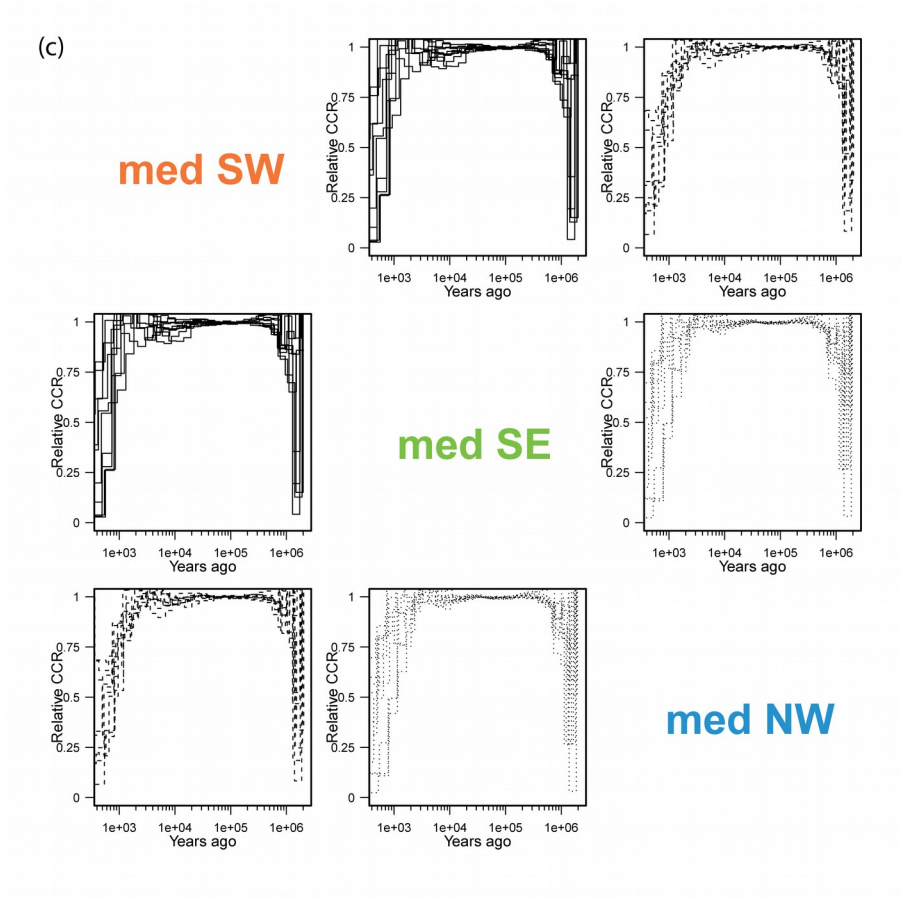

(c)
med SW
med SE
med NW
Relative CCR
Years ago
1e+03
1e+04
1e+05
1e+06
0
0.25
0.5
0.75
1

Fig. S6. Complimentary to Fig. 3b showing genome-wide local estimates PBS for the remaining populations.

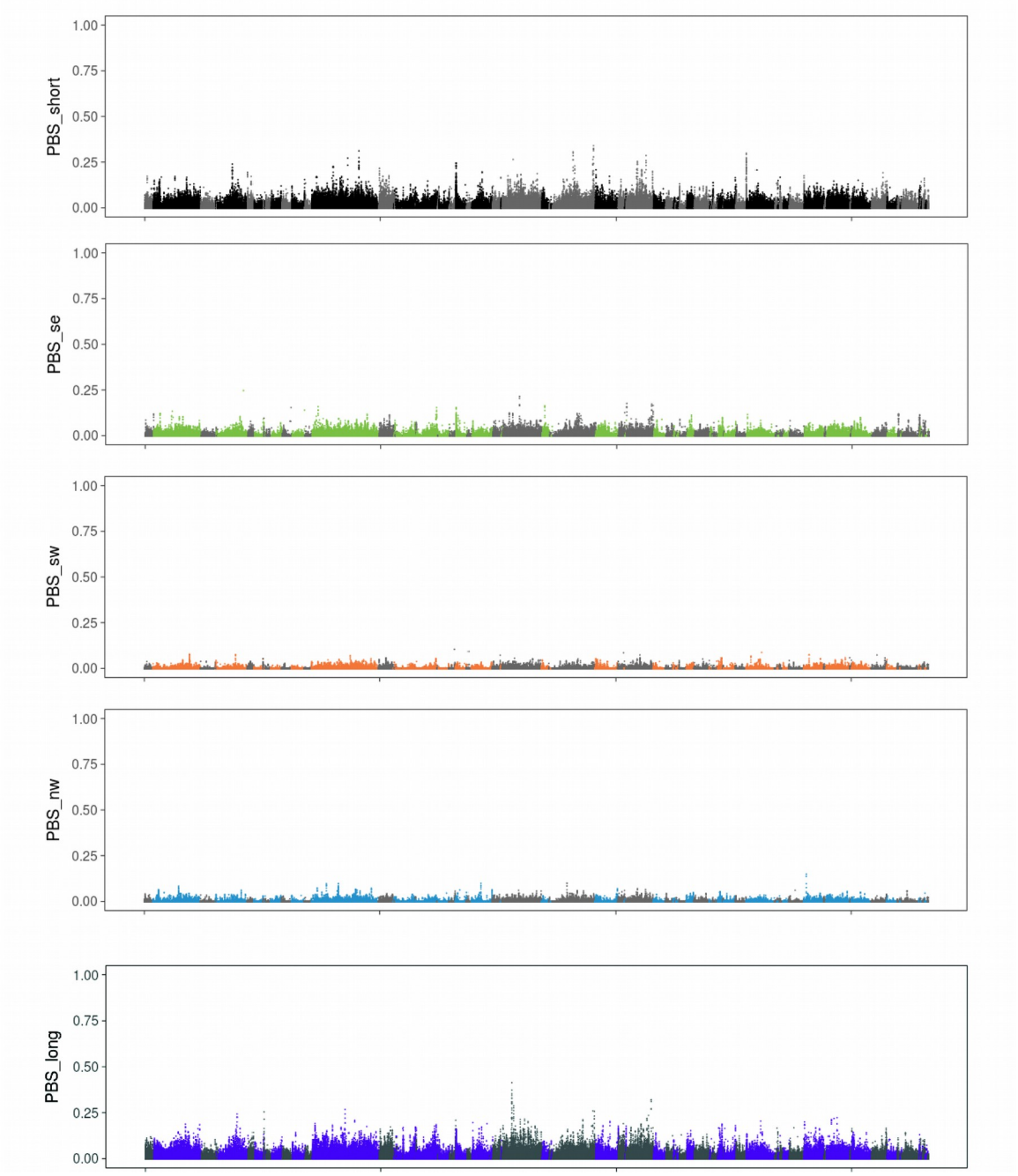
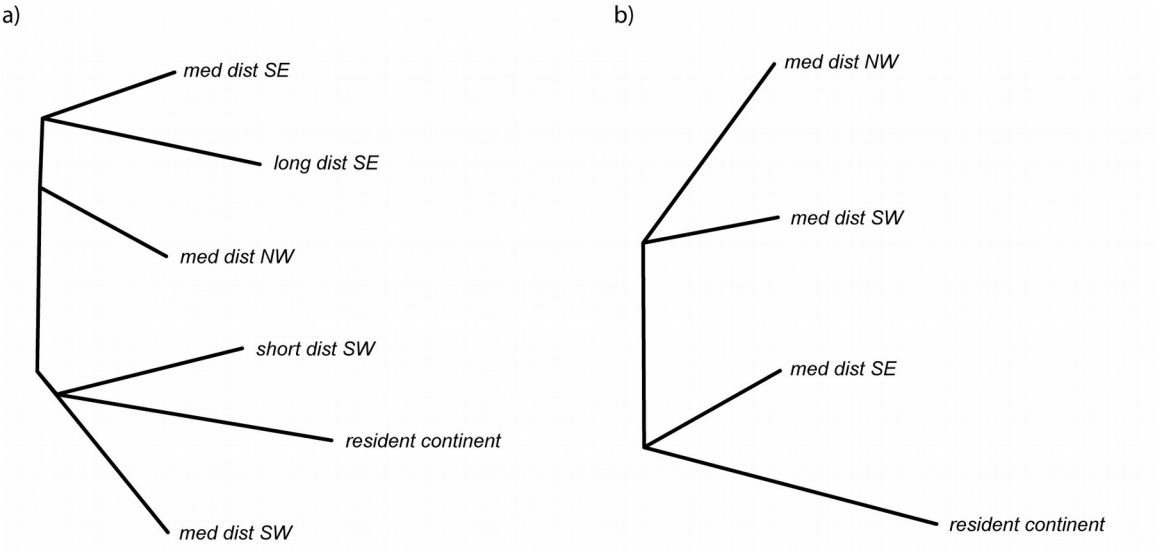


Fig S7. Global neighbor joining trees built hapFLK and data from all genomic regions for comparison with local trees showing positive selection in Fig. 4 (a for the analysis including all phenotypes and b for the analysis limited to medium distance migrants).The resident continent group is only included to root the tree in panel b (i.e., it was not included in the analysis which only focused on medium distance migrants).



CHAPTER 5

Controlling bird migration behaviour through cis-regulatory elements

Juan Sebastian Lugo Ramos*, Gillian Durieux*, Lindsey Caitlin‡, Gregory Gedman‡, Kira Delmore*, Erich Jarvis‡ and Miriam Liedvogel*.

** Max Planck Research Group Behavioural Genomics. Max Planck Institute for Evolutionary Biology. 24306 Plon, Germany.*

‡ Laboratory of Neurogenetics of Language. Rockefeller University. New York City, United States of America.

Contributions: GD,JSLR and ML designed the study. GD, JSLR and KD performed the sampling. GG, LC, and EJ, contribuited with laboratory protocols. ML planned the methodology development. JSLR carried out all sample preparations, performed all analysis, all visualization and wrote the first draft of the manuscript.

Working manuscript. This chapter is planned to be published as one core part of a co-first authored article with G. Durieux jointly analyzing and integrating ATAC-seq and RNA-seq on migratory birds

Controlling bird migration behaviour through cis-regulatory elements

Juan Sebastian Lugo Ramos*, Gillian Durieux*, Lindsey Caitlin‡, Gregory Gedman‡, Kira Delmore*, Erich Jarvis‡ and Miriam Liedvogel*.

** Behavioural Genomic Laboratory. Max Planck Institute for Evolutionary biology. Plon, Germany.*

‡ Laboratory of neurogenetics of Language. Rockefeller university. New York City, United States of America.

ABSTRACT

Bird migration is the quintessential phenotype of animal movement behaviour. Despite the clear demonstration of its strong inherited component, the molecular basis that allow some species to migrate or not have remained entirely unknown. Here, we used an innovative genome-wide approach to investigate differences in chromatin accessibility contrasting migratory birds tested during and outside the migratory, making use of the fact that the migratory behaviour is exclusively exhibited during migratory season. We use ATAC-seq in three focal brain areas that potentially play an assumed role regulating processes related to migratory behaviour: Cluster N, Hippocampus and the anterior ventral region of the Hypothalamus. Intriguingly, our study showed that chromatin accessibility reduced during the migratory season compared to a control condition outside the migratory season. This pattern was particularly noticeable in the Cluster N region, suggesting relevant regulatory processes. We identified potential cis-regulatory elements characterizing the differentially accessible regions in the chromatin. Furthermore, we inferred potential Transcription factors (TFs) changing the gene regulatory landscape during migration. Finally, we leverage this information with population genomic resequencing data to refine identified cis-regulatory elements with features of evolution in our focal species, the Eurasian blackcap.

INTRODUCTION

Migratory animals have the incredible to cope with the challenge of an intense and long journey as part of their annual life cycle. Some sources of evidence support a heritable component for migratory traits like distance and orientation in birds[1],[2]. In insects like monarch butterflies, migration is multigenerational meaning that an individual starts the journey, but it is the fifth offspring generation that complete the travel back. Despite the quantitative genetics demonstrating the heritability of migration, the molecular and genetic mechanisms underlying this, remain unknown so far.

Whole genome resequencing approaches identify highly differentiated markers in non-coding regions between bird populations with contrasting migratory traits. Populations of migratory bird species exhibiting different orientations in Swainson's thrushes *(Catharus ustulatus)* have broad, divergent genomic regions - so-called islands of differentiation -, without affecting protein-coding sequences[3]. In willow warblers *(Phylloscopus trochilus)*, populations with different patterns of orientation strategies have fixed genomic structural variation. Inside the structural variants, the SNPs with the highest differentiation map upstream or downstream genes, but not within protein-coding regions[4]. Genetic associations found that upstream genes related to behaviour and cell signalling, are strongly associated with environmental variables in breeding areas in the yellow warbler[5]. More recently, in golden warblers (*Vermivora spp.*), a single region harbouring the gene VPS13A has been suspected of selection between two different migratory subspecies[6]. Similarly, in the Eurasian blackcap *(Sylvia atricapilla)* a species with a large repertoire in migratory behaviour, and thus ideal to study the genetics of migration, we have previously explored the evolutionary history and genetic associations using a population genomics approach (Chapter 4). Despite the overall low population structure in the migratory populations, most SNPs associated with migratory traits map to non-coding regions of the genome.

Many complex traits –like migration- may rely on cis-regulatory elements. Open Chromatin Regions (OCRs) are cis-regulatory elements where in a given

circumstance, the DNA is completely naked without nucleosomes and thus accessible for Transcription factors (TFs) [7]. OCRs have a special place in regulation because they might harbour relevant regulatory sequences[8]. Correlational evidence shows that TF bind more often to these regions compared to regions of condensed chromatin. Furthermore, there is mounting evidence of the significant contribution of non-coding variants in OCRs to complex diseases phenotypes like schizophrenia and autoimmune responses[8]–[10]. In non-model organisms, the study of chromatin accessibility has allowed to investiate the evolution of traits like flight loss[11] or limb loss [12] and their genetic components and mechanisms. Even more recently, this approach has started to unveil evolutionary mechanisms of adaptation. Some studies describe the pleiotropic effects of different cis-regulatory elements and their role shaping the landscape of evolutionary adaptation in broadly distributed regions of cis-regulatory elements[13], [14].

To investigate the role of cis-regulatory elements shaping migratory behaviour, we characterize the dynamics of OCRs to identify genomic sequences and their potential role controling the expression of migratory behaviours. Furthermore, we explore the population genetics of these elements to narrow down potential cis-regulatory elements with SNPs potentially disrupting the TF-DNA interaction. Specifically, we have performed ATAC-seq in three focal brain regions to contrast chromatin accessibility of birds during migratory season when migratory behaviour is exhibited, and birds outside the migratory season as a control condition during which the migratory behaviour is not shown. We found that all the brain regions present loci of Differentially Accessible Regions (DARs) (i.e. regions changing accessibility in either ON or OFF season samples) when birds migrate and when birds do not migrate. One of our key results is that the Cluster N region harbours most of the differences between the migratory states. This region is known to be involved in magnetic compass orientation in nocturnal migratory birds, such as the blackcap. The genes found close to DARs play a role in pathways of axon guidance, generation of neurons and energy expenditure regulation. A further examination of the potential motifs for TF binding to those regions identified Rev-erb alpha and THR beta are enriched motifs in the DARs of ON season birds. Lastly, we combined the DAR information and motif TF inference with polymorphisms in the genomic sequence to narrow down and

identify migration specific cis-regulatory modules and the evolutionary processes that shape migration.

RESULTS

Chromatin accessibility profiles in the brain of a migratory bird.

We aim to characterise the Open Chromatin Regions (OCRs) that are potentially controlling migratory behaviours in the Eurasian blackcap. We approach that characterisation performing ATAC-seq in brain regions using a "common garden" experiment to classify birds during the migratory season (hereafter, ON) and a control condition with birds out of the migratory season (hereafter, OFF)(see Fig 1A). We characterised OCRs in three brain regions with potential roles on migratory behaviour: Hippocampus (HC); Cluster N (CN) and the Ventral Anterior Hypothalamus (VAH) (Fig1B). With this design, we end up with six different groups that classify the three brain regions in ON or OFF migratory states. We generated and sequenced libraries for a total of 36 samples, as well as two input controls (naked DNA). All samples were sequenced to an average of 50 million reads (Table 1). After removal of PCR duplicates and reads mapping to multiple locations, the average number of reads was 32 million. Table 1 describes other quality measurements like Fraction of reads Inside Peaks (FrIP), PCR Bottleneck (PBC) and reads mapping to the mitochondrial genome (Mitoreads). These previous measurements are quality control of filters commonly used in the ENCODE standards [15] . The quality control filtering reduced the final number of samples to 6 ON and OFF samples for Hippocampus, 7 ON and OFF for Cluster N, and 3 ON and OFF samples for VAH. The quality of our dataset, is comparable to other similar preparations of ATAC-seq in humans, employing the same quality and characteristics of flash-frozen bulk brain tissue [9]. (See table 1).

Due to the limited number of samples from the VAH area available for this working manuscript, and hence lower power in this region to reliably detect OCR, the subsequent results and discussion presented in this manuscript will mainly be focussed on data from the Cluster N and the Hippocampal samples that to allow for more robust conclusions. This work is ongoing and more samples of the

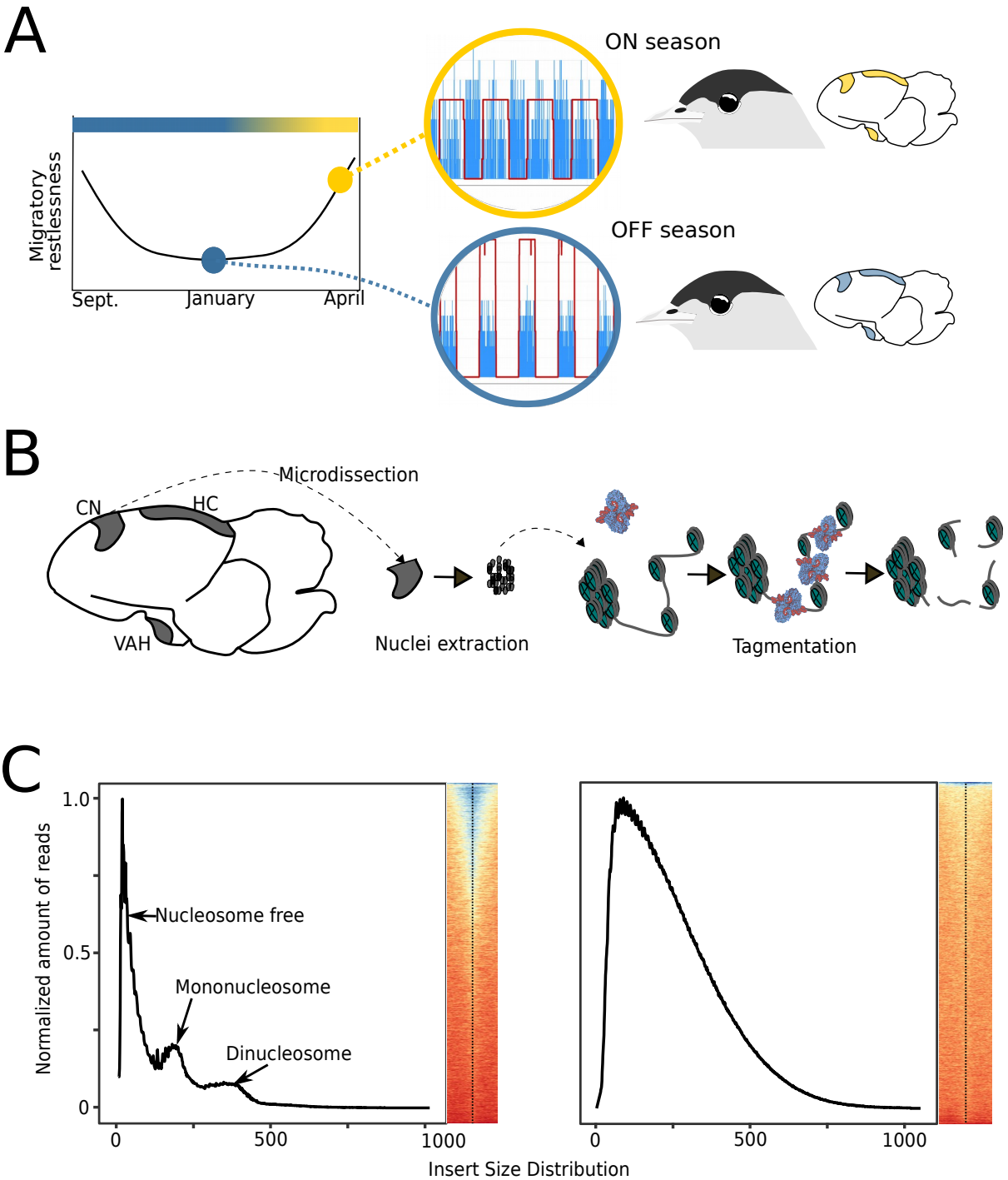


Figure 1. Experimental overview. A) Outline of the experimental design. Birds were sampled OFF (outside) and ON (during) the migratory season (the focus here is in spring migration). Each bird was classified into either OFF or ON season, using nocturnal migratory restlessness activity as proxy for migratory activity (blue bars in circular insets represent activity profiles, the red line indicates light intensity, i.e. valleys represent night while peaks represent day). Migratory restlessness is expressed only during the migratory season. Outside migratory season blackcaps are diurnal and in consequence, birds OFF season show little to no activity during the night (blue inset). Once spring progresses and

the migratory season approaches, birds start to show nocturnal migratory restlessness activity, allowing to classify them as ON season (yellow inset). B) Birds were sampled 1 to 4 hours after onset of darkness, brain tissue was microdissected and immediately flash frozen. Collected tissue was subsequently processed with ATAC-seq to perform the enzymatic tagmentation. C) Quality assessment of samples. Left panel is from a chromatin sample maintaining the nucleosome structure. Right panel is for a sample on naked DNA without nucleosome structure (hereafter, background/BACK). Plots show the insert size distribution indicating the approximate location of mono- and dinucleosomes. The heatmaps to the right of each plot show the frequency of reads around Transcription Start Sites (TSS, dashed line) +/- 1000 bp. Each row in the heatmap corresponds to a single gene. In the heatmap, blue colors indicate high frequency, while red colors indicate low frequency of reads.

VAH will be processed in the near future to complete the dataset for final analyses and publication.

The distribution of insert sizes in each sample passing QC shows the conservation of at least the mono nucleosome portion of the chromatin (see Fig 1C and Fig 1S1). Similarly, a higher frequency of reads near to Transcription Start Sites (TSS, commonly known as highly open chromatin regions), confirms the success of ATAC-seq in the samples (Fig 1D, Fig 1S1). In the naked DNA samples, the same procedure outputs a distribution of insert sizes and read frequency in TSS entirely different to those expected for a high-quality ATAC-seq sample; in these naked DNA samples, reads do not accumulate around TSS, further confirming the success of the ATAC-seq in our samples (Fig 1S1).

A correlation analysis of read numbers in windows of 10 kbs across the whole genome, evidence that the ATAC-seq procedure is reproducible witihn the samples. The average Pearson's correlation value of samples is r^2=0.81 (Fig 1S3). This value is significantly higher compared to samples correlated with the background, naked DNA (r^2=0.62 average of red distribution in Fig 1S2). The correlations are neither significantly different between the samples nor different brain regions or migratory states (Fig 1S2), which could suggests an overall degree of similarity between all the different groups.

Table 1. Quality assessment of each sample.

	Sample	Sequenced read	reads aligned	Frip	PBC	Read Mito	Peaks
CN_OFF	CN_OFF_1	34599118	23803294	0.2175	0.87	9228490	21056
	CN_OFF_20	60101530	42305706	0.2473	0.9	16585608	35293
	CN_OFF_24	52034204	34238380	0.3409	0.87	16227278	33308
	CN_OFF_33	54029188	32233364	0.3280	0.95	20228489	42532
	CN_OFF_35	31845383	29049559	0.2052	0.84	1585607	20261
	CN_OFF_6	92172657	64776833	0.2254	0.98	25827277	43131
	CN_OFF_86	45393558	27597734	0.3928	0.97	16228488	49182
HC_OFF	HC_OFF_13	42174899	24379074	0.2304	0.84	16585606	30320
	HC_OFF_26	36161639	32365815	0.2277	0.85	2227276	36049
	HC_OFF_41	40154274	23358455	0.2609	0.96	15228487	40945
	HC_OFF_44	55125942	32830118	0.2285	0.93	21085605	38134
	HC_OFF_45	39442958	21647134	0.2542	0.87	16227275	31451
	HC_OFF_46	41650978	23855154	0.2562	0.89	16228486	35764
VAH_OFF	VAH_OFF_00	34870207	27074333	0.2285	0.81	6585604	22209
	VAH_OFF_39	48080209	30284385	0.2055	0.77	16227274	19108
	VAH_OFF_89	39088100	33292276	0.2310	0.8	4228485	21090
CN_ON	CN_ON_21	92094669	60798845	0.2056	0.87	29728484	16908
	CN_ON_3	43923768	26127944	0.2476	0.92	16585602	25948
	CN_ON_30	41284377	24488550	0.2441	0.87	15227272	19852
	CN_ON_4	45135113	27339289	0.2422	0.99	16228483	25583
	CN_ON_43	46291211	27795391	0.2611	0.83	17285601	20915
	CN_ON_49	60624797	40828973	0.3094	0.94	18227271	29674
	CN_ON_56	96006551	50910727	0.2903	0.92	43528482	29532
HC_ON	HC_ON_12	49973844	30178020	0.2238	0.83	18585600	17973
	HC_ON_14	49735092	31939268	0.2725	0.84	16227270	23191
	HC_ON_23	42028791	24232967	0.2830	0.91	16228481	24994
	HC_ON_28	45394127	26798303	0.3148	0.96	17385599	25786
	HC_ON_31	36350944	24855125	0.3097	0.87	9927269	25992
	HC_ON_7	49253058	30457234	0.2872	0.90	17228480	25812
VAH_ON	VAH_ON_87	61076614	39280790	0.2159	0.78	20585598	16860
	VAH_ON_9	49936667	32140843	0.2114	0.72	17795824	15248
	VAH_ON_99	44465111	27669295	0.2341	0.77	16795824	18437
BACKGROUND	BACK_62	40257858	27468291	0.0456	0.81	9421363	455
	BACK_58	35650915	24456271	0.0987	0.74	8143677	207
NO_QC	CN_OFF_72	45147358	34769233	0.1615	0.64	10376640	9201
	HC_OFF_22	51473548	37816221	0.1754	0.71	13655842	8596
	HC_OFF_42	41215699	28193345	0.1642	0.68	13020869	10622
	VAH_OFF_59	35129844	30184566	0.1249	0.52	4943793	9711

Total of sequenced reads, and reads uniquely aligned are shown. Fraction of reads in Peaks (FRIP), PCR bottleneck coefficient (PBC). Reads aligning to the mitochondrial genome and the total of peaks identified. Four samples do not pass the quality controls and therefore they are not included in the study. Our criteria include FrIP>0.2 and more than 15000 peaks identified.

The identification of OCRs shows an average of 22716 and 32481 OCRs for ON and OFF season, respectively (see table 1). The sum of OCRs length accounts on average for 5% of the genome in both groups combined. Approximately 80% of

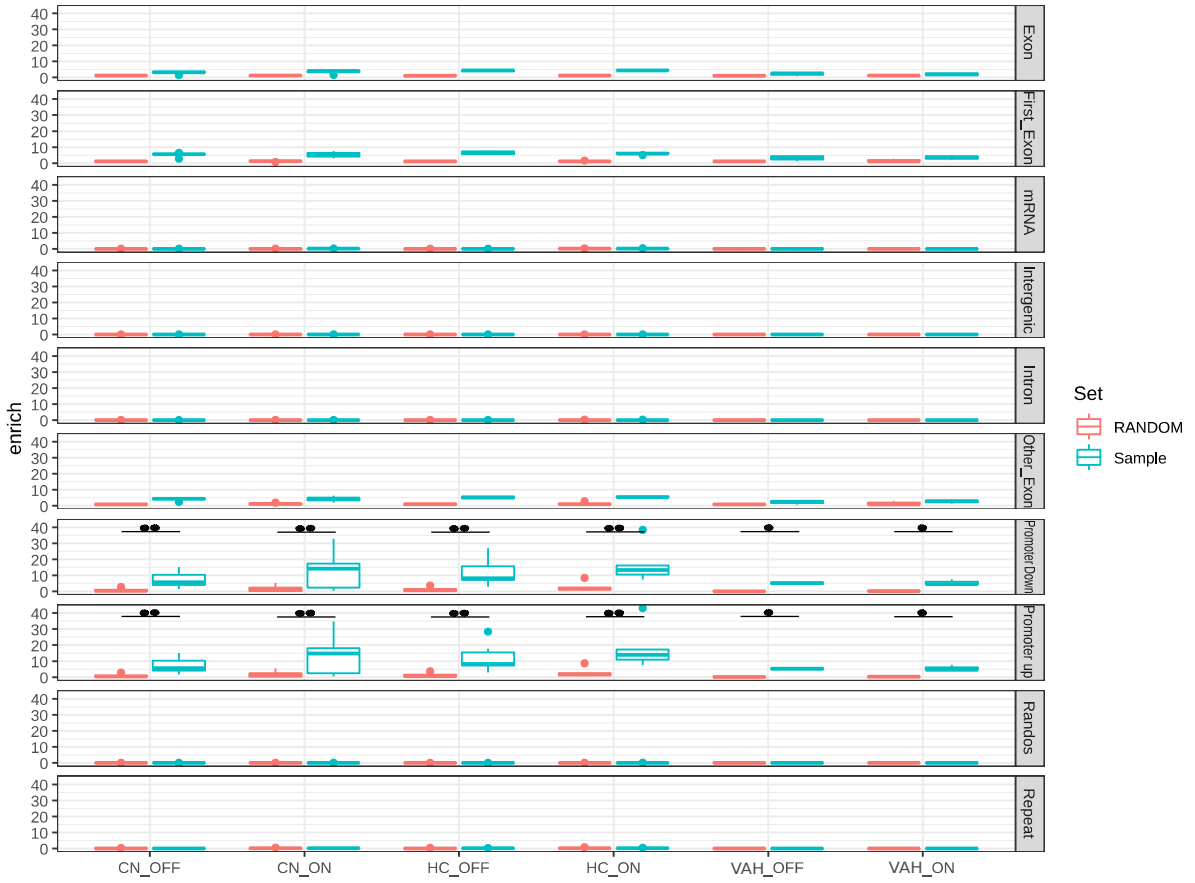
OCRs in each sample are present in 3 or more samples. The presence in other samples and the absence in the naked DNA samples accounts for the reproducibility of each OCR (Fig 2S1A). The proportion of overlap in three or more samples within the same migratory state and brain region is 50% for each OCR (for VAH groups, the 50% are found in two samples Fig 2S1B). The overlap with other samples supports the scenario that the chromatin accessibility landscape has shared features among the different brain areas.

To investigate whether the OCRs identified are potential cis-regulatory regions. Therefore, we evaluated the annotation of OCRs and their conservation. For the former, we compared the overlap of OCRs and Genomic annotations for each sample with a set of the same number of OCRs, shuffled around the genome (Random regions). In our samples, the enrichment of OCRs is significantly different from randomly placed regions, in promoters (defined as 500bp up and downstream the TSS) except for VAH regions with a marginal significance (recall the low sample sie in this region) (See Figure 2).

Migration induces widespread chromatin changes in the brain of migratory birds.

We were interested in identifying the differences in chromatin accessibility that migratory behaviours might induce in the focal brain regions. We can describe the set of OCRs in a sample as an OCR landscape. To assess the similarity of OCR landscapes between all samples, we obtained a matrix of similarity using the Jaccard distance as a metric of overlap between two samples. Clustering analysis of the similarity matrix reveals three main clusters: one with mostly samples from the OFF season, one with mainly ON season samples, and one with mixed samples mainly from the VAH region (Fig 3). A bootstrap analysis supports the 3 clusters mentioned above with values of 90 or higher (Fig 3S1). The degree of similarity does not cluster samples by brain region suggesting that several OCRs might be shared between them(See also Fig 2S1).

It is important to note that the differences might not be limited to the presence/absence of OCRs. The differences might be rather influenced by the degree of accessibility. We quantified the number of reads in each OCR as a proxy for the



Fiigure 2. Peak (Open Chromatin Regions (OCRs)) annotation with features of the genome. The amount of overlap was addressed in the different groups of samples (blue). Additionally, a random sample generated shuffling the peaks around the whole genome was generated for comparison (red). Each genomic feature is noted to the right side of each row. The degree of overlap takes into account the number of bp overlapping for each feature, normalized by the length of such feature in the genome. The genomic features evaluated are: Exons, first exon only, gene bodies(exon+intron), intergenic regions (excluding RepeatMasker regions), introns, other exons, promoter upsteram, promoter downstream (defined by +/- 500 bp from TSS, respectively) random regions in the genome and repeatable elements (identified with Repeatmasker). Significance compared against the random samples is indicated with a Mann-Whitney U test **pval<0.01 *pval<0.1.

intensity of accessibility of the OCR. As a measure to reduce dimensionality, we use a Principal Component Analysis (PCA) of all the samples. We found that the main driver of variance is the difference between OFF and ON migratory season. PC1 is the component that explains most of the variance of ON or OFF samples

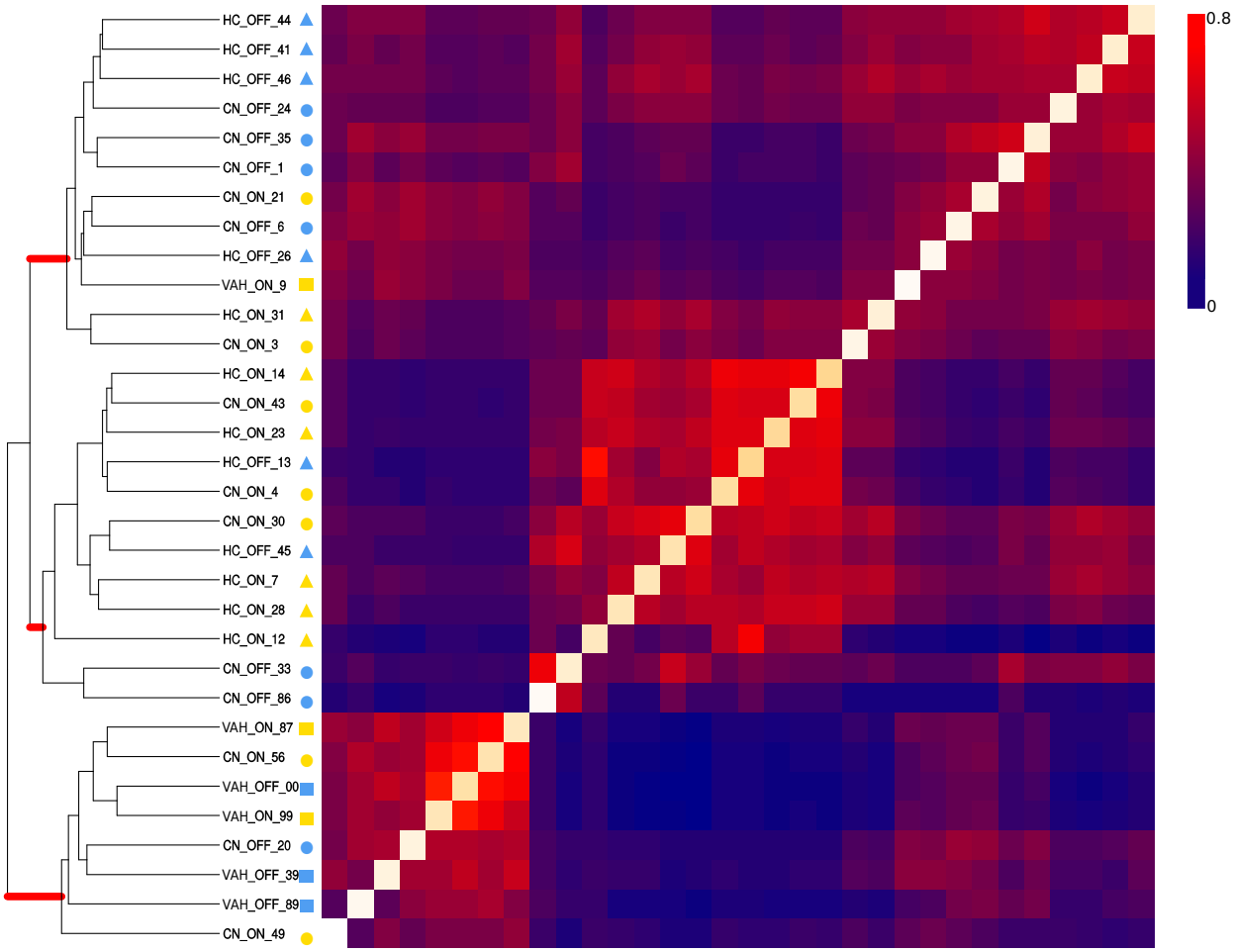


Figure 3. Chromatin accessibility landscape similarity. The heatmap represents the degree of overlap of all the peaks between two samples. High values of similarity are represented in red. A pair of samples having exactly all the same peaks would be a value of 1 (diagonal). The tree to the left represents the clustering pattern of samples, obtained with a bootstrap analysis with 1000 re samplings to support the groups. Red lines show the support of clusters of more than 90%. The different regions are represented with shapes. Hippocampus (triangle), Cluster N (circles) and VAH (squares). The migratory status is represented with colors: blue (outside migratory season OFF), yellow (during migratory season ON)

(Fig 4). To evaluate whether the difference between ON and OFF came about randomly, or by an accurate biological signal in the identified OCRs, we performed the same PCA analysis in randomly shuffled OCRs around the genome (Fig 4S1). The results confirm that the variance obtained in the identified OCRs is not due to any systematic bias between the samples. To address the likelihood of random associations of the ON/OFF migratory states being the main drivers of

the variance, we permuted 100 times the ON/OFF state of the samples and performed the PCA analysis as before. The expected distribution of values in the PC1 has lower variance than the observed distribution (Fig 4S2), confirming that the migratory state mainly drives the variance in our data set.

The gene regulatory landscape of migration induces a general chromatin repression.

The chromatin remodelling happening during migration is changing the gene regulation mechanisms in the three focal brain areas. To obtain a detailed picture of the potential mechanisms, we obtained a consensus set of OCRs (see Methods) for each brain region. Subsequently, we quantified the differences between birds OFF and ON migratory seasons. We identified differentially accessible Regions (DARs) in either of the migratory states. The observed pattern shows higher chromatin accessibility in OFF season birds for the Cluster N and Hippocampus regions (Fig 5A and B), which means potential repression of chromatin in the ON season samples. We performed permutations of the OFF and ON states of the samples to verify that the DARs are not a result of random associations of the migratory phenotype. The distribution of p-values in the real dataset shows a skew towards small p-values, while permutated associations have an even distribution of p-values (Fig.5S1). This shows that the real data set contains significant DARs while in the permutated dataset there are no DARs or those that are significant are portentially false positives.

Similarly, we confirmed that the observed trend is not due to an outlier sample distorting the overall accessibility pattern. We used a Jack-knife approach (also known as Leave one out (LOO) procedure) to confirm that the distribution of Fold change is not significantly different to the whole dataset (Mann-Whitney U pval > 0.1) (Fig 5S2). We again compared our data with randomly picked regions along the genome, to test if the accessibility pattern is systematic in the samples or if it is intrinsic to the OCRs found. This analysis shows that the fold change distribution is only skewed towards the OFF season samples in the consensus OCRs set but not in the randomly picked genomic regions. This confirms that the change on accessibility is due primarily to an effect on the OCRs and not the whole sample (FIG5S3).

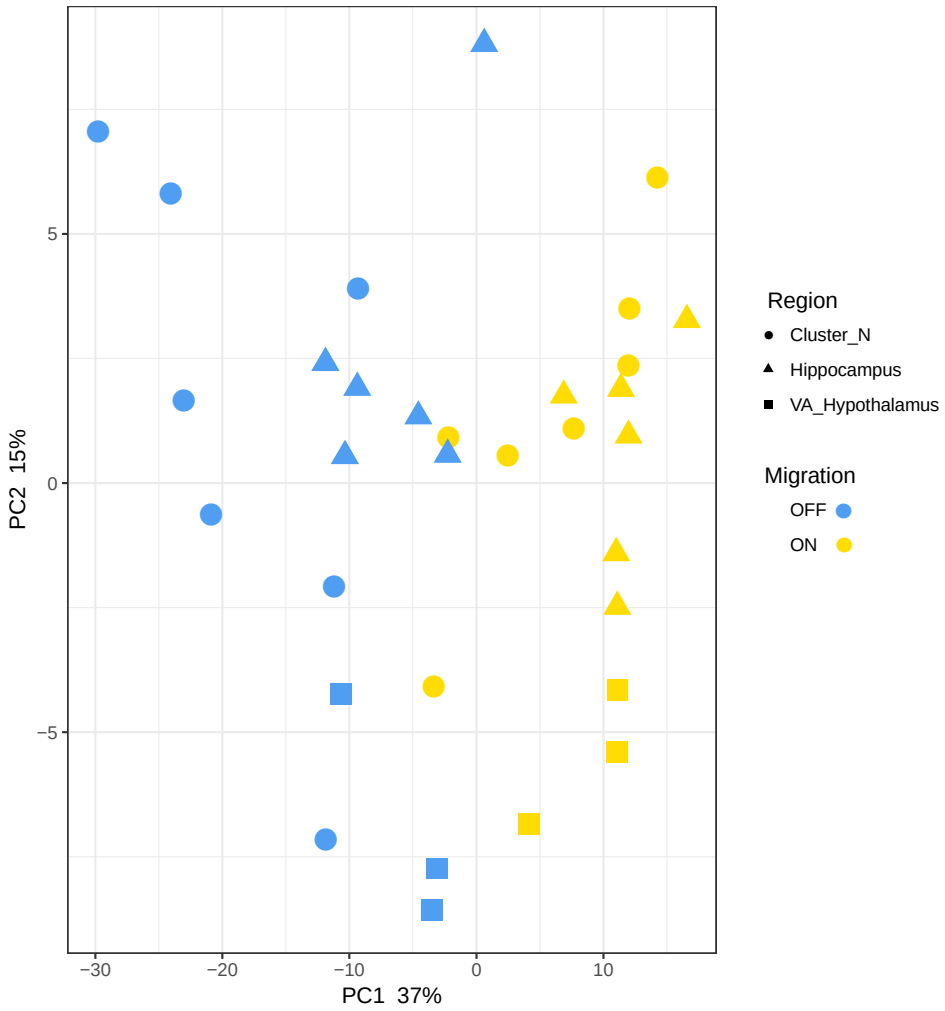


Figure 4. Principal components analysis of read counts in a consensus set of all peaks in all regions and migratory states. The main variability among the samples is dictated by differences in migratory state. Shapes and migratory status are represented as in Figure 3.

It could be possible that a lack of resolution could influence the change in the ON season samples due to degradation or any other factor affecting the quality of the sample. However, we also confirmed that the samples behave similarly compared to a naked DNA sample (which would resemble a completely degraded sample) [7](FIG 5S4). We performed a differentially accessible analysis in the OCRs identified, comparing ON or OFF season samples with the naked DNA samples. Samples from the ON and OFF season have similar patterns of differential accessibility. This confirms that the pattern of differential accessibility in the brain regions is not an artefact of the status of the sample.

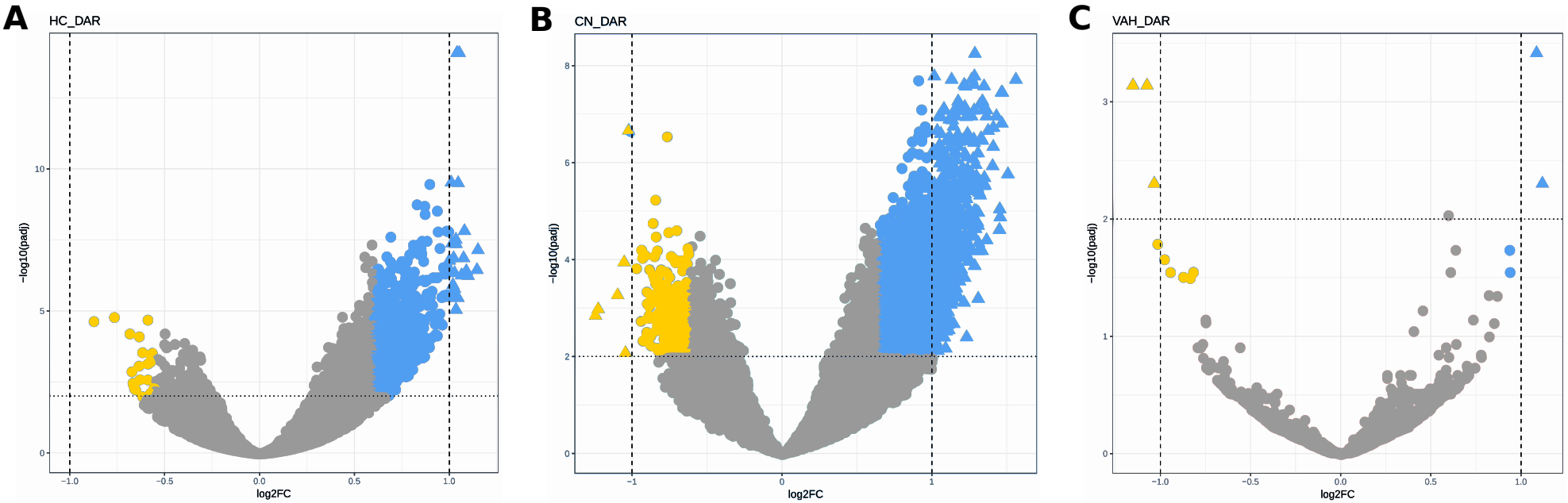


Figure 5. Volcano plots identifying differential accessible regions (DAR) fore each region. A) Hippocampus B) Cluster N, C) Ventral anterior hypothalamus (VAH). In each panel yellow colored characters are DARs significantly more accessible in the experimental group of birds tested during migratory season (ON). Blue characters signify DARs more accessible in control birds tested outside the migratory season (OFF). Significance of DARs are represented with two thresholds settings: triangles denote the strict threshold (padj<0.01 and log2foldchange >1, also denoted by dashed lines) while dots indicate an empirical threshold (see text for details).

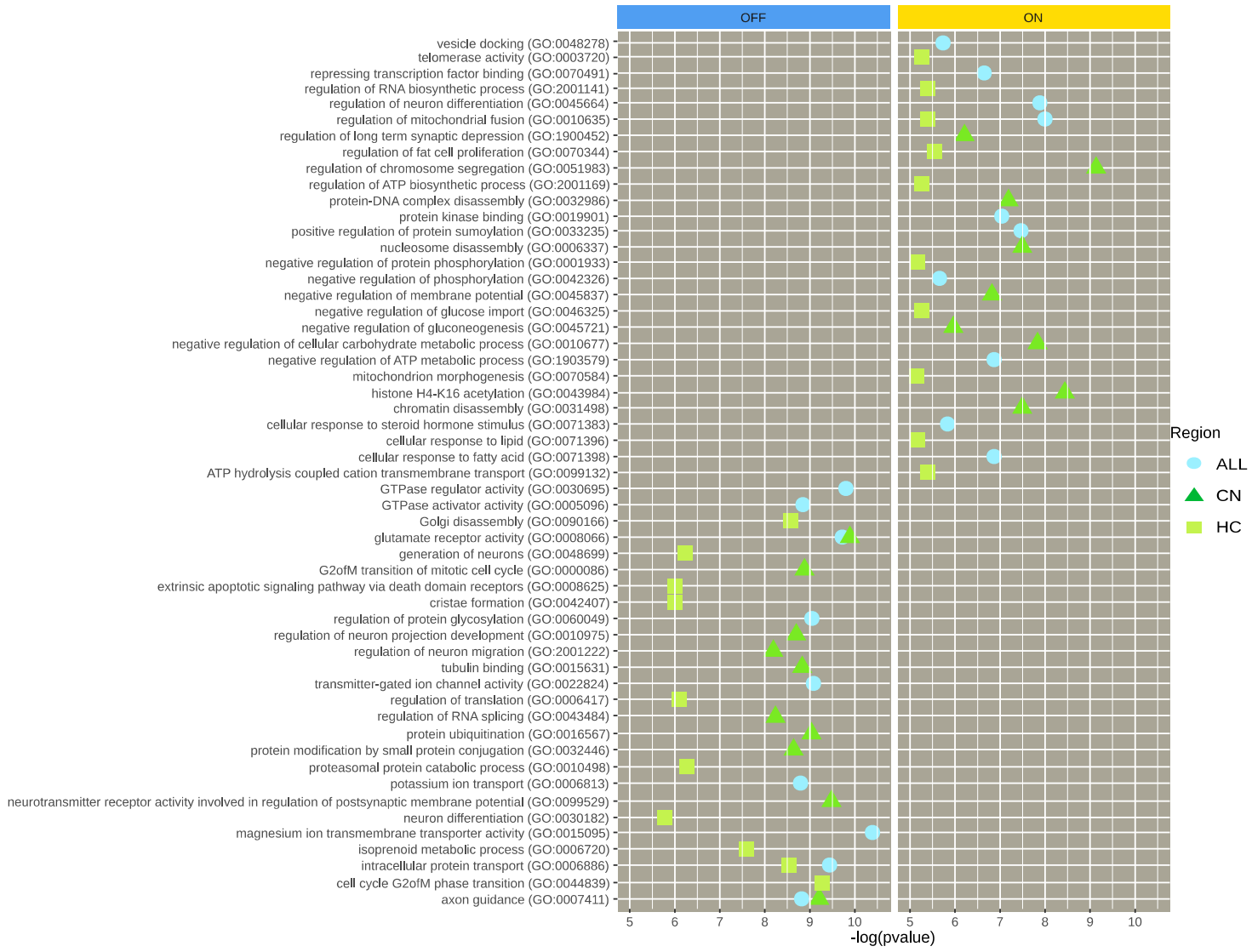


Figure 6. Gene ontology (GO) enrichment of closest genes to each DAR by brain region. Each migratory state is in different panels. Each shape, indicates a different brain region (VAH not significant due to limited sample size) We included a set of genes merging the DARs in all brain regions.

It is surprising that we find a general pattern of repression in the ON season samples. Only few chromatin regions are selectively open only during the migratory season samples but not in OFF season samples, which agrees with the idea that many physiological processed are shut down or downregulated during the energetically demanding process of migration. [2], [16]. It is important to note that this skew was most prominent in the Cluster N region, which suggests that this region goes under an extreme change of regulation during the migratory season and highlights its potential role in controling migration behaviour (Fig 5B).

By assigning the closest gene to any DAR, we can infer patterns of regulation changing at the gene level. With an enrichment of Gene Ontologies (GO) of those genes close to regions with differential accessibility, we confirm the pattern of repression and negative regulation of many processes in the brain of migratory birds (Fig 6). The hits that are more accessible between off and on are mainly part of opposite pathways. The OFF samples have GOs enriched for normal processes enhancing neuronal development, and regular cell division (Fig 6). The opposite pattern is present in ON samples presenting considerable repression of energy expenditure and negative regulation of membrane potential and regulation of long term synaptic depression a phenomenon involved in the formation of memory and focal sensory mechanisms. These findings suggest that there are few DARs in ON migratory samples. The functions of nearby genes are likely relevant to the birds' needs during migration.

Transcription factor usage potentially represses chromatin during migration.

The role of DARs in the expression of nearby genes is due to the transcription factors potentially binding inside these specific regions. It is known that TFs bind more often to regions of accessible chromatin [8]. Thus, it is possible to narrow down the identity of potential regulators by analysing the motifs that each transcription factor binds. We hypothesise that some of them will be responsible for the differences between migratory states.
Using the information of Position Weight Matrices (PWM), we searched motifs of potential transcription factor binding sites (TFBS) inside DARs. We scanned the Position Weight Matrix of 414 TF with known binding motifs, from the database JASPAR[17]. We analysed the enrichment of TFBS in the DARs for a given brain

Table 2 Motif Enrichment for transcription factors.

		ON DARs		OFF DARs		
Brain region	**TF(family)**	**fdr value**	**Enrichment score**	**fdr value**	**enrichment score**	
CN	ERE(NR),IR3	1.00E-12	2.371	0.1	0.241	ON
	p53(p53)	1.00E-03	3.322	1	0	
	RAR:RXR(NR),DR5	1.00E-06	4.129	0.1	0.795	
	Reverb(NR),DR2	1.00E-07	2.184	0.1	0.35	
	THRa(NR)	1.00E-13	1.746	1	0.027	
	THRb(NR)	**1.00E-18**	**1.527**	**0.1**	**0.136**	
	VDR(NR),DR3	1.00E-04	1.434	1	0.043	
	ZEB1(Zf)	**1.00E-08**	**0.737**	**1**	**0.035**	
	ZEB2(Zf)	1.00E-11	1.092	1	0.044	
	Nur77(NR)	1	0.191	1.00E-33	1.664	OFF
	Eomes(T-box)	1	0.041	1.00E-52	0.382	
	DUX4(Homeobox)	1	0.24	1.00E-14	2.624	
	NF1-halfsite(CTF)	1	0.025	1.00E-38	0.323	
	BORIS(Zf)	1	0.143	1.00E-108	1.866	
	Ronin(THAP)	1	0.322	1.00E-87	4.368	
	ISRE(IRF)	1	0.12	1.00E-20	2.177	
	Rfx2(HTH)	**1**	**0.1**	**1.00E-72**	**2.189**	
	Nkx2.2(Homeobox)	1	0.004	1.00E-09	0.169	
	PGR(NR)	1	0.023	1.00E-92	2.062	
HC	E2A(bHLH),near_PU.1	1.00E-06	2.781	1	0.019	ON
	EBF2(EBF)	1.00E-04	2.492	0.1	0.272	
	HEB(bHLH)	1.00E-04	2.081	1	0.07	
	Ptf1a(bHLH)	1.00E-06	2.042	0.1	0.159	
	Tbx5(T-box)	1.00E-05	1.515	0.1	0.105	
	THRb(NR)	**1.00E-10**	**3.585**	**1**	**0.114**	
	ZEB1(Zf)	**1.00E-04**	**1.974**	**1**	**0.046**	
	Gata6(Zf)	1	0.159	1.00E-03	0.581	OFF
	RFX(HTH)	1	0.414	1.00E-11	1.558	
	X-box(HTH)	1	0.737	1.00E-29	3.108	
	Rfx2(HTH)	**1**	**0.361**	**1.00E-11**	**1.557**	
	Lhx3(Homeobox)	1	0.152	1.00E-10	0.75	
	KLF5(Zf)	1	0.059	1.00E-19	0.543	
	FOXA1(Forkhead)	1	0.043	1.00E-02	0.505	
	CHR(?)	1	0.073	1.00E-05	0.996	
	Pit1(Homeobox)	1	0.09	1.00E-09	1.303	
	KLF6(Zf)	1	0.01	1.00E-10	0.402	
VAH	Atoh1(bHLH)	1.00E-03	2.175	1	0.155	ON
	Tlx?(NR)	1.00E-03	3.164	1	1.708	
	PRDM1(Zf)	1.00E-03	3.029	1	1.577	
	TCF4(bHLH)	1.00E-03	2.039	1	0.021	
	Hand2(bHLH)	1.00E-03	3.907	1	2.322	
	MyoD(bHLH)	1.00E-04	3.644	1	2.322	
	EBF1(EBF)	0.1	2.322	1.00E-03	3.907	OFF
	AP-2gamma(AP2)	0.1	1.474	1.00E-02	2.907	

Each brain region is indicated. The columns ON and OFF DARs indicates the enrichment and fdr corrected value for a given TF in a specific region.

region and migratory state, by comparing to a background that includes all OCRs from a single brain region. Several motifs are similarly enriched in DARs for ON and OFF migratory season samples. To differentiate those motifs that might be differentially used in ON or OFF migratory samples, we selected TFs significantly enriched in either migratory state (fdr pval<0.01). Additionally, we checked if the

enrichment is more significant in ON or OFF DARs and selected those that had a ratio of Enrichment score > 1.5. (Table 2). Given the number of DARs in Cluster N and Hippocampus in OFF season samples, many TFs were enriched. We report the top 10 of these groups.

Each brain region has its own set of enriched motifs. We found in both Cluster N and Hippocampus the motifs for THR beta and ZEB1 enriched in the ON migratory season birds. Similarly, motifs for RFX2 were found in both CN and HC for OFF season samples. This finding suggests that these TFs have general purposes for migration (ZEB1 and THR beta) or off migratory season (RFX2). With the exception of motifs for p53, the enrichment in Cluster N of ON season birds is in accordance with the repression pattern found in the differential accessibility analysis. The enrichment of motifs for ZEB1 and ZEB2 suggest repression functions[18]. Similarly, nuclear receptors like Rev-erb, THR alpha and beta, VDR, and RAR:RXR can act as transcriptional repressors, depending on the availability of their ligand(NCOR, TRH and retinoic acid, respectively)[19]. All of the Nuclear receptors with enriched motifs are involved in the regulation of circadian and circannual rhythms[20], [21]. The transcription factors enriched in the Hippocampus, are involved in general pathways of cell proliferation and immune responses. The TFs in the VAH, are enriched in functions related to neuronal development and particularly TLX (also known as NR2E1) has functions related to retinal cells[22].

The different motifs found in the DAR for each brain region and migratory state shows that the elements controlling changes in the regulatory landscape when birds enter migration are related to functions of neuronal development and circadian entrainment. Although the enrichment of motifs shows general patterns of factors influencing accessibility, there might be individual elements that harbour differences relevant to the onset of migration.

Identification of Cis-regulatory modules involved in migration

TFBS are usually highly conserved, but also sources that contribute to adaptation via cis-regulatory changes. Mutations in TFBS can potentially induce differences in the dynamics of gene expression[23]. We looked for potential SNPs inside

DARs with highly confident TFBS. To select those confident sites, we established a threshold using the distribution of TFB scores in random regions around the genome. We set the 5% percentile for each transcription factor as the threshold to use in the DARs found. In those regions over the threshold, we searched for SNPs overlapping the TFBS and evaluating if the SNPs can potentially disrupt the motif. For this, we used a whole-genome resequencing dataset previously generated in the Behavioural Genomics Laboratory (Chapter 5). We only included SNPs that are in highly conserved positions of the TFBS using custom generated scripts.

We found 233 SNPs disrupting 130 motifs of TFs (many TF share binding sites) in 183 cis-regulatory elements (Table 3). The GO enrichment of the closest genes to the cis-regulatory elements identified are involved in many of the ones previously identified (Fig 6). Most of the identified cis-regulatory elements are present in cluster N OFF season (TABLE S1). We also identified cis-regulatory elements for the other brain regions OFF season (HC = 9, VAH=1). Cis-regulatory elements of the ON season are mainly in CN (7). VAH and HC have 1 and 2 cis-regulatory elements, respectively. One module is shared for CN and HC, and another shared for CN and VAH. A GO analysis of the ON season cis-regulatory elements does not give a significant result, due to the limited number of genes.

The integration of DAR, TFBS and SNPs on these regions, results in a region located inside an intron of the Gene VATL1 (Fig. 7). It is one of the top DARs in migratory birds for Cluster N and VAH brain regions. This region was not differentially accessible in the Hippocampus. This region comprises an extended region of approximately 8 kb. Within this region, there are several potential motifs for TFs. The low density of SNPs suggests that this is a conserved region. These findings suggest that the region could be a cis-regulatory module, a cluster of cis-regulatory elements. CRM are sites for integration of multiple TF and therefore make a complex combinatorial regulation of gene expression [24].

Furthermore, this region comprises the majority of motifs for potential TFBS with SNPs. A total of four SNPs disrupt the predicted motif. This makes this region a potential source for adaptation at the regulatory level and at the evolutionary level.

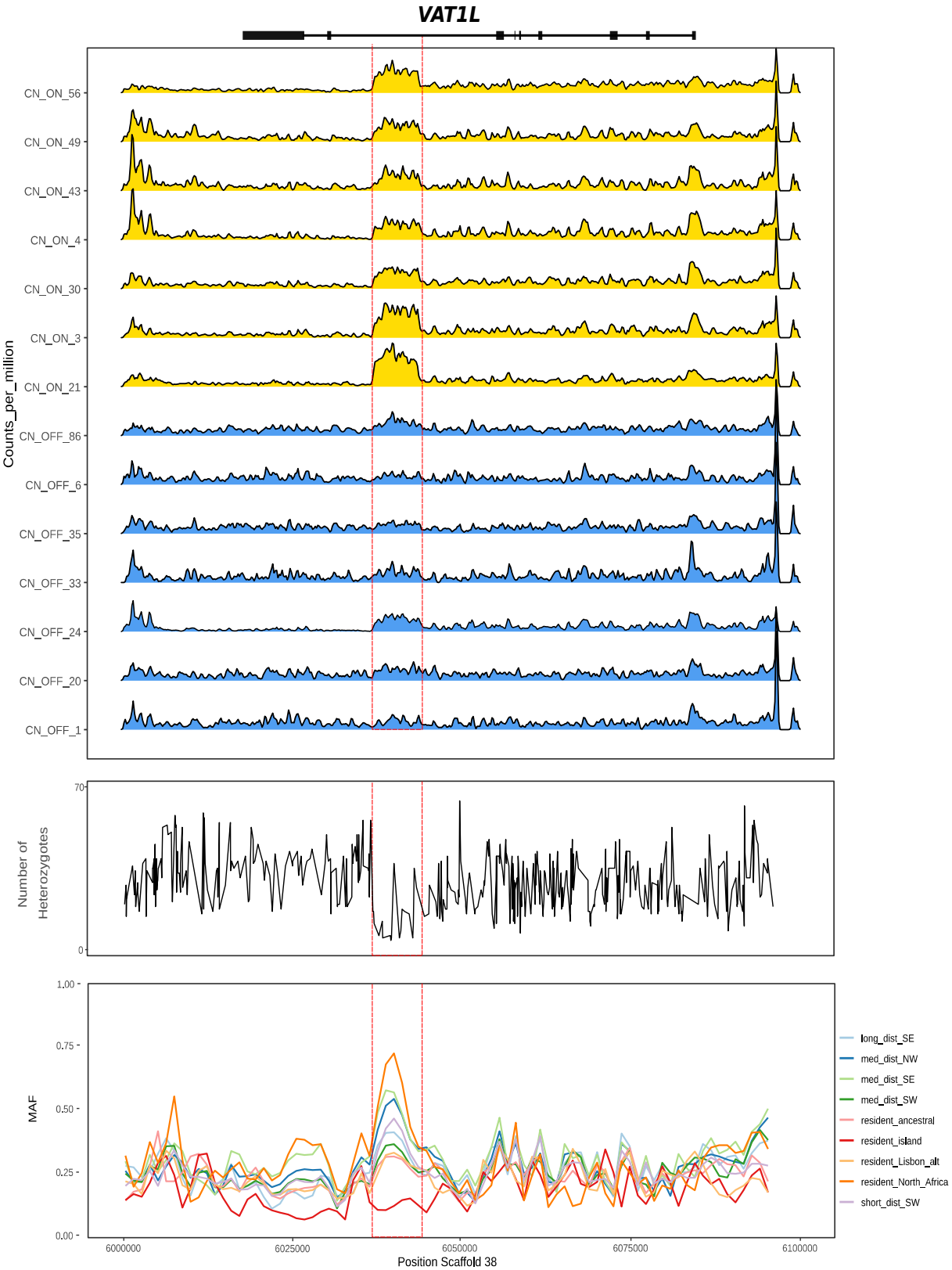


Figure 7. Identification of a potential cis-regulatory module involved in migration. A) Accessibility landscape of Cluster N samples (OFF and ON season blue and yellow respectively). The diagram on top illustrates the location of exons of the gene *VAT1L*. The height of each landscape represents the number of reads per million in every position of

the genome. The red line indicates the location of a highly differentiated DAR (also present in VAH, but absent in Hippocampus). B) Heterozygocity of a panel of 110 whole genome resequenced individuals across different migratory phenotypes (Data from Chapter 4). The number of heterozygotes for each SNP identified is indicated in the y axis. C) Minor allele frequency in the whole genome resequencing panel. Each line represents a different population with different migratory phenotypes (Migratory populations in blues and greens: long distance, medium distance with North West, South East and South West orientations. Resident populations in warm colors: Southern Spain, Atlantic islands, Lisbon and North Africa. Short distance South west migratory populations in pale purple).

When analysing the population data, we found that most of the genotypes in the SNPs in the TFBS were homozygotes. Indeed, we found that this specific region has a drop on heterozygosity compared to its surroundings (Fig 7B). However, the allele frequencies do not show fixation in any population. All the populations, except the residents from Islands and residents from the continent, have similar allele frequencies (Fig 7C). These results suggest that this cis-regulatory module is undergoing a process of underdominance or any other process where the heterozygotes have very low fitness.

DISCUSSION

It is widely accepted that cis-regulatory elements can play core roles that contribute to adaptation or regulation of complex traits. Here, we have explored the contribution of these elements in a complex trait as bird migration. We observed that in birds during the migratory season the chromatin accessibility changes for all the regions and that this difference is the primary driver of variability between our sampling groups. However, we could not identify OCRs that are exclusive for a particular brain region in either the ON or OFF migratory states. Many elements might actually be shared between brain regions and between different migratory states.

Importantly, we found that for two of the three brain regions from our study, our identified patterns highlight repressed accessibility and thus, activity during migration. The pattern is most evident in the Cluster N where many DARs are

more accessible in the OFF season birds. This finding was supported by the fact that the regions more accessible in birds tested during the migratory season, are enriched for binding motifs of repressors, like ZEB1 and ZEB2 and depending on the availability of the ligands, Rev-erb a, RAR:RXR, VDR and THR alpha and beta. This changes the view of migratory behaviour as an activating process, into a mechanism where many elements must be tightly controlled and repressed to allow for successful execution and focus of this demanding behaviour. Possibly, all the negatively regulated metabolic processes and regulation of cell proliferation and neurogenesis found in the GO enrichment must be controlled to allow the bird to allocate energy for the demanding process of migration.

One of the most exciting results of our study is the integration of regulatory elements with genomic and evolutionary patterns. With the combination of these two components of the regulatory network, we were able to identify a cis-regulatory module with potential roles on migratory behaviour. It is a region highly accessible only during migration that contains potential motifs for TF with SNPs potentially disrupting the interaction DNA-TF. More importantly, this region has a drop in heterozygosity levels, which could indicate a process of heterozygote disadvantage or underdominance. Usually, this scenario would lead to disruptive selection [25]. However, this is not the case here, as none of the alleles are entirely fixated in any of the populations studied. It could also suggest that a potential role for structural rearrangements might be taking place around this region. The region found in the VAT1L gene makes part of chromosome 11, a chromosome that has a structural chromosome variant in blackcaps, potentially an inversion (data not shown). Very little is known about VAT1L. It is involved in oxidation-reduction processes, and it is expressed in the brain. A recent single-cell expression study found this gene as a marker for von Economo neurons, a specific type of neurons present is some clades of mammals. The presumable function of this neurons is to allow communication across regions in large brains. Nothing is yet known about this cell type in birds. However, this cis-regulatory module could be controlling transcription of not only VAT1L but of many other genes via long-range interactions. More analysis are needed to shed light into possible processes in that regard.

To our knowledge, this is the first study that investigates migratory behaviour and gene regulation at the cis-regulatory level. Previous studies had a focus on gene expression in various tissues (liver, heart and blood) of several migratory bird species[26]–[29]. Two studies focused on changes of gene expression in the brain using the whole organ[28] or focal regions of the hypothalamus[29]. In general, these studies identify genes that are differentially expressed during migratory season in birds of the same population that are required in pathways related to neurogenesis and synapsis formation. Our GO analysis also suggests the regulation of cell proliferation and neuron generation. Johnston et al. [29] found that CRABP1 was a hub gene for a hypothalamic region in birds tested during migration. This gene is regulated by T3 a thyroid hormone. This could suggest that although we found a motif enrichment of Thyroid regulators (THR alpha and beta) in Cluster N during migration, the thyroid pathway might have an important role in the onset of migration[19]. CRABP1 is also a regulator of retinoic acid signalling, which again is in line with similar findings of motif enrichment for VDR and RAR:RXR transcription factors[20], [21]. This might call attention to the study of Thyroid and retinoic acid pathways for migratory studies, as it is known the role of these pathways in circadian entrainment and seasonal neural plasticity. (REFs from paper!!). We did not find any element related to candidate genes previously associated with bird migration like ADCYAP1[30] or CLOCK[31]. Because the associations of migratory traits with candidate gene approaches are purely sequence based, might not be detected with our approach.

It is also important to note that our approach has limitations. Our experimental setup is the best to maintain controlled conditions; however, these might not confidently represent the whole environmental inputs that birds encounter in the wild. We are not including other cues that birds might use[32], like polarized light and other compasses. We are also aware that most of the functions assigned to the elements found in our study are based on human/mice derived studies. This might have an implicit bias on our results because the genes found might have different functions in avian animals. This is also very important for the motif analysis performed. All the position weight matrices (PWM) are derived from human or mice backgrounds. Therefore many elements might be missed or prone to false positives. The study of chromatin accessibility must thus be tightly

coupled with gene expression analyses to get a better assessment of the mechanisms involved in migration. This aids to confirm how the accessibility changes influence gene expression. For this reason, future approaches should integrate studies of gene expression using RNA-seq or similar methods.

Overall, our study gives new insight into the study of genetics, neurobiology and molecular elements of bird migration behaviour. First, our study gives different interpretations of the processes of how migration might not indicate the activation of elements, but rather the repression of elements that might not be essential for migration. Secondly, our study supports the analysis of region based approaches in the brain, because many of the signals found could be masked when the whole brain is studied. Here, we integrated functional approaches of chromatin accessibility and resequencing data. This aid to narrow down potential elements that are going through evolutionary processes that shaped the migratory behaviour and could give insight on how it appears or disappears along with the avian clade. Finally, the integration of many sources of evidence from multiple levels of the regulatory machinery will have a profound impact on the detection of migratory elements.

METHODS

Bird capture

> Eurasian Blackcap (*S. atricapilla*) juvenile males, i.e. without migratory experience were caught with mist nets at the end of the breeding season and after the post-juvenile moult to allow sex phenotyping (July/August). We caught birds in two distinct locations Freiburg, Germany (47° 59'49.9 "N 7° 45'58.2 "E) and Hartberg, Austria. Birds were kept and monitored for weight and health status at the catching site in cloth cages for one night prior to transporting them to our animal housing facilities at the MPI for Evolutionary Biology in Ploen, Germany. At the facilities, all birds were kept in individual cages, were fed a controlled diet, and experienced a light regime approximated to the region of origin.

Migratory Phenotyping

>In order to classify individuals as "during migration" (ON) or "outside the migratory season" (OFF), we quantified migratory restlessness (MR, Zugunruhe) activity as a proxy for migration. MR is a nocturnal behaviour that caged migratory birds exhibit during the time their con-specifics in the wild carry out actual migration. It consists of nocturnal motor activity with specific behavioural elements, such as nocturnal wing whirring, directed flight and hopping around in the cages. MR is exclusively exhibited during the migratory season, and the timing of MR is in good accordance with the actual timing of migration in wild con-specifics, therefore, serves as a good proxy to characterize migratory behaviour in captive conditions [33].

Birds were monitored with individual motion sensors mounted in their individual cages and behaviour was confirmed through infrared video camera observations. Birds were classified as OFF-season if they did not show nocturnal movement activity for at least five days previous to sacrifice during winter. ON-season birds were identified if the bird showed consistent nocturnal MR activity for at least 3-4 days after a period of inactivity during nights before sacrifice during spring, identified through change point analysis. (SEE FIG 1). Both, OFF and ON season birds were sacrificed 1-3 h after light offset simulating the onset of the night as this species migrates during the night.

Brain regions: Rationale.
>We dissected three focal brain areas assumed to play a central role in processing information relevant to migratory behaviour: the hippocampus (HC), the ventral anterior hypothalamus (VAH), and the Cluster N (CN). It has been extensively indicated that the hippocampus plays a role in navigation and spatial memory in mammals and birds [34], [35]. Cluster N is a brain region involved in processing magnetic compass orientation in migratory birds during night vision [32], [36], [37]. As migration is a behaviour coupled to changes in daylight length and circadian cycle, several nuclei of the hypothalamus are related to gene expression changes during the circadian and circannual cycle. Arguably, the Suprachiasmatic nuclei (SCN), located in the (VAH), is the leading region showing gene expression changes throughout the circadian cycle[38].

Brain microdissection:

> The brain regions were microdissected right after sacrifice with the following approach: the right hemisphere of the brain was placed on a tissue slicer (Stoeling) with Vetbond 3M tissue adhesive and the medial part of the brain facing the blade. Sagittal sections of 400 um were obtained and placed in a clockwise fashion on a Petri dish with ice-cold PBS with complete proteinase inhibitor cocktail. From the most medial sections, the VAH and the HC were obtained. The VAH was obtained from the tissue between the optical decussations, and the hippocampus was obtained from the tissue above the lateral ventricle and 1 to 2 mm before the end of the lateral ventricle. Cluster N was obtained from more lateral sections (approximately 1.2 cm inside the brain) obtaining a 3 mm portion of the hyperpallium and the Dorsal mesopallium after 2mm of the lateral ventricle. Dissections were photo-documented to allow anatomical assessment. The dissected regions were placed in separate Eppendorf tubes, immediately flash frozen in dry ice and stored at -80 C until ATAC-seq procedure. The average time of the obtained brain regions is around 25 minutes since sacrifice time.

ATAC-seq

>The procedure of ATAC-seq was performed following the protocol for frozen sections described on Corces, et al. 2017, based on the principles of Buenrostro 2013[7] with slight modifications. Individual brain sections were slowly thawed in 50 ul of homogenization Buffer on ice during 10 minutes. The tissue was mechanically disrupted with plastic Eppendorf pestles doing 5 to 8 soft strokes. The sample was diluted on washing buffer and strained. The pellet was resuspended in a buffer with glucose to perform a gradient to separate and selectively obtain nuclei. The glucose gradient with the samples were centrifugated at 4500 rpm during 45 minutes in a swinging bucket centrifuge. After centrifugation, we counted approximately 25000 nuclei, and the pellet was resuspended in a tagmentation reaction using 25ul of TD buffer and 2ul of transposase. The procedure was carried out on a thermomixer at 37C degrees for 30 minutes and stopped with proteinase k. DNA was obtained using the minelute PCR kit. The number of PCR cycles to amplify the library was optimised for each region using qPCR. A total of 12 cycles was used for hypothalamus sections and 11 cycles for Cluster N and Hippocampus sections. Library size selection was performed with AMPure beads, and the final library size distribution was

observed to obtain the expected pattern of distribution. The libraries were sequenced on a Next-seq or Hi-seq 4000 machine.

Analysis of ATAC-seq

>Raw sequences were trimmed with trimmomatic with the following specifications NexteraPE-PE.fa:2:30:10 MINLEN:35. Trimmed reads were aligned to the Blackcap 1.0 reference Genome with bwa mem, using default parameters. Reads aligning to the mitochondrial genome and reads with a mapping quality lower than 30 were removed. Optical and PCR duplicate reads were removed using Picard MarkDuplicates. The final file only kept uniquely aligned paired reads with a mapping quality higher than 30.

OCR identification was performed with MACS2[39] with the following specifications: -g 1e9 --nomodel --shift -100 --extsize 200 --keep-dup all -q 0.05 --min-length 80. Regions of the genome where enzyme transpositions are more common than the background, identify those potential regions with more accessible chromatin. The reproducibility of those OCRs was performed with the IDR procedure. Briefly, subsamples for each sample were taken using DownsampleSam from Picard tools with a probability of 0.6. OCRs were called in this subsamples with the same settings as above. The IDR procedure was performed with a false discovery rate < 0.1.

The quality control for each sample was addressed observing the insert size distribution requiring that the mononucleosome fraction was present. Reads inside the OCRs were counted and obtaining a FRiP score > 0.2 was required to pass QC. The distribution of reads around TSS was also used as a criterion for QC: the highest point of the read distribution must be close to the TSS and must be at least one third higher than the lowest point of the distribution. The PCR bottleneck Coefficient (PBC) and Relative strand cross-correlation (RSC) were also used as filters for a QC of the samples.

Differentially accessible chromatin Analysis

>To obtain a consensus set of OCRs we obtained only those OCRs overlapping in 3 or more samples of the same brain region and same migratory state. We obtained a set for each region and one merging the sets for all regions. Then, we

counted reads for each individual sample on each of those OCRs. We obtained a matrix of Each OCR and the raw counts for each sample. We used this as an input for DESeq2.

>The differential accessibility analysis was carried on in DESeq2. First, we explored covariates for the whole dataset, including all samples from different brain areas. We used PCA approach to find high correlations between the two most important PCAs with the covariates: we performed 100 permutations finding that the correlation between the sequencing machine Hi-seq or Next-seq(platform) and PC1 was significant. We modelled the effect of this batch effect, adding this as a covariate in the design of the fitting:

Differential accessibility ~ Platform + Migratory_phenotype.

After the modelling, a regular binomial test was performed paired contrasts between the OFF and ON groups of the same brain region. After this, we set two thresholds to define Differentially Accessible Regions. (DARs): a strict threshold using padjusted value < 0.05 and a Fold change bigger than 2. For the second threshold, we obtained a distribution of log2Foldchange for random regions of the genome and selected the 1% top percentile as the Foldchange threshold value.

To estimate the rate of false positives in this approach, we permuted the migratory status of the samples 100 times and performing the same analysis as before. We estimated the significance of the permutation following the equation:

times the number of sig genes higher than real data / number of permutations.

Similarly, we performed the same analysis for randomly shuffled regions as random OCRs to find the rate of what would be regions expected by chance to be differentially accessed. To eliminate any possible effect of outliers, we performed a Jackknife (or Leave one out – LOO -) approach and analysed the distribution of Foldchanges afterwards, checking that the distribution is not different by performing a Mann-Whitney-U test.

Annotation

For any given set of OCRs, we annotated where in the genome are the OCRs located and if there is any enrichment of the prefered location in the genome. To do this, we obtained the sum of bases overlapping between the OCR and the genomic feature OVER the length of the genomic feature. The features included are exon, intron, gene body(introns + exons), Repeats, Promoters(TSS – 2kb) and intergenic regions.

Gene ontology (GO) enrichment

The functional annotation of any OCR was associated to the closest gene downstream. The GO enrichment was performed with enricher[40] having the human genome as a background. All the enrichments are based on GO annotation for biological function, molecular function and cell compartment.

Transcription factor scanning:

We used the motifs from JASPAR [17] to scan for potential TFBS inside any given set of OCRs. We obtained a null distribution of TFBS scores along the genome by scanning 100000 randomly placed OCRs and establishing an alfa of 0.05 for each Transcription factor. Only the OCRs that had a TFBS score inside the 0.05 alfa were taken as potential TFBS. To evaluate if our OCRs have significant differences in the potential TFBS we performed a scan of 414 TFs in the differentially accessible regions of each brain region and migratory state compared to a background of the OCRs from the same region. We assessed the significance with a binomial test implemented in HOMER[41] .

Genetic associations with Regulatory regions.

We used the resequencing data from 110 individuals, including the different migratory phenotypes, previously published in our group. We identified which SNPs were overlapping with the significant TFBS previously identified (see TF scanning section). We used bedtools and a custom script to obtain only those SNPs that affected very conserved regions of the motifs. For the focal study of the region in VAT1L, we used PLINK 1.90 [42]to calculate heterozygosity and Minor Allele Frequencies (MAF).

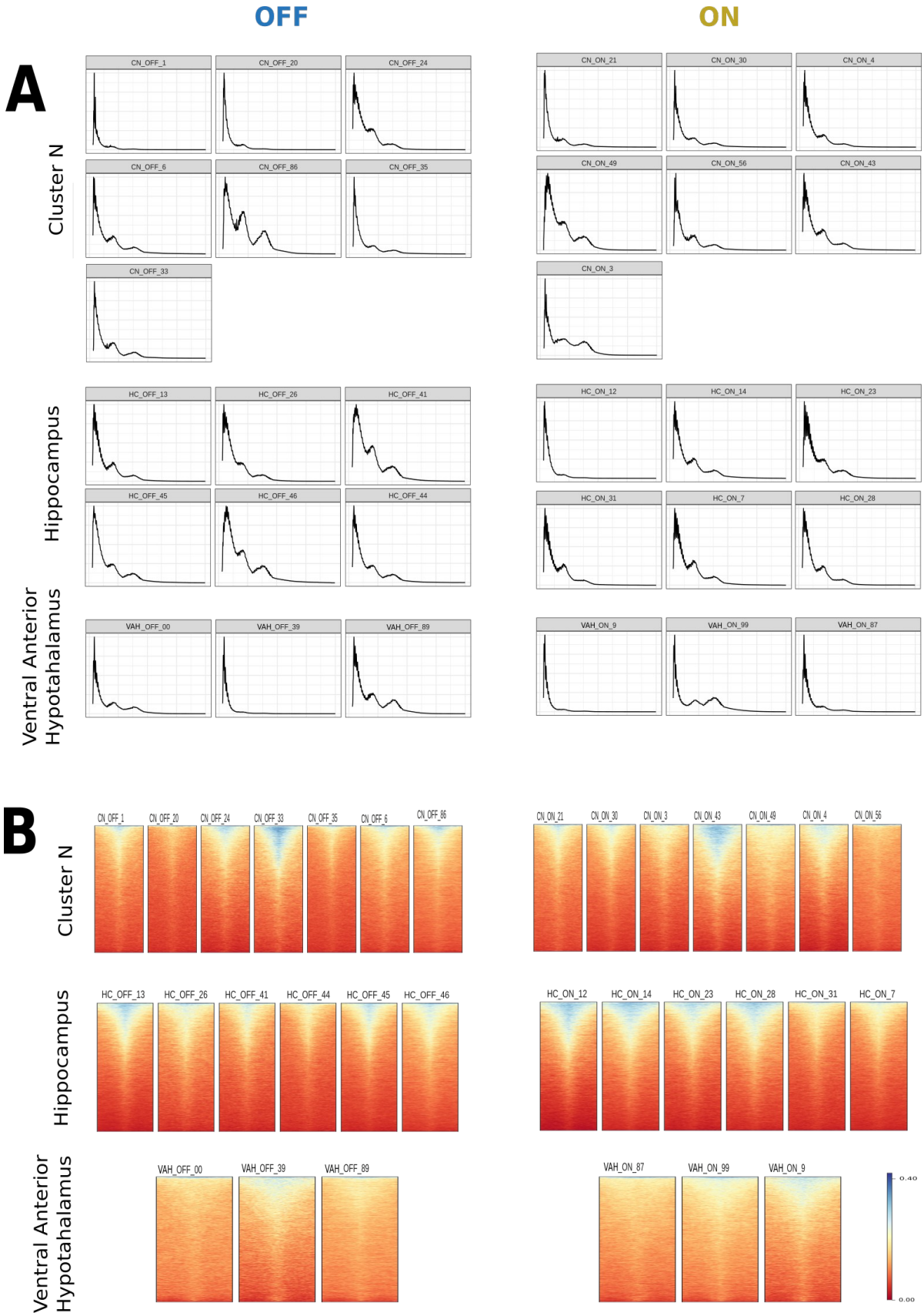

OFF
ON
A
Cluster N
CN_OFF_1
CN_OFF_20
CN_OFF_24
CN_OFF_6
CN_OFF_86
CN_OFF_35
CN_OFF_33
CN_ON_21
CN_ON_30
CN_ON_4
CN_ON_49
CN_ON_56
CN_ON_43
CN_ON_3
Hippocampus
HC_OFF_13
HC_OFF_26
HC_OFF_41
HC_OFF_45
HC_OFF_46
HC_OFF_44
HC_ON_12
HC_ON_14
HC_ON_23
HC_ON_31
HC_ON_7
HC_ON_28
Ventral Anterior Hypotahalamus
VAH_OFF_00
VAH_OFF_39
VAH_OFF_89
VAH_ON_9
VAH_ON_99
VAH_ON_87
B
Cluster N
CN_OFF_1
CN_OFF_20
CN_OFF_24
CN_OFF_33
CN_OFF_35
CN_OFF_6
CN_OFF_86
CN_ON_21
CN_ON_30
CN_ON_3
CN_ON_43
CN_ON_49
CN_ON_4
CN_ON_56
Hippocampus
HC_OFF_13
HC_OFF_26
HC_OFF_41
HC_OFF_44
HC_OFF_45
HC_OFF_46
HC_ON_12
HC_ON_14
HC_ON_23
HC_ON_28
HC_ON_31
HC_ON_7
Ventral Anterior Hypotahalamus
VAH_OFF_00
VAH_OFF_39
VAH_OFF_89
VAH_ON_87
VAH_ON_99
VAH_ON_9
0.40
0.00

Figure 1S1. Insert size distribution and frequency of reads around TSS for every sample passing quality control in this work. A) Insert size distribution showing the presence of mono-nucleosomes (and dinucleosomes). B) Heatmaps showing the frequency of reads +/- 1000 bp around the TSS for each brain region and experimental condition. All the samples of the same brain region are set to scale.

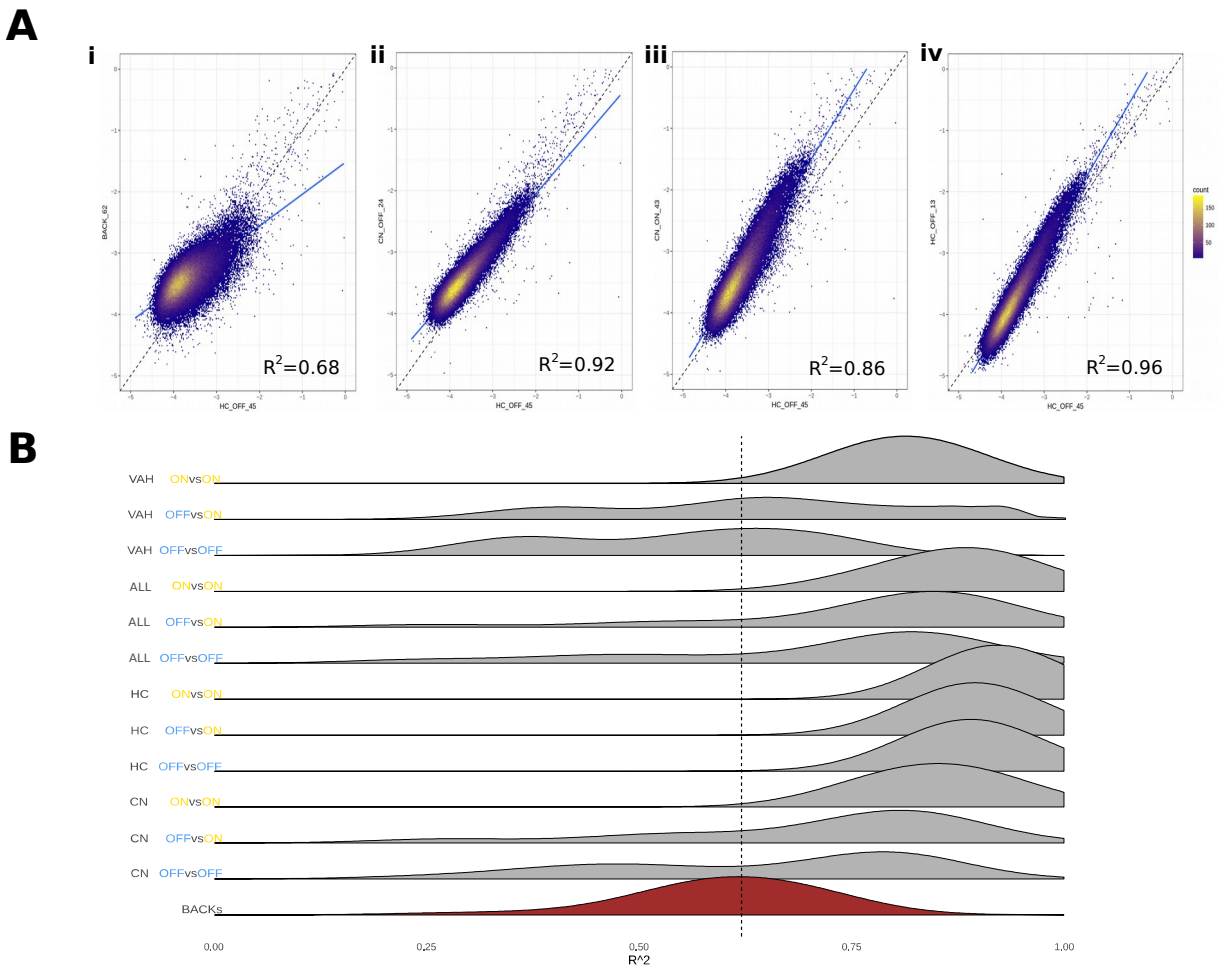


Figure 1S2.Sample replicability. Correlation of counts per million (CPM) normalized reads in windows of 10kb along the genome. A) Shows panels of different pairs of samples. i)correlation of a sample with background (BACK_62). Correlation of reads of samples of different brain region (hippocampus) with different experimental gorups ii) OFF and iii) ON). iv) Correlation of Samples of the same brain region and same experimental group. B) Density plot of pairwise correlations between samples of the groups indicated on the left. Comparisons of samples against the background/naked DNA are indicated in the red distribution. The dashed line indicates the average of correlation of samples against the Backgrounds (mean =0.62)

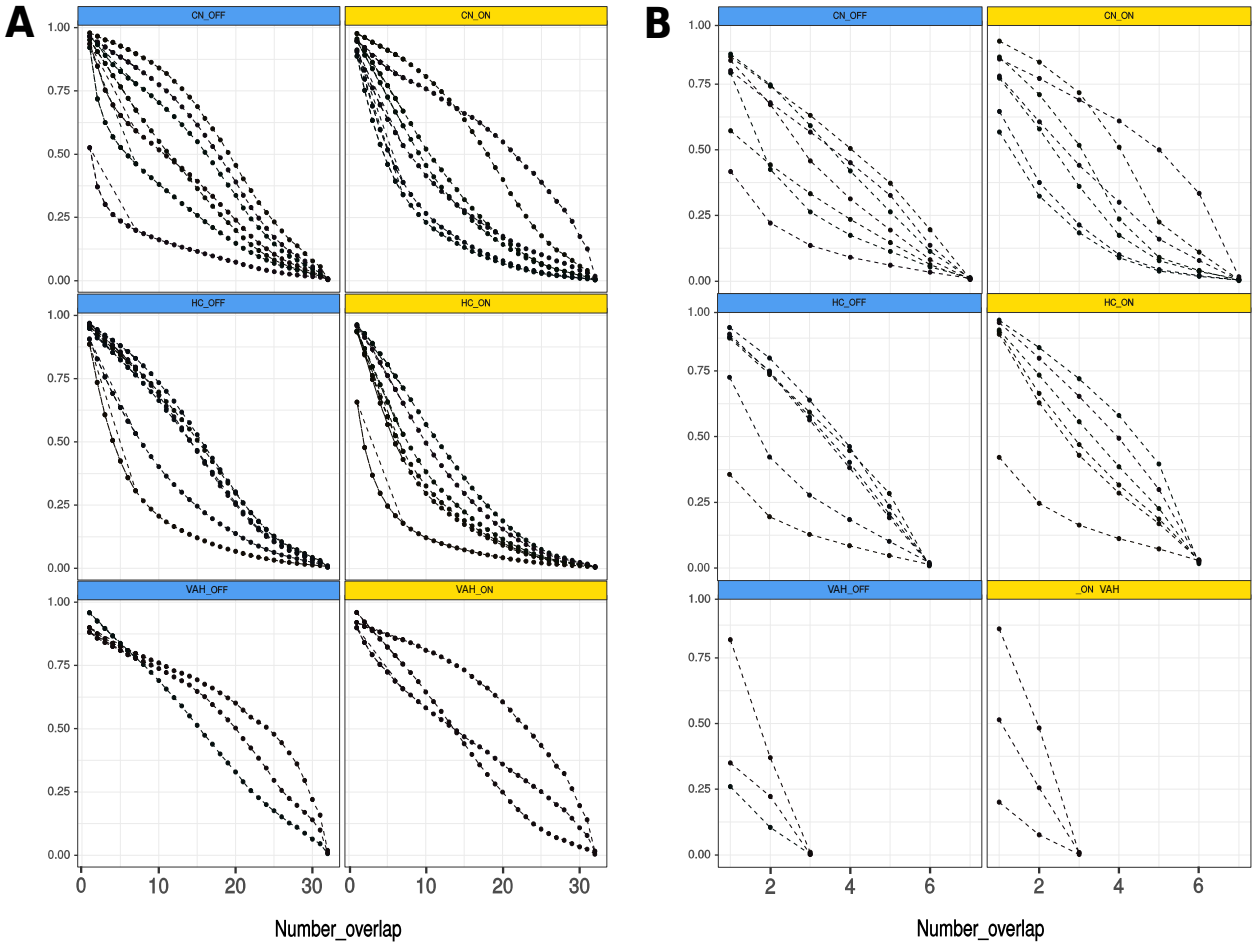


Figure 2S1. Replicability of peaks between samples. Every point indicates the proportion of peaks that are replicated in "n" number of samples (x axis) A value of 1 indicates that 100% of the peaks in one sample are replicated in at least "n" samples. The overlap is performed with the total number of samples included in the study (left panels, n=32), and the sample inside the groups (right panel. Number of samples per experimental condition (ON and OFF) and brain region: VAH n=3, CN n=7 HC n=6).Blue and yellow panels indicate OFF and ON season samples, respectively.

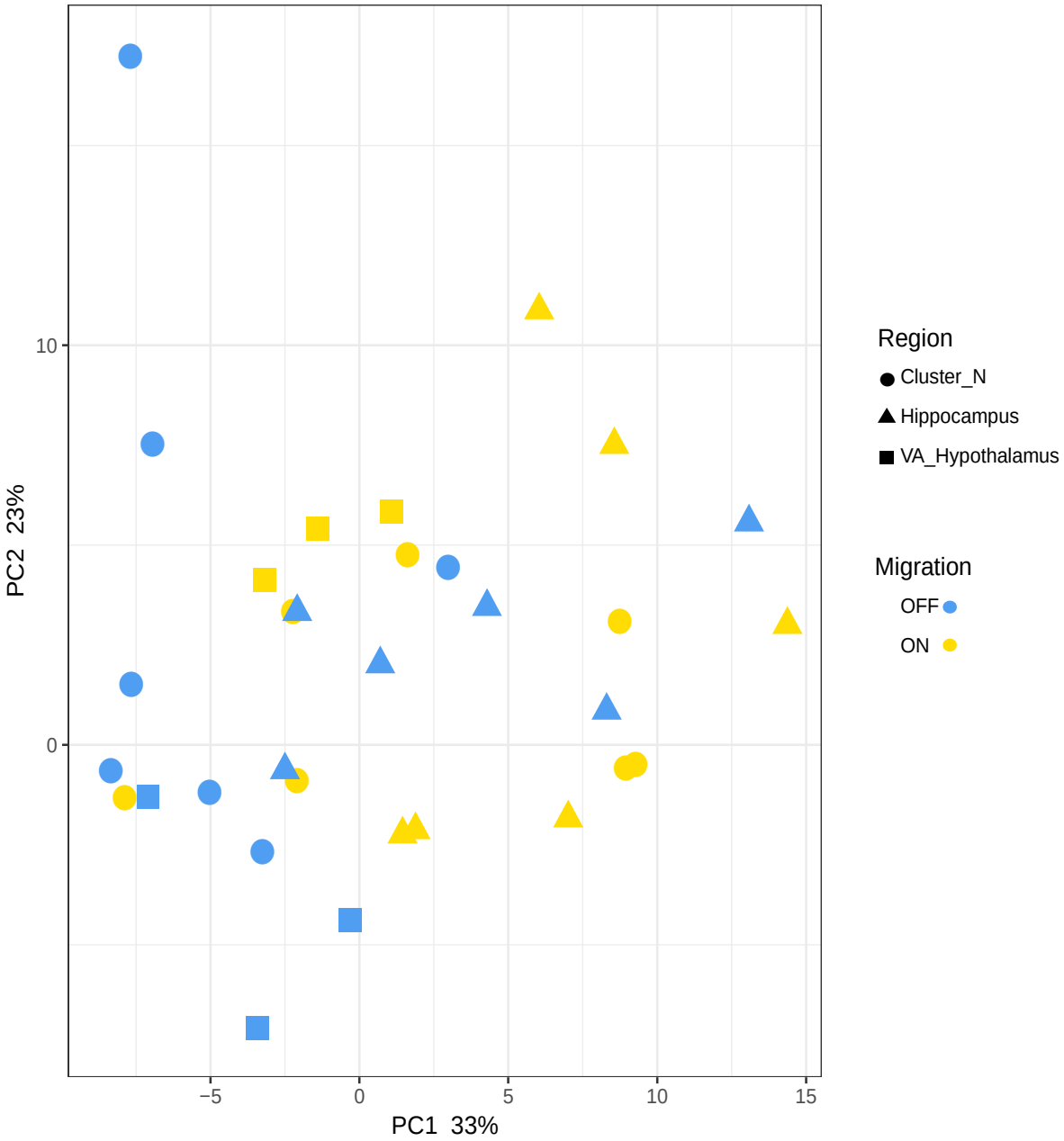


Figure 4S1. Principal Component Analysis for a set of 30000 randomly placed peaks. The same procedure as in Fig 4 was performed.

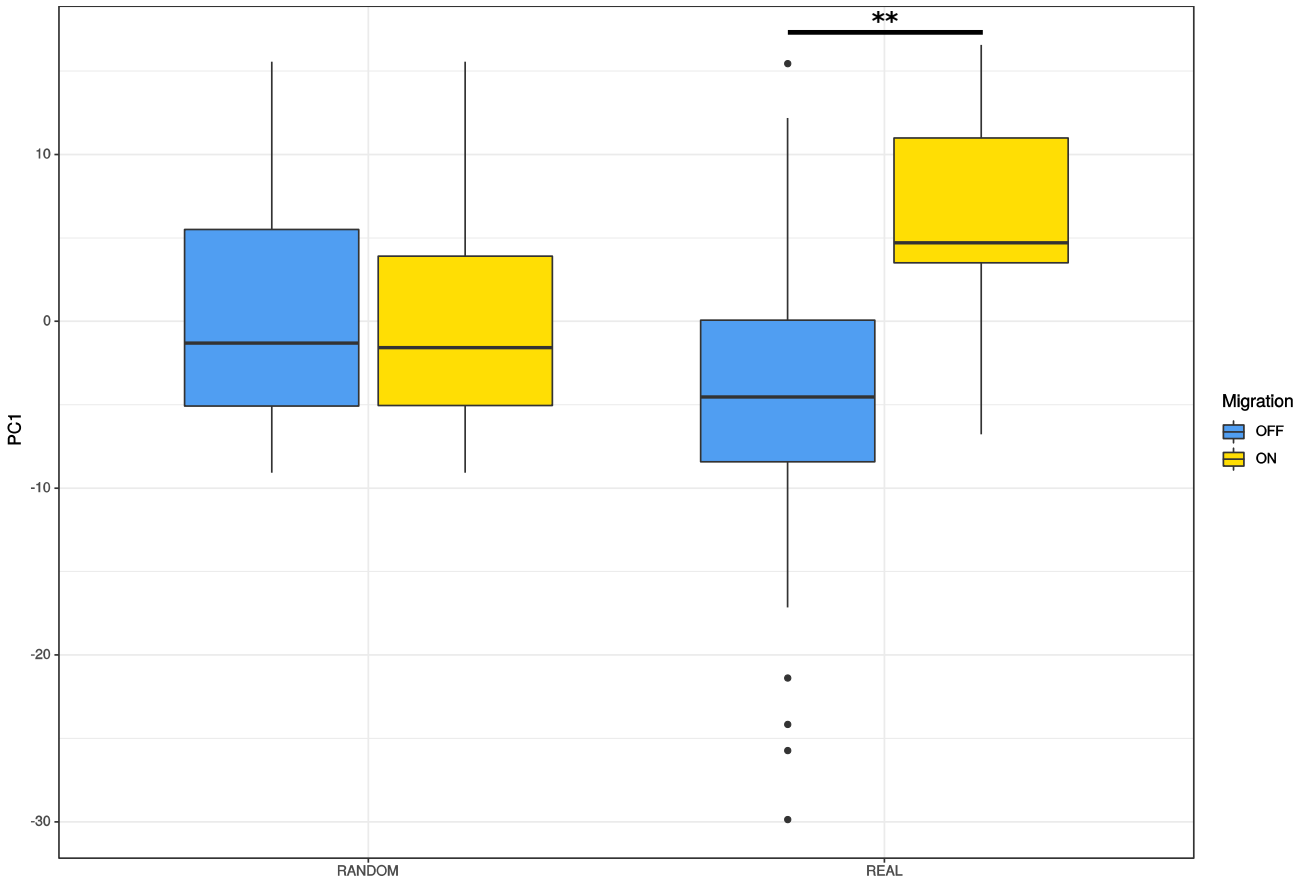


Figure 4S2. Permutation analysis of PC1. Permutations of the migratory state were performed 100 times to obtain a distribution random of values on the PC1. Boxplots show the distribution of PC1 values for each sample (includes all samples n = 32). Values of PC1 for the permutations and the ON or OFF experimental groups are represented in the boxes on the left. Boxes on the right, represent the distribution of the real dataset. The ON and OFF assignments of the real data has significantly different variances (**Levene test p val <0.05).

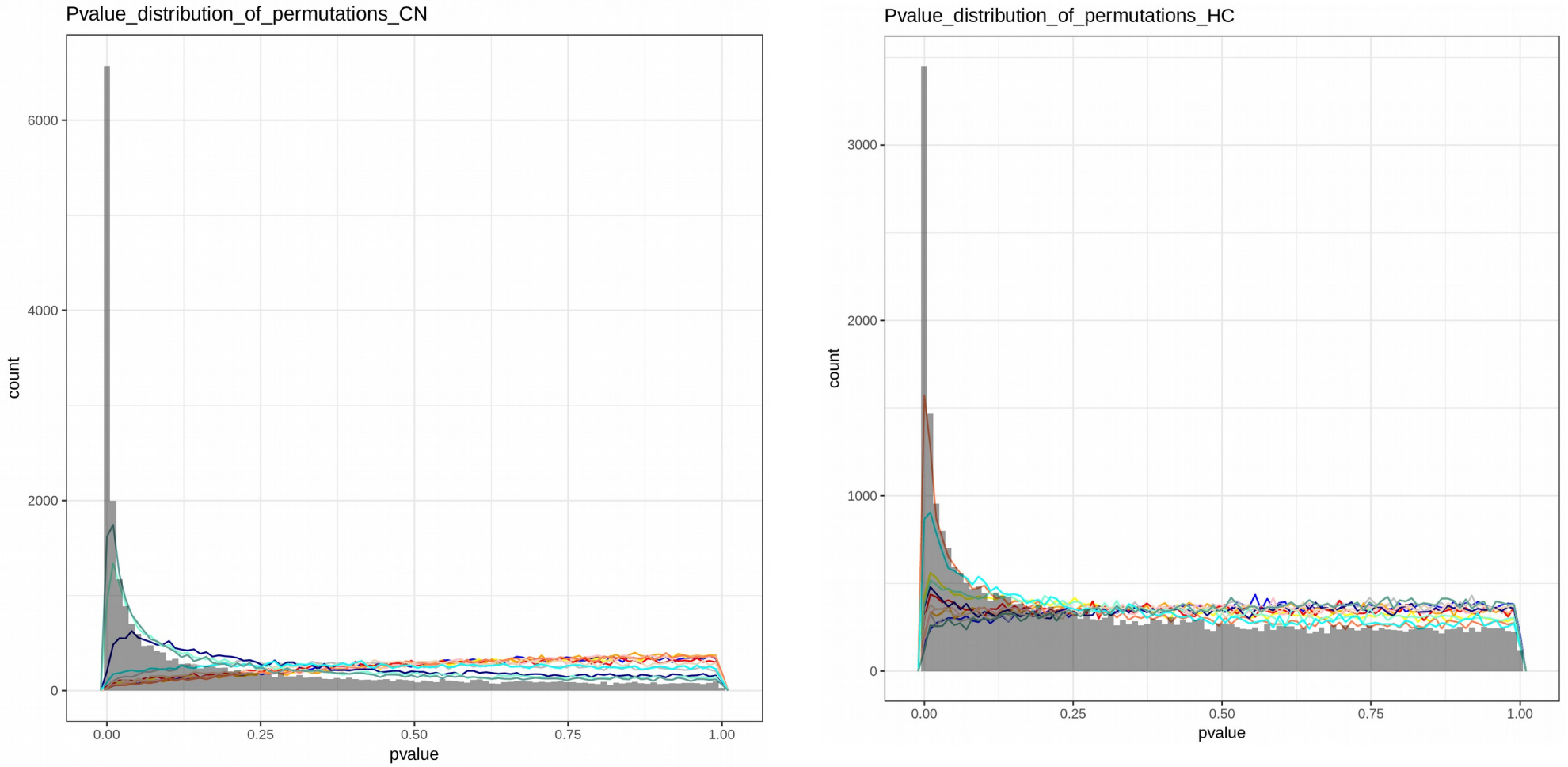


Figure 5S1. Distribution of P-values for 20 permutations of the migratory status in samples of Cluster N and Hippocampus (VAH does not have a distribution of p-values, not many DARs). The gray histogram represents the distribution of the real dataset. All colored lines indicate density plots of permutations.

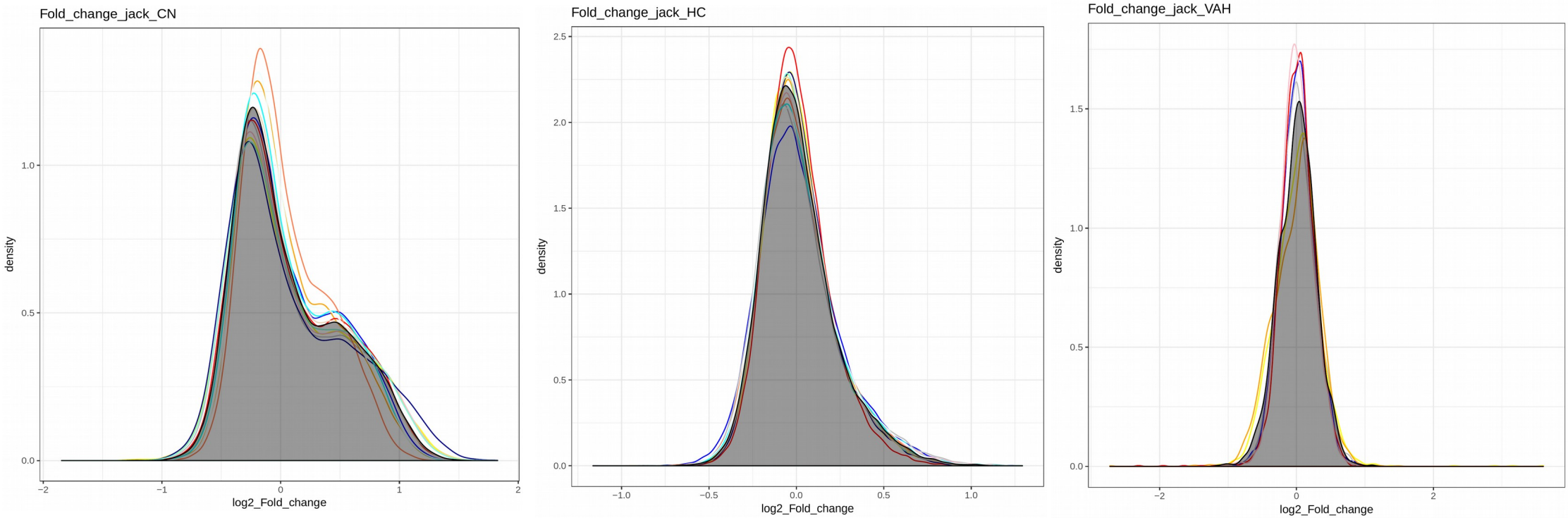


Figure 5S2. Distribution of Foldchanges for jackknifed (Leave One Out - LOO) and real datasets of Cluster N, Hippocampus and VAH. Gray distribution indicates real dataset, and lines indicate permutations. Mann whitney U test pval > 0.1 in all comparisons.

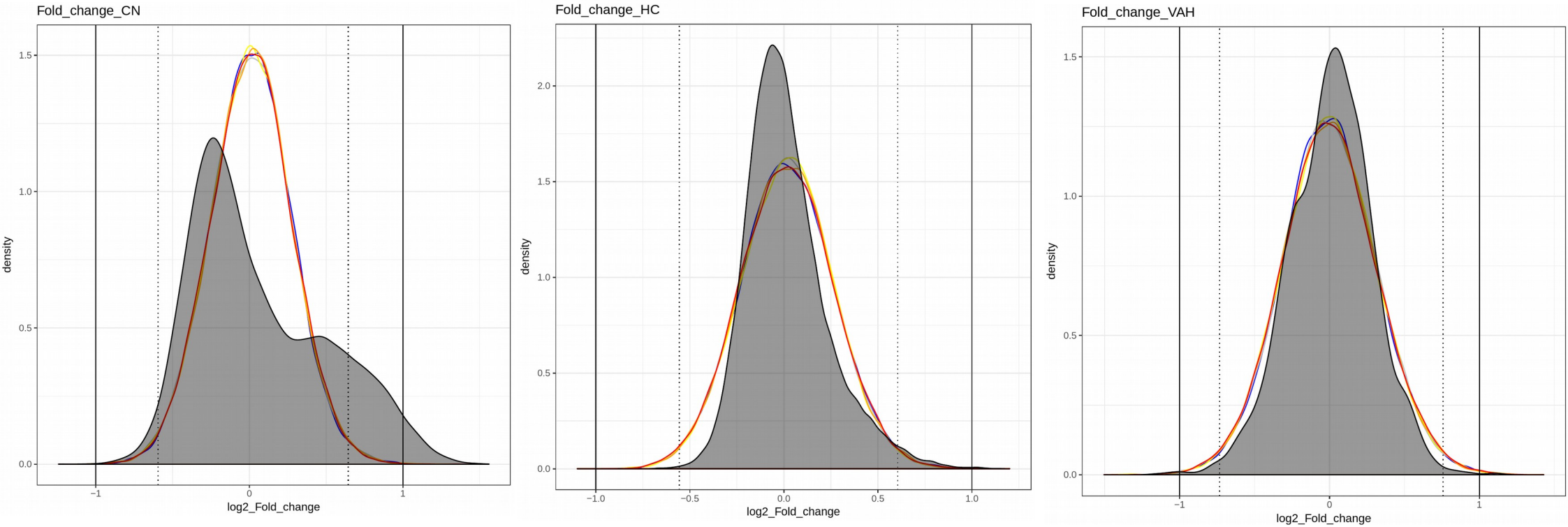


Figure 5S3 Distribution of Fold change for Random regions and the real dataset. The gray histogram represents the distribution of Fold change in the OCRs identified. All colored lines indicate density plots of Randomly placed regions. The dashed line indicates the empirical threshold for each brain region. It is the top and bottom 2.5% percentiles, based on the distribution of fold changes for random regions.

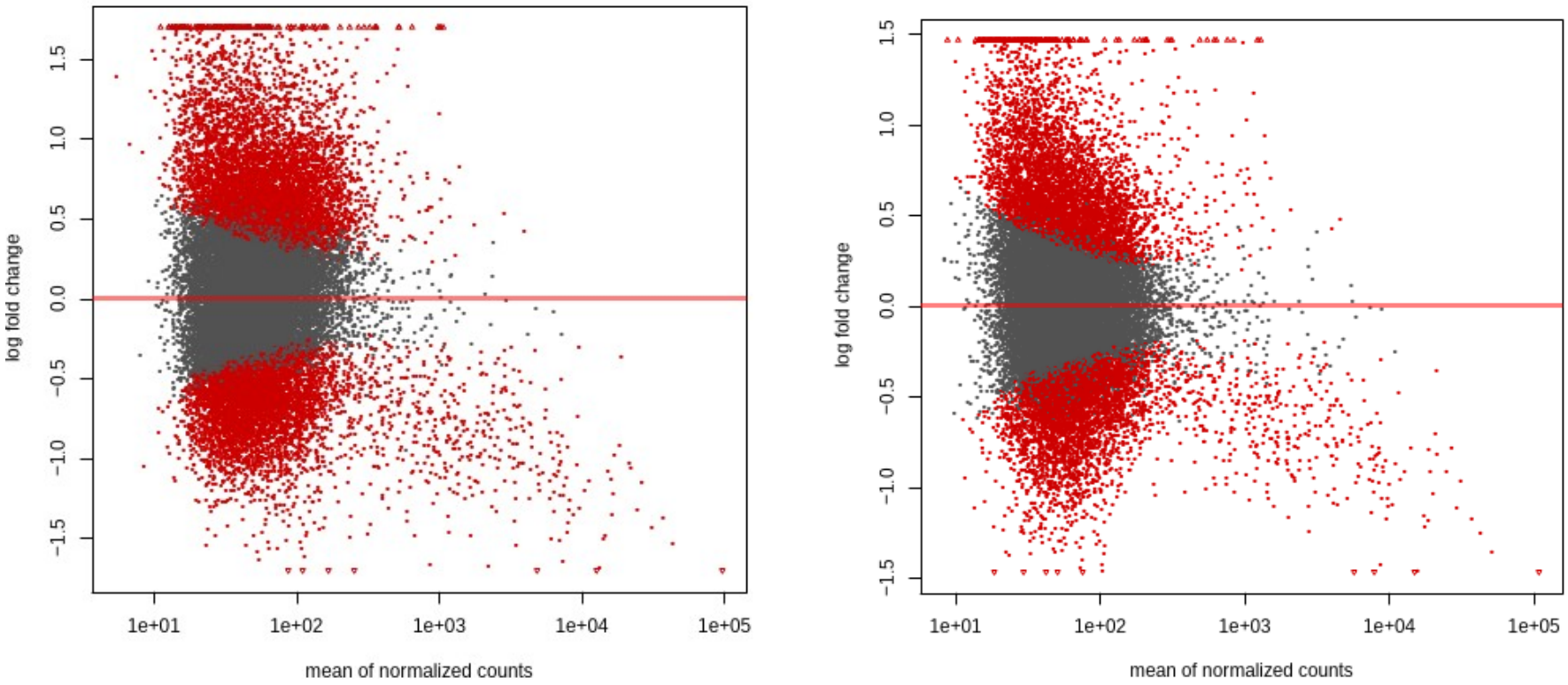


Figure 5S4. Maplots for Cluster N samples compared to Background samples. The left panel shows the MA plot for a differential Accessibility analysis comparing all Cluster N OFF season samples versus background samples. The left panel shows the same procedure for All Cluster N ON season samples versus background samples.

Supplementary Table 1. significant disrupted TFBS. Motifs located in DARs with Single nucleotide Polymorphisms potentially disrupting the interaction. Each SNP indicated is located in a DAR of the group indicated in the column “Region”. The closest gene to the DAR and the potential TF for which the motif is potentially disrupted is indicated

Scaffold	SNP	Putative Transcription factor affected	Closest gene	Region
Super-Scaffold_67	27808353	Bapx1_Homeobox_VertebralCol-Bapx1	ADAM12	CN_OFF
Super-Scaffold_67	27809077	Nkx3.1_Homeobox_LNCaP-Nkx3.1	ADAM12	CN_OFF
Super-Scaffold_35	4079108	Sox10_HMG_SciaticNerve-Sox3	AFAP1	CN_OFF
Super-Scaffold_10	12472208	EWS_FLI1-fusion_ETS_SK_N_MC-EWS_FLI1	AFF1	CN_OFF
Super-Scaffold_22	3665246	ERG_ETS_VCaP-ERG	AGFG1	CN_OFF
Super-Scaffold_37	30298158	STAT6_Stat_Macrophage-Stat6	AKAP17A	CN_OFF
Super-Scaffold_67	6064204	Oct6_POU	ANXA11	CN_OFF
Super-Scaffold_67	6064214	Brn1_POU	ANXA11	CN_OFF
Super-Scaffold_67	6064912	Tbx21_T-box_GM12878-TBX21	ANXA11	CN_OFF
Super-Scaffold_65	9191764	RBPJ_Ebox_?_bHLH_Panc1-Rbpj1	APBB2	CN_OFF
Super-Scaffold_65	9191777	Meis1_Homeobox_MastCells-Meis1	APBB2	CN_OFF
Super-Scaffold_65	9192128	PU.1-IRF_ETS_IRF_Bcell-PU.1	APBB2	CN_OFF
Super-Scaffold_65	9233032	Sox9_HMG_Limb-SOX9	APBB2	CN_OFF
Super-Scaffold_17	19291613	EHF_ETS_LoVo-EHF	ARHGAP18	CN_OFF
Super-Scaffold_41	2260148	NFkB-p65_RHD_GM12787-p65	ARHGAP29	CN_OFF
Super-Scaffold_41	2260683	Eomes_T-box_H9-Eomes	ARHGAP29	CN_OFF
Super-Scaffold_80	10434060	ZNF519_Zf_HEK293-ZNF519.GFP	ARL6	CN_OFF
Super-Scaffold_32	12142163	EBF2_EBF_BrownAdipose-EBF2	ARNT2	CN_OFF
Super-Scaffold_65	8608184	Nkx6.1_Homeobox_Islet-Nkx6.1	ATP8A1	CN_OFF
Super-Scaffold_41	2445071	bZIP:IRF_bZIP	BCAR3	CN_OFF
Super-Scaffold_80	3955219	ZEB1_Zf_PDAC-ZEB1	C21orf91	CN_OFF
Super-Scaffold_80	3955790	PRDM10_Zf_HEK293-PRDM10.eGFP	C21orf91	CN_OFF
Super-Scaffold_30	44227057	Meis1_Homeobox_MastCells-Meis1	C6orf62	CN_OFF
Super-Scaffold_40	22296239	RARa_NR_K562-RARa	CALCR	CN_OFF
Super-Scaffold_30	58722338	E2A_bHLH_proBcell-E2A	CDH2	CN_OFF
Super-Scaffold_67	4006960	Hoxd13_Homeobox_ChickenMSG-Hoxd13.Flag	CDHR1	CN_OFF
Super-Scaffold_56	11783085	RARa_NR_K562-RARa	CELF2	CN_OFF
Super-Scaffold_56	12002481	WT1_Zf_Kidney-WT1	CELF2	CN_OFF
Super-Scaffold_37	38117898	bZIP:IRF_bZIP	CHAMP1	CN_OFF
Super-Scaffold_30	1227698	E2A_bHLH_proBcell-E2A	CLASP2	CN_OFF
Super-Scaffold_30	1256903	GATA3_Zf_iTreg-Gata3	CLASP2	CN_OFF
Super-Scaffold_8	3087701	Sox9_HMG_Limb-SOX9	CLCN4	CN_OFF
Super-Scaffold_40	40030806	Nkx2.2_Homeobox_NPC-Nkx2.2	CMTM7	CN_OFF
Super-Scaffold_67	8786660	COUP-TFII_NR_K562-NR2F1	COMTD1	CN_OFF
Super-Scaffold_30	81237623	Lhx2_Homeobox_HFSC-Lhx2	CPQ	CN_OFF
Super-Scaffold_80	10784816	FOXK2_Forkhead_U2OS-FOXK2	CREG1	CN_OFF
Super-Scaffold_17	69984440	MafB_bZIP_BMM-Mafb	CTSB	CN_OFF
Super-Scaffold_37	31672534	Nanog_Homeobox_mES-Nanog	CYFIP1	CN_OFF
Super-Scaffold_48	6908699	Sp2_Zf_HEK293-Sp2.eGFP	CYP2D49	CN_OFF
Super-Scaffold_10	25911293	Tbx20_T-box_Heart-Tbx20	CYP2U1	CN_OFF
Super-Scaffold_30	91941943	Egr1_Zf_K562-Egr1	DERL1	CN_OFF
Super-Scaffold_30	91942035	Eomes_T-box_H9-Eomes	DERL1	CN_OFF
Super-Scaffold_22	426707	Hoxa13_Homeobox_ChickenMSG-Hoxa13.Flag	DLG1	CN_OFF
Super-Scaffold_30	54694477	AR-halfsite_NR_LNCaP-AR	DLGAP1	CN_OFF
Super-Scaffold_37	4426325	EBF1_EBF_Near-E2A	DNAJC15	CN_OFF
Super-Scaffold_30	47799321	STAT6_Stat_CD4-Stat6	DOK6	CN_OFF
Super-Scaffold_30	47931107	Sox10_HMG_SciaticNerve-Sox3	DOK6	CN_OFF
Super-Scaffold_30	47931356	CArG_MADS_PUER-Srf	DOK6	CN_OFF
Super-Scaffold_40	6015636	Tgif2_Homeobox_mES-Tgif2	DPP6	CN_OFF
Super-Scaffold_41	1331100	bZIP:IRF_bZIP	DPYD	CN_OFF
Super-Scaffold_41	1528868	Arnt_Ahr_bHLH_MCF7-Arnt	DPYD	CN_OFF
Super-Scaffold_30	61358525	bZIP:IRF_bZIP	DTNA	CN_OFF
Super-Scaffold_17	69809075	ZNF264_Zf_HEK293-ZNF264.GFP	EFHC1	CN_OFF
Super-Scaffold_24	10488641	Bcl6_Zf_Liver-Bcl6	EHBP1	CN_OFF
Super-Scaffold_22	5671428	Tbx5_T-box_HL1-Tbx5.biotin	EPHA4	CN_OFF
Super-Scaffold_32	12325129	CDX4_Homeobox_ZebrafishEmbryos-Cdx4.Myc	FAH	CN_OFF
Super-Scaffold_67	1663940	Unknown_Homeobox_Limb-p300	FAM13C	CN_OFF
Super-Scaffold_94	1042215	Hoxa11_Homeobox_ChickenMSG-Hoxa11.Flag	FBXO7	CN_OFF
Super-Scaffold_10	29945123	CRX_Homeobox_Retina-Crx	FHDC1	CN_OFF

Super-Scaffold_8	4203934	ZEB2_Zf_SNU398-ZEB2	FRMPD4	CN_OFF
Super-Scaffold_65	7451386	CEBP_CEBP_bZIP_MEF-Chop	GABRA2	CN_OFF
Super-Scaffold_65	7454413	Rfx1_HTH_NPC-H3K4me1	GABRA2	CN_OFF
Super-Scaffold_65	7504361	Hoxa13_Homeobox_ChickenMSG-Hoxa13.Flag	GABRA2	CN_OFF
Super-Scaffold_47	2044053	Hoxa13_Homeobox_ChickenMSG-Hoxa13.Flag	GALC	CN_OFF
Super-Scaffold_47	2044202	bZIP:IRF_bZIP	GALC	CN_OFF
Super-Scaffold_10	36693962	CRX_Homeobox_Retina-Crx	GALNT7	CN_OFF
Super-Scaffold_10	36694366	TATA-Box_TBP_Promoter	GALNT7	CN_OFF
Super-Scaffold_10	36694470	Sox10_HMG_SciaticNerve-Sox3	GALNT7	CN_OFF
Super-Scaffold_26	780791	Zic3_Zf_mES-Zic3	GAPVD1	CN_OFF
Super-Scaffold_17	66375016	Brn1_POU	GCFC2	CN_OFF
Super-Scaffold_17	66375074	Brn1_POU_Homeobox_NPC-Brn1	GCFC2	CN_OFF
Super-Scaffold_72	2480791	Hoxa13_Homeobox_ChickenMSG-Hoxa13.Flag	GET4	CN_OFF
Super-Scaffold_12	8895011	p63_p53_Keratinocyte-p63	GPC3	CN_OFF
Super-Scaffold_46	23586736	AR-halfsite_NR_LNCaP-AR	GRIP1	CN_OFF
Super-Scaffold_46	23632104	Tbx21_T-box_GM12878-TBX21	GRIP1	CN_OFF
Super-Scaffold_46	23632109	Tbet_T-box_CD8-Tbet	GRIP1	CN_OFF
Super-Scaffold_56	9959304	Brn1_POU	GRM3	CN_OFF
Super-Scaffold_76	16135605	ARE_NR_LNCAP-AR	GRM5	CN_OFF
Super-Scaffold_56	5482772	Fox_Ebox_Forkhead_bHLH_Panc1-Foxa2	GSAP	CN_OFF
Super-Scaffold_17	49936095	Tbx5_T-box_HL1-Tbx5.biotin	GSTA1	CN_OFF
Super-Scaffold_1	5224092	Bcl6_Zf_Liver-Bcl6	GTF2H1	CN_OFF
Super-Scaffold_47	11267449	Isl1_Homeobox_Neuron-Isl1	HEATR5A	CN_OFF
Super-Scaffold_47	11267691	Brn1_POU	HEATR5A	CN_OFF
Super-Scaffold_30	4367357	Sox9_HMG_Limb-SOX9	HECW1	CN_OFF
Super-Scaffold_17	21050214	THRb_NR_Liver-NR1A2	HEY2	CN_OFF
Super-Scaffold_67	23849978	Nanog_Homeobox_mES-Nanog	HSPA12A	CN_OFF
Super-Scaffold_56	10270876	NFAT_AP1_RHD_bZIP_Jurkat-NFATC1	HSPA14	CN_OFF
Super-Scaffold_17	49909016	GLIS3_Zf_Thyroid-Glis3.GFP	ICK	CN_OFF
Super-Scaffold_17	49909176	GATA3_Zf_iTreg-Gata3	ICK	CN_OFF
Super-Scaffold_89	6585905	Brn1_POU	IP6K2	CN_OFF
Super-Scaffold_30	71597866	Nkx2.1_Homeobox_LungAC-Nkx2.1	KCNB2	CN_OFF
Super-Scaffold_10	48377895	KLF3_Zf_MEF-Klf3	KDM3A	CN_OFF
Super-Scaffold_10	34023853	RUNX2_Runt_PCa-RUNX2	KLHL2	CN_OFF
Super-Scaffold_73	4330760	Nkx2.5_Homeobox_HL1-Nkx2.5.biotin	LGMN	CN_OFF
Super-Scaffold_17	44539163	E2F3_E2F_MEF-E2F3	LMBRD1	CN_OFF
Super-Scaffold_103	11303693	Gata1_Zf_K562-GATA1	LOC100859020	CN_OFF
Super-Scaffold_8	13900637	PR_NR_T47D-PR	LOC101816675	CN_OFF
Super-Scaffold_30	39784298	Oct6_POU	LPCAT1	CN_OFF
Super-Scaffold_56	15795412	GFY_?_Promoter	LRGUK	CN_OFF
Super-Scaffold_55	1429916	Mef2b_MADS_HEK293-Mef2b.V5	LRRC4C	CN_OFF
Super-Scaffold_37	39052930	Pitx1_Homeobox_Chicken-Pitx1	MCF2L	CN_OFF
Super-Scaffold_17	69766567	E2F1_E2F_Hela-E2F1	MCM3	CN_OFF
Super-Scaffold_73	17217046	Arnt_Ahr_bHLH_MCF7-Arnt	Meis2a.1	CN_OFF
Super-Scaffold_103	5859145	Pitx1_Homeobox_Chicken-Pitx1	MGAT5	CN_OFF
Super-Scaffold_103	5859367	Hoxa11_Homeobox_ChickenMSG-Hoxa11.Flag	MGAT5	CN_OFF
Super-Scaffold_8	3361606	GATA3_Zf_iTreg-Gata3	MID1	CN_OFF
Super-Scaffold_10	34091858	Tgif1_Homeobox_mES-Tgif1	MSMO1	CN_OFF
Super-Scaffold_67	20689748	Smad4_MAD_ESC-SMAD4	MXI1	CN_OFF
Super-Scaffold_80	841383	Nkx2.1_Homeobox_LungAC-Nkx2.1	MYH15	CN_OFF
Super-Scaffold_80	841604	AR-halfsite_NR_LNCaP-AR	MYH15	CN_OFF
Super-Scaffold_103	18772121	CDX4_Homeobox_ZebrafishEmbryos-Cdx4.Myc	MYO3B	CN_OFF
Super-Scaffold_103	18809436	Tgif1_Homeobox_mES-Tgif1	MYO3B	CN_OFF
Super-Scaffold_32	8475886	BMAL1_bHLH_Liver-Bmal1	MYO5A	CN_OFF
Super-Scaffold_54	4015620	Fox_Ebox_Forkhead_bHLH_Panc1-Foxa2	NCAPD3	CN_OFF
Super-Scaffold_10	35451995	Bcl6_Zf_Liver-Bcl6	NEK1	CN_OFF
Super-Scaffold_10	30851564	PAX5_Paired_Homeobox_GM12878-PAX5	NPY2R	CN_OFF
Super-Scaffold_10	14138597	CDX4_Homeobox_ZebrafishEmbryos-Cdx4.Myc	NR3C2	CN_OFF
Super-Scaffold_47	1031335	Nanog_Homeobox_mES-Nanog	NRDE2	CN_OFF
Super-Scaffold_30	92842491	Znf263_Zf_K562-Znf263	NSMCE2	CN_OFF

Super-Scaffold_40	18198969	Oct6_POU	NSUN6	CN_OFF
Super-Scaffold_67	27164366	Atf1_bZIP_K562-ATF1	OAT	CN_OFF
Super-Scaffold_10	35284543	Rfx6_HTH_Min6b1-Rfx6.HA	PALLD	CN_OFF
Super-Scaffold_1	7313940	Oct6_POU	PDE3B	CN_OFF
Super-Scaffold_40	21882437	Znf263_Zf_K562-Znf263	PEX1	CN_OFF
Super-Scaffold_67	10770104	ZNF467_Zf_HEK293-ZNF467.GFP	PIK3AP1	CN_OFF
Super-Scaffold_48	7422537	ELF5_ETS_T47D-ELF5	PLBD1	CN_OFF
Super-Scaffold_103	27105304	AR-halfsite_NR_LNCaP-AR	PLCL1	CN_OFF
Super-Scaffold_46	4489290	CRX_Homeobox_Retina-Crx	PLEKHA5	CN_OFF
Super-Scaffold_1	6186311	SCL_bHLH_HPC7-Scl	PLEKHA7	CN_OFF
Super-Scaffold_1	6186895	NeuroG2_bHLH_Fibroblast-NeuroG2	PLEKHA7	CN_OFF
Super-Scaffold_35	3108361	bZIP:IRF_bZIP	PPP2R2C	CN_OFF
Super-Scaffold_48	3930150	ZNF7_Zf_HepG2-ZNF7.Flag	PTN	CN_OFF
Super-Scaffold_24	3708663	KLF14_Zf_HEK293-KLF14.GFP	PUS10	CN_OFF
Super-Scaffold_35	867260	HIF-1b_HLH_T47D-HIF1b	RAB28	CN_OFF
Super-Scaffold_10	32302962	TEAD1_TEAD_HepG2-TEAD1	RAPGEF2	CN_OFF
Super-Scaffold_10	32362595	Bcl11a_Zf_HSPC-BCL11A	RAPGEF2	CN_OFF
Super-Scaffold_67	5563401	Oct2_POU	RASGEF1A	CN_OFF
Super-Scaffold_10	48377895	KLF3_Zf_MEF-Klf3	REEP1	CN_OFF
Super-Scaffold_55	7868804	Foxa2_Forkhead_Liver-Foxa2	RGS6	CN_OFF
Super-Scaffold_22	3908653	Gata1_Zf_K562-GATA1	RHBDD1	CN_OFF
Super-Scaffold_17	39308802	THRa_NR_C17.2-THRa	RWDD2A	CN_OFF
Super-Scaffold_73	3586782	Oct2_POU	SERPINA10	CN_OFF
Super-Scaffold_103	22280670	Tgif1_Homeobox_mES-Tgif1	SESTD1	CN_OFF
Super-Scaffold_10	35361750	Olig2_bHLH_Neuron-Olig2	SH3RF1	CN_OFF
Super-Scaffold_10	35361752	Olig2_bHLH_Neuron-Olig2	SH3RF1	CN_OFF
Super-Scaffold_65	6775973	Gata4_Zf_Heart-Gata4	SLAIN2	CN_OFF
Super-Scaffold_73	364112	AR-halfsite_NR_LNCaP-AR	SLC25A29	CN_OFF
Super-Scaffold_41	122424	Hoxa13_Homeobox_ChickenMSG-Hoxa13.Flag	SLC44A5	CN_OFF
Super-Scaffold_46	16658093	Bapx1_Homeobox_VertebralCol-Bapx1	SLC6A15	CN_OFF
Super-Scaffold_46	16658094	Bapx1_Homeobox_VertebralCol-Bapx1	SLC6A15	CN_OFF
Super-Scaffold_10	15959490	SPDEF_ETS_VCaP-SPDEF	SMARCA5	CN_OFF
Super-Scaffold_10	15961393	Otx2_Homeobox_EpiLC-Otx2	SMARCA5	CN_OFF
Super-Scaffold_46	13479029	DMRT1_DM_Testis-DMRT1	SOCS2	CN_OFF
Super-Scaffold_30	92717024	Zac1_Zf_Neuro2A-Plagl1	SQLE	CN_OFF
Super-Scaffold_20	1683974	Rfx6_HTH_Min6b1-Rfx6.HA	SS18L1	CN_OFF
Super-Scaffold_10	1120268	THRa_NR_C17.2-THRa	STXBP5L	CN_OFF
Super-Scaffold_35	8394775	PU.1-IRF_ETS_IRF_Bcell-PU.1	TBC1D19	CN_OFF
Super-Scaffold_17	3911872	GLIS3_Zf_Thyroid-Glis3.GFP	TBCE	CN_OFF
Super-Scaffold_40	25815159	NFIL3_bZIP_HepG2-NFIL3	THSD7A	CN_OFF
Super-Scaffold_10	13819879	Zfp809_Zf_ES-Zfp809	TMEM184C	CN_OFF
Super-Scaffold_10	13820314	EBF1_EBF_Near-E2A	TMEM184C	CN_OFF
Super-Scaffold_73	19911957	Atf3_bZIP_GBM-ATF3	TMEM229B	CN_OFF
Super-Scaffold_73	19912646	Smad3_MAD_NPC-Smad3	TMEM229B	CN_OFF
Super-Scaffold_31	3923621	OCT_OCT_POU_Homeobox_NPC-OCT6	TSPAN12	CN_OFF
Super-Scaffold_31	3924296	Olig2_bHLH_Neuron-Olig2	TSPAN12	CN_OFF
Super-Scaffold_31	3924640	TEAD1_TEAD_HepG2-TEAD1	TSPAN12	CN_OFF
Super-Scaffold_1	1933844	Znf263_Zf_K562-Znf263	TSPAN4	CN_OFF
Super-Scaffold_103	29277234	Oct6_POU	TSSK6	CN_OFF
Super-Scaffold_103	29285265	NFkB-p65_RHD_GM12787-p65	TSSK6	CN_OFF
Super-Scaffold_30	19850211	PAX3_FKHR-fusion_Paired_Homeobox_Rh4-PAX3_FKHR	TUBB2B	CN_OFF
Super-Scaffold_31	725876	Pitx1_Homeobox_Chicken-Pitx1	TUBGCP6	CN_OFF
Super-Scaffold_65	3755292	Tcf7_HMG_GM12878-TCF7	TUSC3	CN_OFF
Super-Scaffold_30	62019145	LRF_Zf_Erythroblasts-ZBTB7A	UBE2V2	CN_OFF
Super-Scaffold_30	62019074	Elk1_ETS_Hela-Elk1	UBE2V2	CN_OFF
Super-Scaffold_10	22297337	Fox_Ebox_Forkhead_bHLH_Panc1-Foxa2	UNC5C	CN_OFF
Super-Scaffold_67	8823484	Smad4_MAD_ESC-SMAD4	VDAC2	CN_OFF
Super-Scaffold_67	8823575	PBX2_Homeobox_K562-PBX2	VDAC2	CN_OFF
Super-Scaffold_76	5165505	Oct4_Sox17_POU_Homeobox_HMG_F9-Sox17	WASF3	CN_OFF
Super-Scaffold_10	45962192	Egr1_Zf_K562-Egr1	WHSC1	CN_OFF

REFERENCES

[1] P. Berthold and A. Helbig, "The genetics of bird migration: stimulus, timing, and direction," *Ibis (Lond. 1859).*, vol. 134, pp. 35–40, 1992.

[2] P. Berthold, A. Helbig, G. Mohr, and U. Querner, "Rapid microevolution of migratory behaviour in a wild bird species," *Nature*, vol. 2, no. 3, pp. 173–179, 1992.

[3] K. E. Delmore, D. P. L. Toews, R. R. Germain, G. L. Owens, and D. E. Irwin, "The Genetics of Seasonal Migration and Plumage Color," *Curr. Biol.*, vol. 26, no. 16, pp. 2167–2173, 2016.

[4] M. Lundberg *et al.*, "Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks," *Evol. Lett.*, pp. 155–168, 2017.

[5] R. A. Bay, R. J. Harrigan, V. Le Underwood, H. L. Gibbs, T. B. Smith, and K. Ruegg, "Genomic signals of selection predict climate-driven population declines in a migratory bird," *Science (80-.).*, vol. 359, no. 6371, pp. 83 LP – 86, Jan. 2018.

[6] D. P. L. Toews, S. A. Taylor, H. M. Streby, G. R. Kramer, and I. J. Lovette, "Selection on VPS13A linked to migration in a songbird," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 37, pp. 18272–18274, 2019.

[7] J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, and W. J. Greenleaf, "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position," *Nat. Methods*, 2013.

[8] R. E. Thurman *et al.*, "The accessible chromatin landscape of the human genome," *Nature*, 2012.

[9] J. F. Fullard *et al.*, "An atlas of chromatin accessibility in the adult human brain," *Genome Res.*, 2018.

[10] D. Calderon *et al.*, "Landscape of stimulation-responsive chromatin across diverse human immune cells," *Nat. Genet.*, vol. 51, no. 10, pp. 1494–1505, 2019.

[11] T. B. Sackton *et al.*, "Convergent regulatory evolution and loss of flight in paleognathous birds," *Science (80-.).*, vol. 364, no. 6435, pp. 74 LP – 78, Apr. 2019.

[12] J. G. Roscito *et al.*, "Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution," *Nat. Commun.*, vol. 9, no. 1, 2018.

[13] J. J. Lewis *et al.*, "Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 48, pp. 24174–24183, 2019.

[14] K. R. L. van der Burg, J. J. Lewis, A. Martin, H. F. Nijhout, C. G. Danko, and R. D. Reed, "Contrasting Roles of Transcription Factors Spineless and EcR in the Highly Dynamic Chromatin Landscape of Butterfly Wing Metamorphosis," *Cell Rep.*, vol. 27, no. 4, pp. 1027-1038.e3, 2019.

[15] ENCODE(Consortium), "ATAC-seq Data Standards and Prototype Processing Pipeline." [Online]. Available: https://www.encodeproject.org/atac-seq/. [Accessed: 01-Mar-2020].

[16] M. Wikelski, E. M. Tarlow, A. Raim, R. H. Diehl, R. P. Larkin, and G. H. Visser, "Costs of migration in free-flying songbirds," *Nature*, vol. 423, no. 6941, p. 704, 2003.

[17] O. Fornes *et al.*, "JASPAR 2020: update of the open-access database of transcription factor binding profiles," *Nucleic Acids Res.*, vol. 48, no. D1, pp. D87–D92, 2020.

[18] S. Leskela *et al.*, "Molecular basis of tumor heterogeneity in endometrial carcinosarcoma," *Cancers (Basel).*, vol. 11, no. 7, pp. 1–21, 2019.

[19] T. M. Ortiga-Carvalho, A. R. Sidhaye, and F. E. Wondisford, "Thyroid hormone receptors and resistance to thyroid hormone disorders," *Nat. Rev. Endocrinol.*, vol. 10, no. 10, pp. 582–591, 2014.

[20] H. Dardente, D. G. Hazlerigg, and F. J. P. Ebling, "Thyroid hormone and seasonal rhythmicity," *Front. Endocrinol. (Lausanne).*, vol. 5, no. FEB, pp. 1–11, 2014.

[21] K. D. Shearer, T. H. Goodman, A. W. Ross, L. Reilly, P. J. Morgan, and P. J. McCaffery, "Photoperiodic regulation of retinoic acid signaling in the hypothalamus," *J. Neurochem.*, vol. 112, no. 1, pp. 246–257, 2010.

[22] T. Miyawaki *et al.*, "Tlx, an orphan nuclear receptor, regulates cell numbers and astrocyte development in the developing retina," *J. Neurosci.*, vol. 24, no. 37, pp. 8124–8134, 2004.

[23] A. D. Johnston, C. A. Simões-Pires, T. V. Thompson, M. Suzuki, and J. M. Greally, "Functional genetic variants can mediate their regulatory effects through alteration of transcription factor binding," *Nat. Commun.*, vol. 10, no. 1, pp. 1–16, 2019.

[24] D. Shlyueva, G. Stampfel, and A. Stark, "Transcriptional enhancers: From properties to genome-wide predictions," *Nature Reviews Genetics*. 2014.

[25] J. Gillespie, *Population Genetics: A Concise Guide*. Maryland: JHU Press., 2010.

[26] P. Franchini *et al.*, "Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species," *Mol. Ecol.*, 2017.

[27] M. Lundberg *et al.*, "Characterisation of a transcriptome to find sequence differences between two differentially migrating subspecies of the willow warbler Phylloscopus trochilus.," *BMC Genomics*, vol. 14, no. 1, p. 330, 2013.

[28] J. Boss *et al.*, "Gene expression in the brain of a migratory songbird during breeding and migration," *Mov. Ecol.*, pp. 1–11, 2016.

[29] R. A. Johnston, K. L. Paxton, F. R. Moore, R. K. Wayne, and T. B. Smith, "Seasonal gene expression in a migratory songbird," *Mol. Ecol.*, vol. 25, no. 22, pp. 5680–5691, 2016.

[30] J. C. Mueller, F. Pulido, and B. Kempenaers, "Identification of a gene associated with avian migratory behaviour," *Proc. R. Soc. B Biol. Sci.*, vol. 278, no. 1719, pp. 2848–2856, 2011.

[31] G. Bazzi *et al.*, "Clock gene polymorphism and scheduling of migration: a geolocator study of the barn swallow Hirundo rustica," *Sci. Rep.*, vol. 5, no. January, p. 12443, 2015.

[32] H. Mouritsen, D. Heyers, and O. Güntürkün, "The Neural Basis of Long-Distance Navigation in Birds," *Annu. Rev. Physiol.*, vol. 78, no. 1, pp. 133–154, 2016.

[33] B. M. V. Doren, M. Liedvogel, and B. Helm, "Programmed and flexible: long-term Zugunruhe data highlight the many axes of variation in avian migratory behaviour," *J. Avian Biol.*, vol. 48, no. 1, pp. 155–172, 2017.

[34] D. F. Sherry and S. A. MacDougall-Shackleton, "Seasonal change in the avian hippocampus," *Front. Neuroendocrinol.*, vol. 37, pp. 158–167, 2015.

[35] V. V. Pravosudov, A. S. Kitaysky, and A. Omanska, "The relationship between migratory behaviour, memory and the hippocampus: An intraspecific comparison," *Proc. R. Soc. B Biol. Sci.*, vol. 273, no. 1601, pp. 2641–2649, 2006.

[36] H. Mouritsen, G. Feenders, M. Liedvogel, K. Wada, and E. D. Jarvis, "Night-vision brain area in migratory songbirds," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 23, pp. 8339–8344, 2005.

[37] M. Zapka, D. Heyers, M. Liedvogel, E. D. Jarvis, and H. Mouritsen, "Night-time neuronal activation of Cluster N in a day- and night-migrating songbird," *Eur. J. Neurosci.*, 2010.

[38] V. M. Cassone and D. F. Westneat, "The bird of time: Cognition and the avian biological clock," *Front. Mol. Neurosci.*, vol. 5, no. MARCH, pp. 1–8, 2012.

[39] Y. Zhang *et al.*, "Model-based analysis of ChIP-Seq (MACS).," *Genome Biol.*, vol. 9, no. 9, p. R137, 2008.

[40] M. V Kuleshov *et al.*, "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update.," *Nucleic Acids Res.*, vol. 44, no. W1, pp. W90-7, Jul. 2016.

[41] S. Heinz *et al.*, "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities.," *Mol. Cell*, vol. 38, no. 4, pp. 576–589, May 2010.

[42] C. C. Chang, C. C. Chow, L. C. A. M. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee, "Second-generation PLINK: rising to the challenge of larger and richer datasets," *Gigascience*, vol. 4, no. 1, Feb. 2015.

PART III

PREFACE

Hybrid zones have been a long-standing feature of interest to study evolution because they represent a window in time of to evaluate what are the processes that differentiate populations. At the same time also represent the setup to study characteristics gene flow between populations and fitness consequences for hybrids. With the increasing availability of genomic resequencing studies of hybrid zones, there is evidence of the role of introgression patterns for barrier loci or adaptations and the potential evolutionary outcomes of gene flow across hybrid zones (hybrid speciation, adaptive introgression or extinction via hybridzation).

Nonetheless, all the approaches so far, take only into account pairs of populations in the same or very closely related species. But could we draw any patterns of parallelism in the genomic differentiation of different hybrid zones in different species? We researched this question looking for correlated patterns of divergence in eight pairs of songbird populations in hybrid zones. Songbirds are an ideal model for this question, because the genomes of the avian clade is highly conserved in features like chromosome numbers, recombination and macro synteny. We calculated windowed estimates of divergence (F_{ST} and d_{XY}) and its correlation in the different pairs. Our results showed that the degree of repeatability depend on two main factors: the divergence estimator (i.e, F_{ST} or d_{XY}) and the pair location along the speciation continuum.

CHAPTER 6

Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds

Kira E. Delmore, 1,2 Juan S. Lugo Ramos, 1 Benjamin M. Van Doren, 3 Max Lundberg, 4 Staffan Bensch, 4 Darren E. Irwin, 5 and Miriam Liedvogel 1

1Max Planck Institute for Evolutionary Biology, Behavioural Genomics, 24306, Plön, Germany

2E-mail: delmore@evolbio.mpg.de

3 Edward Grey Institute, Department of Zoology, University of Oxford, OX1 3PS Oxford, United Kingdom

4 Lund University, Department of Biology, 223 62 Lund, Sweden

5 Biodiversity Research Center, University of British Columbia, V6T 1Z4 Vancouver, British Columbia, Canada

Contributions: I performed the whole genome alignments, annotated the consensus alignments and analysed pairwise *dN/dS* patterns to obtain the mutation rate *(dS)* . I analysed the repeatability patterns of *dS*. I reviewed and edited draft versions of the manuscript.

Publication:
Delmore, K.E., Lugo Ramos, J.S., Van Doren, B.M., Lundberg, M., Bensch, S., Irwin, D.E. and Liedvogel, M. (2018), Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. Evolution Letters, 2: 76-87.

LETTER

doi:10.1002/evl3.46

Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds

Kira E. Delmore,[1,2] Juan S. Lugo Ramos,[1] Benjamin M. Van Doren,[3] Max Lundberg,[4] Staffan Bensch,[4] Darren E. Irwin,[5] and Miriam Liedvogel[1]

[1]*Max Planck Institute for Evolutionary Biology, Behavioural Genomics, 24306, Plön, Germany*

[2]*E-mail: delmore@evolbio.mpg.de*

[3]*Edward Grey Institute, Department of Zoology, University of Oxford, OX1 3PS Oxford, United Kingdom*

[4]*Lund University, Department of Biology, 223 62 Lund, Sweden*

[5]*Biodiversity Research Center, University of British Columbia, V6T 1Z4 Vancouver, British Columbia, Canada*

Received March 1, 2017
Accepted February 12, 2018

Heterogeneous patterns of genomic differentiation are commonly documented between closely related populations and there is considerable interest in identifying factors that contribute to their formation. These factors could include genomic features (e.g., areas of low recombination) that promote processes like linked selection (positive or purifying selection that affects linked neutral sites) at specific genomic regions. Examinations of repeatable patterns of differentiation across population pairs can provide insight into the role of these factors. Birds are well suited for this work, as genome structure is conserved across this group. Accordingly, we reestimated relative (F_{ST}) and absolute (d_{XY}) differentiation between eight sister pairs of birds that span a broad taxonomic range using a common pipeline. Across pairs, there were modest but significant correlations in window-based estimates of differentiation (up to 3% of variation explained for F_{ST} and 26% for d_{XY}), supporting a role for processes at conserved genomic features in generating heterogeneous patterns of differentiation; processes specific to each episode of population divergence likely explain the remaining variation. The role genomic features play was reinforced by linear models identifying several genomic variables (e.g., gene densities) as significant predictors of F_{ST} and d_{XY} repeatability. F_{ST} repeatability was higher among pairs that were further along the speciation continuum (i.e., more reproductively isolated) providing further insight into how genomic differentiation changes with population divergence; early stages of speciation may be dominated by positive selection that is different between pairs but becomes integrated with processes acting according to shared genomic features as speciation proceeds.

KEY WORDS: **Bird, genomic differentiation, genomic hitchhiking, islands of differentiation, population divergence, speciation.**

The integration of genomic data into research on population differentiation and speciation has led to the observation that genomic differentiation between closely related populations is often highly variable across the genome, with areas of elevated differentiation interspersed with areas of low differentiation (e.g., Nadeau et al. 2013; Renaut et al. 2013; Han et al. 2017; Vijay et al. 2017). One of the main conclusions from this observation is that speciation can proceed through a few focal changes and does not require divergence across the entire genome. This conclusion conforms to the genic view of speciation proposed by Wu (2001), but there is still considerable controversy concerning the factors that generate variation in estimates of genomic differentiation. This controversy has led to the development and extensive evaluation of two models.

The first model is termed divergence with gene flow (or speciation with gene flow; Nachman and Payseur 2012) and invokes

both selection and gene flow to explain heterogeneous patterns of differentiation. Specifically, this model holds that divergent selection at loci involved in reproductive isolation protects some regions of the genome from gene flow, elevating an otherwise homogenized landscape of differentiation (Nosil et al. 2009; Nosil and Feder 2012). The second model proposes that selection alone can generate variation in differentiation by accelerating lineage sorting at some regions of the genome. In other words, genomic regions that do not show elevated differentiation simply continue to share ancestral polymorphism (Noor and Bennett 2009; Turner and Hahn 2010; Cruickshank and Hahn 2014). We refer to this model as selection in allopatry and note that there are variants on this model related to when selection acts (Cruickshank and Hahn 2014; Delmore et al. 2015; Irwin et al. 2016).

One common feature of both the divergence with gene flow and selection in allopatry models is that features of the local genomic landscape should contribute to variation in differentiation. For example, genomic regions with lower rates of recombination, higher rates of mutation and elevated gene densities can promote linked selection, defined as any form of selection that influences variation at nearby neutral sites (Charlesworth et al. 1993; Lohmueller et al. 2011; Charlesworth 2012; Cutter and Payseur 2013; Enard et al. 2014). Linked selection can be positive, acting on new or existing mutations (genetic hitchhiking, Maynard Smith and Haigh 1974) or purifying, removing deleterious mutations from the population (background selection, Charlesworth et al. 1993). Measures of differentiation like F_{ST} include a term for within population variation and can be inflated by the reductions in variation that often accompany linked selection (Charlesworth 1998). Low recombination rates make it difficult for linked neutral sites to escape the effects of new mutations via recombination. Higher mutation rates and gene densities provide more targets for selection.

It was recently suggested that patterns of genomic differentiation will reflect features of the local genomic landscape more at later stages of speciation, as drift and selection at these features will take time to influence differentiation (Burri 2017). Comparative analyses examining genomic differentiation across multiple population pairs are ideal for both implicating features of the local genomic landscape in generating genomic differentiation and examining their temporal effects. For example, if genomic variables are conserved across pairs, constraints imposed by processes like linked selection in these regions should generate correlated or repeated patterns of genomic differentiation. Comparative analyses are beginning to accumulate but are often limited to a closely related group of species or populations, precluding an evaluation of temporal effects and introducing statistical nonindependence if a limited number of pairs are included (e.g., sticklebacks, Jones et al. 2012; sunflowers, Renaut et al. 2014; guppies, Fraser et al. 2015; songbirds, Irwin et al. 2016; Van Doren et al. 2017; copepods, Pereira et al. 2016). In addition, working at broader taxonomic scales may eliminate the role shared selective regimes play in generating repeatable patterns of differentiation, isolating the effects of genomic constraints.

Here, we overcome these limitations using new and archived genomic data to estimate genomic differentiation between eight pairs of birds that span a broad taxonomic range (the most recent common ancestor to them all was ~52 MYA, http://www.timetree.org/). We look for (1) correlated patterns of genomic differentiation across these pairs (referred to as "repeatability") and (2) an association between repeatability and the location of pairs along the speciation continuum (i.e., their level of reproductive isolation). We also (3) use linear models to implicate specific genomic features in generating repeatable patterns of genomic differentiation; these features include proxies for both recombination and mutation rates, gene density, chromosome size and proximity to chromosome ends and centromeres. We quantified the speciation continuum using hybrid zone width and genetic distance between pairs. Chromosome size, proximity to chromosome ends and centromeres may influence genomic differentiation as they have shown associations with recombination rates (Butlin 2005; Smukowski and Noor 2011). We also include linkage disequilibrium (LD) as a predictor in linear models; if linked selection is generating repeatable patterns, LD should be higher where repeatability is higher. Birds are well suited for this work as a considerable amount of information is known about speciation in this group (Price 2008) and genomic features are highly conserved across this group; birds have stable chromosome numbers, low rates of interchromosomal rearrangements, high synteny, and similar recombination landscapes (Dawson et al. 2007; Griffin et al. 2007; Backström et al. 2008; Stapley et al. 2008; Ellegren 2010; Kawakami et al. 2014; Zhang et al. 2014; Singhal et al. 2015; Kawakami et al. 2017).

Thus far we have only discussed how the local genomic landscape can affect F_{ST}, a relative measure of differentiation that is inflated by reductions in within population variation. Many studies are beginning to include d_{XY} in their analyses. This is an absolute measure of differentiation that does not include a term for within population variation. Under the divergence with gene flow model of speciation, those regions that contribute to reproductive isolation should have elevated d_{XY} compared to background levels of absolute differentiation. Under the selection in allopatry model of speciation, linked selection should have no effect on d_{XY} or reduce it compared to background levels (Nachman and Payseur 2012). The latter reductions could occur in response to recurrent linked selection in ancestral populations (which eventually results in reduced genetic distance between populations, Cruickshank et al. 2014) and/or selective sweeps of globally adaptive alleles (e.g., Delmore et al. 2015; Irwin et al. 2016). Given increasing interest in d_{XY} and its potential to reflect the local genomic landscape,

we include it in our analyses as well. All of the population pairs included in the present study occur in the temperate region where they have likely experience periods of allopatry with glacial expansions (Hewitt 2000). Accordingly, in analyses for objective 3 where we identify specific regions that show repeatable patterns we will focus on those at the bottom of the d_{XY} distribution.

Results

The eight pairs of birds in our study include European blackcaps (*Sylvia atricapilla*), subspecies of greenish warbler (*Phylloscopus trochiloides viridanus* and *P. t. plumbeitarsus*), subspecies of willow warbler (*Phylloscopus trochilus trochilus* and *P.t. acredula*), species of stonechat (European *Saxicola rubicola* and Siberian *S. maurus*), subspecies groups of the Swainson's thrush (coastal *Catharus ustulatus ustulatus* and inland *C. u. swainsoni*), species of flycatcher (collared *Ficedula albicollis* and pied *F. hypoleuca*), species of crow (hooded *Corvus cornix* and carrion *C. corone*), and species of wood warbler (blue- *Vermivora chrysoptera* and golden-winged *V. cyanoptera* warblers; Fig. 1). To gain an overview of the relationship between these pairs, we constructed a phylogeny for the group using whole-genome sequence data from all autosomal chromosomes (Fig. 1). The crow is the most distantly related species; all the remaining species cluster into two groups. One group includes the greenish warbler, willow warbler and blackcap while the other includes the flycatcher, stonechat, thrush, and blue/golden-winged warblers. Disregarding sister-pair relationships, the most closely related species are flycatchers and stonechats, and greenish and willow warblers. This topology is what we expected based on previous phylogenetic studies (e.g., Jetz et al. 2012, birdtree.org).

REPEATABILITY IN PATTERNS OF GENOMIC DIFFERENTIATION

To estimate repeatability in patterns of genomic differentiation across pairs we organized scaffolds from each species' reference into chromosomes using synteny with the flycatcher and estimated F_{ST} and d_{XY} between populations in each pair using the same 100 kb windows (Fig. 1). An initial comparison across pairs suggests that patterns may only be modestly consistent but stronger when considering d_{XY}. We evaluated this observation by correlating windowed estimates of F_{ST} and d_{XY} across pairs. In accordance with our observation, correlation coefficients varied from –0.02 to 0.18 for F_{ST} and 0.04 to 0.51 for d_{XY} (Table 1). Squaring the highest coefficients for F_{ST} and d_{XY}, these results suggest that up to 3 and 26% of the variation can be explained by correlations of F_{ST} and d_{XY} between pairs respectively.

We used a second method to quantify repeatability between pairs based on the overlap of outlier windows. We identified outlier windows for each pair as those in the top 5 percentile of the F_{ST} distribution and bottom 5 percentile of the d_{XY} distribution and compared the number of outlier windows that were shared (or overlapped) across pairs to the expected number based on the hypergeometric distribution (see Methods). Similar to results from correlations, estimates of overlap were higher and more significant for d_{XY} (Table S3A). This was also the case when we combined strings of outlier windows into peaks, acknowledging the fact outlier windows may not be independent of one another (Table S3B).

ASSOCIATION BETWEEN REPEATABILITY AND THE SPECIATION CONTINUUM

We looked for an association between repeatability and the speciation continuum using four different measures for the speciation continuum: hybrid zone width, the percentage of hybrids in these zones and genetic distance from both *cytb* and autosomal sequences. Starting with hybrid zone width and the percentage of hybrids, we obtained estimates for each pair from the literature and assumed reproductive isolation is greater in narrow hybrids zones with fewer hybrids (Barton and Hewitt 1985). Hybrid zone width ranged from 0 km for greenish warblers to 600 km for blue/golden-winged warblers; the percentage of hybrids ranged from 0% for greenish warblers to 70% for willow warblers (Table S4). We transformed these values into distance matrices and compared them with the correlation matrices generated using windowed estimates F_{ST} and d_{XY} above. Controlling for genetic distance between pairs, both hybrid zone width and the percentage of hybrids were negatively correlated with the with the correlation matrix based on F_{ST} but not d_{XY}. In other words, pairs with more narrow hybrid zones and fewer hybrids showed higher repeatability in F_{ST} (Partial Mantel tests, hybrid zone width, $R = -0.48$, CI = –0.67-(–0.21), $P = 0.02$; percentage of hybrids, $R = -0.42$, CI = –0.85-(–0.0082), $P = 0.02$) but not d_{XY} (Partial Mantel test, hybrid zone width, $R = 0.18$, CI = –0.21–0.39, $P = 0.23$; percentage of hybrids, $R = -0.18$, CI = –0.44 to 0.47, $P = 0.45$).

There are some caveats associated with using hybrid zone width and the percentage of hybrids as a proxies for reproductive isolation (e.g., differences in dispersal distance may affect estimates of hybrid zone width and there is considerable variation in how the percentage of hybrids is estimated, see Discussion). Accordingly, we reran these analyses using genetic distance between populations within each pair as a measure of reproductive isolation and assuming pairs that are more reproductively isolated from one another exhibit greater genetic distances. We estimated genetic distance using both *cytb* and autosomal sequences and similar to results using parameters from hybrid zones, we found a significant relationship between genetic distance within pairs and the correlation matrix based on F_{ST} but not d_{XY}; repeatability increased with genetic distance between pairs for F_{ST}

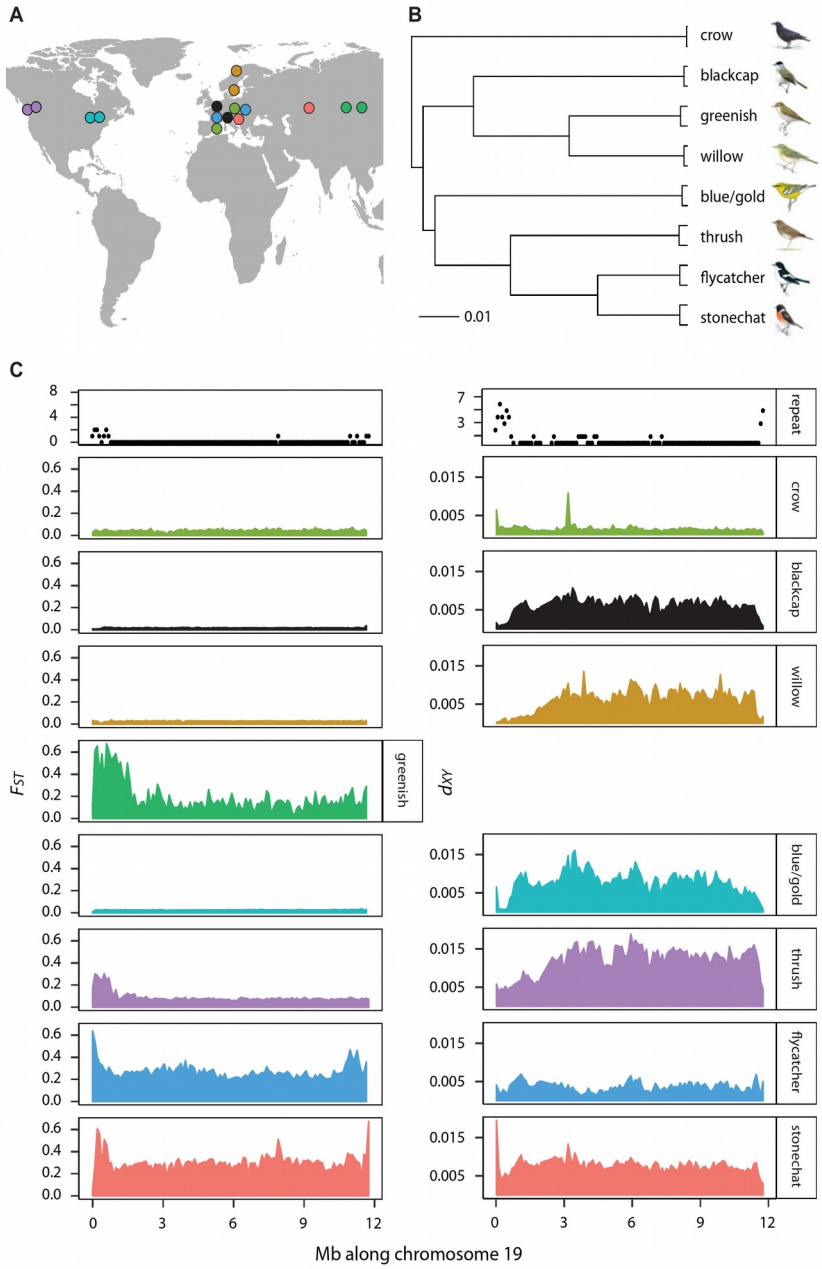


Figure 1. Population pairs, their (A) geographic ranges (circles as the center of sampling distributions) and (B) phylogenetic relationships (each branch has 100 bootstrapped support). Panel (C) shows windowed (100 kb) estimates of relative (F_{ST}) and absolute (d_{XY}) differentiation across chromosome 19 for population pairs along with repeatability at the top, measured as the number of pairs each windows as considered an outlier in (top 5 percentile of F_{ST} distribution and bottom 5 percentile of d_{XY} distribution; d_{XY} not estimated for greenish warblers).

Table 1. Repeatability in genomic differentiation across population pairs of birds.

	Flycatchers	Crows	Willows	Blackcaps	Greenish	Stonechats	Thrushes	Blue/gold
Flycatchers		0.17^{***}	0.21^{***}	0.23^{***}		0.51^{***}	0.11^{***}	0.17^{***}
Crows	0.058^{***}		0.09^{***}	0.14^{***}		0.21^{***}	0.04^{**}	0.22^{***}
Willows	0.020	0.028^{*}		0.29^{***}		0.12^{***}	0.19^{***}	0.24^{***}
Blackcaps	–0.017	0.037^{**}	0.029^{**}			0.18^{***}	0.44^{***}	0.42^{***}
Greenish	0.11^{***}	0.054^{***}	–0.020	0.021				
Stonechats	0.18^{***}	0.099^{***}	0.030^{**}	0.045^{***}	0.14^{***}		0.37^{***}	0.11^{***}
Thrushes	0.084^{***}	0.024^{**}	0.12^{***}	0.0099	0.11^{***}	0.13^{***}		0.21^{***}
Blue/gold	0.0077	0.036^{*}	0.055^{***}	0.052^{***}	0.027^{*}	0.032^{*}	0.099^{***}	

P-values corrected for multiple testing ($^{*}0.05$, $^{**}0.01$, $^{***}0.001$).
Values are correlation coefficients comparing windowed estimates of F_{ST} (below diagonal) and d_{XY} (above diagonal) between each set of population pairs. d_{XY} was not estimated for greenish warblers. For results based on outlier status and overlap values see Table S3.

(Partial Mantel test, *cytb*, $R = 0.41$, CI = 0.23–0.53, $P = 0.049$; autosomal, $R = 0.53$, CI = 0.32–0.67, $P = 0.009$) but not d_{XY} (Partial Mantel test, *cytb*, $R = –0.19$, CI = –0.42–0.20, $P = 0.22$; autosomal, $R = 0.016$, CI = (–0.24–0.26), $P = 0.48$).

Note, we reran these analyses replacing repeatability estimated by correlation coefficients (Table 1) with values based on the overlap of outlier windows and found similar associations (Table S3; e.g., results for F_{ST} and hybrid zone width, $R = –0.41$, CI = –0.72-(–0.15), $P = 0.03$; *cytb*, $R = 0.33$, CI = 0.05–0.52, $P = 0.05$; autosomal, $R = 0.41$, CI = 0.22–0.59, $P = 0.006$). This is an important finding, as it suggests the associations we documented are not related to the fact that there is a greater range of F_{ST} values at later stages of speciation.

GENOMIC FEATURES OR PROCESSES AS PREDICTORS OF REPEATABILITY

The repeated patterns of F_{ST} and d_{XY} we documented above suggest that variation in genome-wide estimates of differentiation is influenced by conserved features of the local genomic landscape. We used generalized linear models (GLMs) to evaluate the role-specific genomic features play in generating these patterns. Repeatability was quantified for each window as the number of pairs in which the window was considered an outlier (recall outlier status was determined using the top 5 percentile of the F_{ST} distribution and bottom 5 percentile of the d_{XY} distribution). Separate GLMs were run for each species pair using seven predictor variables (estimated for each pair separately): the proportion of GC bases (proxy for recombination rate), synonymous mutation rate (d_s; proxy for mutation rate), gene count, LD, and three variables related to where windows are located in the genome (micro- or macrochromosomes [chromosome size], proximity to chromosome ends, and centromeres).

Results from these GLMs can be found in Figure 2 (with blackcaps as predictor) and Table S5 (for each case as predictor). GC content, gene density, LD, and proximity to both chromosome ends and centromeres were significant predictors of repeatability in F_{ST} across all species pairs. Each of these variables was positive predictors of repeatability, except GC content that was negative. In the case of position, this means that windows near the center of chromosomes showed higher repeatability. Linkage disequilibrium, proximity to centromeres, and chromosome size were significant predictors of repeatability in d_{XY} across all species pairs. In the case of chromosome size, this means that windows on microchromosomes show more consistent patterns. GC content, proximity to chromosome ends, gene density, and d_s were also significant predictors of repeatability in d_{XY} but not for all species pairs.

Note that while the predictor variables used in these models are species-specific (e.g., gene density estimate for each pair separately) the response variable is the same–-a summary variable quantifying repeatability across species pairs. Accordingly, results from these models are not entirely independent.

Discussion

We used genomic data from eight population pairs of birds that span a broad taxonomic scale to study the contribution of local genomic features to variation in genome-wide estimates of differentiation. We rely on the fact genomic features are conserved across birds to draw inferences from our analyses and discuss these findings below, including the potential temporal effect these features can have of genomic differentiation.

Our first objective was to determine if patterns of genomic differentiation were correlated (or repeated) across population pairs. Our results suggest that up to 3% of the variation in F_{ST} and 26% of the variation in d_{XY} can be explained by correlations across pairs. Shared features of the genomic landscape likely contribute to these correlations. For example, linked selection (positive [genetic hitchhiking] or purifying [background selection]) is influenced by genomic features and tends to reduce within

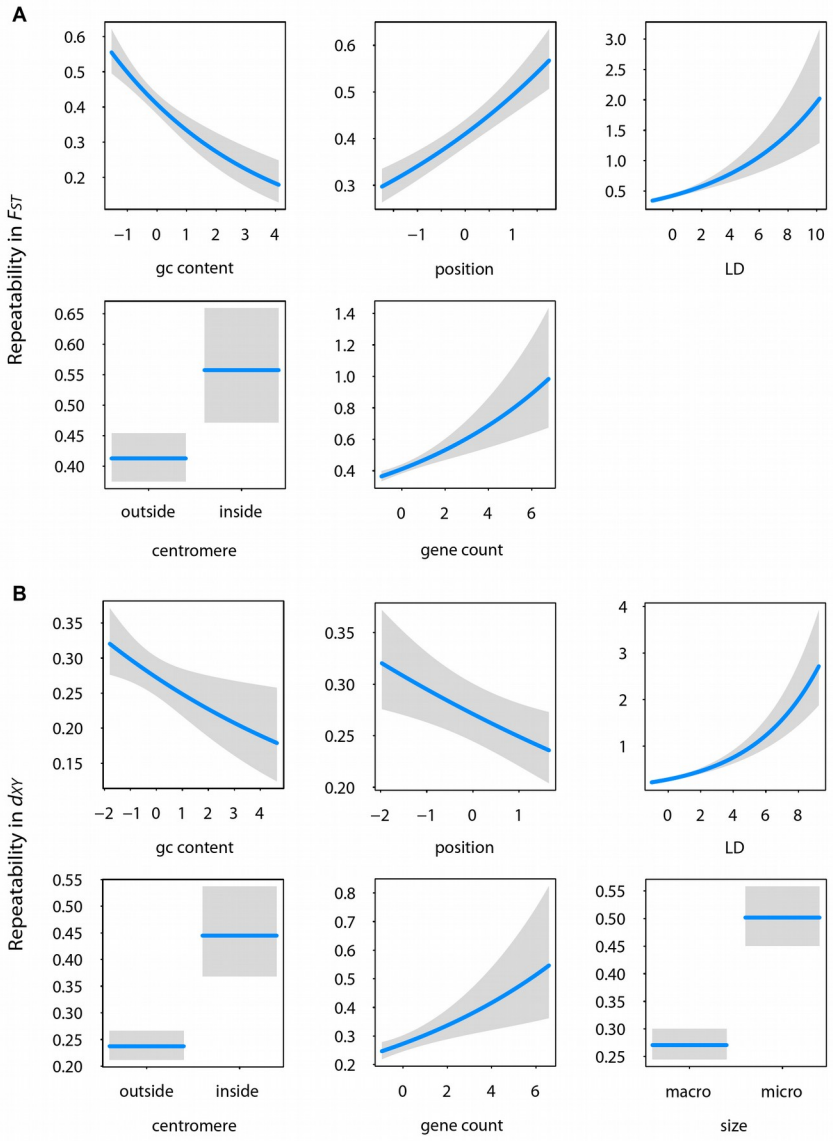


Figure 2. Results from GLMs examining the relationship between repeatability and predictor variables related to features of the local genomic landscape. Relationships shown are limited to significant predictor variables and results from blackcaps (results for nonsignificant predictor variables and the remaining population pairs and can be found in Table S5). Repeatability is estimated as the number of pairs each window was considered an outlier in (outliers are windows in the top 5 percentile of each species pairs' distribution for FST [A] and bottom 5 percentile for dXY [B]). Correlation coefficients for full models are 0.17 for FST and 0.27 for dXY. Parameter estimates are as follows: FST, GC content –0.20 [±0.04, $P < 0.001$], position 0.19 [±0.03, $P < 0.001$], gene count 0.13 [±0.03, $P < 0.001$], linkage disequilibrium 0.15 [±0.02, $P < 0.001$], centromere 0.32 [±0.09, $P < 0.001$]; dXY: GC content –0.09 [±0.04, $P = 0.01$], position –0.08 [±0.03, $P < 0.001$], gene count 0.11 [±0.03, $P < 0.001$], size 0.62 (±0.08, $P < 0.001$], linkage disequilibrium 0.24 [±0.02, $P < 0.001$], centromere 0.63 [±0.11, $P < 0.001$]). Each predictor is scaled and their effects are plotted with other variables held at their medians. A positive association with position indicates increased repeatability at the center of chromosomes.

population variation, inflating F_{ST}. Many of these genomic features are conserved across birds (Dawson et al. 2007; Griffin et al. 2007; Backström et al. 2008; Stapley et al. 2008; Ellegren 2010; Kawakami et al. 2014; Zhang et al. 2014; Singhal et al. 2015; Kawakami et al. 2017) and likely generated the repeatable patterns we observed. In support of this suggestion, recombination rates (approximated by GC content) and gene densities, two genomic features that are preserved across birds and influence linked selection were consistent predictors of repeatability in F_{ST}. Genetic drift may also contribute to the correlations we documented. For example, drift in a low recombination region can cause it to show consistently high or low F_{ST}. Nevertheless, drift is generally expected to reduce genetic diversity genome-wide. Note that the amount of variation explained by correlations across pairs is not as high as other studies (e.g., 49–77% of the variation in F_{ST} explained by correlations across subspecies pairs of greenish warblers in Irwin et al. 2016). Nevertheless, most of these studies focus on pairs that are much more closely related than those in the present study, such that pairs share more genomic features and are subject to similar selective forces.

Our second objective was to determine if there was a positive association between repeatability and the location of pairs along the speciation continuum. We measured the speciation continuum using hybrid zone width, the proportion of hybrids in these zones and genetic distance and found that pairs with narrower hybrid zones and greater genetic distances exhibited more similar patterns of F_{ST}. The latent effects of drift and selection may explain this pattern. Specifically, when population divergence begins, allele frequencies will be roughly equal and F_{ST} will be close to zero. Drift and selection will start acting on standing genetic variation and any new mutations that arise. The effects of these processes, especially drift and background selection, may take time to accumulate (Burri 2017). Accordingly, positive selection and genetic hitchhiking may be the primary forces affecting differentiation early in speciation. If the selective context of speciation is different for the pairs under study, this should lead to less repeatable (or correlated) patterns of differentiation. As speciation proceeds, drift and background selection will begin to affect differentiation as well and, combined with positive selection and genetic hitchhiking, these processes could result in the landscape of differentiation reflecting genomic features more directly. This scenario was described by Burri (2017) and is in line with recent work showing that linked selection (positive or purifying in nature) may generate repeated patterns of differentiation at longer time scales (Phung et al. 2016; Dutoit et al. 2017; Van Doren et al. 2017; Vijay et al. 2017). Note that the beginning stages of speciation may be less repeatable even without different selective pressures. For example, there may be more than one way to respond to selection and the chance positive selection affects the same genomic region may be low.

Our findings related to the speciation continuum will require additional study. To begin with, we used Partial Mantel tests for these analyses that may be prone to Type I errors (Harmon and Glor 2010, but see Diniz-Filho et al. 2013). Alternatives exist (e.g., Redundancy Analyses or correlograms) but require larger sample sizes. Our use of hybrid zone width to quantify the speciation continuum also assumes that dispersal distance is the same across all pairs. This is a common assumption among songbirds as it is difficult to obtain unbiased estimates of dispersal for this group that is based on large sample sizes. Analyses using the percentage of hybrids are not affected by dispersal distance but are associated with another set of assumptions, including that each study had the same resolution to identify hybrids and used similar sampling strategies. Regardless, the association we documented between repeatability and the speciation continuum is intriguing and was documented using not only parameters from hybrid zones but also genetic distance between populations in each pair.

Thus far we have focused mainly on results for F_{ST}; results for d_{XY} require careful explanation. Concerning the repeatable patterns we documented (first objective, up to 26% of the variation in d_{XY} explained by correlations across pairs), at the outset we argued that speciation in the pairs we studied would have been punctuated with periods of allopatry as all pairs occur in the temperate region where glacial advances would have isolated populations in different refugia (Hewitt 2000). Under this scenario (i.e., without gene flow), d_{XY} should reflect the amount of sequence divergence that has been acquired since populations split (along with variation that existed in the common ancestor) and linked selection should have no effect on d_{XY} or reduce it (e.g., if recurrent linked selection in ancestral populations removes variation from populations prior to their split; Nachman and Payseur 2012; Cruickshank and Hahn 2014). Consistent with the above scenario, patterns of d_{XY} were repeatable across pairs and several genomic features including gene densities and chromosome size predicted repeatable patterns at the bottom of the d_{XY} distribution. Nevertheless, it is important to note that during periods of secondary contact gene flow will reduce d_{XY} in some regions, generating peaks of differentiation. Accordingly, some of the repeatable patterns we documented may also be related to elevated d_{XY} (Nachman and Payseur 2012).

Continuing with d_{XY} repeatability (i.e., results for the first objective), correlation coefficients were higher for d_{XY} than F_{ST}, with the highest correlation coefficient for d_{XY} being more than twice that for F_{ST} (0.18 vs 0.51). This finding could be related to the fact that d_{XY} reflects processes that have accumulated over multiple speciation events (see below for additional explanation) while F_{ST} mainly reflects processes in extant populations. Accordingly, if population pairs are sampled too early in speciation, F_{ST} may not reflect local genomic features yet as it will take

time to accumulate (Burri 2017). This suggestion follows from the argument described above about the latent effects of drift and background selection. Nevertheless, additional explanations for increased d_{XY} repeatability are also possible. For example, d_{XY} shows a strong relationship with mutation rates (Geneva et al. 2015; Rosenzweig et al. 2016). Accordingly, much of the pattern we documented may be related solely to variation in mutation rates. It is also important to note that d_{XY} is estimated using far more sites than F_{ST} (variant and invariant for d_{XY} vs just variant for F_{ST}). Accordingly, these estimates may be more precise, leading to stronger correlation coefficients.

Finally, while we found a correlation between the speciation continuum and repeatability in F_{ST} (second objective) we did not document this association for d_{XY}. Again, this finding may be related to the fact that d_{XY} reflects processes that have accumulated over multiple speciation events while F_{ST} mainly reflects process in extant populations, including speciation. For example, if the recombination landscape has remained the same for millions of years, recurrent linked selection in areas of low recombination has likely been reducing variation over the same time period and these reductions will be passed down over speciation events (Burri 2017). Under this scenario, it will not matter what stage of differentiation the population pairs under study are at, these reductions will be reflected in estimates of d_{XY}. As we have already discussed, there are situations where d_{XY} will reflect processes in extant populations (especially if gene flow is occurring) but the underlying effect of ancestral diversity appears to override any effect these processes have on d_{XY} repeatability and the speciation continuum.

We documented modest but significant repeatability in relative and absolute differentiation across eight population pairs of birds and showed that several genomic features predicted this repeatability. As genomic features are conserved across birds, these results suggest that at least moderate amounts of variation in genome-wide differentiation can be attributed to processes acting at genomic features, including linked selection that may derive from both positive and purifying selection. A considerable amount of the remaining variation in genomic differentiation is likely related to processes specific to each episode of population divergence. This is especially true for pairs early in the process of speciation, as our observation that repeatability increases with the location of pairs along the speciation continuum suggests processes acting at shared genomic features become more important later in this processes. To the best of our knowledge, this is the first empirical support for a temporal role of genomic features in structuring genomic differentiation and we encourage future studies incorporating additional pairs to study this association and the genetics of speciation. Studies focused on a single system and points in time provide only a snapshot of this extensive and often prolonged process.

Methods

STUDY SPECIES AND DATASETS

We searched the Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) and European Nucleotide Archive (http://www.ebi.ac.uk/ena) for genomic data collected from birds. We limited our search to species for which a draft reference genome had been assembled for the target species or one that was closely related. This search resulted in the inclusion of eight pairs (Table S1). The only pair we did not have a reference genome for was the *Vermivora* warblers but a reference for the closely related yellow-rumped warblers (*Setophaga coronata*) is available and was used in the original publication for these data (Toews et al. 2016). All pairs are from the order Passeriformes (perching birds or songbirds) and breed in temperate regions (Fig. 1).

GENERATING CONSENSUS REFERENCE GENOMES

The reference genomes we acquired were all assembled into scaffolds, except the collared flycatcher's genome, which was organized into chromosomes based on linkage map for the species and synteny with zebra finch (Ellegren et al. 2012). Accordingly, we used this genome to ensure all windows compared across species were orthologous. To maintain chromosomal synteny, we aligned the scaffolds of each genome individually against the flycatcher genome with SatsumaSynteny (default parameters; Grabherr et al. 2010). We then used bash scripts to parse the output, obtaining information on the order and orientation of query scaffolds and conducted a final alignment with the LASTZ plugin in Geneious (Harris 2007; Kearse et al. 2012). We merged these scaffolds into pseudochromosomes by calling the query base where alignments occurred and Ns in the presence of a gap. Details on consensus genome coverage can be found in Table S2.

CONSTRUCTING PHYLOGENETIC NETWORK

We used ANGSD (Korneliussen et al. 2014) to obtain consensus fasta sequences for populations from each species pair (-doFasta 2 -doCounts 1 -minQ 20 -setMinDepth 10) and IQ-TREE (Nguyen et al. 2016) to construct a maximum-likelihood tree from these sequences. Patristic distance between pairs was estimated using this tree using the cophenetic.phylo function from the R package "ape." To compare relative divergence across species pairs we used the chronos function from the same R package, using the default of correlated rate model and generating an ultrametric tree.

ESTIMATING DIFFERENTIATION

We focused on SNPs for the present study and used a common reference-based bioinformatics pipeline to call them. Details can be found in Delmore et al. (2015, 2016). Briefly, we trimmed reads with trimmomatic (TRAILING:3 SLIDINGWINDOW:

4:10 MINLEN:30) and aligned them to consensus genomes using bwa *mem* (Li and Durbin 2009) using default settings. We then used GATK (McKenna et al. 2010) and picardtools (http://broadinstitute.github.io/picard) to identify and realign reads around indels (*RealignerTargetCreator*, *IndelRealigner*, default settings) and removed duplicates (*MarkDuplicates*, default settings) for all datasets except greenish warbler that consisted of GBS data.

We used two estimates of differentiation in our study: F_{ST} and d_{XY}. We estimated F_{ST} for datasets comprised of individuals using ANGSD, estimating site frequency spectrums for each population separately (-dosaf 1, -gl 1, -remove_bads, -unique_only, -minMapQ 20, -minQ 20, -only_proper_pairs 1, -trim 0) and using these to obtain joint frequencies spectrums for population pairs. These joint frequency spectrums were then used as priors for allele frequencies at each site to estimate F_{ST}. For datasets comprised of pools we estimated F_{ST} using Popoolation2 (Kofler et al. 2011; -min-coverage 30 for Swainson's thrushes and 10 for stonechats, -min-count 3, -minq 20). We summarized F_{ST} into windows of 100 kb and limited analyses to windows with data from all pairs. We excluded the Z chromosome from all analyses as some of the pairs included females where systematic biases related to coverage could affect estimates of differentiation.

We estimated d_{XY} for datasets comprised of individuals using ANGSD as well. First, we estimated allele frequencies at each SNP for both populations of each pair combined -doMajorMinor 4, -doMaf 2, -gl 1, -doCounts 1, -remove_bads, -unique_only, -minMapQ 20, -minQ 20, -only_proper_pairs 1, -trim 0, -SNP_pval 1e-6). We then reran the program by population using only the SNPs that passed the previous step, to ensure SNPs fixed in one population were not lost. Once we had these allele frequencies, we estimate d_{XY} at each SNP using a script provided with ANGSD (https://github.com/mfumagalli/ngsPopGen//scripts/calcDxy.R) and as (p1*(1−p2))+(p2*(1−p1)) where p is the allele frequency of a given allele in populations 1 and 2, respectively and averaged these values in the same 100 kb windows used for F_{ST}. Estimates of d_{XY} by SNP have to be normalized by the number of sites (variant and invariant) in a window. We obtained these values using ANGSD to estimate read depth at all sites (-doCounts 1, -dumpCounts 1, -remove_bads, -unique_only, -minMapQ 20, -minQ 20, -only_proper_pairs 1, -trim 0) and excluded sites with coverage less than three times the sample size and more than three times the average coverage to ensure roughly three reads per individual and exclude sites that may have mapping problems resulting from copy number variants. Analyses were limited to windows with data from all pairs and windows with more than 5000 callable sites, as d_{XY} can be highly variable with small sample sizes and coverage (e.g., Clarkson et al. 2014). This filter precluded the use of greenish warblers in analyses of d_{XY} as data for this pair were derived from reduce-representation sequencing and did not have high coverage in windows of 100 kb. This was not a problem for the original publication (Irwin et al. 2016) as windows were defined by SNPs rather than base pairs.

For stonechats and thrushes, for which we used pooled sequencing data, we calculated d_{XY} with a custom script. We excluded sites with coverage below 10 for stonechats and below 30 for thrushes. We estimated d_{XY} by multiplying allele frequencies for each base as above and averaging across sufficiently covered bases in each window.

ESTIMATING REPEATABILITY IN PATTERNS OF GENOMIC DIFFERENTIATION

We estimated overall repeatability between pairs by correlating windowed-estimates of differentiation across pairs. We also estimated repeatability using information on outlier status and overlap. Specifically, we identified outlier windows for each species pairs as those above the top 5% quantile for F_{ST} and below the bottom 5% quantile for d_{XY}. For each comparison across pairs (e.g., flycatcher to stonechat, flycatcher to greenish warbler, etc.), we counted the number of outlier windows that were shared (or overlapped) and compared this to the expected number of overlapping windows using the hypergeometric distribution, assuming that each window had an equal probability of being considered an outlier. We used these values (observed and expected) to calculate z scores for each comparison and calculated one-sided *P*-values (i.e., the probability of obtaining an overlap value as extreme or more extreme than our observed value). Z scores are effect sizes that correspond to the number of standard deviations above the average expectation in each comparison, allowing for direct comparison across studies. Note that it is also possible that the outlier windows we identified are not independent of one another. Accordingly, we reran this analysis modifying our overlap approach by combining outlier windows into peaks if they occurred next to one another and using the number of peaks as our estimate of overlap. We used a permutation test to quantify the significance of these values, holding the number and size of outlier regions constant while randomly permuting their location 1000 times and calculating one-sided *P*-values again.

PLACING PAIRS ALONG THE SPECIATION CONTINUUM

We used four methods to place pairs along the speciation continuum (i.e., to quantify the level of reproductive isolation), starting with the width of hybrid zones and percentage of hybrids in these zones, which we obtained from the literature. The more narrow a zone and the fewer hybrids present the higher the reproductive isolation (Barton and Hewitt 1985; Moore and Dolbeer 1989; Paradis et al. 1998). If hybridization is extremely rare (e.g., with

the greenish warblers), we set the width as zero. We also used genetic distance within pairs, which we obtained by aligning *ctyb* sequences (downloaded from NCBI https://www.ncbi.nlm.nih.gov/) using MEGA as *p*-distance (the proportion of nucleotide sites that differ between groups) and the distance matrix generated by IQTREE for autosomal chromosome alignments (see "Constructing phylogenetic network"). We used Mantel tests to compare distance matrices quantifying the speciation continuum with distance matrices quantifying repeatability. We accounted for small sample sizes in these tests by using permutation tests to quantify significance and the non-parametric Spearman rank correlation coefficient.

MEASURING GENOMIC FEATURES AND PROCESSES AND THEIR EFFECT ON REPEATABILITY

We looked at the relationship between repeatability and seven structural features of the genome: recombination rate, mutation rate, gene density, chromosome size, proximity to chromosome ends, and centromeres and linkage disequilibrium. We estimated these features for each species and ran separate generalized linear models (GLMs) with repeatability as the response variable with a Poisson distribution for each species. These models were run with the glm function in base R and the ANOVA function was used to evaluate the significance of each predictor variable. We visualized results with the "visreg" package and estimated correlation coefficients (and confidence intervals) for each model by regressing observed repeatability to repeatability predicted by each model.

We used GC content as a proxy for recombination. Recombination affects the patterns of local base composition via the unbalanced transmission of "strong" (GC) over "weak" (AT) alleles at double-strand breaks (Mugal et al. 2015). This process is termed GC-biased gene conversion and direct support was recently presented in birds (Smeds et al. 2016). Positive correlations between recombination and GC content have also been documented in birds (Kawakami et al. 2014; Burri et al. 2015). Synonymous mutations occur in the exon of genes but have no effect on the sequence of amino acids. The use of d_s for mutation rate analysis assumes these sites do not experience selection (Eyre-Walker and Keightley 1999). We used a phylogenetic framework to obtain these estimates; details can be found in the Supplementary Methods. Briefly, we annotated each consensus genome with MAKER and identified potential homologues for high quality transcripts (AED $<$ 0.05) using a Blastn search against all transcripts from the flycatcher (flycatcher was searched against zebra finch). We aligned codons from each pair of sequences using PRANK (http://wasabiapp.org/software/prank) and calculated d_N/d_s with PAML v4.8. All d_N/d_s calculations were performed pairwise, comparing all the species with the flycatcher and this in turn, compared to zebra finch. Estimates of d_s were extracted for GLMs. We used PLINK 1.9 (Chang et al. 2015) to estimate linkage disequilibrium for one population from each pair (the same population for which we had a reference genome), as the squared correlation coefficient (r^2) between pairs of SNPs. SNPs were output from ANGSD using the same filters described above for F_{ST} and d_{XY} and including an additional filter for minor allele frequency, requiring SNPs have minor allele frequencies greater than 0.05. PLINK was run with the command line with the command line "–ld-window 100 –ld-window-kb 100 –ld-window-r2 0" to limit the analyses to SNPs with fewer than 100 variants between them and no more than 100 kilobases apart and report pairs with r^2 values below 0.2 as well. We determined the midpoints for all SNP pairs, binned them into the same 100 kb windows used for F_{ST} and d_{XY} and calculated average values for each window.

Avian genomes are composed of micro- and macrochromosomes. We considered microchromosomes those that are less than 20 Mb (Ellegren 2013) and macrochromosomes those that are greater than 40 Mb. We identified the position of each window along the chromosome by dividing chromosomes in half and generating a variable that increased by a value of 1 for each window from the end of the chromosome to its center. We standardized this measure by dividing these values by half the number of windows on each chromosome, so values increased to 1 at the center of chromosomes. We inferred the location of centromeres using methods employed by Ellegren et al. (2012) and Delmore et al. (2015). Specifically, we identified FISH probes from Warren *et al.* (2010) on either side of centromeres in the zebra finch genome and used NCBI's blastn (Altschul et al. 1990) to find their location in our genome. We considered sequences between FISH probes as "centromeric regions" and note this method only gives us a rough approximation for the location of centromeres.

AUTHOR CONTRIBUTIONS

K.D. and M.Li. conceived and designed the study, all authors contributed data. K.D. conducted all analyses excluding the estimation of d_{XY} for pooled datasets and repeatability using peaked overlap values (BV). J.L. helped with the construction of consensus genomes and estimates of d_s and M.Lu. with the estimation of d_{XY} for individual datasets. K.D. wrote the manuscript with comments from all authors.

ACKNOWLEDGMENTS

We acknowledge funding from the Max Planck Society and NSERC and discussions with Diethard Tautz, Bernhard Haubold, Sam Yeaman, Reto Burri, Sam Flaxman, José Alexandre Felizola Diniz Filho, Diana Rennison, Kieran Samuk, Greg Owens, Sara Miller, Barbara Helm, Matthias Weissensteiner, and Niclas Backström.

DATA ACCESSIBILITY

Accession numbers for all genomic data used in this paper can be found in Table S1.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Backström, N., N. Karaiskou, E. H. Leder, L. Gustafsson, C. R. Primmer, A. Qvarnström, et al. 2008. A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. Genetics 179:1479–1495.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Burri, R. 2017. Interpreting differentiation landscapes in the light of long-term linked selection. Evol. Lett. 1:118–131.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, et al. 2015. 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.

Butlin, R. K. 2005. Recombination and speciation. Mol. Ecol. 14:2621–2635.

Charlesworth, B. 1998. Measures of divergence between populations and the effect of forces that reduce variability. Mol. Biol. Evol. 15:538–543.

———. 2009. Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10:195–205.

———. 2012. The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the *Drosophila* X chromosome. Genetics 191:233–246.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Clarkson, C. S., D. Weetman, J. Essandoh, A. E. Yawson, G. Maslen, M. Manske, et al. 2014. Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. Nat. Commun. 5(ncomms5): 248.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

Cutter, A. D., and B. A. Payseur. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14:262–274.

Dawson, D. A., M. Åkesson, T. Burke, J. M. Pemberton, J. Slate, and B. Hansson. 2007. Gene order and recombination rate in homologous chromosome regions of the chicken and a passerine bird. Mol. Biol. Evol. 24:1537–1552.

Delmore, K. E., S. Hübner, N. C. Kane, R. Schuster, R. L. Andrew, F. Câmara, et al. 2015. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol. Ecol. 24:1873–1888.

Delmore, K. E., D. P. L. Toews, R. R. Germain, G. L. Owens, and D. E. Irwin. 2016. The genetics of seasonal migration and plumage color. Curr. Biol. 26:2167–2173.

Diniz-Filho, J. A. F., T. N. Soares, J. S. Lima, R. Dobrovolski, V. L. Landeiro, M. P. de Campos Telles, et al. 2013. Mantel test in population genetics. Genet. Mol. Biol. 36:475–485.

Dutoit, L., N. Vijay, C. F. Mugal, C. M. Bossu, R. Burri, J. Wolf, et al. 2017. Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. Proc. R. Soc. B. 284: 20162756.

Ellegren, H. 2010. Evolutionary stasis: the stable chromosomes of birds. Trends Ecol. Evol. 25:283–291.

———. 2013. The evolutionary genomics of birds. Annu. Rev. Ecol. Evol. Syst. 44:239–259.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, et al. 2012. The genomic landscape of species divergence in *Ficedula flycatchers*. Nature 491:756–760.

Enard, D., P. W. Messer, and D. A. Petrov. 2014. Genome-wide signals of positive selection in human evolution. Genome Res. 24:885–895.

Eyre-Walker, A., and P. D. Keightley. 1999. High genomic deleterious mutation rates in hominids. Nature 397:344–347.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8:610–618.

Fraser, B. A., A. Künstner, D. N. Reznick, C. Dreyer, and D. Weigel. 2015. Population genomics of natural and experimental populations of guppies (*Poecilia reticulata*). Mol. Ecol. 24:389–408.

Geneva, A. J., C. A. Muirhead, S. B. Kingan, and D. Garrigan. 2015. A new method to scan genomes for introgression in a secondary contact model. PloS One 10:e0118621.

Grabherr, M. G., P. Russell, M. Meyer, E. Mauceli, J. Alföldi, F. Di Palma, et al. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26:1145–1151.

Griffin, D. K., L. B. W. Robertson, H. G. Tempest, and B. M. Skinner. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. Cytogenet. Genome Res. 117:64–77.

Han, F., S. Lamichhaney, B. R. Grant, P. R. Grant, L. Andersson, and M. T. Webster. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. Genome Res. 27:1004–1015.

Harmon, L. J., and R. E. Glor. 2010. Poor statistical performance of the Mantel test in phylogenetic comparative analyses. Evol. Int. J. Org. Evol. 64:2173–2178.

Harris, R. S. 2007. Improved pairwise alignment of genomic DNA. Pennsylvania State University.

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

Irwin, D. E., M. Alcaide, K. E. Delmore, J. H. Irwin, and G. L. Owens. 2016. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. Mol. Ecol. 25:4488–4507.

Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. Nature 491:444–448.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Kawakami, T., C. F. Mugal, A. Suh, A. Nater, R. Burri, L. Smeds, et al. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. Mol. Ecol. 26:4158–4172.

Kawakami, T., L. Smeds, N. Backström, A. Husby, A. Qvarnström, C. F. Mugal, et al. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol. Ecol. 23:4035–4058.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kofler, R., R. V. Pandey, and C. Schlötterer. 2011. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics 27:3435–3436.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15:1.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760.

Lohmueller, K. E., A. Albrechtsen, Y. Li, S. Y. Kim, T. Korneliussen, N. Vinckenbosch, et al. 2011. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLOS Genet. 7:e1002326.

Maynard Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genet. Res. 23:23–35.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Moore, W. S., and R. A. Dolbeer. 1989. The use of banding recovery data to estimate dispersal rates and gene flow in avian species: case studies in the red-winged blackbird and common grackle. Condor 91:242–253.

Mugal, C. F., C. C. Weber, and H. Ellegren. 2015. GC-biased gene conversion links the recombination landscape and demography to genomic base composition: GC-biased gene conversion drives genomic base composition across a wide range of species. BioEssays News Rev. Mol. Cell. Dev. Biol. 37:1317–1326.

Nachman, M. W., and B. A. Payseur. 2012. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philos. Trans. R. Soc. B Biol. Sci. 367:409–421.

Nadeau, N. J., S. H. Martin, K. M. Kozak, C. Salazar, K. K. Dasmahapatra, J. W. Davey, et al. 2013. Genome-wide patterns of divergence and gene flow across a butterfly radiation. Mol. Ecol. 22:814–826.

Noor, M. A., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439–444.

Nosil, P., and J. L. Feder. 2012. Genomic divergence during speciation: causes and consequences. Philos. Trans. R. Soc. B Biol. Sci. 367:332–342.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Paradis, E., S. R. Baillie, W. J. Sutherland, and R. D. Gregory. 1998. Patterns of natal and breeding dispersal in birds. J. Anim. Ecol. 67:518–536.

Pereira, R. J., F. S. Barreto, N. T. Pierce, M. Carneiro, and R. S. Burton. 2016. Transcriptome-wide patterns of divergence during allopatric evolution. Mol. Ecol. 25:1478–1493.

Phung, T. N., C. D. Huber, and K. E. Lohmueller. 2016. Determining the effect of natural selection on linked neutral divergence across species. PLoS Genet. 12:e1006199.

Price, T. 2008. Speciation in birds. Roberts and Company Publishers, Colorado.

Ravinet, M., A. Westram, K. Johannesson, R. Butlin, C. André, and M. Panova. 2016. Shared and nonshared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. Mol. Ecol. 25:287–305.

Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, et al. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. Nat. Commun. 4:1827.

Renaut, S., G. L. Owens, and L. H. Rieseberg. 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. Mol. Ecol. 23:311–324.

Rosenzweig, B. K., J. B. Pease, N. J. Besansky, and M. W. Hahn. 2016. Powerful methods for detecting introgressed regions from population genomic data. Mol. Ecol. 25:2387–2397.

Singhal, S., E. M. Leffler, K. Sannareddy, I. Turner, O. Venn, D. M. Hooper, et al. 2015. Stable recombination hotspots in birds. Science 350:928–932.

Smeds, L., C. F. Mugal, A. Qvarnström, and H. Ellegren. 2016. High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. PLOS Genet. 12:e1006044.

Smukowski, C. S., and M. A. F. Noor. 2011. Recombination rate variation in closely related species. Heredity 107:496.

Stapley, J., T. R. Birkhead, T. Burke, and J. Slate. 2008. A linkage map of the zebra finch *Taeniopygia guttata* provides new insights into avian genome evolution. Genetics 179:651–667.

Toews, D. P., S. A. Taylor, R. Vallender, A. Brelsford, B. G. Butcher, P. W. Messer, et al. 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr. Biol. 26:2313–2318.

Turner, T. L., and M. W. Hahn. 2010. Genomic islands of speciation or genomic islands and speciation? Mol. Ecol. 19:848–850.

Vallender, R., R. J. Robertson, V. L. Friesen, and I. J. Lovette. 2007. Complex hybridization dynamics between golden-winged and blue-winged warblers (*Vermivora chrysoptera* and *Vermivora pinus*) revealed by AFLP, microsatellite, intron and mtDNA markers. Mol. Ecol. 16:2017–2029.

Van Doren, B. A., L. Campagna, B. Helm, J. Illera, I. J. Lovette, and M. Liedvogel. 2017. Correlated patterns of genetic diversity and differentiation across an avian family. Mol. Ecol. 26:3982–3997.

Vijay, N., M. Weissensteiner, R. Burri, T. Kawakami, H. Ellegren, and J. B. Wolf. 2017. Genome-wide patterns of variation in genetic diversity are shared among populations, species and higher order taxa. Mol. Ecol.

Warren, W. C., D. F. Clayton, H. Ellegren, A. P. Arnold, L. W. Hillier, A. Künstner, et al. 2010. The genome of a songbird. Nature 464:757–762.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Zhang, G., E. D. Jarvis, and M. T. P. Gilbert. 2014. A flock of genomes. Science 346:1308–1309.

Associate Editor: Z. Gompert

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Datasets included in the present study.
Table S2. Coverage of consensus genomes.
Table S3. Repeatability measured using outlier status and overlap values.
Table S4. Data used to measure the speciation continuum.
Table S5. Results from GLMs examining the relationship between repeatability and predictor variables related to genomic features.

Table S1. Datasets included in the present study.

Species	*Ficedula* flycatchers	*Corvus* crows	*Phylloscopus* greenish warblers	*Catharus* thrushes	*Phylloscopus* willow warblers	*Saxicola* stonechats	*Sylvia* blackcaps	*Vermivora* warblers
Pair	Collared and pied (*F. albicollis* and *hypoleuca*)	Hooded and carrion (*C. corone orone* and *C. c. cornix*)	Eastern and western (*P. tochiloides plumbeitarsus* and *P. t. viridanus*)	Coastal and inland (*Catharus ustulatus ustulatus* and *C. u. swainsoni*)	Southern and northers (*P. trochilus trochilus* and *P. t. acredula*)	European and Siberian (*S. rubicola* and *S. maurus*)	German and Austrian populations (*Sylvia atricapilla*)	Blue- and golden-winged (*V. cyantoptera* and *V. chrysoptera*)
Resequencing data	WGS data from four populations (n=20/population)	WGS data from two populations (n=15/population)	GBS data from 34 and 19 birds, respectively	WGS poolseq data from 2 populations (n=10/population)	WGS data from two populations (n=9/population)	WGS poolseq data (n=49 [European] and 52[Siberian])	WGS data from two populations (n=15/population)	
Accession numbers	ENA PRJEB7359; Genbank AGT002000000	SRA PRJNA192205	SRA SRR1176844; DDBJ/ENA/GenBank LYPA00000000	SRA PRJNA275819	SRP074112	ENA PRJEB19452; PRJEB19453	https://www.zoology.ubc.ca/~kdelmore/; Dryad upon acceptance.	PRJNA325126; PRJNA325157
References	Burri et al. 2015 Kawakami et al. 2014	Poelstra et al. 2014	Alcaide et al. 2014; Irwin et al. 2016	Delmore et al. 2015	Lundberg et al. 2017	Van Doren et al. 2017	Delmore et al. in prep	Toews et al. 2016

Table S2. Summary of final consensus sequences included in the present study. The length of each chromosome is provided along with the percentage of the flycatcher chromosome it covers. Lengths are the number of base pairs in the consensus without gaps of Ns. Macrochromosomes are those greater than 40 Mb and microchromosomes those less than 20 Mb.

chr	**flycatcher**	blackcap		crow		blue/gold		willow		thrush		greenish		stonechat	
	length	length	%	length	%	length	%	Length	%	length	%	length	%	length	%
1	**120002344**	90136271	75.1	59620835	49.7	95993241	80.0	54058012	45.0	97347748	81.1	97683349	81.4	102066987	85.1
1A	**74947036**	48879787	65.2	40202034	53.6	57498691	76.7	57749718	77.1	58550819	78.1	61028700	81.4	63356200	84.5
2	**157563209**	132248109	83.9	73562279	46.7	122799733	77.9	116957659	74.2	129397975	82.1	130746161	83.0	117858622	74.8
3	**115844353**	96698794	83.5	75853557	65.5	90728830	78.3	92584803	79.9	96761318	83.5	97746071	84.4	97985779	84.6
4	**70439523**	52016224	73.8	60250282	85.5	56882071	80.8	37147958	52.7	58149848	82.6	59354464	84.3	62340083	88.5
4A	**21182716**	16026903	75.7	17022485	80.4	15762793	74.4	15154164	71.5	15488001	73.1	14334247	67.7	18296289	86.4
5	**64724594**	49907695	77.1	52762582	81.5	48862770	75.5	54367822	84.0	51849320	80.1	53034140	81.9	56805964	87.8
6	**37227452**	31384756	84.3	25164402	67.6	30082679	80.8	29498648	79.2	30044214	80.7	30396279	81.7	33408554	89.7
7	**39412007**	34094738	86.5	33361110	84.6	31303551	79.4	34263494	86.9	32013904	81.2	33078282	83.9	34858256	88.4
8	**32100816**	22935871	71.4	27830922	86.7	26078388	81.2	28020258	87.3	25395769	79.1	26648552	83.0	28513982	88.8
9	**26793321**	22576347	84.3	22910380	85.5	21086208	78.7	21779656	81.3	20931760	78.1	21800432	81.4	21898138	81.7
10	**21346708**	17924858	84.0	16798936	78.7	16978521	79.5	17865715	83.7	16663500	78.1	17498509	82.0	18776919	88.0
11	**21727166**	18258423	84.0	12507783	57.6	17216928	79.2	18746169	86.3	16691697	76.8	17523717	80.7	19428494	89.4
12	**21938106**	18236025	83.1	18666245	85.1	17207319	78.4	18665559	85.1	17090405	77.9	17265384	78.7	19429616	88.6
13	**18641552**	15235186	81.7	14485192	77.7	14290123	76.7	14820107	79.5	13540864	72.6	14345885	77.0	16149134	86.6
14	**17374186**	14026179	80.7	14416366	83.0	13296177	76.5	13987910	80.5	12888791	74.2	13233457	76.2	15145780	87.2
15	**14943019**	12117646	81.1	9640549	64.5	11301992	75.6	12105690	81.0	10698802	71.6	10643530	71.2	12836525	85.9
17	**12378331**	9789432	79.1	9758513	78.8	7512843	60.7	9969948	80.5	8307378	67.1	8957804	72.4	10137087	81.9
18	**13163162**	9605582	73.0	7496181	56.9	8062768	61.3	9676338	73.5	8554042	65.0	8166354	62.0	10117730	76.9
19	**11933672**	9561990	80.1	9593105	80.4	8425778	70.6	9775189	81.9	8406962	70.4	8779264	73.6	10234172	85.8
20	**15675940**	12527461	79.9	12566710	80.2	11973761	76.4	12967530	82.7	11239349	71.7	12053729	76.9	12804409	81.7
21	**8073070**	5742695	71.1	5944624	73.6	4230261	52.4	5818547	72.1	4709579	58.3	5100060	63.2	6288065	77.9
22	**5733621**	2396835	41.8	3162411	55.2	1871261	32.6	2718128	47.4	2053051	35.8	1540315	26.9	3592435	62.7
23	**7944683**	5308793	66.8	5126807	64.5	4132107	52.0	4646607	58.5	3739029	47.1	3522204	44.3	5559624	70.0
24	**8009359**	5796045	72.4	5988272	74.8	5037427	62.9	5997533	74.9	4496957	56.1	4707706	58.8	6516443	81.4
25	**2802420**	1025451	36.6	963319	34.4	721357	25.7	978166	34.9	682990	24.4	369145	13.2	1215130	43.4
26	**7653694**	4938326	64.5	4968167	64.9	3648646	47.7	4565093	59.6	3510613	45.9	3577204	46.7	3938873	51.5
27	**5572044**	3195560	57.3	3092758	55.5	2563901	46.0	3153518	56.6	2546310	45.7	2545477	45.7	3727994	66.9
28	**6182350**	3690716	59.7	3728152	60.3	2782648	45.0	3995888	64.6	2898680	46.9	2931216	47.4	4436153	71.8
Z	**59856998**	45300926	75.7	37259879	62.2	46372801	77.5	45029809	75.2	47742596	79.8	48790136	81.5	48563346	81.1
Total	**1041187452**	**811583624**	**77.9**	**684704837**	**65.8**	**794705574**	**76.3**	**757065636**	**72.7**	**812392271**	**78.0**	**827401773**	**79.5**	**866286783**	**83.2**

Table S3. Summarizing repeatability in genomic differentiation across pairs using outlier status of windows and observed values of overlap. Values in (a) are z-scores, comparing the number of outlier windows that were shared across each comparison to the mean expected number obtained using the hypergeometric distribution. Z-scores are effect sizes that indicate how many standard deviations the observed value is beyond the mean expected value. Values in (b) are based on the same outlier windows, but combining strings of outliers into peaks and expressing estimates of overlap as the total number of shared peaks in the comparison divided by the total number of unique peaks in the comparison. Results for F_{ST} are shown below the diagonal and d_{XY} above (not estimated for greenish warbelrs).

a)

	Flycatcher	Crows	Willows	Blackcaps	Greenish	Stonechats	Thrushes	Blue/gold
Flycatcher		0.92	0.40	1.23		3.56***	1.30	0.62
Crows	0.22		0.25	1.60*		1.60*	1.60*	0.55
Willows	-0.44	-0.39		2.35*		0.77	1.98*	2.20**
Blackcaps	-0.22	0.06	0.72			2.96***	4.24***	4.01***
Greenish	2.16*	0.39	-0.06	0.22				
Stonechats	1.88*	0.61	0.17	-0.11	1.22		4.76***	1.07
Thrushes	1.00	-0.17	1.33	0.39	0.78	1.50		2.58***
Blue/gold	-0.55	0.17	1.33	0.66	0.55	-0.66	1.66*	

b)

	Flycatcher	Crows	Willows	Blackcaps	Greenish	Stonechats	Thrushes	Blue/gold
Flycatcher		0.08*	0.15**	0.14*		0.2***	0.16***	0.18***
Crows	0.04		0.09***	0.10**		0.06*	0.06*	0.08**
Willows	0.04	0.04		0.22***		0.10***	0.20*	0.20*
Blackcaps	0.04	0.06	0.07			0.14***	0.22***	0.32**
Greenish	0.09*	0.05	0.09	0.07				
Stonechats	0.15***	0.08	0.04	0.05	0.11**		0.10***	0.12*
Thrushes	0.05	0.05	0.07	0.06	0.08	0.08		0.23***
Blue/gold	0.03	0.06	0.1	0.09	0.06	0.04	0.08*	

p-values corrected for multiple testing (* 0.05, ** 0.01, *** 0.001)

Table S4. Variables used to estimate the speciation continuum, including genetic distance based on cytb and autosomal sequences, hybrid zone width and the proportion of hybrids in each zone (the latter variable is missing for European blackcaps).

	Cytb	Autosomal	Width (km)	Proportion of hybrids
Flycatchers	0.036	0.0035	20 [a]	3[a]
Crows	0.029	0.00044	67[b]	12 [a]
Willows	0.003	0.0017	350 [a]	70[g]
Blackcaps	0.027	0.0021	340[c]	
Greenish	0.051	0.0032	0[d]	0 [d]
Stonechats	0.035	0.0039	0[e]	0[e]
Thrushes	0.048	0.0044	50[f]	20.5 [f]
Blue/ golden-winged	0.027	0.0029	600 [a]	0[h]

[a] Price 2008 [b] Haas And Brodin 2005 [c] Berthold et al. 1990 [d] Alcaide et al. 2014 [e] Helm 2009 [f] Ruegg 2009 [g] Lundberg et al. 2017 [h] Vallender et al. 2007

Supplementary Methods – Additional details on how d_s was estimated.

We performed gene prediction for each consensus genome with the MAKER pipeline which included four rounds of gene prediction as follows. The first round included gene prediction with EXONERATE using cDNA transcripts retrieved from Ensembl for zebra finch, chicken and flycatchers. This round also included repeat masking using the library of “aves” included in REPEATMASKER. For the second round, an HMM model was obtained from all gene predictions to use as input for the gene predictor SNAP. An additional round of repeat masking was run as described before. Third and fourth rounds of MAKER included two gene predictors: SNAP using HMM models from the previous round and the “chicken” HMM model available in AUGUSTUS. In every iteration we accepted only models with start and finish codons and genes > 50 amino acid (AA) length. Once we had annotated each consensus genome, we identified potential homologues for high quality transcripts (AED < 0.05) using a Blastn search against all transcripts from the flycatcher (flycatcher was searched against zebra finch). In this search we obtained the best hit of a transcript with at least 60% of identity and coverage of at least 50% of the flycatcher transcript. We then aligned codons from each pair of sequences using PRANK to calculate d_N/d_s with PAML v4.8 package. All d_N/d_s calculations were performed pairwise, comparing all the species with the flycatcher and this in turn, compared to zebra finch. We extracted only d_s values from this analysis and to avoid false positives and (potential mistakes in alignments) we filtered out results with d_s values bigger than 2 SD.

Table S5. Results from GLMs examining the relationship between repeatability and predictor variables related to genomic factors for each species pair. Repeatability is estimated as the number of pairs each window was considered an outlier in (outliers = windows in the top 5 percentile of each species pairs' distribution for F_{ST} and bottom 5 percentile of d_{XY}). d_{XY} was not estimated for greenish warblers. Parameter estimates, standard errors, test statistic (z value) and significance (p value) are shown for each predictor along with correlation coefficients (and confidence intervals) for each model. The coefficients were obtained by regressing observed repeatability to repeatability predicted by each model. Information on centromeres is only known for six macrochromosomes. Accordingly, we reran models with centromere included and size excluded and show parameter estimates for centromeres from these models in the last row for each pair and measure of genomic differentiation. A positive association with position indicates increased repeatability at the center of chromosomes.

A) F_{ST} repeatability

	Blackcap cor = 0.17 (0.14 - 0.21)				Crow cor = 0.17 (0.14 - 0.21)				Stonechat cor = 0.18 (0.15 - 0.21)				Willow cor = 0.18 (0.15 - 0.22)			
	estimate	std error	z value	p value	estimate	std error	z value	p value	estimate	std error	z value	p value	estimate	std error	z value	p value
(Intercept)	-0.85	0.04	-23.4	0.001	-0.89	0.04	-21.15	0.001	-0.86	0.04	-22.79	0.001	-0.96	0.04	-22.65	0.001
proportion of GC bases	-0.2	0.04	-5.39	**0.001**	-0.14	0.05	-2.98	**0.001**	-0.18	0.04	-4.79	**0.001**	-0.29	0.04	-6.82	**0.001**
dS	0	0.02	0.1	0.92	-0.23	0.06	-3.66	**0.001**	0.01	0.02	0.23	0.82	-0.26	0.1	-2.53	**0.01**
position	0.19	0.03	6.89	**0.001**	0.14	0.03	4.63	**0.001**	0.14	0.03	5.02	**0.001**	0.1	0.03	3.5	**0.001**
chromosome size	0	0.07	-0.07	0.94	-0.02	0.07	-0.25	0.8	0.06	0.07	0.83	0.41	0.13	0.07	1.94	**0.05**
gene count	0.13	0.03	4.55	**0.001**	0.12	0.03	4.02	**0.001**	0.11	0.03	3.5	**0.001**	0.15	0.03	4.84	**0.001**
LD	0.15	0.02	7.02	**0.001**	0.09	0.02	5.4	**0.001**	0.11	0.03	4.44	**0.001**	0.22	0.02	10.02	**0.001**
proximity to centromeres	0.32471	0.08741	3.715	**0.001**	0.30279	0.08947	3.384	**0.001**	0.300976	0.089094	3.378	**0.001**	0.36529	0.09179	3.98	**0.001**

	Swainsons cor = 0.29 (0.26 - 0.32)				Greenish cor = 0.22 (0.16 - 0.28)				Flycatcher cor = 0.17 (0.14 - 0.20)				Yellow cor = 0.17 (0.12 - 0.21)			
	estimate	std error	z value	p value	estimate	std error	z value	p value	estimate	std error	z value	p value	estimate	std error	z value	p value
(Intercept)	-1.04	0.04	-26.34	0.001	-1.06	0.09	-12.39	0.001	-0.86	0.03	-25.68	0.001	-0.9	0.05	-18.24	0.001
proportion of GC bases	-0.09	0.04	-2.69	**0.01**	-0.17	0.08	-2.17	0.03	-0.19	0.04	-5.11	**0.001**	-0.11	0.06	-1.95	0.05
dS	0	0.03	0.13	0.89	-0.03	0.11	-0.24	0.81	-0.05	0.04	-1.29	0.2	0.01	0.04	0.34	0.73
position	0.16	0.03	5.92	**0.001**	0.23	0.06	3.84	**0.001**	0.16	0.03	6.21	**0.001**	0.21	0.04	5.72	**0.001**
chromosome size	0.24	0.07	3.46	**0.001**	-0.21	0.15	-1.45	0.15	0.05	0.06	0.78	0.44	0.01	0.09	0.12	0.9
gene count	0.15	0.03	5.68	**0.001**	0.15	0.05	2.76	**0.01**	0.1	0.03	3.63	**0.001**	0.14	0.04	3.36	**0.001**
LD	0.3	0.02	15.56	**0.001**	0.13	0.04	3.47	**0.001**	0.07	0.01	7.17	**0.001**	0.08	0.03	2.67	**0.01**
proximity to centromeres	0.232794	0.090644	2.568	**0.001**	0.30089	0.08552	3.518	**0.001**	0.29497	0.08143	3.622	**0.001**	0.20763	0.09321	2.228	**0.026**

B) d_{XY} repeatability

	Blackcap cor = 0.27 (0.23 - 0.30)				Crow cor = 0.29 (0.25 - 0.32)				Stonechat cor = 0.24 (0.20 - 0.28)				Willow cor = 0.27 (0.23 - 0.30)			
	estimate	std error	z value	p value	estimate	std error	z value	p value	estimate	std error	z value	p value	estimate	std error	z value	p value
(Intercept)	-1.24	0.05	-23.62	0.001	-1.34	0.05	-24.6	0.001	-1.26	0.05	-23.01	0.001	-1.32	0.06	-23.42	0.001
proportion of GC bases	-0.09	0.04	-2.48	**0.01**	-0.01	0.04	-0.38	0.7	-0.16	0.04	-4.26	**0.001**	-0.22	0.04	-5.55	**0.001**
dS	-0.04	0.04	-1.07	0.28	-0.29	0.05	-5.79	**0.001**	0	0.03	0.14	0.89	-0.28	0.09	-3.08	**0.001**
position	-0.08	0.03	-2.82	**0.001**	-0.16	0.03	-5.4	**0.001**	-0.16	0.03	-5.05	**0.001**	-0.18	0.03	-6.02	**0.001**
chromosome size	0.62	0.08	7.84	**0.001**	0.67	0.08	8.76	**0.001**	0.77	0.08	9.81	**0.001**	0.77	0.08	9.9	**0.001**
gene count	0.11	0.03	3.37	**0.001**	0.11	0.03	3.56	**0.001**	0.11	0.03	3.52	**0.001**	0.1	0.03	3.36	**0.001**
LD	0.24	0.02	12.87	**0.001**	0.13	0.01	9.54	**0.001**	0.08	0.03	3.06	**0.001**	0.2	0.02	9.91	**0.001**
proximity to centromeres	0.63	0.11	5.66	**0.001**	0.78	0.1	7.54	**0.001**	0.95	0.11	8.9	**0.001**	0.63	0.12	5.36	**0.001**

Chapter 7

DISCUSSION AND PERSPECTIVES

Seasonal adaptations such as animal migration are part of the life cycle of many animals. It is expected that there are some mechanisms behind the prevalence of this behaviour. With different sources of evidence it is now accepted that migration has a genetic component. Such components must have been optimised through the scope of selection and can be sources of adaptation. In this thesis, I have presented studies that evaluate the potential mechanisms of genomic elements involved in migration, and how evolution has potentially shaped the variability of this behaviour.

Migratory tracks of blackcaps in the wild confirm old experimental findings, but challenge their interpretation.

A precise description and characterisation of the migratory phenotype is an essential prerequisite to istudy the genetic basis of migratory behaviour. Different techniques -all of which are indirect measures- were used in the past to infer the migratory routes of small songbirds: ringing recoveries, isotopes, and funnel experiments aided to assess the distance of migration and orientation patterns. Relevant findings from these approaches provided evidence for a genetic basis of some migratory traits [1], [2]. Crossbreeding experiments and phenotypic evaluation using funnel experiments showed that individuals crossbred from opposite orientation patterns have an intermediate orientation pattern. However, nothing was known about wether this phenomenon could be happening the wild or what are the fitness consequences for putative hybrids .

In a collaborative effort I was part durying my PhD research (Chapter 2), we demonstrated for the first time that birds from the area of the migratory divide do take intermediate routes, and most importantly, successfully return to their natal breeding grounds. Eurasian blackcaps have a migratory divide, a geographical region where two breeding populations with opposing orientations meet and potentialy mate and hybridise. Using light-level geolocators allowed us to reconstruct the migratory journeys of individuals withing the hybrid zone across the migratory divide. Our results confirm the previous findings of an intermediate phenotype that has been shown in experimental cross-breeding

settings. In essence, individuals breeding nearer to the edges of the divide, follow Southeast(SW) or Southwest(SE) orientation patterns while the individuals in the middle, i.e. those potentially hybridizing, follow an intermediate route. These results challenge the previous assumption about selection against hybrids. It was previously thought that individuals with an intermediate orientation would be selected against because the route make them go through difficult geographical barriers like the alps, the Mediterranean Sea and possibly wide stretches of the Sahara desert [1]. However, the recovery tracks we were able to retieve from individuals in the middle of the migratory divide suggest that these birds can overcome such barriers and return to their breeding grounds. We acknowledge that this could be a product of a biased sampling. As geolocators are archival tags, and thus we rely on birds to return to their natal breeding grounds in order to recapture them to allow for downloading the data. From the total of individuals that were tagged, we are recovering only those that came back to roughly the same areas. This could mean that we are obtaining only those that survived the journey and possibly some of those birds do not survive the journey followed the intermediate orientation pattern. Consequently, those individuals are not included in any analysis. However, it is importatn to note that the return rates of individuals is similar (20-25%) in populations inside and outside the migratory divide.

Another novelty in this chapter is the revealed origin of wintering blackcaps in the UK. Since the 1960s, observations of blackcaps overwintering across the UK has been documented[3], [4]. However, the origins of the wintering blackcaps remained a mistery until now. In this chapter, we tracked individuals from the UK from their wintering grounds in the UK to reveal their breeding destinations. We found that the breeding areas of the birds do not belong to a single population but instead, they come from all over central Europe suggesting that the UK wintering adaptation is in low frequency across most European populations.

With this study, we also found an essential element for the genetics of migration. High repeatability of the travelling routes of individuals tracked in consecutive years indicates the innate nature of the behaviour. The degree of repeatability in individuals of a species with a wide range of migratory phenotypes like the blackcaps, shows that the variability of routes taken by each individual are repeatable, but in a population-wide view, the species still has high variability.

Limitations of the candidate gene approach.

Candidate gene approaches rely on the known function of certin genes in model species to probe the feasibility of an effect in the non-model focal species. Some of the early genes suggested to be associated with migration are *ADCYAP1* and *CLOCK*. Variability in these genes correlates with migratory distance [5], [6] and breeding latitude [7]. Other candidates are genes associated with phenotypes like morning - evenings and sleep patterns like Period (PER1, PER2, and PER3)[8].

A complex behaviour like migration does not likely involve just a few elements. Instead, it probably include hundreds of genes acting in concert to execute the final behaviour. In chapter 3, we analyse the molecular evolution of 25 candidate genes of migration. We found that the associations of migratory traits with structural variation are not significant in a macroevolutionary scale. Lenght polymorphisms do not correlate with variability of migratory behaviour between and within species. Overall, the gene candidate approach gives unconclusive results for a single gene associated with migration. Because the genetic associations of candidate genes are analysed individually, their effect sizes are usually not estimated. This is important because it might mean ignoring the contribution of other genes or epistatic effects on the candidate gene. This disregard of effect sizes might be the reason why many genome-wide studies of migration, whether genomic or gene expression-based, do not find associations between migration and the candidate genes.

We argue for a broad investigation of the genetic elements of migration. Currently, there are several datasets of whole-genome resequencing of migratory species [9]–[13]. Additionally, more and more studies like the ones presented in this thesis, are proposing new candidates that can be analysed on the available genomic datasets as a first approximation. Moreover, some candidates have been around for some time, like Cryptochrome-4 [14], [15] might influence sensorial adaptations relevant for migration (i.e. magnetoreception) and deserve more attention.

Blackcap genomics reveal variability in migratory genotype with low population structure.

Due to the extensive variability on migratory traits in a single species, blackcaps are the ideal system to study the genetics of migration. It includes the complete

range of variation from long-distance migratory animals to entirely resident populations. It is also ideal for studying the patterns of evolution of migration in order to know what evolutionary events can lead to the variability of the phenotype. Inthis way, we might be able to explain the appearance and disappearance of this behaviour.

Previous studies on the evolution of this species have shown that migration variability started very recently, from 4 000 to 13000 years ago [16]. More importantly, no traces of genomic differentiation between the populations have been found. Only the comparison between migrants and residents seems to show a consistent, albeit low, genetic difference. The search for genetic difference in the migratory divide has been inconclusive. While Perez-Tris et. al [16] argue for a genetic differentiation inside the migratory divide, Mettler et.al [17] finds no genetic differences between individuals with different patterns of orientation.

In the collaboration I did in Chapter 4, we used whole-genome resequencing data of individuals ranging all the phenotypes to describe the patterns of evolution and potential population structure. Our results confirm that there is low genetic differentiation within migratory populations, supporting Mettler et al [17] result of low genetic difference between individuals with migratory orientations. The gene flow among migratory populations supports the idea of potential interbreeding between populations of opposite orientations, inside the migratory divide. This result could support that the intermediate phenotypic orientation found in Chapter 2 comes from the interbreeding of opposite orientations, but to get conclusive results in this regard more analysis need to be conducted.

The patterns of population structure found in our study show that the highest genomic differentiation is present between individuals of migratory and resident populations. This difference allows us to look for a genetic basis of migration. We observed that there are a few, small genomic regions going under selection. Most of them selected in resident populations. The SNPs with the strongest assotiation to the changes in phenotype are close to genes encoding a G protein-coupled receptor regulating Neuropeptide Y (NPY) and a glycosylation enzyme. A similar analysis with the phenotype of orientation in focus, find some regions selected in the population of North West orientation (i.e. UK wintering blackcaps). Importantly, those genomic regions suggest that this recently adapted orientation phenotype, derives from standing variation. It is accepted that

standing variation has potential for rapid adaptation, which would support the very recent expansion of the UK overwintering phenotype.

A gene regulatory characterisation of migratory behaviour, suggests a general shut down and tight control for energy expenditure during migration.

Most of the SNPs with the highest differentiation in the genome-wide study of blackcaps are located in non-coding regions of the genome. This suggests that certain cis-regulatory regions are essential for the regulation of gene expression of migration. Previous studies had analysed gene expression changes in the brain of migrant animals, but my core PhD research chapter (chapter five), is the first addressing the characterisation of the cis-regulation role in bird migration. Here, we document for the first time how chromatin accessibility changes in three brain regions are relevant for migration. One of the most surprising results is that our data show that the expression of migratory behaviour seems to impose a general chromatin repression in the cells of the focal brain regions. Notably, in Cluster N, a region involved in magnetic sensing, the repression of chromatin is stronger compared to birds tested outside the migratory season. Such repression suggests that the bird on migration reduces and tightly controls all metabolic processes that involve unnecessary energy expenditure.

Supporting the difference in chromatin accessibility, the potential Transcription Factors (TF) binding to the open chromatin regions (OCRs) in migrating birds are enriched for repressors like ZEB1 and ZEB2[18] and TF like RAR:RXR that can act as repressors depending on the availability of their ligand (retinoic acid)[19].

One of the most significantly accessible regions in migratory birds is a large region of approximately 8kb in an intron of the gene VAT1L. This region contains potential TF binding sites with SNPs that could disrupt the DNA-TF interaction. Furthermore, the SNPs overlapping with TF binding sites have characteristics of underdominance or heterozygotic disadvantage.

Is speciation following the same patterns in songbird species?

In previous chapters, we suggest that migratory divides could play a role in the genetic differentiation of phenotypes in nearby populations. These migratory divides are a form of hybrid zones, a natural experiment to study the process of speciation. Several studies have shown that populations of hybrid zones differentiate at the genomic level [20], [21].

Nevertheless, until our large scale cross-species comparative approach in hybrid zones (Chapter 6), we did not know how this process is happening at the macroevolutionary scale. In the study we analyse the degree of repeatability in genomic differentiation. Our results suggest that the degree of repeatability depends on the measure of genomic differentiation (namely, F_{ST} or d_{xy}) and the location of the population pair in the continuum of speciation. The measures of F_{sT} and d_{xy} are interpreted differently in the extent that F_{ST} reflects patterns of undergoing processes, while d_{xy} shows patterns of sequence divergence reflecting several past processes of speciation. Repeatability since divergence (d_{xy}) might be small because each population goes under specific selective pressures after a split. Similarly, the relative differentiation (F_{sT}) might have small repeatability if the extant populations have not been under enough time for differentiation. In both cases, more time since differentiation will make the differentiation landscape clearer. The consequences of a more apparent differentiation could mean that populations under recent divergence might not have accumulated enough differences while populations with long times of differentiation have more significant differences and therefore influences more correlated differentiation between species. The origin of high repeatability comes presumably from linked selection, selecting indirectly, regions across the genome creating correlations in differentiation landscapes.

Concluding and looking forward

Overall, three elements are relevant to emphasise from this thesis:

1) In a world going under climate change, there might be a constant pressure for rapid adaptation threatening migratory bird species. The recent expansion of individuals in the UK is an example of rapid adaptation where standing variation might be playing an important role. Two findings support the influence of standing variation: individuals wintering in the UK, come from breeding areas all over Europe (Chapter 2), and standing variation of genomic regions is positively selective in birds that migrate Northwest orientation (Chapter 4). Evidence suggests that high frequency of determinate haplotypes can give advantage to some individuals to adapt in different environments. Individuals wintering in the UK might be using this as a source of flexibility in orientation while maintaining

all the necessary elements for migration. An experimental setup designed to analysise allele-specific gene expression or chromatin accessibility would be one way to confirm the functional input of each allele and each haplotype. Such experimental setting could be an exciting roadmap for the general understanding of ecological and behavioural adaptation.

2) Tracking birds and associating genomic variants to migratory behaviour is one of the best ways to identify molecular elements for migration. Recently, [10], [11] has identified genetic associations to the migratory behaviour of Golden warblers and Swainson's thrush, respectively. However, these species have morphotypes that suggest and undergoing the process of speciation. Because migratory populations of blackcaps have overall low genomic differentiationwhile diverging in migratory traits like timing, distance, and orientation, the genotype associations with those migratory traits will have a high degree of confidence.

3) Nowadays there are increasing genomic sequencing and functional approaches to study migratory behaviours. Comparative approaches to identify potential new candidate genes (see Chapter 3) or genome-wide patterns of differentiation (Chapter 4), will generate and support hypotheses about how migration appears and disappears through the avian clade.

Finally, the integration of several sources of evidence will be a key to understand complex traits like migration. In recent years, plenty of studies have been published that integrate several data sources to improve our understanding about the mechanisms of autoimmune diseases, height, cancer and other complex phenotypes. We now have the opportunity to implement such approaches in the study of migration. I would argue to whenever possible, performing "common garden" experiments or using highly controlled setups to measure genetic (whole genome sequencing) and phenotypic features (gene expression, chromatin accessibility) in the same individuals. The integration of all information controlling for behavioural variability will accelerate the findings towards an understanding of bird migration. The studies presented in this thesis would be greatly complemented by such experiments, leading us to abetter understanding of this complex and interesting behaviour.

REFERENCES

[1] a Helbig, "Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palaearctic warblers (Aves: Sylviidae)," *J. Exp. Biol.*, vol. 199, no. Pt 1, pp. 49–55, 1996.

[2] P. Berthold and A. Helbig, "The genetics of bird migration: stimulus, timing, and direction," *Ibis (Lond. 1859).*, vol. 134, pp. 35–40, 1992.

[3] P. Berthold, A. Helbig, G. Mohr, and U. Querner, "Rapid microevolution of migratory behaviour in a wild bird species," *Nature*, vol. 2, no. 3, pp. 173–179, 1992.

[4] S. Bearhop *et al.*, "Assortative Mating as a Mechanism for Rapid Evolution of a Migratory Divide," *Science (80-.).*, vol. 310, no. 5747, pp. 502–504, 2005.

[5] J. C. Mueller, F. Pulido, and B. Kempenaers, "Identification of a gene associated with avian migratory behaviour," *Proc. R. Soc. B Biol. Sci.*, vol. 278, no. 1719, pp. 2848–2856, 2011.

[6] J. Ralston *et al.*, "Length polymorphisms at two candidate genes explain variation of migratory behaviors in blackpoll warblers (Setophaga striata)," *Ecol. Evol.*, vol. 9, no. 15, pp. 8840–8855, 2019.

[7] A. Bourret and D. Garant, "Candidate gene-environment interactions and their relationships with timing of breeding in a wild bird population," *Ecol. Evol.*, vol. 5, no. 17, pp. 3628–3641, 2015.

[8] C. Steinmeyer, J. C. Mueller, and B. Kempenaers, "Search for informative polymorphisms in candidate genes: Clock genes and circadian behaviour in blue tits," *Genetica*, vol. 136, no. 1, pp. 109–117, 2009.

[9] M. Alcaide, E. S. C. Scordato, T. D. Price, and D. E. Irwin, "Genomic divergence in a ring species complex.," *Nature*, vol. 511, no. 7507, pp. 83–5, 2014.

[10] K. E. Delmore, D. P. L. Toews, R. R. Germain, G. L. Owens, and D. E. Irwin, "The Genetics of Seasonal Migration and Plumage Color," *Current Biology*. 2016.

[11] D. P. L. Toews, S. A. Taylor, H. M. Streby, G. R. Kramer, and I. J. Lovette, "Selection on VPS13A linked to migration in a songbird," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 37, pp. 18272–18274, 2019.

[12] R. A. Bay, R. J. Harrigan, V. Le Underwood, H. L. Gibbs, T. B. Smith, and K. Ruegg, "Genomic signals of selection predict climate-driven population declines in a migratory bird," *Science (80-.).*, vol. 359, no. 6371, pp. 83 LP – 86, Jan. 2018.

[13] M. Lundberg *et al.*, "Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks," *Evol. Lett.*, pp. 155–168, 2017.

[14] P. J. Hore and H. Mouritsen, "The Radical-Pair Mechanism of Magnetoreception," *Annu. Rev. Biophys.*, vol. 45, no. 1, pp. 299–344, 2016.

[15] M. Liedvogel and H. Mouritsen, "Cryptochromes - A potential magnetoreceptor: What do we know and what do we want to know?," *J. R. Soc. Interface*, vol. 7, no. SUPPL. 2, 2010.

[16] J. Pérez-Tris, S. Bensch, R. Carbonell, A. J. Helbig, and J. L. Tellería, "Historical Diversification of Migration Patterns in a Passerine Bird," *Evolution (N. Y).*, vol. 58, no. 8, pp. 1819–1832, Mar. 2004.

[17] R. Mettler *et al.*, "Contrasting Patterns of Genetic Differentiation among Blackcaps (," vol. 8, no. 11, pp. 1–12, 2013.

[18] S. Leskela *et al.*, "Molecular basis of tumor heterogeneity in endometrial carcinosarcoma," *Cancers (Basel).*, vol. 11, no. 7, pp. 1–21, 2019.

[19] K. D. Shearer, T. H. Goodman, A. W. Ross, L. Reilly, P. J. Morgan, and P. J. McCaffery, "Photoperiodic regulation of retinoic acid signaling in the hypothalamus," *J. Neurochem.*, vol. 112, no. 1, pp. 246–257, 2010.

[20] N. J. Nadeau *et al.*, "Population genomics of parallel hybrid zones in the mimetic butterflies, H. melpomene and H. erato.," *Genome Res.*, vol. 24, no. 8, pp. 1316–1333, Aug. 2014.

[21] S. Renaut, G. L. Owens, and L. H. Rieseberg, "Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers.," *Mol. Ecol.*, vol. 23, no. 2, pp. 311–324, Feb. 2014.

Acknowledgements.

First and foremost, I would like to thank Dr Miriam Liedvogel for being so supportive and understanding. This thesis is the way it is, due to the freedom I had to explore topics, for her empathy in my hardest moments and her giving me impulse to thrive. Thank you for all of this. And thanks for being a good role model in and beyond science.

Thank you to all current and past members of the Behavioural Genomics Laboratory. Tania, Andrea, Karen, Corinna, Jun, Yennifer and Peter it is exciting to see you in the group, that cleverness, inquiring, and criticism is so refreshing. Special thanks to Gillian; I could not have asked for a better lab mate. We went through long, hard and arduous hours to keep our project going. In between, I had the most fantastic company to discuss the science, morals, analytics and bad jokes about Georgia O'keefe. Thanks to Susanne and Camilo, for teaching and helping us to take care of the birds we brought for the migratory restlessness experiments.

I appreciate the help of several groups where I took part during the development of this thesis. Thanks to the members of my thesis Advisory Committee: Prof. *Eva Stukenbruck* and Dr. *Julien Duthiel*, for keeping everything on track, for your advice and productive discussions. Most of the knowledge about evolution I acquired during my PhD was due to taking part in the lab seminars and journal clubs of the Evolgen Department. Thank you for the input I received on my presentations, it improved this thesis significantly. Special thanks to *Dr Guy Reeves*, for offering fair criticism and help when needed. The neurobiology parts of this thesis, and the first time I heard of ATAC-seq, was at the Jarvis lab at Duke and Rockefeller University. Thank you for being such a welcoming laboratory. Every time I visited, you made me feel like one of the team. Special thanks to *Professor Erich, Lindsey and Greg*. Thank you to all the members of the *Biological clocks group*. It has been a pleasant experience to learn about midges and the moon, sharing ideas, analysis, limitations and novelties of non-model organisms. Special thanks to *Jule Neumann* for translating my Zusammenfassung. I am grateful to the International Max Planck Research School for funding my PhD and to the IMPRS coordinator Kerstin Mehnert for the help through all the process.

This thesis was significantly improved thanks to the input I got from my colleagues. Thank you Gillian Durieux, Johana Fajardo, Maria Bargues, Filipa Moutinho and members of the Behavioural Genomics Laboratory

The Ploen blues is not an easy song to play. Luckily, friends make it sound smooth. *Johana*, I wholeheartedly appreciate these almost ten years of friendship. You are a pillar in the hardest moments I have been through. Thank you for being there and being so phenomenal in science. I totally admire your discipline and creativity. *Loukas*, thank you for being so supportive. I am grateful for the high and lows of our friendship. I will also remember the long nights and breaks for skateboarding in the institute. I hope we have more moments to share. *Maria*, not even sparing vowels in Catala' I could be thankful for all, but, thanks for all the care and for bearing all my stupidity. *Gill*, thank you for the kitty therapy and the many conversations beyond science. *Filipa*, thank you for welcoming me into your house, your home and your family, I will not forget my 31st birthday. *Alejandro*, thank you for all the talks about the Colombian abhorrent political class and for taking me out for coffees. Thanks to *Devika* and *Anuradha* for sharing coffees and delicious food with me. *Jule*, thank you for the coffees and for the hard immersion in German with "El principito". The wrinkly spreaders (Michael, Maria, Loukas and Andrew) give me the space to get back into what used to be my half-life, thank you so much. Thanks to the basket and pool crew. It always gave me the endorphins and oxytocin I needed. Sorry for the injuries.

Thank you to my friends from other latitudes. My science friends, Johanna C., Camila, Alejandra, Gerardo, Julian, Veronica y Fabio. Ustedes son la adicion perfecta para querer seguir haciendo ciencia, porque admiro lo que hacen todos y cada uno y ademas son de la gente mas divertida que conozco. No hay momento en que no falten las risas cada vez que nos vemos.

My pueblo pop friends. Andres, aunque no hablamos seguido sabemos que contamos el uno con el otro en cualquier situacion. Lo amo mucho. Molly, siempre has sido de mi admiracion, pero ademas ha sido bonito vernos madurar y crecer juntos. Cristian, gracias por todo el apoyo y la espiritualidad que me

comparte, eso me alivia el corazon. Paola, gracias por quererme ver todas las veces que vuelvo a Colombia y por apoyarme en los momentos mas dificiles.

In spite the distance, my family has been always there to make me feel their company. Thank you to my mother Luz Marina and my father Oscar. Ustedes han estado siempre ahi conmigo y para mi incondicionalmente dando todo el amor y paciencia que he necesitado. Thanks to my brother Oscar Javier and my sister Angelica. Gracias por ser inspiracion y ejemplo. Crecer con y bajo su guardia es gran parte de lo que me ha traido hasta aqui. A todos gracias tambien por apoyarme en mi carrera. Es bien sabido que no todos tienen el privilegio de tener respaldo y empatia con las carreras cientificas, pero ustedes me lo han dado todo sin preguntar.

Maria Paula. Your premature departure from this world made my ground shake. Still does. It was a massive threat to my own stability and to finish this PhD. You were the person I loved the most. Those 6 years together make a significant part of what I am now as a person. Your tender love and brilliant mind were always a lighthouse in my darkest times, even sailing from this part of the world. I was happy to know you were somewhere in this world, spreading happiness and light. Now, your legacy lives in all those who shared the most incredible times with you. Thank you for everything. Everything.

CONTRIBUTIONS

Chapter 2

"Versatile migratory strategies and evolutionary insights revealed by tracks of wild Eurasian blackcaps" Submitted to PloS Biology.

Personal Contributions: I was a core member of the fieldcrew across the the migratory divide from beginning to end. i.e. four consecutive breeding seasons. I provided comments on interpretations of the results and comments on the manuscript draft.

Author contributions

Conceptualization: KD, BMVD, BCS, ML; Methodology: KD, BMVD; Formal Analysis: KD, BMVD; Fieldwork: KD, BMVD, TC, TGG, RRG, TH, DH, HJ, JSLR, IM, BSM, RJP, MR, GCMR, HPJ, WV, ML; Writing –Original Draft: BMVD with input from KD and ML; Writing –Review & Editing: KD, BMVD, GJC, TC, TGG, RRG, JSLR, IM, BSM, MR, BCS, HPJ, WV, ML; Visualization: BMVD; Supervision: GJC, MR, HPJ, BCS, ML; Project Administration: WV, IM, HPJ, GJC, MR, ML; Funding Acquisition: BMVD, MR, HPJ, ML.

Chapter 3

"Candidate genes for migration do not distinguish migratory and non-migratory birds" Published in J Comp Physiol A (2017).

Personal Contributions: The study was designed by me and M. Liedvogel. I conducted all analysis and visualizations. I wrote the first draft and all authors contributed with revisions and result interpretations.

Author contributions:

Conceptualization: JSLR, ML; Methodology: JSLR, KD, ML; Formal Analysis: JSLR; Writing –Original Draft: JSLR with input from KD and ML; Writing –Review & Editing: JSLR, KD, ML; Visualization: JSLR; Supervision: ML; Project Administration: JSLR, ML; Funding Acquisition: ML.

Chapter 4

"The evolutionary history and genomics of European blackcap migration." Accepted at Elife

Personal contributions: I performed the gene prediction and functional annotation of the Eurasian blackcap (Sylvia atricapilla) genome. I also contributed with

analysis of SFS for the standing variation selection. I contributed review and editing of the final draft.
Author contributions:
Conceptualization by KD and ML; Methodology by KD, ML, JP, JCI; Formal analysis by KD, GD, JI and JL; Resources by ML, JP, JC and GS; Writing – original draft preparation by KD; Writing – review & editing all authors; Visualization KD; Funding acquisition KD and ML, JP, JC and GS.

Chapter 5

"Controlling bird migration behaviour through cis-regulatory elements"
Working manuscript
Personal Contributions: I designed the study together with GD and ML. GD, KD and I performed the sampling and carried out all sample preparations. I performed all analysis, all visualization and wrote the first draft of the manuscript.
Author contributions:
Conceptualization: JSLR, GD, ML; Methodology: JSLR, GD, GG, LC, EJ, ML; Formal Analysis: JSLR; Writing –Original Draft: JSLR with input from GD and ML; Writing – Review & Editing: JSLR, GD, ML; Visualization: JSLR; Supervision: EJ, ML; Project Administration: JSLR, GD, ML; Funding Acquisition: ML.

Chapter 6

Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. Published in Evolution Letters (2018)
Personal contributions: I performed the whole genome alignments, annotated the consensus alignments and analysed pairwise dN/dS patterns to obtain the mutation rate (dS) . I analysed the repeatability patterns of dS. I reviewed and edited draft versions of the manuscript.
Author contributions:
K.D. and M.Li. conceived and designed the study, all authors contributed data. K.D. conducted all analyses excluding the estimation of d XY for pooled datasets and repeatability using peaked overlap values (BV). J.L. helped with the construction of consensus genomes and estimates of ds and M.Lu. with the

estimation of d XY for individual datasets. K.D. wrote the manuscript with comments from all authors.

Complete list of co-authors (in alphabetic order after JSLR.

Full name	Abreviation
Juan Sebastian Lugo Ramos	JL
Ben C. Sheldon	BCS
Benjamin van Dooren	BvD
Britta S. Meyer	BM
Darren E Irwin	DI
Dieter Hiemer	DH
Erich Jarvis	EJ
Gernot Segelbrecht	GS
Gillian Durieux	GD
Graham C. M. Roberts	GCMR
Greg J. Conway	GJC
Gregory Gedman	GG
Hannah Justen	HJ
Henk P. van der Jeugd	HPJ
Ivan Maggini	IM
Javier Perez-Tris	JP-T
Juan Carlos Illera	JcI
Jun Ishigohoka	JuI
Kira Delmore	KD
Lindsey Cantin	LC
Magdalena Remisiewicz	MR
Max Lundberg	MaL
Miriam Liedvogel	ML
Robbie J. Phillips	RJP
Ryan R. Germain	RRG
Staffan Bensch	SB
Tania Garrido-Garduño	TGG
Teja Curk	TC
Timo Hasselmann	TH
Wolfgang Vogl	WV

AFFIDAVIT

I hereby declare that:

i. Apart from my supervisor's guidance, the content and design of this thesis is the product of my own work. The co-authors' contributions are listed in the dedicated section;

ii. This thesis has not been already submitted either partially or wholly as part of a doctoral degree to another examination body, and no other materials are published or submitted for publication than indicated in the thesis;

iii. The preparation of the thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation;

iv. Prior to this thesis, I have not attempted and failed to obtain a doctoral degree.

Plön, March 2020:

_____

Juan Sebastian Lugo Ramos