MUSEO NACIONAL DE CIENCIAS NATURALES – CSIC
UNIVERSIDAD AUTÓNOMA DE MADRID

The genetic basis of speciation mechanisms in island birds

TESIS DOCTORAL

María José Recuerda Carrasco

Madrid, 2022

Museo Nacional de Ciencias Naturales
Consejo Superior de Investigaciones Científicas
Departamento de Biodiversidad y Biología Evolutiva

Universidad Autónoma de Madrid
Departamento de Biología

The genetic basis of speciation mechanisms in island birds

Memoria presentada por la licenciada María José Recuerda Carrasco para optar al grado de Doctora en Biología por la Universidad Autónoma de Madrid.

DIRECTOR:
Dr. Borja Milá Valcárcel
Museo Nacional de Ciencias Naturales
Dpto. de Biodiversidad y Biología Evolutiva

CODIRECTOR:
Dr. Rafael Zardoya San Sebastián
Museo Nacional de Ciencias Naturales
Dpto. de Biodiversidad y Biología Evolutiva

TUTOR:
Dr. José Luis Bella Sombría
Universidad Autónoma de Madrid
Dpto. de Biología

TESIS DOCTORAL

2022

María José Recuerda Carrasco ha sido financiada a través de una Contrato Predoctoral para la Formación de Profesorado Universitario (FPU16/05724) concedida por el Ministerio de Ciencia e Innovación.

El desarrollo de la tesis ha sido posible gracias a la financiación del Ministerio de Ciencia a través de los proyectos del Plan Estatal de I+D+i: CGL2015-66381P y PGC2018-098897-B-I00.

Cover:
Common chaffinch (*Fringilla coelebs*) from La Palma; illustration by María J. Recuerda Carrasco

"There is grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms most beautiful and wonderful have been, and are being, evolved."

"But natural selection, as we shall hereafter see, is a power incessantly ready for action, and is immeasurably superior to man's feeble efforts, as the works of nature are to those of art."

Charles Darwin, The Origin of Species (1859).

CONTENTS

AGRADECIMIENTOS

En primer lugar quiero agradecer a mis directores de tesis, Borja Milá y Rafael Zardoya, por brindarme la oportunidad de pedir la beca y realizar este proyecto con ellos y por apoyarme y guiarme en el desarrollo de este trabajo. Han sido un referente a lo largo de todo el proceso y lo han hecho todo más fácil.

A mi tutor Pepe Bella que con su buen humor ha hecho que la burocracia parezca hasta fácil. Muchas gracias por la eficacia, y sobre todo por el apoyo y la confianza durante todos estos años.

Gracias a Guillermo Friis que fue el que me introdujo con paciencia en el mundo de las pantallas negras llenas de letras. Y a todos los que también han formado parte de 'Milab' que han sido un gran apoyo: Javi, Bárbara, Quique y Mercé.

También, a todos los compañeros del Museo Paloma, Martí, Miriam, Silvia, Paula, Samu, Lourdes, Violeta, Yi, Goyo, Lucía, Juanes, Iker, Iván, Karen, Carlos, Joserra, Andrea y Mariana por haberme ayudado y haber formado parte de esta etapa.

A todos los que han colaborado en el trabajo de campo: Oscar Frías, José Luis González, Alex, Pedro, Félix Medina, Pepe Navarrete y Manuel Nogales.

A todos los colaboradores porque sin su ayuda y trabajo no habría sido posible este trabajo: Guillermo Blanco, Joel Vizueta, Julio Rozas, Juan Carlos Illera, Keith Hobson, Andrea Meseguer y Benoit Nabholz. Además, a Benoit y sus compañeros del ISEM, agradecerles también el tiempo de estancia en Montpellier en el que me acogieron y aprendí mucho.

Y en el terreno personal, gracias a mi familia, a mis padres que siempre han estado ahí y me han dado la confianza, el apoyo y el amor para llevar a buen puerto todos mis objetivos, en concreto este, que es el más exigente hasta la fecha. Gracias a mi hermana, simplemente por ser como es, siempre alegre. Y también a Rober, Adri y Gon que hacen todo más fácil y llevadero. Dicen que la familia no se elige, pero no podría haber elegido una mejor.

A los que sí "elegí", mis amigos, muchas gracias por todo, en especial a ellas, Regi, Blanca, Clau, Andrea, Celia, , Rebe, Ana, María y Anabel, gracias por estar ahí en mi montaña rusa constante, soy muy afortunada de teneros. A mis "pinzones" que me

alegran cada semana: Hugo, Iván, Ángela, Nachos, Claver y Quique. A Barde y Alvarito por los datos. Y por supuesto a “mi” Borja, que me ha acompañado desde el principio y ha sufrido y disfrutado conmigo cuando tocaba y me ha hecho reír cada día. Además de ayudarme con Photoshop…
Muchas gracias.

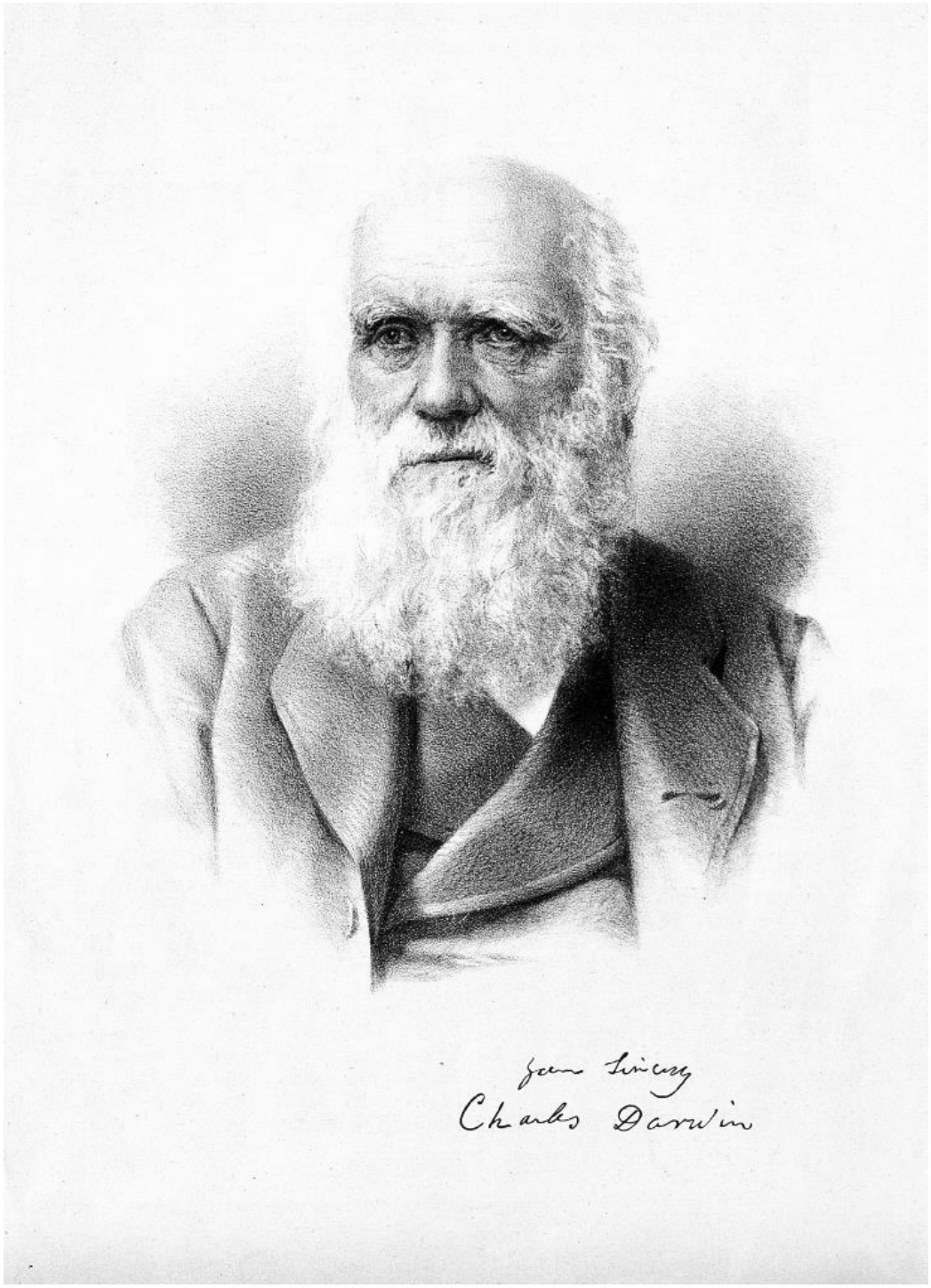
ABSTRACT

Understanding the process of speciation, from incipient population divergence to the formation of reproductively isolated lineages, has been one the main goals of evolutionary biology. Recent advances in high-throughput DNA sequencing technologies have allowed studying the different stages of the speciation process in non-model organisms, providing affordable access to genome-wide data and information on neutral and adaptive population divergence and the genetic basis of relevant fitness traits. In order to contribute to the understanding of the speciation process, the aims of this thesis are: (1) to detect phenotypic divergence and the regions under selection that contribute to avian adaptation to insular environments by comparing the genomic landscapes of four different passerines that have colonized oceanic islands, including the red billed chough (*Pyrrhocorax pyrrhocorax*) and the common chaffinch (*Fringilla coelebs*) of La Palma in the Atlantic Ocean, and the dark-eyed junco (*Junco hyemalis/insularis*) and the house finch (*Haemorhous mexicanus*) of Guadalupe Island in the Pacific Ocean; (2) to study the mechanisms driving the diversification of the common chaffinch as it colonized different archipelagos in the Macaronesian region, and (3) to explore the genetic basis of local adaptation of the common chaffinch in La Palma, and its role in driving evolutionary divergence. The island-mainland four-species comparison shows that even though species have evolved parallel phenotypic changes upon island colonization that are consistent with the island rule, the genomic processes underlying these changes are lineage-specific. In Macaronesia, the common chaffinch has colonized the Atlantic archipelagos sequentially, starting from the continent to Azores, then Madeira and finally the Canary Islands, diverging in phenotype and genotype, and generating a species-level radiation. Within the island of La Palma, the common chaffinch have extremely reduced dispersal, and populations from two contrasting habitats show differences in phenotypic fitness traits and genomic structure associated with habitat variables, suggesting the role of local adaptation in the presence of gene flow and allowing the study of the divergence process at a very small spatial scale.

RESUMEN

Entender el proceso de especiación, desde la divergencia incipiente de poblaciones hasta la formación de linajes aislados reproductivamente, ha sido uno de los principales objetivos de la biología evolutiva. Los recientes avances en las técnicas de secuenciación masiva de ADN han permitido estudiar las diferentes etapas del proceso de especiación en organismos no modelo, permitiendo un acceso asequible a datos genómicos e información en la divergencia neutral y adaptativa, así como de la base genética de los rasgos adaptativos. Con el fin de contribuir a la comprensión del proceso de especiación, los objetivos de esta tesis son: (1) detectar la divergencia fenotípica y las regiones bajo selección que contribuyen a la adaptación de las aves a los ambientes insulares mediante la comparación de los paisajes genómicos de cuatro especies de pareriformes que han colonizado islas oceánicas, incluyendo la chova piquirroja (*Pyrrhocorax pyrrhocorax*) y el pinzón común (*Fringilla coelebs*) de la Palma en el Océano Atlántico, y el junco ojioscuro (*Junco hyemalis/insularis*) y el camachuelo mexicano (*Haemorhous mexicanus*) de Guadalupe en el Océano Pacífico; (2) estudiar los mecanismos que impulsan la diversificación del pinzón común a medida que colonizaba los diferentes archipiélagos de la región Macaronésica, y (3) explorar las bases genéticas de la adaptación local del pinzón común en La Palma, y su papel en promover la divergencia evolutiva. La comparación isla-continente de las cuatro especies muestra que aunque las especies han evolucionado cambios fenotípicos paralelos tras la colonización de islas que son consistentes con la ley insular, los procesos genómicos que subyacen a estos cambios son específicos para cada linaje. En Macaronesia, el pinzón común ha colonizado los archipiélagos Atlánticos de forma secuencial, empezando desde el continente hasta Azores, después Madeira y finalmente las Islas Canarias, divergiendo en fenotipo y genotipo, y dando lugar a una radiación a nivel de especie. Dentro de la isla de La Palma, el pinzón común muestra una dispersión extremadamente reducida, y las poblaciones de dos hábitats muy diferenciados muestran diferencias fenotípicas en rasgos adaptativos y en la estructura genómica asociada con variables de hábitat, sugiriendo el papel de la adaptación local en la presencia de flujo génico y permitiendo el estudio del proceso de divergencia a escalas espaciales muy pequeñas.

GENERAL INTRODUCTION



Charles Robert Darwin. Lithograph.

Speciation, "that mystery of mysteries"

Ever since Darwin, the evolutionary process of species formation, or what he referred to as "that mystery of mysteries", quoting the philosopher John Herschel, has been a topic of major interest for evolutionary biologists. In his seminal work 'On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life' (Darwin, 1859) he threw some light on that "mystery" by proposing natural selection as the mechanism that promotes the formation of new species by gradual change. Since then, many authors have focused on clarifying the evolutionary process of species formation and different models of speciation have been proposed, ranging from models based on geographic patterns to those focused on the mechanisms driving the evolution of reproductive isolation (Wu, 2001;Coyne & Orr, 2004; Butlin et al., 2008; Price, 2008; Feder et al., 2012).

During the 20th century, Ernst Mayr promoted the major role of geographic isolation in speciation (i.e., allopatry, Mayr, 1942,1963) and many studies have provided evidence confirming this model (e.g., Glor et al., 2004; Cunha et al., 2005; Schwarzer et al., 2017; Dong et al., 2020). Alternative models have proposed that speciation can occur without geographical barriers and in the presence of gene flow (i.e., sympatry and parapatry). These models are theoretically challenging because divergence is counteracted by gene flow, which tends to homogenise alleles among populations, precluding divergence and speciation. However, a balance between the opposing forces of directional selection and gene flow makes sympatric speciation theoretically possible (Endler, 1973; Lande, 1982; Rice & Hostert, 1993), and even though it is less common than allopatric speciation (Barraclough & Vogler, 2000) some evidence is found in nature (Bolnick & Fitzpatrick, 2007; Feder et al., 2012; Tigano & Friesen 2016) and an increasing number of examples of divergence with gene flow provide support for this model (Milá et al., 2009; Martin et al., 2013; Supple et al., 2015; Han et al., 2017; Samuk et al., 2017; Poelstra et al., 2018; McLaughlin et al., 2020). The importance of this balance between opposing evolutionary forces led to a change in the emphasis on the degree of geographical isolation and resulted in the appearance of new models that focused on the interaction between selection and gene flow (Rundle & Nosil, 2005; Nosil, 2012).

Research on speciation has been inevitably influenced by the species concept debate. The "Biological Species Concept" (BSC) developed by Ernst Mayr (1942), defines species as "'groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups", and has been the most widely accepted species concept during the 20th century. However, this definition presents intrinsic limitations because it is not applicable to asexual organisms and assessing the degree of reproductive isolation is impossible for allopatric taxa or populations (Cronquist, 1978; Stace, 1991); and the debate about whether or not we should abandon the BSC is still currently under discussion (Butlin & Stankowski, 2020; X. Wang et al., 2020). Thus, several species concepts have been proposed trying to overcome different issues. For instance, the "Evolutionary Species Concept" (Simpson, 1951; Wiley, 1978) considers species as evolutionary independent lineages and can be applied to organisms that reproduce sexually or asexually. The "Phylogenetic Species Concept" (PSC, Eldredge & Cracraft, 1980; Nelson & Platnick, 1981; Donoghue, 1985; Mishler & Brandon, 1987; De Queiroz & Donoghue, 1988; Nixon & Wheeler, 1990; Wheeler, 1999) overcomes the allopatry issue by defining species by the presence of diagnostic traits and their phylogenetic relationship with other species instead of trying to determine the reproductive isolation among them. This concept has gained relevance with the development of sequencing technologies and the emergence of molecular systematics.

Depending of which concept is assumed, the number and boundaries of different species would vary (K. De Queiroz, 2005). In the last decades, some authors as Hey (2001) and Brookfield (2002) supported by Coyne and Orr (2004) have come to the conclusion that the debate of the species concept is out of the boundaries of the scientific method, and propose to apply different species concepts according to the question and the organism of interest (Hey, 2001; Brookfield, 2002; Coyne & Orr, 2004). However, other authors, like K. De Queiroz, have focused on the shared elements of most species concepts to attempt a unified general lineage concept (GLC) which defines species as "separately evolving metapopulation lineages" (K. De Queiroz, 1998, 2005, 2007). In order to eliminate the incompatibilities among alternative species concepts, the main point of this unified species concept is the reinterpretation of the specific properties of each concept (e.g., reproductive isolation

for the BSC, or reciprocal monophyly for the PSC) as neither necessary nor sufficient for the definition of species (K. De Queiroz, 2005). In the context of this thesis we use the BSC species concept, which is compatible with the definition of the speciation continuum (Stankowski & Ravinet, 2021) and due to the reinterpretation of reproductive isolation as a continuum is also compatible with the GLC.

The speciation continuum and genomic islands of divergence

The speciation process can happen relatively fast depending on the generation time, but the most common cases occur gradually over hundreds to thousands of generations (Coyne & Orr, 2004). As mentioned before, the geographical approach to speciation modes has shifted towards the inclusion of gene flow, becoming a continuum ranging from panmixia to complete isolation (Butlin et al., 2008). Therefore, the 'speciation continuum' includes all the stages at which genetic differences between two diverging lineages gradually accumulate until they reach complete reproductive isolation (Hendry et al., 2000; Schluter, 2000; Rundle & Nosil, 2005; Butlin et al., 2008; Mallet, 2008; Nosil, Harmon & Seehausen, 2009). In a recent review Stankowski & Ravinet (2021), following Seehausen et al., (2014), defined the speciation continuum as "a continuum of reproductive isolation". To facilitate the study of the process, Hendry (2009) describes some examples of the states along the speciation continuum and point out that groups can get stuck in one state, move back and forth between states, or skip some stages. The genomic differentiation at the initial stages of divergence is limited to small regions of the genome, while at the end of the speciation process, when gene flow is interrupted, differentiation is spread across the whole genome (Wolf & Ellegren, 2016). At intermediate stages of the divergence process, genomes are semipermeable because only some regions are affected by gene flow (Roux et al., 2016). It is in these intermediate stages that species delimitation is controversial, a period also known as the 'grey zone of speciation' (Roux et al., 2016). However, genetic divergence can be affected by several factors besides reproductive isolation, such as effective population size, bottlenecks and geographic isolation (Stankowski & Ravinet, 2021). An increasing number of studies along the speciation continuum have shown different patterns of heterogeneity across the genome at different stages of the divergence process (Ravinet et al., 2017).

The divergent regions of the genome that may be involved in reproductive isolation have been called 'genomic islands of speciation' (Turner et al., 2005), 'genomic islands of differentiation' (Harr, 2006) and 'genomic islands of divergence' (Nosil, Funk & Ortiz-Barrientos, 2009). Even though several processes could lead to regions of high genomic differentiation (Cruickshank & Hahn, 2014), adaptation and divergent selection along with genetic linkage and reduced recombination are usually involved (Wu, 2001; Turner et al., 2005; Nosil, Funk & Ortiz-Barrientos, 2009; Feder et al., 2012). In order to detect these highly differentiated regions, Lewontin & Krakauer (1973) proposed a genome-scan approach based on $F_{ST,}$ which is a relative measure of genetic differentiation. Since then, several methods have been developed, some based on the original idea (e.g., Beaumont & Nichols, 1996; Foll & Gaggiotti, 2008), and others based on other approaches that consider the physical linkage among loci and detect selective sweeps (e.g., Sabeti et al., 2002; Sabeti et al., 2007). However, when using genome scans to detect outlier regions, several confounding factors must be taken into account. These include demography and genomic features that contribute to the generation of the peaks of relative differentiation (Fig. 1)(Hoban et al., 2016; Ravinet et al., 2017), so that combining multiple summary statistics is necessary to distinguish among different processes.

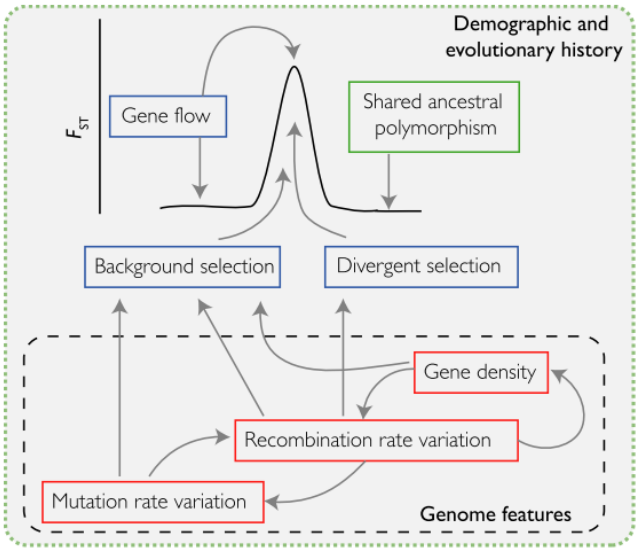


Figure 1. Features shaping genomic landscapes. Both genomic features (red boxes) and external processes (blue boxes) contribute to the formation of peaks of relative differentiation. The effects of all these factors are also influenced and determined by the demographic and evolutionary history. From Ravinet et al., (2017).

In early studies, F_{ST} peaks in genome scans were interpreted as signatures of strong selection and the valleys as regions homogenized by gene flow (Nosil, Funk & Ortiz-Barrientos, 2009). The peaks were supposed to include the sites that were highly differentiated due to selection as well as the neutral loci that were hitchhiked due to linkage (Charlesworth et al., 1997; Feder & Nosil, 2010). However, these peaks of relative differentiation can be caused by different processes such as reductions in within-population diversity (π) generated by background selection, while absolute divergence (d_{xy}) remains unaffected (Cruickshank & Hahn, 2014; Burri et al., 2015). Therefore, understanding the relationship between F_{ST}, d_{xy} and π is essential to differentiate the mechanisms generating the islands of differentiation, and in order to explain their formation, four different models have been proposed (Cruickshank & Hahn, 2014; Irwin et al., 2016; Irwin et al., 2018).

These models differ in the degree of isolation of populations, and include two divergence-with-gene-flow models and two allopatric models (Fig. 2). In order to discriminate between the speciation-with-gene-flow model (Fig. 2A) and the selection in allopatry model (Fig. 2B), Nachman & Payseur (2012) and Cruickshank & Hahn (2014) suggested to evaluate the pattern of absolute divergence. In the presence of gene flow, d_{xy} is expected to be high in regions of high relative differentiation when these regions are promoting reproductive isolation between populations, whereas in the allopatric model, d_{xy} is expected to remain similar in high and low F_{ST} regions, and F_{ST} peaks are produced because selection reduces nucleotide diversity in each population, which does not affect absolute divergence. The recurrent selection model (Fig. 2C) was developed in order to explain the pattern of low d_{xy} associated with F_{ST} peaks (Nachman & Payseur, 2012; Cruickshank & Hahn, 2014). Those regions were selected in the common ancestor, reducing genetic diversity previous to the split into the two daughter populations which afterwards, in allopatry, suffered a reduction in genetic diversity in the same regions due to selection, resulting in a F_{ST} peak and low d_{xy} caused by selection in the common ancestor. The same pattern of low d_{xy} and high F_{ST} could be observed in the presence of gene flow, so the fourth model, "sweep-before-differentiation" (Fig. 2D, Irwin et al., 2016) accounts for this scenario between hybridizing populations. The spread of advantageous alleles across the hybrid zone is faster than that of neutral regions, which results in the reduction of d_{xy} in those

regions relative to the rest of the genome. Further selection in the same regions in local populations will reduce genetic diversity and cause an increase in F_{ST}.

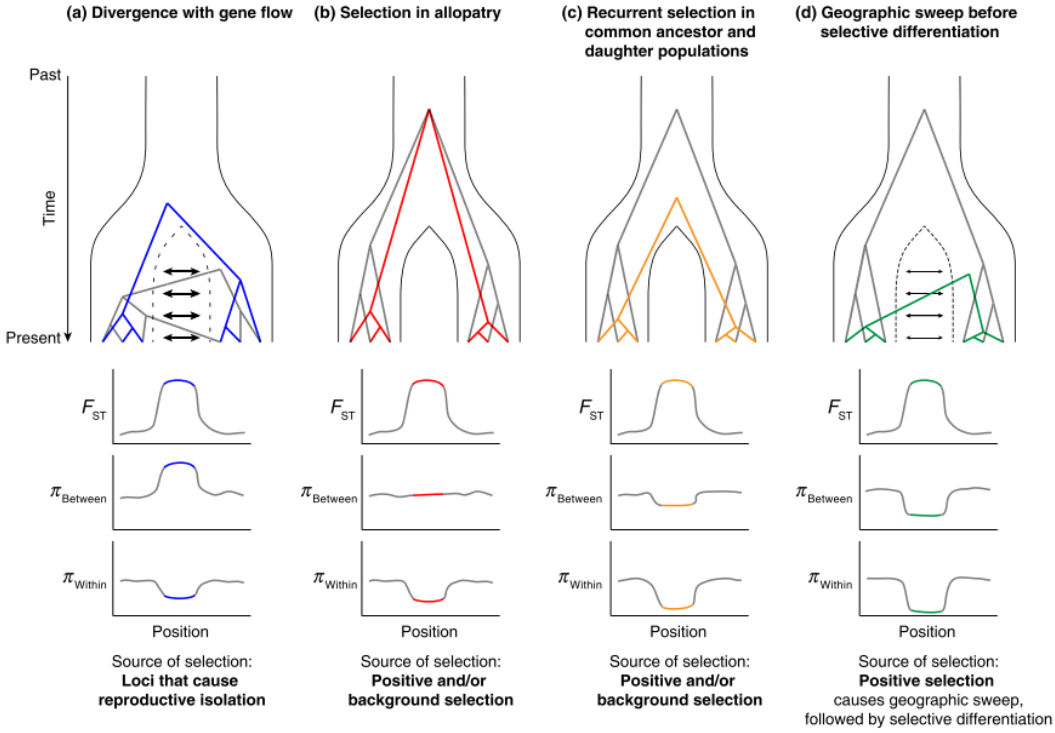


Figure 2. Schemes of the four models for the formation of genomic islands of differentiation showing the relationships among relative differentiation ($\boldsymbol{F}_{ST}$), absolute divergence ($\pi_{Between}$ also known as d_{xy}) and genetic diversity (π_{Within} or π) and also including the sources of selection for each model. There are two scenarios with gene flow including: (a) divergence with gene flow and (d) geographic sweep before selective differentiation. There are also two allopatric models assuming no gene flow, corresponding to (b) selection in allopatry and (c) recurrent selection in common ancestor and daughter populations. From (Irwin et al., 2018).

Islands as natural laboratories

Oceanic islands have proven to be excellent scenarios for the study of the speciation process and the underlying mechanisms. Several island features such as their simplified ecosystems and restricted area and isolation, provide ideal research settings for understanding the role of natural selection and drift in evolution and speciation (Coyne & Orr, 2004; Losos & Ricklefs, 2009). Moreover, island populations are more likely to be affected by both evolutionary forces due to new selective pressures, reduced dispersal, and the founder effect caused by a small colonizing group (e.g., Prentice et al., 2017; Funk et al., 2016; Armstrong et al., 2018). Therefore,

islands are “natural laboratories” of evolution playing an important role in the progress of evolutionary and ecological theories (Losos & Ricklefs, 2009; Warren et al., 2015).

Colonization events are usually associated with genetic drift (Wright, 1931) due to founder events, which cause random changes in allele frequencies because the original colonizers carry only a portion of the genetic diversity found in the genetic pool of the source population. The stochasticity associated with a new founder population can also result in the loss or fixation of rare variants and/or the reduction in frequency of alleles that were common in the ancestral population (Excoffier & Ray, 2008). Therefore, the founding population differentiates genetically from the source almost instantly, and speciation can take place relatively quickly upon colonization (A. De Queiroz, 2005; Cowie & Holland, 2006; Gillespie et al., 2012). Moreover, the loss of genetic diversity associated with founder events has major implications. For instance, the reduction in effective population size (N_e) negatively affects the efficacy of selection (Leroy et al., 2021) and the adaptive potential of a species may be hampered due to the loss of adaptive alleles (Bijlsma & Loeschcke, 2012; Lande & Barrowclough, 1987) which can also lead to inbreeding depression (Charlesworth & Charlesworth, 1987; Newman & Pilson, 1997) increasing the probability of population extinction (Newman & Pilson, 1997; Nieminen et al., 2001; Frankham, 2005). Thus, the extent of diversity loss depends on the number of founders, the growth rate of the founded populations, and the number of colonization steps (Clegg, Degnan, Kikkawa, Moritz, Estoup & Owens, 2002). When several colonization events occur sequentially, the effects of founder events are even more dramatic because each colonization step is being founded from a population which is already genetically impoverished (Clegg, Degnan, Kikkawa et al., 2002; Chapter II, Recuerda et al., 2021a).

Insular systems are useful for understanding the process of species diversification. However, island colonization can be a complex process and combining an exhaustive sampling with phylogenetic analyses is important to determine the closest relative (either from the mainland or a neighboring island) and to infer the timing and the colonization sequence (Whittaker & Fernández-Palacios, 2007; Losos & Ricklefs, 2009; Warren et al., 2015; Whittaker et al., 2017). Many of the best-known adaptive radiations have occurred on archipelagos, including that of Darwin’s finches (Grant & Grant, 2002; Lamichhaney et al., 2016), Hawaiian honeycreepers (Lerner et al., 2011),

Anolis lizards (Glor et al., 2004), and even cichlid fishes (Wagner et al., 2013; Salzburger, 2018) if we consider lakes as islands within the mainland. Adaptive radiations are defined as the rapid diversification of species through divergence in morphology, physiology and/or behavior due to ecological opportunity (Gillespie et al., 2001), that is provided by insular environments upon colonization. Gillespie et al.,(2020) have tried to identify the commonalities about adaptive radiations and they conclude that "can proceed along multiple distinct evolutionary trajectories". Moreover, since (Gittenberger, 1991) introduced the concept of non-adaptive radiations as "evolutionary diversification from a single ancestor, not accompanied by relevant niche differentiation", there has been a debate about how to distinguish between adaptive and nonadaptive radiations (Rundell & Price, 2009; Czekanski-Moir & Rundell, 2019). Rundell & Price (2009) define nonadaptive radiation as allopatric or parapatric species that occupy ecologically similar habitats. Usually nonadaptive radiations are associated with the divergence of sexual signalling traits that diverge in allopatry or reduced dispersal (Rundell & Price, 2009; Czekanski-Moir & Rundell, 2019; Lambert et al., 2019). Examples of this kind of radiation are also found in archipelagos. For instance, *Anolis* lizards that have been mentioned as an example of adaptive radiation in this same paragraph, can also represent a nonadaptive radiation of several species that diverged within the same ecomorph (Glor et al., 2003).

In addition, insular environments impose common selective pressures, such as reduced predation and shifts among inter and intraspecific competition due to the absence of competitors (Blondel, 2000; Losos & Ricklefs, 2009) which can promote parallel phenotypic changes. For instance, the island rule defined by Foster (1964) states that upon island colonization, small and large mammal species tend to increase and decrease their size, respectively. This rule has been confirmed generally for vertebrates (Benítez-López et al., 2021) and specifically in birds, where it also affects bill length, resulting in bill elongation in short-billed species and bill shortening in long-billed species (Clegg & Owens, 2002). Moreover, the loss of flight is another common pattern of evolution in island birds (e.g., Burga et al., 2017a), which is usually explained by the lack of predators (Wright et al., 2016). Either divergent selection or genetic drift, or a combination of both, could be driving these phenotypic changes (Kolbe et al., 2012). However, it is less likely that the random nature of drift can

account for the repeated pattern of morphological change found among insular and mainland counterparts, being selection a more plausible explanation (Clegg, 2009).

Parallel and convergent evolution

The repeatability of evolution has been widely discussed and has also been empirically studied by many evolutionary biologists (Gould, 1990; Beatty, 2006; Morris, 2010; Blount et al., 2018). The repeatable results obtained in laboratory replays are common, especially when the founding populations are similar and repeatability is considered in a broad way, for instance finding that selection acts on the same genes and/or pathways, but not on the same mutations at the nucleotide level (Blount et al., 2018). As reviewed by (Bolnick et al., 2018), parallel and convergent evolution concepts have been usually mixed. The main difference among them is that parallel evolution refers to the occurrence of similar phenotypic or genomic changes that appeared independently in groups with a common ancestry (Simpson, 1961), whereas convergent evolution involves similar evolution in more distantly related taxa (Gould, 2002). However, this distinction is problematic, because the limit between "common ancestry" and "more distantly related" is imprecise and relative (Bolnick et al., 2018).

Famous examples of parallel phenotypic evolution in species that colonize new environments several times have been found in nature, including cichlid fish (Elmer et al., 2014), sticklebacks (Magalhaes et al., 2020), *Anolis* lizards (Mahler et al., 2013) and Darwin's finches (Lamichhaney et al., 2016), suggesting that parallel evolution is driven by natural selection triggering the same adaptive responses repeatedly, since is unlikely that genetic drift generates the same adaptive changes in several independent lineages (Schluter, 2000; Simpson, 1953).

Moreover, it has been shown that convergent phenotypic traits could be generated by similar (e.g., Zhen et al., 2012) but also different molecular solutions (e.g., Steiner et al., 2009). Thanks to the advances in high-throughput sequencing, the genetic basis of parallel evolution in phenotypic traits is now being widely studied, providing insights into the mechanisms of ecological speciation (Schluter et al., 2004; Rosenblum et al., 2014; Morales et al., 2019). At the molecular level, the difference between parallel and convergent evolution is that the former consists on the independent evolution of

similar phenotypes in different taxa by changes in orthologous genes (Rosenblum et al., 2014), or at the amino acid level, obtaining the same derived amino acid from the same ancestral amino acid (i.e., same ancestral state) (Storz, 2016). In contrast, convergent evolution occurs through non-orthologous genes, and within orthologous genes is caused by substitutions leading to the same derived amino acid from a different ancestral amino acid (i.e., different ancestral state). The degree of similarity at the molecular level can range from the same fixed mutations, to the same genes but different mutations, to the same pathways but different genes (Manceau et al., 2010; Sackton & Clark, 2019). There are several factors that influence the probability and similarity of parallel and convergent molecular evolution, and among them four general themes stand out, including: natural selection dynamics, demography, genetic constraints and phylogenetic history (Rosenblum et al., 2014). First, if selective pressures are shared among different lineages, natural selection may favor the appearance of similar adaptive phenotypes, which could also increase the probability of molecular parallelism (e.g., Magalhaes et al., 2020). Secondly, demographic history may play an important role due to several factors. For instance, stochastic processes determine the interaction among selection and drift, favoring or impeding the parallelism at the molecular level, and also gene flow can allow for the introduction of adaptive alleles, promoting molecular parallelism (e.g., Grant et al., 2004). The latter themes entail that closely related taxa are more correlated phylogenetically, and are likely to share similar genetic architecture and standing genomic variation, and therefore their probability of using the same genes to obtain similar phenotypes is higher than among more distantly related taxa (Conte et al., 2012). Moreover, there are traits showing lower genotype-phenotype degeneracy, meaning that several molecular pathways exist to obtain the same phenotype (Rosenblum et al., 2014). For instance, polygenic traits typically involved in local adaptation are likely to be modified through several alternative pathways, increasing the adaptive potential of organisms, yet at the same time reducing the probability of parallel evolution (Boyle et al., 2017; Szukala et al., 2022).

Local adaptation and polygenic selection

The genomic basis of local adaptation and the counteracting effect of gene flow have been topics of major interest for evolutionary biology due to their role in generating phenotypic and genomic divergence (Kawecki & Ebert, 2004; Blanquart et al., 2013; Savolainen et al., 2013; Hoban et al., 2016; Tiffin & Ross-Ibarra, 2014). Local adaptation results when divergent selection associated with environmental variables acts on individuals of the same species that inhabit differentiated environments, leading to habitat-specific modified traits (Kawecki & Ebert, 2004). When these habitats are contiguous, local adaptation can happen even in the face of gene flow if selection is strong enough (e.g., Porlier et al., 2012; Dubay & Witt, 2014; Cezard et al., 2015; Supple et al., 2015; Dennenmoser et al., 2017). However, there must be a balance between gene flow and selection because gene flow is an opposing evolutionary force that can homogenize the genomic differentiation generated by selection, precluding local adaptation at small spatial scales (Kawecki & Ebert, 2004). Genetically distinct ecotypes could appear when gene flow is non-random or restricted among adjacent habitats (Supple et al., 2015; Parchman et al., 2016). Limited gene flow in highly vagile organisms could arise due to several factors, including geographical barriers (Milá et al., 2009); a reduction in dispersal, as observed upon island colonization in some organisms (Bertrand et al., 2014), also defined as the island syndrome (Blondel, 2000; Losos & Ricklefs, 2009); a reduction in migrant fitness (Hendry, 2004; Nosil et al., 2005); assortative mating (Kirkpatrick & Ravigné, 2002; Servedio, 2016); natal habitat preference, as birds select their breeding site that is similar to their natal habitat (Stamps et al., 2009), and matching habitat choice, where individuals settle in the habitat that better suits their traits to maximize their fitness (Edelaar et al., 2008; Edelaar & Bolnick, 2012). Even though local adaptation is mainly driven by divergent natural selection, genetic drift can also affect this process. Local adaptation can be confounded by genetic drift due to the random fixation of differentiated genotypes, and moreover, drift can constrain local adaptation by reducing the genetic variance which is the substrate for selection (Yeaman & Otto, 2011; Blanquart et al., 2012).

Local adaptation implies a higher fitness of individuals in their respective environment. A way to test for local adaptation is to associate differences in

phenotypic traits with fitness among individuals inhabiting different environments by performing common garden and reciprocal transplant experiments (Savolainen et al., 2013). However, performing this kind of experiments is challenging, especially for non-model organisms in the wild (Fig. 3), therefore, usually it is only possible to infer local adaptation indirectly. Assuming that selection pressures vary among habitats, and if phenotypic changes are detected at small spatial scales, genomic signatures of selection are searched across the genome in order to find the genomic evidence of local adaptation (Barrett & Hoekstra, 2011). Several approaches have been developed in order to detect genes involved in local adaptation. Genome-wide association studies (GWAS, Korte & Farlow, 2013) and quantitative trait locus (QTL) mapping (Stinchcombe & Hoekstra, 2008) combine genotypic and phenotypic data to identify loci related to specific phenotypes; and genome-environment association analysis (GEA, Rellstab et al., 2015) are designed to detect correlations between environmental variables and the genotype (Fig. 3).

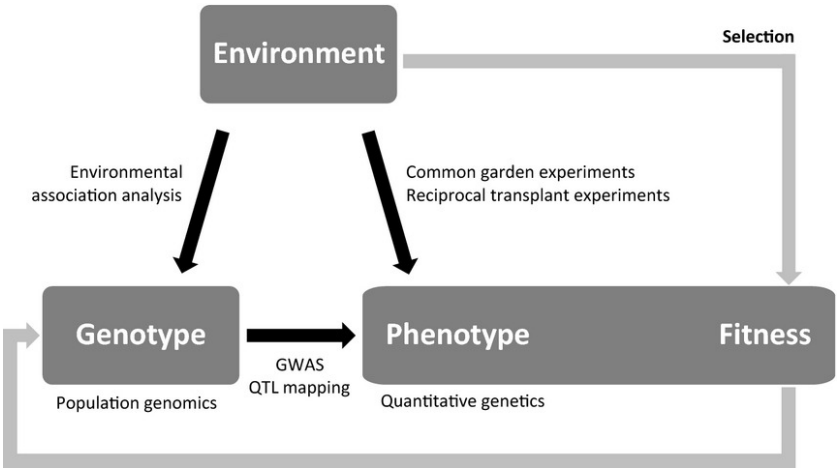


Figure 3. Approaches to detect signatures of natural selection involved in local adaptation. From Rellstab et al., (2015).

Identifying the loci responsible for adaptive trait variation is challenging because most of such traits are complex polygenic traits, which means that selection affects many loci via subtle changes in allele frequencies across the genome, making their detection difficult (Pritchard & Di Rienzo, 2010). Both GWAS and GEA are powerful approaches to detect polygenic selection (Forester et al., 2018; Kess & Boulding, 2019) and the advances in sequencing technologies have allowed their implementation in non-model organisms (e.g., Friis et al., 2018; Knief et al., 2019). As

these methods demonstrate, most adaptive traits in nature are polygenic (Santure & Garant, 2018), and even for model organisms, characterizing their genetic basis remains challenging (Pritchard & Di Rienzo, 2010; Rockman, 2012). Nowadays, many studies are focusing on filling this gap and are emphasizing the role of gene regulatory networks (GRNs) in the genomic basis of polygenic traits because it has been shown that regulatory regions are more strongly affected by polygenic selection than coding regions (reviewed in Fagny & Austerlitz, 2021). GRNs integrate evidence from different levels (genomics, epigenomics, transcriptomics, and proteomics) to characterize the regulatory associations between genes, proteins and their regulators (Fagny & Austerlitz, 2021). Since nowadays generating this kind of data is becoming logistically feasible for many species, the implementation of these methods can be generalized.

Multidisciplinary approaches in the genomic era

The advances and cost reduction in Next-Generation Sequencing (NGS) have provided the opportunity to carry out genome-wide comparative studies along the whole speciation continuum even in non-model organisms. These techniques allow to generate thousands of single-nucleotide polymorphism (SNP) loci across the genome and investigate the role of specific genes, the genomic architecture and gene flow on the evolution of divergence and reproductive isolation (Stapley et al., 2010, Feder et al., 2012; Seehausen et al., 2014; Riesch et al., 2017). Just in the last decade there has been a dramatic increase in genomic data available from all taxa along with high quality reference genomes (e.g., Cornetti et al., 2015; Ekblom et al., 2018; Shao et al., 2018; Li et al., 2019; Zhang et al., 2019; Ducrest et al., 2020; Humble et al., 2020; Yang et al., 2020; Ye et al., 2020; Li et al., 2021; Peñaloza et al., 2021; Westfall et al., 2021). This context has provided the opportunity of significant advances in all genomic fields and the development of powerful bioinformatic tools for the analysis of genomic data (Seehausen et al., 2014; Bleidorn, 2017; Casillas & Barbadilla, 2017; Campbell et al., 2018; Foote, 2018; Sackton & Clark, 2019; Bourgeois & Warren, 2021). Reduced-representation genome sequencing techniques such as Restriction-site Associated DNA (RAD-seq, Baird et al., 2008) and genotyping-by-sequencing (GBS, Elshire et al., 2011) are being commonly used for studying population structure due to

their high resolution even for recently diverged populations or taxa (e.g., Geraldes et al., 2014; Alter et al., 2017; Pearman et al., 2020). These data can also be used in phylogenomics (e.g., Near et al., 2018; Rancilhac et al., 2019; Loureiro et al., 2020), detection of adaptive loci (e.g., Funk et al., 2016; Hohenlohe et al., 2017; Xu et al., 2017; DeSilva & Dodd, 2020), hybridization (e.g., Quattrini et al., 2019; Appelhans et al., 2020; Bernhardt et al., 2020) and evolutionary history reconstruction (e.g., Friis et al., 2016; Okuyama et al., 2020; Albaladejo et al., 2021). Compared with reduced-representation methods, the resolution with whole-genome sequencing data improves (Szarmach et al., 2021) and as sequencing technologies become more affordable, the number of studies using whole-genome sequencing data in non-model organisms is increasing (e.g., Árnason et al., 2018; Ekblom et al., 2018; Malinsky et al., 2018a; Sato et al., 2020; Zhu et al., 2021).

With all these new techniques and available data, evolutionary biology is moving towards an integrative approach, evaluating interactions among the individual phenotypes, their genotypes and their environment (Losos et al., 2013). The field of 'speciation genomics' includes analysis that study the relationships among phenotypic and ecological data with genomic markers which are particularly useful for understanding the evolutionary process (Campbell et al., 2018). Similarly, the field of taxonomy and systematics is shifting to an integrative framework, through the documentation and assessment of taxonomic hypotheses by combining several lines of evidence (Tietze, 2018), which always improves hypothesis corroboration (Winker, 2021). Therefore, taxonomy and systematics are also taking advantage of the genomic era, with a major impact on conservation and management of wild populations (Hohenlohe et al., 2020; Latch, 2020; Russello et al., 2020).

Study system

This thesis studies the process of evolutionary divergence by combining high-throughput sequencing data with phenotypic and ecological information at three different spatial scales using avian species as model systems. Birds are convenient for studying the molecular basis of divergence because their genomes have relatively simple and conserved architecture (Singhal et al., 2015), are relatively small (~1Gb, Gregory, 2002) and birds are well studied organisms (Grant, 2001; Price, 2008;

Winker, 2021). Moreover, we focus on species that have colonized insular environments, which provide the opportunity to study the roles of drift and selection (Barton, 1996), colonization history (Grant, 2001), demographic events (Leroy et al., 2021) and even adaptive radiations (Losos & Ricklefs, 2009).

At the broadest scale, we conduct a comparative analysis of four different species of passerines that have populations both on the mainland and on oceanic islands (Fig. 4). These include two species that have colonized La Palma (Canary Islands, Atlantic Ocean), the red-billed chough (*Pyrrhocorax pyrrhocorax*) and the common chaffinch (*Fringilla coelebs*), and two species that have colonized Guadalupe island in the Pacific Ocean, the dark-eyed junco (*Junco hyemalis.*) and the house finch (*Haemorhous mexicanus*) (Fig. 4). In each geographic group, one species has diverged recently from the mainland (within the last 100,000 yrs, red-billed chough and house finch) while the other represents an older colonization (over 400,000 yrs, common chaffinch and junco). The red billed-chough is a species in the family Corvidae with a wide panpaleartic distribution, ranging from China to southern Europe, including the British islands and an isolated population in La Palma Island (Canary Islands) and North Africa, with another remote population in Ethiopia. The population from La Palma is more closely related to Iberian populations than to montane North African populations, and is thought to have colonized the island from the coast of Morocco when suitable habitat existed there during the late Pleistocene, forming a continuous range with Iberia (Morinha et al., 2020). The common chaffinch has a broad distribution in Eurasia and has also colonized Northern Africa and three Atlantic archipelagos of Macaronesia: the Azores, Madeira and the Canary Islands. The genus *Junco* (family Passerellidae) is distributed from Central through North America, and colonized Guadalupe Island, off the coast of Baja California, Mexico. Here we compare the dark-eyed junco (*Junco hyemalis*) found across North America with the population on Guadalupe, which has been recently split from *Junco hyemalis* and is considered to be a separate species, the Guadalupe junco (*Junco insularis*). Friis et al., (2016) showed that the juncos inhabiting North America recolonized the area through a population expansion after the last glacial maximum (18,000 years ago), whereas the juncos from Guadalupe island represent an older lineage (Aleixandre et al., 2013; Milá et al., 2016). Finally, the house finch is a common and widespread species in the family Fringillidae

and it has colonized Guadalupe Island at a later time than the Guadalupe junco (Power, 1983).

There are evident phenotypic differences such as plumage coloration between insular and mainland counterparts among all the species, except for the red-billed chough. Morphological differences between insular and mainland populations have been studied for the common chaffinch (Grant, 1980; Dennison & Baker, 1991), for the junco (Howell & Cade, 1954; Aleixandre et al., 2013) and for the house finch (Howell & Cade, 1954; Power, 1983). Moreover, differences in song have been detected between insular and mainland populations for the common chaffinch (Lachlan et al., 2013) and the junco (Aleixandre et al., 2013). The red-billed chough phenotypic differences between the island and the continent have been less studied, but some have been documented, for instance in diet (Blanco et al., 2014).

At the geographic scale of archipelagos and islands we focus on one of the previous four species, the common chaffinch (Fig. 5). The chaffinch currently consists of a species complex that includes 15 subspecies widely distributed across Eurasia, north of Africa and the Atlantic archipelagos of Azores, Madeira and the Canary Islands (Collar et al., 2020). The common chaffinch's distribution, especially across the Macaronesian archipelagos, has promted extensive study in order to understand their evolutionary history (Grant, 1980; Marshall & Baker, 1999; Baker & Marshall, 1999: Stensrud, 2013; Lachlan et al., 2013; Rodrigues et al., 2014).

At a medium geographic scale we focus on the colonization history and divergence of the common chaffinch in Macaronesia from mainland populations in Iberia and Northern Africa (Fig. 5A). At the more local within-island scale, we focus on the chaffinch population on the island of La Palma in the Canary archipelago (Fig. 5B). This population of the common chaffinch is of particular interest because there the species inhabits two really differentiated habitats, both in climatic conditions and available resources, the cloud and the pine forest. Grant (1979) found differences in bill width among cloud and pine forest individuals within La Palma. The species is also found in El Hierro, La Gomera, Gran Canaria and Tenerife, where it is found only in cloud forest. Gran Canaria and Tenerife have large tracts of pine forest, but the closely

related Tenerife and Gran Canaria blue chaffinches (*Fringilla teydea* and *F. polatzeki*, respectively) occupy this habitat and apparently exclude the common chaffinch from it. El Hierro is small and chaffinches are rare in the small tracts of pine forest, and La Gomera lacks pine forest. The eastern islands of Lanzarote and Fuerteventura are not inhabited by the common chaffinch because there is no suitable forest habitat there.

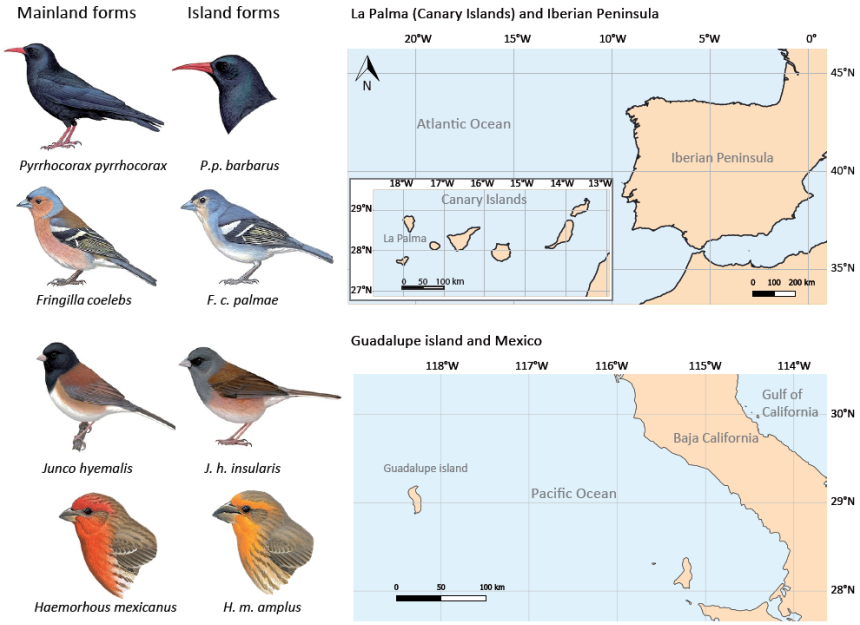


Figure 4. Model systems at a large geographic scale, comparing insular and mainland counterparts from four passerine species that have colonized oceanic islands in two different geographic regions. The red-billed chough (*Pyrrhocorax pyrrhocorax*) and the common chaffinch (*Fringilla coelebs*) from Iberia and the island of La Palma, and the house finch (*Haemorhous mexicanus*) and the junco (*Junco hyemalis/insularis*) from North America and Guadalupe island, in the Pacific Ocean. Bird illustrations from del Hoyo et al., (2018).

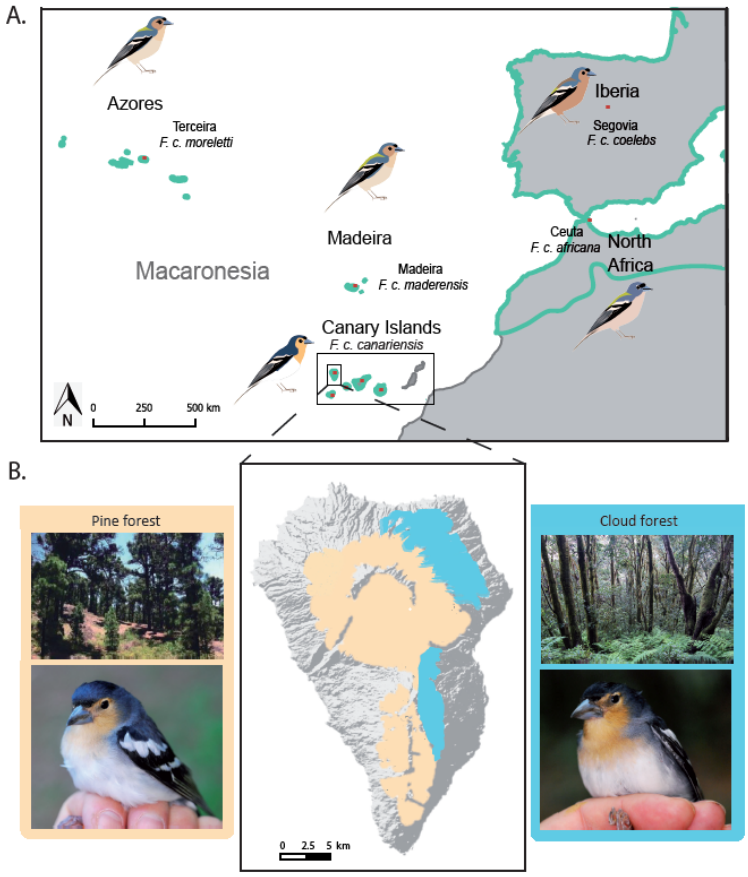


Figure 5. (A) Geographic map depicting the common chaffinch populations in Iberia (*F. c. coelebs*), North Africa (*F. c. africana*) and the Macaronesian region, including the archipelagos of Azores (*F. c. moreletti*), Madeira (*F. c. maderensis*) and the Canary Islands, which include four subspecies (for simplicity we just include *F. c. canariensis* in the figure). (B) Small-scale study of the habitat-driven divergence of the common chaffinch within the island of La Palma, where it inhabits two contrasting habitats: the humid cloud forest, represented by the blue area in the map, and the drier pine forest, represented by the orange area in the map.

GENERAL OBJECTIVES

The main goal of this thesis is to study the genetic basis of speciation in different stages of the speciation continuum and to detect the signatures of selection across the genome due to colonization of insular environments using avian species on oceanic islands as a model system. We implemented high-throughput sequencing techniques to generate single-nucleotide polymorphism (SNP) markers from non-model organisms to study the mechanisms underlying the divergence process.

In **Chapter I**, we generated whole-genome sequencing data from four species of passerines that have colonized oceanic islands in order to compare mainland and insular counterparts to detect the genomic regions under selection upon island colonization. Because all four species have colonized oceanic islands and therefore have suffered similar selective pressures and similar phenotypic changes (i.e., changes in body size and bill shape), we also assess if the same genomic regions are involved in the adaptive process. We also try to identify the mechanisms shaping the genomic landscapes of differentiation and infer the demographic history of each species using coalescent-based methods.

In **Chapter II**, we generated genotyping-by-sequencing data (GBS) in order to obtain thousands of SNPs to reconstruct the evolutionary and colonization history of the common chaffinch in Macaronesia using phylogenomic and biogeographic analyses. We also combined phenotypic and genomic data and performed species delimitation analysis in order to determine if current subspecies should be upgraded to species level.

In **Chapter III**, we combine GBS and whole-genome resequencing data (WGS) to assess the genomic signatures of local adaptation to contrasting habitats in the common chaffinch within the island of La Palma. Using the GBS dataset, we performed genome-environment association analysis (GEA) and genome-wide association studies (GWAS) to detect loci related to the environmental and phenotypic differences found among habitats. In order to study the genomic signatures of recent selection in more detail, we used WGS data from one population per habitat to perform genome scans and detect selective sweeps.

In **Chapter IV**, we generated a high-quality, chromosome-level genome assembly for the common chaffinch to use it in the rest of the chapters as reference for mapping and variant calling. The genome assembly was generated combining Illumina shotgun sequencing and Chicago and Hi-C libraries. We also characterized the completeness, the repetitive elements and the structural and functional annotation of the common chaffinch reference genome.

GENERAL METHODS

Field Sampling

Individual birds were captured in the field using mist nets. When possible, we captured adult males holding territories by placing a single mist net inside the territory and using song playbacks to attract the male to the net. In some areas (Los Tilos and Fuencaliente on La Palma, and all house finches and some juncos on Guadalupe), birds were captured at watering holes or feeding stations at picnic areas, so that both territorial and non-territorial birds were potentially included in the sample. For the red-billed chough, a non-territorial species, birds were captured at communal roosting sites, and some blood samples were obtained from nestlings at about 15–40 days of age in both La Palma (2003–2008) and the Iberian Peninsula (1988–2012). All captures were conducted with the required collecting and animal handling permits, and birds were released unharmed at the site of capture after processing.

From each bird captured we obtained morphological measurements, feathers for isotopic composition analysis and a blood sample for DNA extraction. A wing ruler was used to measure unflattened wing length to the nearest 0.5 mm, and dial calipers of 0.1-mm precision were used to measure tail length, tarsus length, bill culmen, exposed bill culmen, and bill width and depth, following Milá *et al.,* (2008). All measurements were taken by a single observer (Borja Milá). Blood samples were extracted by venipuncture of the brachial vein and were stored in absolute ethanol at -20ºC in the laboratory, until DNA extraction. All individuals were aged, sexed and marked with a uniquely numbered aluminium band. For the common chaffinch we also extracted feather samples from four different patches (nape, back, breast and rump) to measure coloration using a spectrophotometer in the laboratory. The common chaffinch samples from Azores, Madeira, La Gomera and Tenerife were collected by collaborator Juan Carlos Illera (University of Oviedo) and as the measurements were taken by two different observers (Borja Milá and Juan Carlos Illera) we did not perform comparative analysis of morphological measurements and used published data from Grant (1979).

Diet characterization

We used the analysis of feather isotope ratios of carbon (^{13}C) and nitrogen (^{15}N) isotopes as a proxy of diet to detect diet shifts among insular and mainland counterparts from the four species (Chapter I) and differences among habitats between common chaffinch individuals from the cloud and pine forest within La Palma (Chapter III). We used $1cm^2$ from a rectrix feather to extract the isotopic composition following the methods in (Hobson & Clark, 1992) and then compared among populations using an analysis of variance performed in R v.3.5 (R Core Team 2017). The ratio of stable isotopes of carbon ($^{13}C/^{12}C$) and nitrogen ($^{15}N/^{14}N$) in consumer tissues can be related to the isotopic composition of diet and therefore are useful to understand dietary patterns in wild populations (Peterson & Fry, 1987). The results are "delta" (δ) values expressed in units of permil (‰) which are presented relative to universal international standards to make these measures comparable across studies. The $\delta^{13}C$ allows determining the relative contributions of C_3 and C_4 plants to avian diets in areas where these two plant types coexist (Von Schirnding et al., 1982). Stable isotopes of nitrogen have been used to test the food web position but high levels of $\delta^{15}N$ can also indicate protein catabolism caused by nutritional stress (Hobson et al., 1993).

DNA extraction and Genotyping-by-sequencing

Genomic DNA was extracted from blood samples using a Qiagen DNeasy kit (QiagenTM, Valencia, CA) following the manufacturer's protocol. In order to obtain genome-wide data we used genotyping-by-sequencing (GBS, Elshire et al., 2011) to sequence four plates, each containing 94 samples and two blanks. We obtained 376 samples of which 83 were used in chapter II (see Appendix I) and 200 in chapter III (see Appendix II), of which 9 from La Palma were shared by both datasets as controls. The restriction enzyme used for digestion was PstI and sequencing was carried out on an Illumina Hiseq X Ten platform obtaining paired-end reads.

Resequencing

We obtained resequencing data for 9-12 individuals from insular and mainland populations from the four species for Chapter I, and for 12 individuals per habitat of the common chaffinch within the island of La Palma for Chapter III. Resequencing was conducted at Novogene Co. (UK) commercial services in a SE50 Illumina platform with 18x coverage.

Variant calling

GBS data were demultiplexed using the axe-demux command from axe v. 0.3.2-5 (Murray & Borevitz, 2018). For both GBS and resequencing datasets the read quality was assessed using FASTQC and then reads were trimmed to remove barcodes, Illumina adapters, and low-quality ends using TrimGalore! V, 0.4.4 (Krueger, 2015). Reads were mapped against the reference genome using Burrows-Wheeler Aligner (BWA, Li & Durbin, 2009) with the "-mem" algorithm. In chapters I, II and III we used as reference genome for the common chaffinch, the chromosome-level assembly from Chapter IV (GCA_015532645.2, Chapter IV: Recuerda et al., 2021b). In chapter I for the house finch we used also the common chaffinch reference genome (Chapter IV: Recuerda et al., 2021b), for the Junco we used the *Junco hyemalis* reference genome (GCA_003829775.1, Friis et al., 2022) and for the red-billed chough we used the *Corvus moneduloides* reference genome (GCA_009650955.1, bCorMon1.pri). For the GBS dataset, the variant calling was performed with GATK 3.6 HaplotypeCaller and GenotypeGVCFs tools (McKenna et al., 2010). For the resequencing from chapters I and III the variants were called using BCFTOOLS v.1.3.1 (Danecek et al., 2021) due to computational time limitations. We included invariant sites in all the resequencing datasets (Chapter I and III) and in the GBS dataset from chapter III. All datasets were filtered using VCFtools version 0.1.15 (Danecek et al., 2011). Specific details and filtering parameters are specified in each chapter.

CHAPTER I: The genetic basis of evolutionary divergence of oceanic island birds: a comparative-genomics approach

María Recuerda, Guillermo Blanco, Keith Hobson, Benoit Nabholz, Borja Milá

Pictures by:
P. pyrrhocorax: Aurélien Audevard
P. pyrrhocorax barbarus : Lau Claes
F.c.coelebs coelebs: Miguel Montoro Peinado.
F.c.coelebs palmae: Fred Hoorn. License: CC-BY-NC-ND.
H.mexicanus: Andy Reago & Chrissy McClarren/Flickr/CC by 2.0
H.mexicanus amplus: Eric VanderWerf. From (ebird.org).
Junco hyemalis: Sandy Stewart/Flickr/
Junco insularis: Eric VanderWerf. From (ebird.org).

Abstract

The colonization of oceanic islands by plants and animals has often led to the formation of new species and provides a valuable research model to study processes like evolutionary divergence and local adaptation. Understanding the factors driving phenotypic and genomic differentiation of insular populations is of major interest to gain insights into the process of divergence and speciation. Comparing patterns across different insular taxa subjected to similar selective pressures upon colonizing oceanic islands provides the opportunity to study parallel evolution and identify shared patterns in their genomic landscapes of differentiation. Birds are excellent model systems because many species have colonized oceanic islands and their genomes have a highly conserved genomic structure. We sequenced whole genomes from mainland and insular counterparts from four different species of passerines (*Fringilla coelebs*, *Pyrrhocorax pyrrhocorax, Haemorhous mexicanus* and *Junco hyemalis/insularis)* to infer their demographic history and study the pattern of genomic differentiation and the factors shaping them. We estimated the relative (F_{ST}) and absolute (d_{xy}) differentiation, nucleotide diversity (π), Tajima's D, gene density and recombination rate. Furthermore, we searched for selective sweeps and possible inversions along the genome. Insular individuals were larger than their mainland counterparts in most morphological traits, consistent with the island rule. Species differed in their demographic histories, yet all of them showed a marked decrease in effective population size (N_e) upon island colonization. However, patterns of differentiation and signals of selection were found to be species-specific. Despite the highly conserved structure of bird genomes and the similar selective factors involved, we find that patterns of genomic divergence in the four species were distinct, and each was shaped by different processes, including selective sweeps, chromosomal inversions and recurrent selection.

Introduction

The colonization of oceanic islands by mainland individuals has been a major engine of biological diversification, resulting in the evolution of thousands of new species across the world (Schluter, 2000; Grant, 2001; Warren et al., 2015; Gillespie et al., 2020). These colonization events have provided valuable research models to study processes like evolutionary divergence and local adaptation (Grant & Grant, 2002; Losos & Ricklefs, 2009; Brown et al., 2013). Upon colonization of oceanic islands, individuals across taxonomic groups have often been subjected to similar demographic and selective factors, like population bottlenecks and strong selection for local adaptation and reduced dispersal (Woolfit & Bromham, 2005; Whittaker et al., 2017). Shared patterns of phenotypic evolution of insular populations across taxonomic groups has led to general biogeographic rules, like Foster's rule, also known as the "island rule", which postulates that on islands small animals tend to become larger, and large animals tend to become smaller (Foster, 1964; Benítez-López et al., 2021). These patterns suggest the possibility of parallel evolution across species, and provide the opportunity to test whether the selective mechanisms acting during island colonization are shared across species, and whether selection acts on the same or different genomic loci.

The genomic underpinnings of island colonization are poorly understood, even though an increasing number of studies are addressing this topic thanks to the recent high-throughput advances in sequencing (reviewed by Sackton & Clark, 2019). Both selection and drift can drive phenotypic changes in islands; however when patterns of parallel phenotypic changes are observed, they are more likely to be driven by selection rather than random drift (Clegg, 2009; Rosenblum et al., 2014). Parallel phenotypic changes could be promoted in islands by similar selective pressures due to their particular features compared to the mainland, such as simplified ecosystems and reduced trophic resources (Losos & Ricklefs, 2009). In addition, in insular environments species encounter new ecological niches, a reduction in predation and an increase in intraspecific competition, while interspecific competition is reduced (Blondel, 2000; Losos & Ricklefs, 2009). These insular selective pressures usually result in size changes (Benítez-López et al., 2021), usually attributed to the absence of predators and the shifts in competition, and also result in diet shifts in order to adapt to the new resources, leading to behavioral (Sayol et al., 2018; Lapiedra et al., 2021),

morphological (Glor et al., 2004; Campana et al., 2020) and physiological adaptations (Blanco et al., 2014; Tattersall et al., 2018). The molecular basis of convergent phenotypic traits could be shared, but could also be the result of different genomic pathways. The degree of molecular parallelism depends on whether the same mutation occurs on the same gene, changes in different nucleotides within the same gene, or changes in different genes within the same pathway (Manceau et al., 2010; Sackton & Clark, 2019). The probability of molecular parallelism is determined by several factors, increasing when selective pressures are similar and genomic constraints such as demography and phylogenetic history are shared (Rosenblum et al., 2014). For polygenic traits that can be modified through multiple different pathways, molecular parallelism is less likely (Rosenblum et al., 2014; Boyle et al., 2017) resulting instead in heterogeneous patterns of differentiation.

Understanding the factors that generate heterogeneous patterns of differentiation across the genome is one of the main goals of population genomics (Cruickshank & Hahn, 2014; Burri, 2017; Ravinet et al., 2017; Chase et al., 2021). The main factors shaping differentiation patterns are drift and selection, but demographic history and genome features such as recombination rate and gene content also affect the distribution of the differentiated regions (Ravinet et al., 2017). Recent advances in sequencing technologies have allowed studying the genomic landscapes, which show the distributional pattern of specific genomic variables across the genome (Ellegren et al., 2012; Nadeau et al., 2012; Poelstra et al., 2014; Meier et al., 2018). Regions that are highly divergent from the genomic background are known as "islands of differentiation" (Turner et al., 2005; Ellegren et al., 2012) and are usually detected as regions of high relative divergence (F_{ST}, Weir & Cockerham, 1984). The initial genome scans interpreted F_{ST} peaks as signatures of strong selection surrounded by valleys homogenized by gene flow (Nosil, Funk & Ortiz-Barrientos, 2009). Those F_{ST} peaks were caused by marked differences in allele frequencies at locally adapted sites and the neutral loci linked to them (Charlesworth et al., 1997; Feder & Nosil, 2010). However, when considering patterns of absolute divergence (d_{xy}) and within population diversity (π) besides F_{ST}, new interpretations of how these islands of differentiation originate have been put forward. F_{ST} peaks could also appear when population diversity is low in either of the populations compared, while d_{xy} is less affected by this pattern. Several processes such as positive and/or background

selection can reduce within population nucleotide diversity and generate the "islands" of relative divergence, while absolute divergence remains unchanged (Cruickshank & Hahn, 2014; Burri et al., 2015; Irwin et al., 2018). Four models have been proposed to explain the cause of the islands of differentiation (Irwin et al., 2016; Irwin et al., 2018) and in order to differentiate these models is crucial to understand the relationship between F_{ST}, d_{xy} and π (Cruickshank & Hahn, 2014; Irwin et al., 2016; Han et al., 2017; Irwin et al., 2018). Two of those models account for speciation in the presence of gene flow ("Divergence-with-gene-flow" and "sweep-before-differentiation") and the other two for allopatric speciation ("Selection in allopatry" and "Recurrent selection")(Irwin et al., 2016). Moreover, other factors such as demographic history, heterogeneity of mutation rate and recombination rate across the genome, as well as gene density, could modify the genomic landscape (Ravinet et al., 2017). Therefore, to correctly interpret the genomic landscapes of differentiation it is important to understand the demographic and evolutionary history of the target species (Ravinet et al., 2017). Variations in effective population size (N_e) can produce different genomic signatures. For instance, marked reductions in N_e such as those caused by population bottlenecks can modify levels of background selection and therefore the baseline for the detection of outlier loci (Ferchaud & Hansen, 2016; Leroy et al., 2021).

Covariation of genomic patterns of differentiation among different avian species has been shown across wide evolutionary timescales (Van Doren et al., 2017; Delmore et al., 2018; Vijay et al., 2017) and the coincidence of differentiation peaks has been of special interest to understand the process of convergent molecular evolution where similar loci evolve independently in several species (Seehausen et al., 2014). Bird genomes show high synteny (Zhang, 2014), a stable number of chromosomes (Ellegren, 2010), similar recombination landscapes (Singhal et al., 2015; Kawakami et al., 2017), and across species microchromosomes show higher density in gene content than macrochromosomes (Dutoit et al., 2017; Singhal et al., 2015). The similarity in genomic landscapes of differentiation across closely related and diverged avian species could be due to the non-random distribution of gene content across the genome and the coincidence of low recombination areas along with linked selection (Van Doren et al., 2017; Irwin et al., 2018), since it has been shown that the recombination landscape in birds can be maintained across species over long evolutionary times (Singhal et al., 2015).

Here we use a comparative genomic approach to examine patterns of genome-wide differentiation in avian species that have colonized oceanic islands, with the goal of assessing the relative roles of demographic history, time of divergence, and directional selection in driving divergence upon island colonization. Our model system is composed of four passerine species that have mainland populations and have colonized oceanic islands; two species from mainland Europe that have colonized the island of La Palma in the Canary Islands, Atlantic Ocean, the common chaffinch (*Fringilla coelebs)* and the red-billed chough (*Pyrrhocorax pyrrhocorax)*, and two species from North America that have colonized Guadalupe Island on the Pacific Ocean, the house finch *(Haemorhous mexicanus*) and the dark-eyed junco (*Junco hyemalis*). The red-billed chough and the house finch have diverged from mainland populations within the last 100,000 years, whereas the common chaffinch and the junco have been separated from their mainland relatives for over 500,000 years (Aleixandre et al., 2013; Morinha et al., 2020; Chapter II: Recuerda et al., 2021a). Given that all four species have colonized oceanic islands and have been subjected to similar selective pressures, we first analysed if the differences in diet and phenotype between insular and mainland counterparts affected the same traits across species. Changes in morphological traits and diet are expected upon colonization of the new insular environment (Warren et al., 2015; Whittaker et al., 2017) and those changes are likely to have detectable genomic signatures. Therefore, we also asked if the genomic landscapes of differentiation are similar among species when taking divergence time into account. In order to interpret the genomic landscapes and given the marked geographic isolation of the islands in our system, we focus on models without gene flow and assume that in all four species insular populations have been isolated since colonization (Power, 1983; Aleixandre et al., 2013; Morinha et al., 2020; Chapter II: Recuerda et al., 2021a). The "selection in allopatry" model explains regions of high relative differentiation between populations by a reduction in within-population diversity in the regions under selection on each population (Nachman & Payseur, 2012; Cruickshank & Hahn, 2014; Burri et al., 2015; Vijay et al., 2017), whereas selection has no effect on between-population divergence and therefore d_{xy} is expected to be similar between regions regardless of F_{ST} levels (Cruickshank & Hahn, 2014; Han et al., 2017)**.** In contrast, the "recurrent selection" model (Nachman & Payseur, 2012; Cruickshank & Hahn, 2014) was developed to explain the pattern

where regions showing high relative divergence showed low absolute divergence (Cruickshank & Hahn, 2014; Van Doren et al., 2017). In this model without gene flow, selection reduces standing variation in those regions in the common ancestor before it splits into the current populations generating a pattern of low d_{xy}, and after the split those regions experience selection that reduces within-population genetic diversity, resulting in high F_{ST}.

We performed whole genome resequencing (WGS) of 9-12 individuals per treatment per species in order to determine whether the four species showed similar patterns of differentiation in their genomic landscapes, and whether these patterns have been shaped by similar processes. We studied the demographic history and performed genomic scans of F_{ST}, d_{xy}, π, Tajima's D, recombination rate, gene content and selective sweeps. We also scanned the genomes looking for putative chromosomal inversions, which have been shown to underlie major phenotypic polymorphisms in birds (Tuttle et al., 2016). We detected regions under selection among insular and mainland counterparts as F_{ST} outliers and selective sweeps, and identified shared candidate genes among the four species.

Comparing the genomic signatures of the divergence process in four different species that have been exposed to similar selective pressures (by colonizing oceanic islands) and that differ in colonization time (which can be considered as a proxy for different stages along the speciation continuum), can provide useful understanding for the mechanisms shaping the genomic landscapes through the divergence process over time.

Methods

Study Area and fieldwork

We sampled the common chaffinch and the red billed chough continental populations in the Iberian Peninsula at Segovia and Monegros, respectively. The insular populations from both species were sampled in La Palma, the most north-western island of the Canary Islands archipelago (Fig. 1A, Table 1). The continental populations of the house finch and dark-eyed junco were sampled in California, and two house finch individuals were captured in Sierra Juarez (Baja California, Mexico).

Both insular populations were sampled in Guadalupe Island in the Pacific Ocean (Fig. 1B, Table 1).

Table 1. Sample sizes and sampling population for the four species from mainland and insular populations: red-billed chough, common chaffinch, dark-eyed junco and house finch.

Species	$\mathbf{N_{cont}}$	**Mainland Population**	$\mathbf{N_{is}}$	**Island Population**
Red-billed chough	12	Monegros, Iberia	12	La Palma
Common chaffinch	9	Segovia, Iberia	12	La Palma
Dark-eyed junco	12	California	12	Guadalupe
House finch	12	California/Sierra Juarez	12	Guadalupe

All four species were captured in the field using mist nets, and mesh traps were also used to capture red-billed choughs. All individuals were marked with uniquely numbered aluminium bands, sexed, aged and measured. A blood sample was obtained by venipuncture of the sub-brachial vein and stored in absolute ethanol at -20°C in the laboratory for DNA extraction. A tail feather was collected to study the diet by isotope analysis. After processing, birds were released unharmed at the site of capture. We used molecular sexing of choughs by the amplification of the *Chd1* gene, which in females, as the heterogametic sex, is present on the W and Z chromosomes resulting in differences in intron length and renders two distinct bands in the agarose gel whereas in males, that are homogametic, just one band is detected (Griffiths et al., 1996).

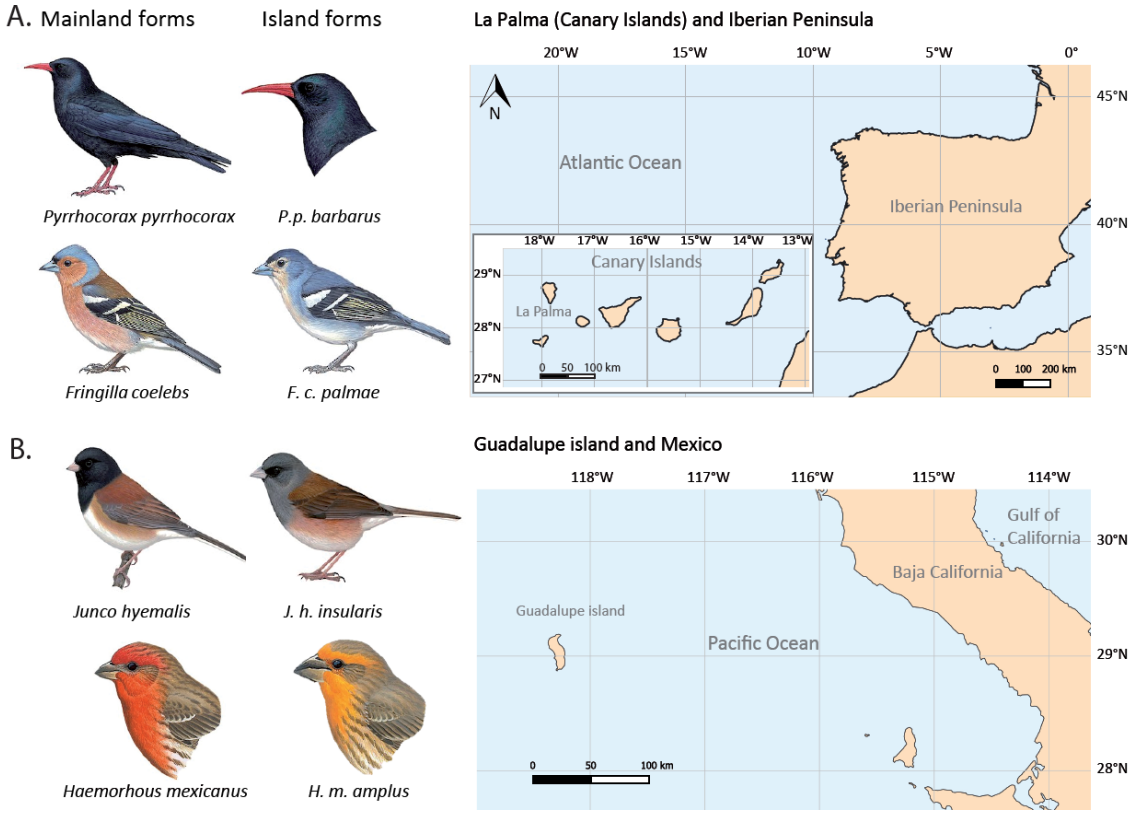


Figure 1. Target taxa for the comparative genomics analysis. (A) Species that have colonized La Palma in the Atlantic Ocean: the red-billed chough and the common chaffinch. (B) Species that have colonized Guadalupe island in the Pacific Ocean: the dark-eyed junco and the house finch. Bird illustrations from del Hoyo et al., (2018).

Morphological data and analysis

We compared the morphological traits of adult males from mainland and insular populations for all species except for the red-billed chough. We used univariate ANOVA to compare the means among treatments for each species. A wing ruler was used to measure unflattened wing length to the nearest 0.5 mm, and dial callipers of 0.1-mm precision were used to measure tail length, tarsus length, bill culmen, total bill length, bill width and bill depth, following (Milá et al., 2008). All measurements were taken by a single observer (BM).

Diet characterization and analysis

Isotopic analysis of a rectrix feather fragment (1 cm^2) from 11-13 individuals per treatment of each species was conducted following the methods in Hobson & Clark (1992). The results are "delta" (δ) values which are expressed in units of permil (‰).

The ratios of carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) isotopes were obtained for each population and compared using an ANOVA analysis. The $\delta^{13}C$ allows determining the relative contributions of C_3 and C_4 plants to avian diets in areas where these two plant types coexist. Stable isotopes of nitrogen have been used to test the food web position but high levels of $\delta^{15}N$ can also indicate protein catabolism caused by nutritional stress (Hobson et al., 1993). Analyses were performed in R v. 3.5 (Team, 2017).

Genome resequencing

DNA was extracted with a QIAGEN Blood and Tissue kit following the manufacturer's protocol. Resequencing of 24 individuals per species (12 per treatment, but only 9 for the mainland common chaffinch) with 18x coverage was conducted with Novogen commercial services. Sequencing was performed on a SE50 (Illumina). Reads were trimmed with Trim galore (Krueger, 2015) and mapped on their respective reference genome using BWA (Burrows-Wheeler Aligner, Li & Durbin, 2009). For the common chaffinch and the house finch we used the common chaffinch reference genome (GCA_015532645.2, Chapter IV: Recuerda et al., 2021b), for the Junco we used the *Junco hyemalis* reference genome (GCA_003829775.1, Friis et al., 2022) and for the red-billed chough we used the *Corvus moneduloides* reference genome (GCA_009650955.1, bCorMon1.pri) SNPs were called using BCFTOOLS v.1.3.1 (Danecek et al., 2021) including invariant sites. Filtering was performed with VCFTOOLS v. 0.1.15 (Danecek et al., 2011) separately for variant and invariant sites, using the following criteria for variant sites: (i) Indels and sites with more than two alleles were removed, (ii) a number of reads per site between 10 and 40; (iii) a minimal genotype quality of 30; (iv) a minor allele frequency of 0.01); and (v) 25% maximum of missing data and a minimal genotype quality of 30. Then, variant and invariant sites were merged using BCFTOOLS concat. The reference genomes from all four species were aligned to the Zebra finch genome (*Taeniopygia guttata*, bTaeGut2.pat.W.v2) using nucmer from the MUMmer package (v.4.0, '-b 400' and filtering with 'delta-filter -1'; Marçais et al., 2018) and chromosomes were numbered accordingly (see Table 2, Fig.2).

Table 2. Correspondence between the chromosomes of the Zebra finch (*Taeniopygia guttata*) genome assembly (bTaeGut2.pat.W.v2), including the chromosome name and Refseq ID and the scaffolds of the reference genomes for the common chaffinch (*Fringilla coelebs*), the junco (*Junco hyemalis*) and the New Caledonian crow (*Corvus Moneduloides*).

Chrom.	RefSeq Zebra finch	Common chaffinch assembly	Junco assembly	New Caledonian crow assembly
1	NC_044998.1	ScDMRwT_581_HRSCAF_676	ScoVZU6_963__HRSCAF___984	NC_045477.1
1A	NC_044999.1	ScDMRwT_3155_HRSCAF_3635	ScoVZU6_1238__HRSCAF___1266	NC_045479.1
2	NC_045000.1	ScDMRwT_473_HRSCAF_551	ScoVZU6_2668__HRSCAF___2714	NC_045476.1
3	NC_045001.1	ScDMRwT_2155_HRSCAF_2491	ScoVZU6_1491__HRSCAF___1522	NC_045478.1
4	NC_045002.1	ScDMRwT_1738_HRSCAF_2009	ScoVZU6_2655__HRSCAF___2701	NC_045480.1
5	NC_045004.1	ScDMRwT_3240_HRSCAF_3767	ScoVZU6_3544__HRSCAF___3601	NC_045489.1
4A	NC_045003.1	ScDMRwT_1904_HRSCAF_2197	ScoVZU6_1225__HRSCAF___1252	NC_045481.1
6	NC_045005.1	ScDMRwT_2188_HRSCAF_2528	ScoVZU6_982__HRSCAF___1003	NC_045483.1
7	NC_045006.1	ScDMRwT_1364_HRSCAF_1572	ScoVZU6_3251__HRSCAF___3306	NC_045482.1
8	NC_045007.1	ScDMRwT_804_HRSCAF_937	ScoVZU6_2192__HRSCAF___2234	NC_045484.1
9	NC_045008.1	ScDMRwT_200_HRSCAF_244	ScoVZU6_201__HRSCAF___207	NC_045485.1
10	NC_045009.1	ScDMRwT_1361_HRSCAF_1569	ScoVZU6_4447__HRSCAF___4518	NC_045488.1
11	NC_045010.1	ScDMRwT_1149_HRSCAF_1324	ScoVZU6_3582__HRSCAF___3639	NC_045487.1
12	NC_045011.1	ScDMRwT_2095_HRSCAF_2420	ScoVZU6_4448__HRSCAF___4521	NC_045486.1
13	NC_045012.1	ScDMRwT_1072_HRSCAF_1239	ScoVZU6_2057__HRSCAF___2097	NC_045490.1
14	NC_045013.1	ScDMRwT_2434_HRSCAF_2805	ScoVZU6_3803__HRSCAF___3864	NC_045491.1
15	NC_045014.1	ScDMRwT_1873_HRSCAF_2164	ScoVZU6_4455__HRSCAF___4539	NC_045493.1
17	NC_045015.1	ScDMRwT_2070_HRSCAF_2390	ScoVZU6_1929__HRSCAF___1968	NC_045496.1
18	NC_045016.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_582__HRSCAF___593	NC_045494.1
19	NC_045017.1	ScDMRwT_2017_HRSCAF_2332	ScoVZU6_2973__HRSCAF___3026	NC_045495.1
20	NC_045018.1	ScDMRwT_211_HRSCAF_256	ScoVZU6_1043__HRSCAF___1065	NC_045492.1
21	NC_045019.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_3047__HRSCAF___3100	NC_045497.1
22	NC_045020.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_1118__HRSCAF___1143	NC_045502.1
23	NC_045021.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_300__HRSCAF___308	NC_045498.1
24	NC_045022.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_957__HRSCAF___978	NC_045500.1
25	NC_045023.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_1653__HRSCAF___1687	NC_045504.1
26	NC_045024.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_3291__HRSCAF___3346	NC_045499.1
27	NC_045025.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_3522__HRSCAF___3579	NC_045501.1
28	NC_045026.1	ScDMRwT_1489_HRSCAF_1712	ScoVZU6_4457__HRSCAF___4541	NC_045503.1
Z	NC_045027.1	ScGsoED_1_HRSCAF_2	ScU1KuS_2__HRSCAF___21	NC_045511.1

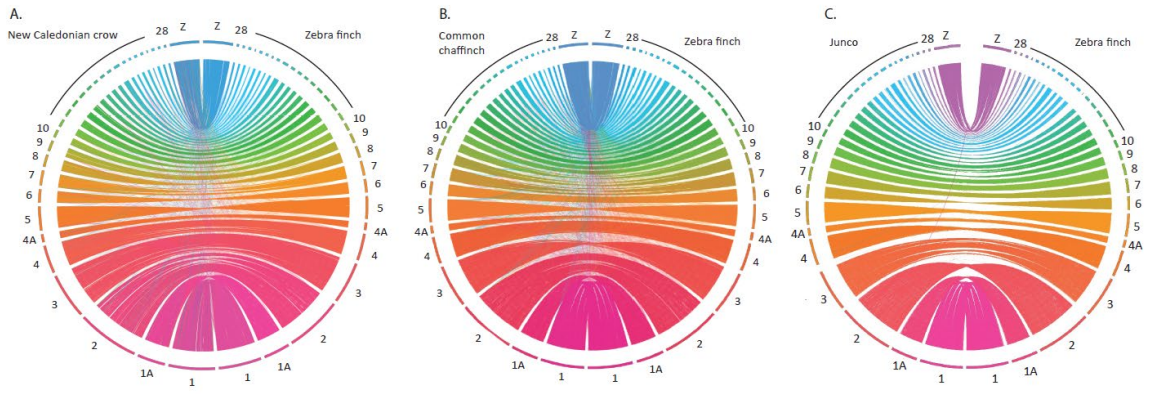


Figure 2. Synteny plots of the three reference genomes used, against the zebra finch genome (*Taeniopygia guttata*, bTaeGut2.pat.W.v2) performed with Circos (Krzywinski et al., 2009). (A) New Caledonian crow (*Corvus moneduloides*), (B) Common chaffinch (*Fringilla coelebs*) and (C) Dark-eyed junco (*Junco hyemalis*). The zebra finch is on the right hemisphere of each plot.

Inference of demographic history

The history of the effective population size of each species was estimated using Pairwise Sequentially Markovian Coalescent (PSMC) analysis (Li & Durbin, 2011). The PSMC model infers the demographic history based on genome-wide heterozygous sequence data. We used SAMTOOLS (Li et al., 2009) to obtain diploid consensus sequences from BAM files generated with BWA-mem (Li & Durbin, 2009). Sites with sequencing depth lower than 10 and higher than 35 were removed. Because sex chromosomes can show different rates and patterns of evolution than autosomes (reviewed by Irwin, 2018; Wright & Mank, 2013), we focused our comparisons of differentiation statistics on autosomes only. We converted the diploid consensus sequence to PSMC input files (psmcfa) using the tool fq2psmcfa included in the PSMC software. Then, the program PSMC was used to infer the population history with the options '-N25 –t5 –r1 –p "4+30*2+4+6+10', except for the mainland common chaffinch, and for both populations of the house finch, where the upper time limit was set to 1 (-t1). We performed 100 bootstraps for one genome per species and treatment. The atomic time interval was set following Nadachowska-Brzyska et al., (2016). We used a mutation rate of 4.6 e-9 mutation/site/generation (Smeds et al., 2016) that has been used in other avian systems for PSMC analysis (e.g., Ericson et al., 2017; Hanna et al., 2017; Sato et al., 2020; Campana et al., 2020). Generation time was set to two years for all the species (Baker & Marshall, 1999; Reid et al., 2003; Møller,

2006; Friis et al., 2016). The plots were generated using the function "plot_psmc" within the PSMC software.

Inference of recombination rate

In order to determine if the genomic landscapes of differentiation were affected by recombination, we estimated recombination rates across the genome for insular and continental populations for the four different species using LDhat software (McVean & Auton, 2007). First, we created a modified likelihood lookup table based on the LDhat precomputed tables using a sample size of 12 per treatment (9 for the continental common chaffinch) and a population mutation rate parameter estimate of 0.001. Then vcf files were split into chunks of 10,000 SNPs and converted to ldhat format using VCFTOOLS v. 0.1.15 (Danecek et al., 2011). The input files generated were used in LDhat "interval" to estimate the effective recombination rate by implementing a Bayesian MCMC sampling algorithm with five million iterations, sampling every 5,000 steps and a block penalty of 10. Finally, the results were summarized using the LDhat module "stat", discarding 20% of the samples as burn-in.

Genome scans and detection of selective sweeps

In order to detect genomic signatures of selection among the island and mainland counterparts from the four different species, we estimated two different statistics, the fixation index (F_{ST}, Weir & Cockerham, 1984) and the Cross-population extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007). First, F_{ST}, d_{xy} and π using were calculated in non-overlapping windows of 10 kb using pixy v. 2 (Korunes & Samuk, 2021). Pixy takes into account the invariant sites for π and d_{xy} calculations, thus overcoming the problem of most programs that use VCF files to calculate those statistics that do not distinguish among invariant and missing sites resulting in deflated estimates (Korunes & Samuk, 2021). We also computed Tajima's D (Tajima, 1989) in non-overlapping 10Kb windows with VCFTOOLS (Danecek et al., 2011). Then, the averaged values of each variable were transformed to Zscores using the "scale" command in R. To detect F_{ST}, outliers we corrected for multiple testing setting the FDR to 0.05 (Benjamini & Hochberg, 1995).

To detect selective sweeps, we computed the Cross-population extended haplotype homozygosity (XP-EHH, Sabeti et al., 2007) using the R package "rehh" (Gautier et al., 2017). First, we phased the vcf files containing only the variant sites in 50Kb windows using Shapeit v2.r904 (Delaneau et al., 2013). The XP-EHH is based on the comparison of haplotype lengths between populations and has more detection power when the selected haplotype is near fixation in one population and still polymorphic in the other. The genomic regions showing a –log10(p-value) ≥ 3 were considered to be under selection. Then, we looked for overlapping regions between the F_{ST} and the XP-EHH outliers. We generated Manhattan plots for all the statistics using the R package qqman (Turner, 2018) in R v. 3.6 (R Core Team, 2018).

Detecting putative chromosomal inversions

We also studied how patterns of population structure varied along the genome in order to detect potential chromosomal inversions using the R package "lostruct" (Li & Ralph, 2019). SNP data for each species including only variant sites was converted to BFC format using BCFTOOLS version 1.9 (Li et al., 2009). We implemented the script provided by Huang et al., (2020), dividing the genome into 1,000 SNPs non-overlapping windows and applying a principal component analysis (PCA) to each window. Euclidian distances between the two first principal components (PCs) between windows were calculated and mapped using multidimensional scaling (MDS) into a 40-dimensional space in order to see the similarity of the relatedness patterns between windows. To identify genomic regions with extreme MDS values, windows with absolute values greater than 4 SD over the mean across all windows were selected for each MDS coordinate. We performed 1,000 permutations of windows over chromosomes to test if outlier regions were randomly distributed across chromosomes. The putative inversion coordinates were the start position of the first outlier window and the end position of the last outlier window. The script included additional analysis to check if the MDS outliers were detecting inversions or instead other processes such as linked selection. First, a PCA was performed using the SNPs from each putative inversion with SNPRelate (Zheng et al., 2012). Inversions in the PCA would split the samples into three different groups (i.e., the two orientations and the heterozygotes in an intermediate cluster). The R function "kmeans" with the

Hartigan and Wong method (Hartigan & Wong, 1979) was used to identify the composition of groups of genotypes by performing clustering on the first PC, setting the initial cluster centres as the maximum, minimum and middle of the PC scores range. Then, another test was performed averaging the individual heterozygosity per group detected by the k-means clustering. Inversions would show the pattern of higher heterozygosity of the central group relative to the other two groups. Finally, only MDS outlier regions that clustered into three groups in the PCA and showed higher heterozygosity in the middle group were considered as putative inversions.

Candidate genes and GO-term enrichment analysis

We extracted the candidate genes of the genomic regions detected to be under selection by both methods separately (F_{ST} and XP-EHH outliers) using bedtools intersect and the annotation of their respective reference genome. We checked their functions in genecards (Rappaport et al., 2017). We obtained the GO terms using the zebra finch dataset in biomaRt in R. Then, we performed a Gene ontology (GO) enrichment analysis for each set of outliers in the category "biological function" using the TopGO R package (Alexa & Rahnenfuhrer, 2016). To estimate the statistical significance, we used the Fisher exact test implementing the weight01 method. As recommended by the TopGO authors, we did not implement corrections for multiple testing and presented raw p-values for the top-10 GO terms related to biological processes.

Results

Morphological differences and diet characterization and analysis

The morphological analysis revealed marked differences in most of the traits between insular and continental populations for the three analysed species. We found the same pattern of significant greater values for most of the traits in the insular populations compared to mainland, except for the junco wing length that is larger in the continent (Table 3-5).

Table 3. Mean values ($\bar{x}$), standard deviation (sd) and ANOVA test including *F* statistic, degrees of freedom (df) and *p*-value for the common chaffinch morphological traits from the insular and mainland populations. The sample size for each population is 25. The compared traits are wing length, tail length, tarsus length, culmen, exposed culmen, bill depth and bill width. Statistically significant *p*-values are in bold.

Trait	$\bar{x}_{island}$ **± sd n=25**	$\bar{x}_{continent}$ **± sd n=25**	***F* value** $_{df}$	***p*-value**
Wing length	86.28 ± 1.52	85.94 ± 1.79	0.52 $_{1,48}$	0.47
Tail length	75.05 ± 1.61	66.77 ± 2.72	171.3 $_{1,48}$	**<.0001**
Tarsus length	21.8 ± 0.54	18.26 ± 0.52	559 $_{1,48}$	**<.0001**
Culmen	11.78 ± 0.44	9.76 ± 0.35	322.7 $_{1,48}$	**<.0001**
Exp. Culmen	15.47 ± 0.61	12.53 ± 0.41	403.4 $_{1,48}$	**<.0001**
Bill depth	8.09 ± 0.26	6.73 ± 0.21	417.9 $_{1,48}$	**<.0001**
Bill width	6.24 ± 0.17	5.39 ± 0.22	242.6 $_{1,48}$	**<.0001**

Table 4. Mean values ($\bar{x}$), standard deviation (sd) and ANOVA test including *F* statistic, degrees of freedom (df) and *p*-value for the junco morphological traits from the insular and mainland populations. The sample size for each population is 12. The compared traits are wing length, tail length, tarsus length, culmen, exposed culmen, bill depth and bill width. Statistically significant *p*-values are in bold.

Trait	$\bar{x}_{island}$ **± sd n=12**	$\bar{x}_{continent}$ **± sd n=12**	***F* value** $_{df}$	***p*-value**
Wing length	68.59 ± 1.31	71.38 ± 1.76	19.27 $_{1,22}$	**<.0005**
Tail length	62.01 ± 2.06	62.77 ± 2.93	0.53 $_{1,22}$	0.47
Tarsus length	21 ± 0.41	20.11 ± 0.41	27.56 $_{1,22}$	**<.0001**
Culmen	9.8 ± 0.42	8.68 ± 0.42	43.13 $_{1,22}$	**<.0001**
Exp. Culmen	13.08 ± 0.46	11.52 ± 0.44	72.1 $_{1,22}$	**<.0001**
Bill depth	6.14 ± 0.16	5.69 ± 0.28	23.08 $_{1,22}$	**<.0001**
Bill width	4.43 ± 0.15	4.56 ± 0.22	2.66 $_{1,22}$	0.12

Table 5. Mean values ($\bar{x}$), standard deviation (sd) and ANOVA test including *F* statistic, degrees of freedom (df) and *p*-value for the house finch morphological traits from the insular and mainland populations. The sample size for each population is 8 and 5 for the insular and mainland populations, respectively. The compared traits are wing length, tail length, tarsus length, culmen, exposed culmen, bill depth and bill width. Statistically significant *p*-values are in bold.

Trait	$\bar{x}_{island}$ **± sd n=8**	$\bar{x}_{continent}$ **± sd n=5**	***F* value** $_{df}$	***p*-value**
Wing length	81.13 ± 1.57	79.4 ± 2.38	2.51 $_{1,11}$	0.14
Tail length	65.15 ± 1.65(n=7)	62.86 ± 4.55	1.56 $_{1,10}$	0.24
Tarsus length	19.29 ± 0.51	17.5 ± 0.46	41.14 $_{1,11}$	**<.0001**
Culmen	9.95 ± 0.33	8.3 ± 0.62	40.06 $_{1,11}$	**<.0001**
Exp. Culmen	12.54 ± 0.37	10.88 ± 0.13	90.56 $_{1,11}$	**<.0001**
Bill depth	10 ± 0.35	7.88 ± 0.20	145.2 $_{1,11}$	**<.0001**
Bill width	7.71 ± 0.24	6.44 ± 0.19	97.74 $_{1,11}$	**<.0001**

The feather isotope analysis revealed that all species except the red-billed chough showed differences between mainland and island counterparts in at least one of the isotopes (Table 6). The common chaffinch showed differences in the $\delta^{13}C$, the Junco in the $\delta^{15}N$ and the house finch in both isotopes (Table 6).

Table 6. Analysis of variance for carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) isotope ratios from rectrix feathers for island and mainland counterparts from each of the four species: Species name, sample size for the island and the continent (N_{is} and N_{cont}, respectively), mean and standard deviation (SD) for $\delta^{13}C$ and $\delta^{15}N$ for the island and the continent, degrees of freedom (Df), *F* statistic value (*F*) and *p*-value (*p*). Significant *p*-values are highlighted in bold.

	Species	**Nis**	**Ncont**	**Mean is. ± SD**	**Mean cont. ± SD**	**Df**	***F***	***p***
$\delta^{13}C$	*F. coelebs*	12	13	-23.24 ± 0.63	-21.06 ± 2.12	1,23	11.72	**<.005**
	H. mexicanus	12	15	-22.47 ± 0.52	-19.88 ± 3.05	1,25	8.41	**<.01**
	J. oreganus	12	11	-22.66 ± 0.63	-23.40 ± 1.31	1,21	3.08	0.09
	P. phyrrocorax	12	12	-22.88 ± 0.96	-22.90 ± 0.51	1,22	0.0055	0.94
$\delta^{15}N$	*F. coelebs*	12	13	4.98 ± 0.75	6.30 ± 3.10	1,23	2.05	0.16
	H. mexicanus	12	15	7.47 ± 2.11	7.07 ± 1.74	1,25	29.91	**<0.0001**
	J. oreganus	12	11	9.24 ± 0.78	4.9 ± 1.28	1,21	98.53	**<0.0001**
	P. phyrrocorax	12	12	11.72 ± 2.28	7.24 ± 1.97	1,22	0.26	0.62

Whole-genome resequencing

The total number of sites obtained in the variant calling was close to the length of the reference genomes, the number of variant sites for all species was similar (40-50 million) except for the red-billed chough, which was lower (~13 million), and the same pattern was maintained after filtering (Table 7). The lower number of variants of the red-billed chough is consistent with its recent divergence, although the house finch shows a high level of polymorphism, comparable to the other two species that diverged longer time ago.

Table 7. Number of sites obtained in the variant calling for each species performed with BCFTOOLS, including the total number of sites (variant and invariant), only the variant sites and the total number of sites and variant sites after filtering.

Species	Total sites	Variant sites	Filtered dataset	Filtered variants
Red-billed chough	1,026,780,642	13,411,090	1,013,391,645	4,626,766
Common chaffnch	994,789,329	52,711,626	979,600,935	39,937,905
Dark-eyed junco	1,001,453,032	42,444,447	987,515,496	31,060,630
House finch	973,127,256	43,257,766	954,187,282	29,423,214

Inference of demographic history

PSMC demographic inference revealed a consistent pattern for the four different species, showing higher and more stable population size for mainland populations and a sharp decrease in effective population size in insular populations following colonization. The divergence time estimates obtained from the PSMC analysis are around 0.9 million years for the common chaffinch, 100,000 years for the house finch, 400,000 years for the dark-eye junco, and 30,000 years for the red-billed chough (Fig. 3). The continental population of the red-billed chough showed the smallest effective population size, and the smallest difference between the continental and insular populations among the target species.

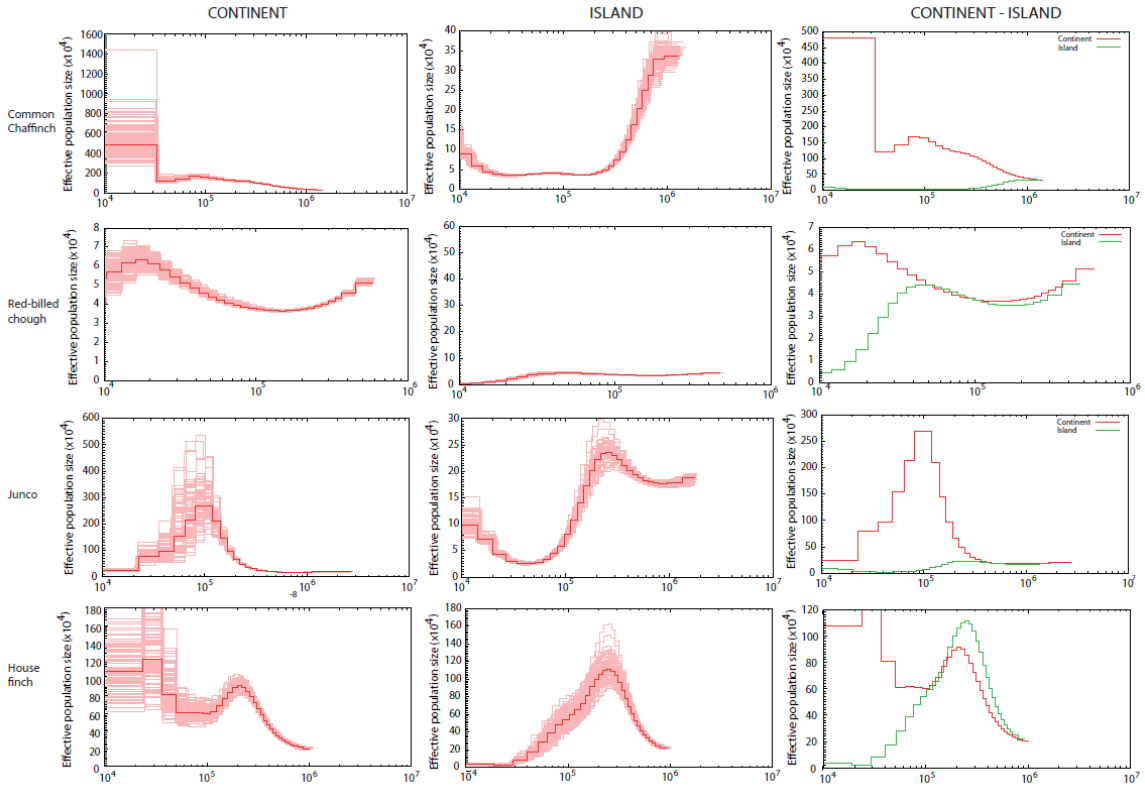


Figure 3. Demographic history of insular and mainland populations. The analysis was performed using Pairwise Sequentially Markovian Coalescent (PSMC). The first two columns represent the demographic inference (dark red line) and 100 bootstrap replicates (lighter red lines) for one individual per treatment and species. The left and right columns correspond to the continent and the island individuals, respectively. In the third column the two plots were combined for a direct comparison, with the red and green lines corresponding to the continental and insular populations, respectively. The point where both lines depart from each other corresponds to the time of colonization, which is around 40,000 y for the red-billed chough, 0.9 my for the common chaffinch, 100,000 y for the house finch and 400,000 for the dark-eye junco. The mutation rate used was of 4.6e-9 mutation/site/generation for all species, and the generation time used in all cases was two years.

Inferring parallel evolution from genome-wide scans

Genome-wide scans showed high heterogeneity across the four target species. The F_{ST} genomic landscapes varied strongly among species (Fig. 4-7). The mean value of F_{ST} was higher in the common chaffinch, followed by the dark-eyed junco, and corresponds to relatively longer divergence times. The red-billed chough showed a slightly higher mean F_{ST} than the house finch (Table 8). The red-billed chough showed a value of genetic diversity that was one order of magnitude lower than the rest, both for insular and mainland populations. The insular common chaffinch population showed the second lowest genetic diversity while the continental population showed the highest diversity value (Table 8). All species showed consistently higher gene content and recombination rates at microchromosomes, and in general, recombination rates were higher at the start and at the end of chromosomes (Fig. 4-7).

Table 8. Divergence and diversity across the genome. Mean values, standard deviation and range of genomic summary statistics for the four species, including: fixation Index ($\boldsymbol{F_{ST}}$), absolute genomic divergence (d_{xy}), and genetic diversity for the insular and the continental populations.

Species	$\boldsymbol{F_{ST}}$ **± sd**	**range**	$\boldsymbol{d_{xy}}$ **± sd**	**range**	$\boldsymbol{\pi_{island}}$ **± sd**	**range**	$\boldsymbol{\pi_{continent}}$ **± sd**	**range**
Red-billed chough	0.21 ± 0.12	[-0.055 - 0.89]	0.0008 ± 0.0004	[0 - 0.15]	0.0005 ± 0.003	[0 - 0.013]	0.0008 ± 0.0004	[0 - 0.020]
House finch	0.14 ± 0.09	[-0.45 - 0.66]	0.006 ± 0.002	[0 - 0.017]	0.0043 ± 0.002	[0 - 0.018]	0.0052 ± 0.002	[0 - 0.016]
Dark-eyed junco	0.26 ± 0.07	[0.006 - 0.68]	0.005 ± 0.002	[0 - 0.023]	0.0022 ± 0.001	[0 - 0.023]	0.0049 ± 0.002	[0 - 0.022]
Common chaffinch	0.40 ± 0.05	[-0.033 - 0.88]	0.009 ± 0.003	[0 - 0.022]	0.0016 ± 0.001	[0 - 0.021]	0.0091 ± 0.003	[0 - 0.023]

Red-billed chough

The red-billed chough genomic landscape shows high levels of relative differentiation across the whole genome with few outlier regions. The mean genetic diversity in both populations is one order of magnitude lower than in the other three species (Table 9), showing very low values across the whole genome except in the microchromosomes, where the genetic diversity and divergence show higher values corresponding with regions of high gene content. However, the XP-EHH analysis revealed evidence of selective sweeps, showing a few clear peaks along the genome that coincide with drops in Tajima's D (Fig.4).

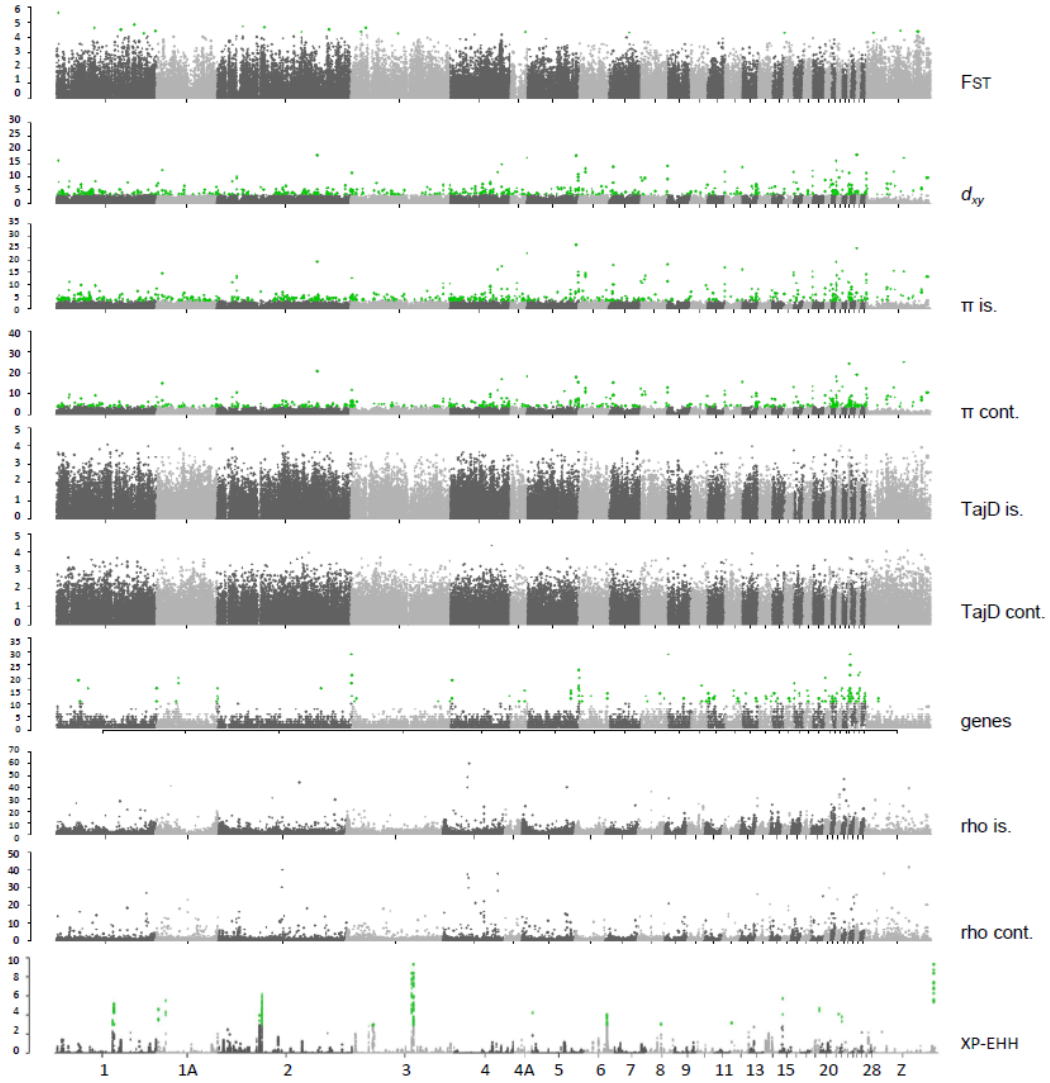


Figure 4. Genomic scans for several summary statistics for the red-billed chough (*Pyrrhocorax pyrrhocorax*). From top to bottom, fixation index (F_{ST}), genomic divergence (d_{xy}), genetic diversity for insular and continental populations (π), Tajima's D for insular and continental populations (TajD), number of genes, recombination rates for insular and mainland populations (rho) and Cross-populations Extended Haplotype Homozygosity (XP-EHH). Chromosome numbers correspond to the Zebra finch genome (*Taeniopygia guttata*, bTaeGut2.pat.W.v2). Green dots represent outliers with the false discovery rate (FDR) set at 0.05 after applying the Benjamini and Hochberg correction, except for the XP-EHH, where the threshold is set at –log10 (*p*-value) ≥ 3.

Common chaffinch

The common chaffinch genome landscape is characterized by several F_{ST} peaks that coincide with valleys in d_{xy} and π, and peaks in Tajima's D mainly in the continent (i.e., peaks in chromosomes 1,1A, 2, 3, 4, 4A, 6, Fig. 5). This pattern is consistent with the

model of recurrent selection, which states that selection in the ancestor previous to the mainland-island split, generates a pattern of low d_{xy} and subsequent selection after divergence reduces genetic diversity generating F_{ST} peaks. The XP-EHH detected selective sweeps mostly concentrated in the microchromosomes and the Z chromosomes; few of them were coincident with F_{ST} peaks.

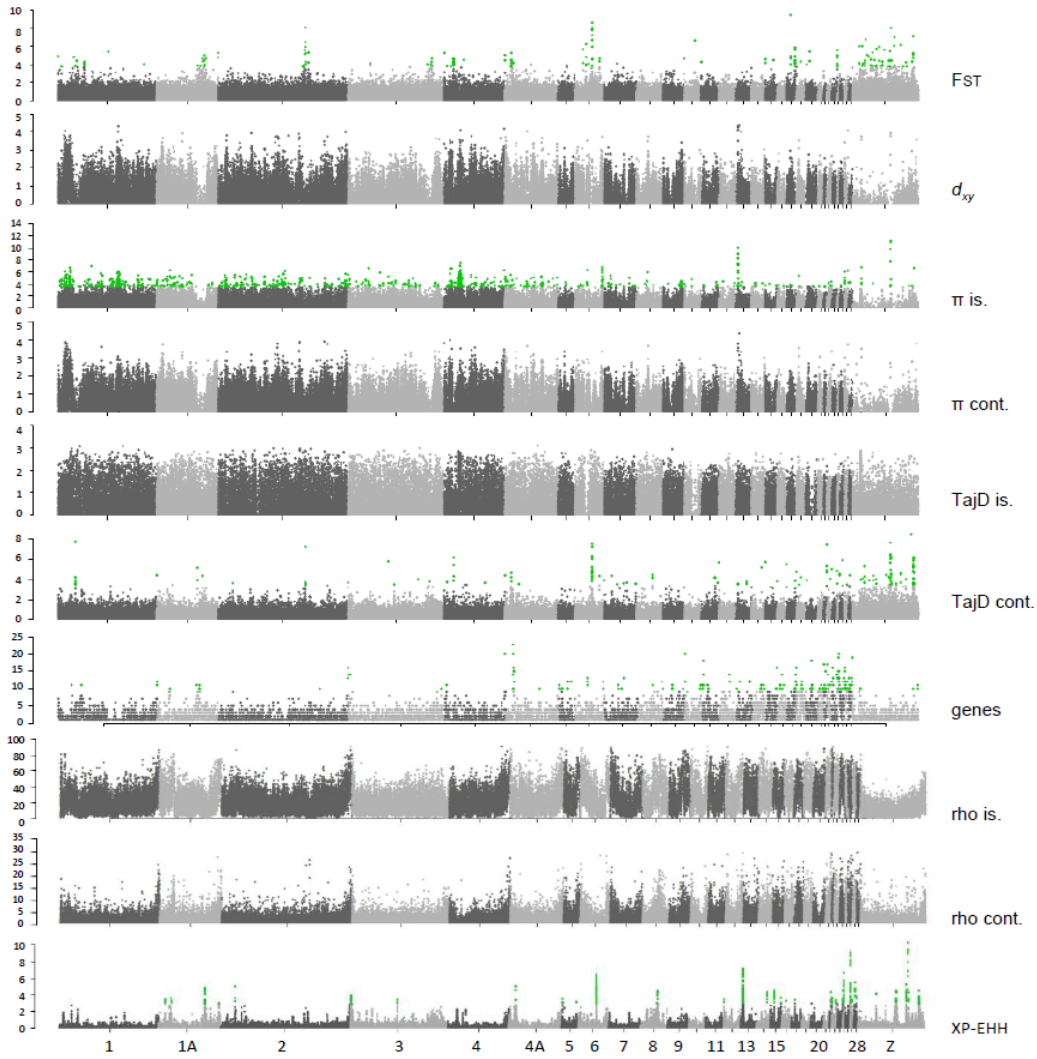


Figure 5. Genomic scans for several summary statistics for the common chaffinch (*Fringilla coelebs*). From top to bottom, Fixation index (F_{ST}), genomic divergence (d_{xy}), genetic diversity for insular and continental populations (π), Tajima's D for insular and continental populations (TajD), number of genes, recombination rates for insular and mainland populations (rho) and Cross-populations Extended Haplotype Homozygosity (XP-EHH). Chromosome numbers correspond to the Zebra finch genome (*Taeniopygia guttata*, bTaeGut2.pat.W.v2). Green dots represent outliers with the false discovery rate (FDR) set at 0.05 after applying the Benjamini and Hochberg correction, except for the XP-EHH, where the threshold is set at –log10 (*p*-value) ≥ 3.

Dark-eyed/Island Junco

The dark-eyed Junco genomic landscape, as in the red-billed chough, is relatively highly differentiated across the whole genome, but there are few outlier genomic regions, which in several cases coincide with chromosomal extremes (Fig. 6). The XP-EHH scans revealed the absence of significant selective sweeps across the genome, with only three small regions detected.

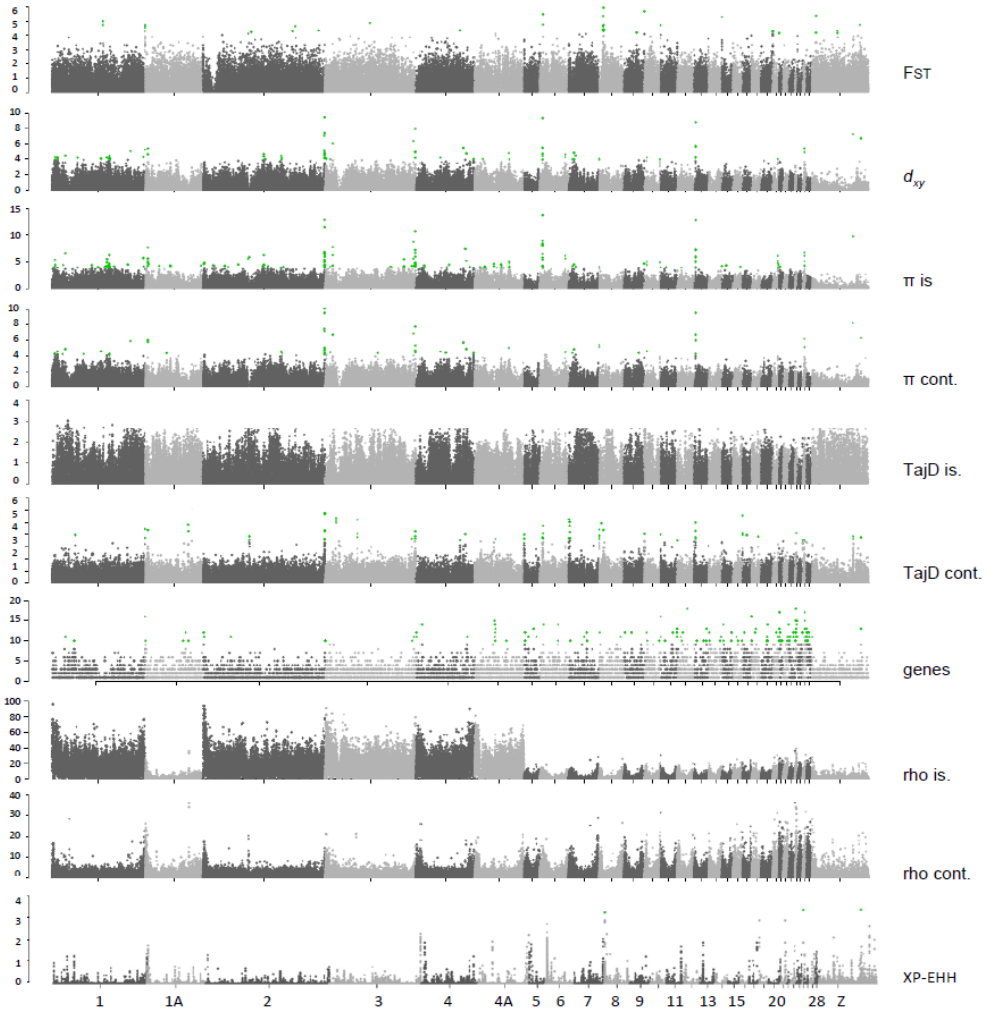


Figure 6. Genomic scans for several summary statistics for the dark-eyed junco (*Junco hyemalis*). From top to bottom, Fixation index (F_{ST}), genomic divergence (d_{xy}), genetic diversity for insular and continental populations (π), Tajima's D for insular and continental populations (TajD), number of genes, recombination rates por insular and mainland populations (rho) and Cross-populations Extended Haplotype Homozygosity (XP-EHH). Chromosome numbers correspond to the Zebra finch genome (*Taeniopygia guttata*, bTaeGut2.pat.W.v2). Green dots represent outliers with the false discovery rate (FDR) set at 0.05 after applying the Benjamini and Hochberg correction, except for the XP-EHH, where the threshold is set at –log10 (*p*-value) ≥ 3.

House finch

The house finch genomic landscape is characterized by a relatively large and highly differentiated region in the middle of chromosome 3, representing 47 million base pairs. It coincides with high values of Tajima's D in the continental population and a region of low recombination, while d_{xy} and π have regular values (Fig. 7). At the end of the same chromosome and at the beginning of chromosome 4, there are two F_{ST} peaks that coincide with a valley in d_{xy} and π, and a peak in Tajima's D. This pattern is consistent with the recurrent selection model. The microchromosomes show high relative differentiation along with high recombination rates and enriched gene content. The XP-EHH scan showed a flat landscape with no evidence for significant selective sweeps.

Detecting putative chromosomal inversions

After combining all possible evidence, the analysis to detect inversions revealed that the red-billed chough genome has no putative inversions. The dark-eyed junco genome showed two possible inverted regions in chromosomes 6 and 7 (Table 9, Fig. S1A) but neither of them was coincided with an F_{ST} outlier region. The common chaffinch genome showed two possible inversions in chromosomes 2 and 4, and both coincided with F_{ST} outlier regions (Table 9, Fig. S1B). The house finch genome revealed five putative inversions, one in chromosome 3, one in chromosome 1A, and three in chromosome 1. (Table 9, Fig. S1C). Only the inversion in chromosome 3 coincides with an F_{ST} outlier region (Fig. 7).

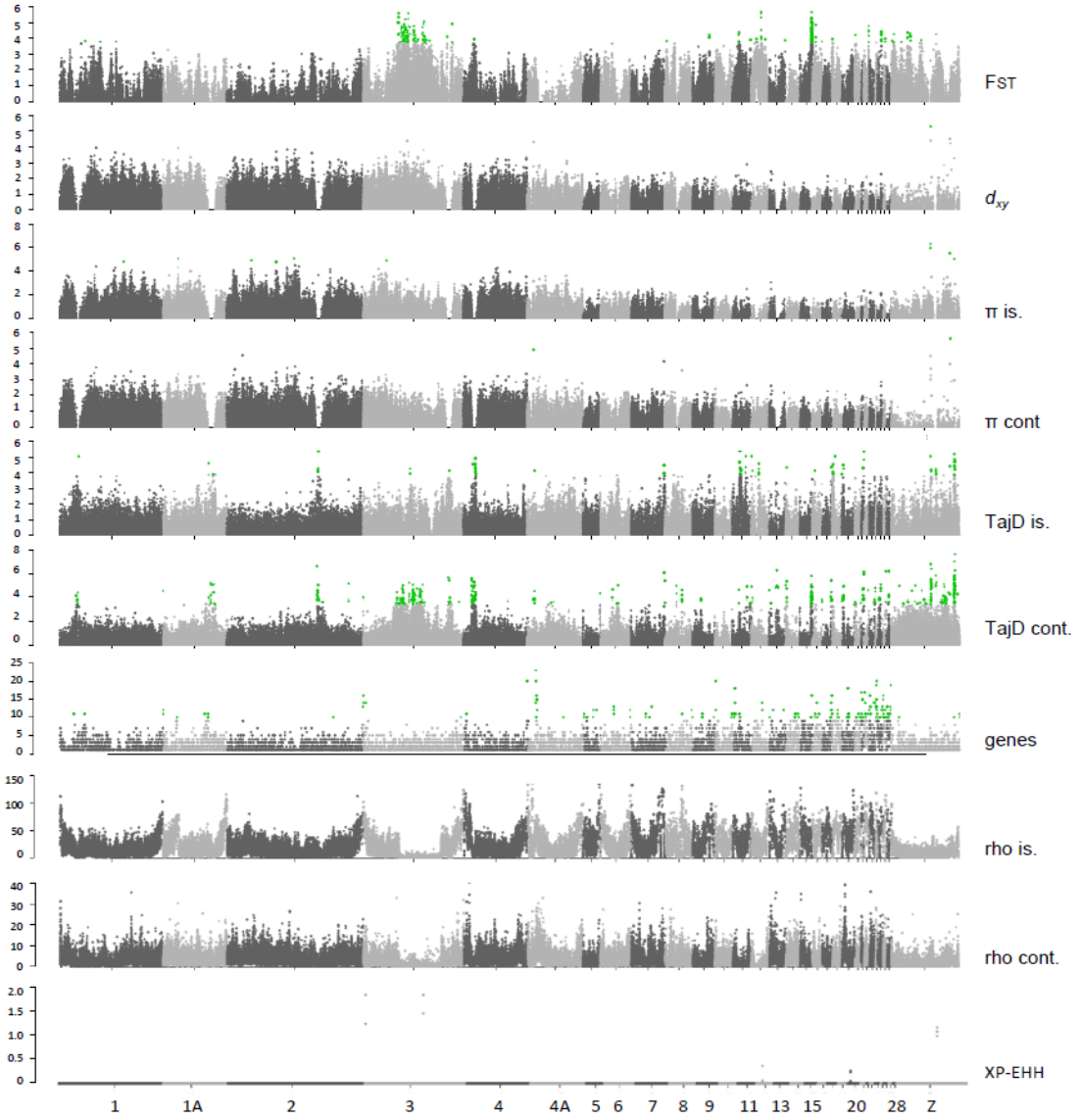


Figure 7. Genomic scans for several summary statistics for the house finch (*Haemorhous mexicanus*). From top to bottom, Fixation index (F_{ST}), genomic divergence (d_{xy}), genetic diversity for insular and continental populations (π), Tajima's D for insular and continental populations (TajD), number of genes, recombination rates por insular and mainland populations (rho) and Cross-populations Extended Haplotype Homozygosity (XP-EHH). Chromosome numbers correspond to the Zebra finch genome (*Taeniopygia guttata*, bTaeGut2.pat.W.v2). Green dots represent outliers with the false discovery rate (FDR) set at 0.05 after applying the Benjamini and Hochberg correction, except for the XP-EHH, where the threshold is set at –log10 (*p*-value) ≥ 3.

Detection of candidate genes and GO-term enrichment analysis

Among all comparisons, there were only two genes putatively under selection that were shared between two species. Among the house finch and the dark-eyed junco, we found the *morc2* gene, and between the dark-eyed junco and the common chaffinch the *spef2* gene. The *morc2* gene is associated with Marie-Tooth Disease, Axonal, Type 2z (CMT2Z) and Developmental Delay, Impaired Growth, Dysmorphic Facies, and Axonal Neuropathy (DIGFAN) diseases. CMT2Z is characterized by distal lower limb muscle weakness and sensory impairment (Vujovic et al., 2021) and DIFGAN by impaired motor and intellectual development, poor overall growth, usually short body height and microcephaly and subtly dysmorphic facial features (Sacoto et al., 2020). The *spef2* gene is involved in sperm development and also plays a role in osteoblast differentiation, being required for normal bone growth (Lehti et al., 2018).

Table 9. Detection of chromosomal inversions. Outlier multidimensional scaling (MDS) values detected by locstruct for the dark-eyed junco, the common chaffinch and the house finch, considered as putative inversions.

MDS	Chr	Scaffold	Start	End	Outlier windows	PC1 Variance (%)	PC2 Variance (%)	Proportion of between clusters sum of squares
Dark-eyed junco								
MDS14	7	ScoVZU6_3251_HRSCAF_3306	217052	1814797	27	47.58	11.78	0.96
MDS30	6	ScoVZU6_982_HRSCAF_1003	11099520	12152631	34	57.15	34.83	0.996
Common chaffinch								
MDS07	4	ScDMRwT_1738_HRSCAF_2009	3482050	17162018	113	41.82	7.56	0.997
MDS15	2	ScDMRwT_473_HRSCAF_551	10279761	125763279	88	44.19	9.27	0.996
House finch								
MDS13	3	ScDMRwT_2155_HRSCAF_2491	32963219	80679350	502	32.34	21.34	0.99
MDS19	1A	ScDMRwT_3155_HRSCAF_3635	20208302	61201276	34	55.76	27.6	0.91
MDS22	1	ScDMRwT_581_HRSCAF_676	17057369	24751704	8	49.45	10.4	0.97
MDS24	1	ScDMRwT_581_HRSCAF_676	87108777	88764953	6	57.37	22.24	0.95
MDS34	1	ScDMRwT_581_HRSCAF_676	74088965	75085134	3	56.51	22.83	0.96

Red-billed chough

The F_{ST} outliers mapped to 19 genes and the XP-EHH outliers detected selective sweeps in 14 genes, without overlap between the two methods. Due to the high relative differentiation across the genome and the absence of clear F_{ST} peaks, we believe that the clear selective sweeps along the genome are a better approach to

detect candidates for the red-billed chough. However, most of the genes among the 14 outlier genes found within sweeps have unknown function, and only five genes have known function and had GO terms associated. From the 20,580 available genes from the *Corvus moneduloides* genome, 8,632 from the gene universe (including the five significant genes) could be used for the analysis. Among the top-10 GO terms for the XP-EHH outliers we found several related to chromatin cohesion (i.e., regulation of cohesion loading and negative regulation of sister chromatid cohesion) (Table 10).

Table 10. Top ten GO terms associated to 5 out of the 14 outlier genes detected by XP-EHH in the island-mainland red-billed chough comparison. The analysis was performed using the topGO package with the weight01 algorithm and the fisher test under the ontology biological process. The table includes the GO term Accession (GO.ID), the term name, the number of annotated genes (Annot.), the number of significant genes (Sign.), the number of expected genes (Exp.) and the *p*-value of the Fisher test using the weight01 algorithm.

GO.ID	Term name	Annot.	Sign.	Exp.	Fisher
GO:0070130	negative regulation of mitochondrial translation	1	1	0	0.00058
GO:0002760	positive regulation of antimicrobial humoral response	1	1	0	0.00058
GO:0071922	regulation of cohesin loading	2	1	0	0.00116
GO:0072573	tolerance induction to lipopolysaccharide	2	1	0	0.00116
GO:0045875	negative regulation of sister chromatid cohesion	3	1	0	0.00174
GO:0060623	regulation of chromosome condensation	3	1	0	0.00174
GO:0035562	negative regulation of chromatin binding	5	1	0	0.00289
GO:0008156	negative regulation of DNA replication	10	1	0.01	0.00578
GO:0042273	ribosomal large subunit biogenesis	23	1	0.01	0.01325
GO:0048146	positive regulation of fibroblast proliferation	25	1	0.01	0.0144

Common chaffinch

The genomic scan detected 85 genes in the F_{ST} outlier regions, and the XP-EHH revealed 1,724 outliers that mapped to 21 genes, of which 3 were shared with the F_{ST} candidates. Among the 16,563 available in the gene universe, the GO term analysis detected 9,065 feasible genes, including 48 out of the 85 significant genes detected as F_{ST} outliers. Among the top-10 GO terms we found several involved in transcription regulation such as "transcription-dependent tethering of RNA polymerase II gene DNA at nuclear periphery" and "histone H3-K4 acetylation" and others affecting translation like "lysyl-tRNA aminoacylation" (Table 11). There were also two terms related to cell

adhesion “regulation of protein localization to cell-cell adherens junction” and “regulation of focal adhesion assembly” and also two terms associated with the organization or cellular components including: “positive regulation of endosome organization” and “lysosome localization”.

Table 11. Top ten GO terms associated with 54 out of the 85 outlier genes detected by the F_{ST} scan in the island-mainland common chaffinch comparison. The analysis was performed using the topGO package with the weight01 algorithm and the fisher test under the ontology biological process. The table includes the GO term Accession (GO.ID), the term name, the number of annotated genes (Annot.), the number of significant genes (Sign.), the number of expected genes (Exp.) and the *p*-value of the Fisher test using the weight01 algorithm.

GO.ID	Term name	Annot.	Sign.	Exp.	Fisher
GO:0007213	G protein-coupled acetylcholine receptor signaling pathway	7	2	0.04	0.00043
GO:0090668	endothelial cell chemotaxis to vascular endothelial growth factor	1	1	0.01	0.00463
GO:0043973	histone H3-K4 acetylation	1	1	0.01	0.00463
GO:1904980	positive regulation of endosome organization	1	1	0.01	0.00463
GO:1904702	regulation of protein localization to cell-cell adherens junction	1	1	0.01	0.00463
GO:0006430	lysyl-tRNA aminoacylation	1	1	0.01	0.00463
GO:0000972	transcription-dependent tethering of RNA polymerase II gene DNA at nuclear periphery	1	1	0.01	0.00463
GO:0032418	lysosome localization	26	2	0.14	0.00635
GO:0051893	regulation of focal adhesion assembly	29	2	0.15	0.00786
GO:0015966	diadenosine tetraphosphate biosynthetic process	2	1	0.01	0.00925

Dark-eyed junco

The F_{ST} genome scan detected few peaks distributed across the genome that mapped to 24 genes, and three regions were detected as sweeps by the XP-EHH scan but did not contain known genes. Among the 24 genes, only 16 had GO terms associated with them. The GO enrichment analysis performed with a gene universe of 17,038 genes found 9,220 feasible genes including 15 significant genes. The top-10 GO terms revealed three terms related to the centrosome, including “negative regulation of protein localization to centrosome”, “protein localization to pericentriolar material” and “positive regulation of mitotic centrosome separation” (Table 12).

Table 12. Top ten GO terms associated with 16 out of the 24 outlier genes detected by the F_{ST} scan in the island-mainland dark-eyed junco comparison. The analysis was performed using the topGO package with the weight01 algorithm and the fisher test under the ontology biological process. The table includes the GO term Accession (GO.ID), the term name, the number of annotated genes (Annot.), the number of significant genes (Sign.), the number of expected genes (Exp.) and the *p*-value of the Fisher test using the weight01 algorithm.

GO.ID	Term name	Annot.	Sign.	Exp.	Fisher
GO:1905793	protein localization to pericentriolar material	2	1	0	0.0028
GO:0006104	succinyl-CoA metabolic process	2	1	0	0.0028
GO:0048692	negative regulation of axon extension involved in regeneration	2	1	0	0.0028
GO:1904780	negative regulation of protein localization to centrosome	2	1	0	0.0028
GO:0045869	negative regulation of single stranded viral RNA replication via double stranded DNA intermediate	2	1	0	0.0028
GO:0090309	positive regulation of methylation-dependent chromatin silencing	2	1	0	0.0028
GO:0061034	olfactory bulb mitral cell layer development	2	1	0	0.0028
GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline	3	1	0	0.0042
GO:0046604	positive regulation of mitotic centrosome separation	3	1	0	0.0042
GO:0051968	positive regulation of synaptic transmission, glutamatergic	3	1	0	0.0042

House finch

The F_{ST} genome scan detected 111 genes putatively under selection, while the XP-EHH scan detected no outliers. From the genes identified under selection, 20 were clustered in the middle region of chromosome 3, and two were at the end of the same chromosome. The remaining genes were mainly clustered at the microchromosomes. From the 16,563 available genes in the gene universe, 9,065 including 83 significant genes could be used in the GO enrichment analysis. Within the top 10 significant GO terms (Table 13) we find "growth plate cartilage chondrocyte morphogenesis" which is involved in the skeletal development and morphogenesis. Likewise, involved in morphogenesis we find the term "zonula adherens maintenance" which is related to cell-cell adhesion. We also find several terms associated with transcription, including "negative regulation of telomerase RNA reverse transcriptase activity", "glutaminyl-tRNA aminoacylation" and two histone acetylations (H2-K14 and H3-K23).

Table 13. Top ten GO terms associated with 79 out of the 111 outlier genes detected by the F_{ST} scan in the island-mainland house finch comparison. The analysis was performed using the topGO package with the weight01 algorithm and the fisher test under the ontology biological process. The table includes the GO term Accession (GO.ID), the term name, the number of annotated genes (Annot.), the number of significant genes (Sign.), the number of expected genes (Exp.) and the *p*-value of the Fisher test using the weight01 algorithm.

GO.ID	Term	Annot.	Sign.	Exp.	Fisher
GO:0035873	lactate transmembrane transport	5	3	0.05	0.00032
GO:0044154	histone H3-K14 acetylation	9	2	0.08	0.00193
GO:0043972	histone H3-K23 acetylation	1	1	0.01	0.0075
GO:0045218	zonula adherens maintenance	1	1	0.01	0.0075
GO:0072274	metanephric glomerular basement membrane development	1	1	0.01	0.0075
GO:2000845	positive regulation of testosterone secretion	1	1	0.01	0.0075
GO:0003429	growth plate cartilage chondrocyte morphogenesis	1	1	0.01	0.0075
GO:1905662	negative regulation of telomerase RNA reverse transcriptase activity	1	1	0.01	0.0075
GO:0006425	glutaminyl-tRNA aminoacylation	1	1	0.01	0.0075
GO:0090258	negative regulation of mitochondrial fission	1	1	0.01	0.0075

Discussion

Our comparative analysis of mainland and insular populations of four passerine species yielded shared patterns of phenotypic divergence and demographic history, in contrast to species-specific patterns of genome-wide variation. Relative to the mainland, all insular populations showed increases in body size, and suffered reductions in effective population size and lower levels of genetic diversity, which is consistent with previous findings (Frankham, 1997; Leroy et al., 2021; Benitez et al., 2021). Island colonizations are usually initiated by a small group of individuals, and that drift, combined with the small size of the island's geographic area, leads to a small effective population size and low genetic diversity (Frankham, 1995). Among the four species, the red-billed chough showed the smallest effective population size in both insular and mainland populations, which corresponds to the lowest levels of genetic diversity. In the mainland, this species has shown marked levels of genetic structure in the absence of geographic barriers, suggesting that social barriers due to complex behavioral interactions may constrain gene flow and the effective size of local populations (Morinha et al., 2017); the insular population is unlikely to be an exception (Morinha et al., 2020).

Considering mean differences in tarsus length, a proxy for structural body size in birds (Rising & Somers, 1989; Freeman & Jackson, 1990; Senar & Pascual, 1997), we detected that the three small passerines increase in size upon island colonization. This is consistent with the island rule, which posits that small birds evolve towards a larger size and large birds towards a smaller size (Benitez et al., 2021). However, the difference in the house finch tarsus length among insular and mainland populations was not significant probably due to the small sample size. Regarding beak size, we find that insular individuals from the small sized and short-billed species show longer bills than their mainland counterparts. The beak is both a feeding and thermoregulatory structure with great evolutionary potential that allows birds to quickly adapt to new environmental conditions (Grant & Grant, 2008) and therefore plays a fundamental role in avian fitness (Boag & Grant, 1981; Price et al., 1984; Grant, 1986; Gibbs & Grant, 1987; Tattersall et al., 2017; Gamboa et al., 2022). Generally, island colonization is accompanied by dietary shifts, as individuals have to adapt to the available, and usually scarcer, insular trophic resources (Whittaker et al., 2017; Price, 2008). Stable isotope ratios of ^{13}C and ^{15}N from the feathers, which we used as proxies of diet, indicate that the three smaller species show differences in at least one of the isotopes between mainland and insular populations, suggesting dietary shifts. We did not find significant differences in feather isotope ratios in the red-billed chough, although it has been already documented that the insular red-billed choughs feed mainly on vegetal food and fruits, whereas the continental red-billed choughs feed mostly on insects (Blanco et al., 2014).

A major question in evolutionary biology is how recurrently natural selection ends in the same adaptive phenotypic changes, and whether these changes are the result of similar or different genetic routes (Morris, 2010; Blount et al., 2018). Finding shared patterns of genomic variation either at the intra- or interspecific levels has been of major interest to understand the mechanisms underlying local adaptation (Burri et al., 2015; Van Doren et al., 2017; Delmore et al., 2018). These shared regions across taxa are particularly interesting when the differentiation happened independently and not due to common genomic features (Seehausen et al., 2014). Background selection plays an important role in generating parallel genomic landscapes across species (Burri, 2017), especially in taxa like birds, where gene content distribution and

recombination landscapes are well conserved across broad evolutionary times (Singhal et al., 2015). A striking result of our comparative analysis of four passerine species is the lack of parallelism in their respective genomic landscapes. We found highly differentiated genomic regions in all four species that were associated with reduced genetic diversity, suggesting the role of selection. Yet, the lack of congruence in the location of these regions indicates that the four species adapted to insular environments in different ways, through changes in different loci. Moreover, patterns of recombination rate in these regions suggest that the mechanisms generating these patterns differ in each of the four species, and include selective sweeps caused by directional selection, chromosomal inversions, and historical factors like recurrent selection, even among species with similar divergence times.

The two species that diverged more recently showed very different genomic patterns. The red-billed chough is characterized by the inflated values of F_{ST} along the genome, while showing very low absolute divergence and genetic diversity. In contrast, the house finch showed the lowest relative divergence and normal levels of absolute divergence and genetic diversity. The divergence between the island and mainland populations in the house finch seems to be driven mainly by a chromosomal inversion in chromosome 3, and lacks signatures of selective sweeps. In contrast, the red-billed chough revealed several selective sweeps along the genome, which are consistent with recent divergence. The pair of species that had longer divergence times also showed very different landscapes that appear to have been shaped by different processes. The common chaffinch genomic landscape is characterized mainly by the model of recurrent selection (Irwin et al., 2016), whereas the dark-eyed junco landscape does not reveal a clear pattern consistent with a particular model because the sweeps are not significant. However, there is a trend of peaks of differentiation at the chromosome extremes. Delmore et al., (2015) found a similar pattern between inland and coastal Swainson's thrushes and concluded that they probably corresponded to telocentric chromosome centromere regions, which show highly repetitive DNA.

According to our demographic analysis, the divergence between red-billed choughs on La Palma and the Iberian Peninsula took place around 30,000 years ago, considering a generation time of two years. A previous study (Morinha et al., 2020) estimated the

divergence event in a similar time range, within the last 10,000 years using mitochondrial data and around 30,000 years using iMA2, however they used a generation time of 6 years. If we apply that value, the divergence time estimate changes to around 110,000 years. The red-billed chough also shows the smallest effective population size that also corresponds with the low genetic diversity, which is one order of magnitude lower than the rest of the analysed species. This reduced genetic diversity also results in an inflation of the relative divergence (Charlesworth, 1998; Cruickshank & Hahn, 2014), causing a high baseline to detect outliers while the absolute divergence remains low. The recent divergence of the red-billed chough is apparent due to the low divergence along the genome with a mean d_{xy} value of $8.2 \cdot 10^{-4}$. The regions of higher divergence and genetic diversity are located in the microchromosomes, which have relatively higher recombination rates and higher gene content (Burt, 2002). However, the scan for selective sweeps, which is more efficient in detecting recent divergence, revealed clear peaks along the genome. The red-billed chough is the species showing clearer significant selective sweeps, which is also consistent with the low genetic diversity of the species due to genomic hitchhiking of the sites around the selected loci (Kaplan et al., 1989). The GO term analysis of the genes detected within the selective sweeps revealed several terms among the top ten related with chromatin cohesion. Specifically, among the five known outlier genes, the WAPL gene negatively regulates the association of cohesin with chromatin, having an opposing function to the NIPBL gene. Mutations in the NIPBL gene cause Cornelia de Lange syndrome (CdLS), therefore, mutations in WAPL gene could generate similar developmental deficits to CdLS (Dorsett & Krantz, 2009). CdLS can affect most organ systems, but typical characters include craniofacial structures, upper extremities, eyes and the gastrointestinal system (Jackson et al., 1993; Bhuiyan et al., 2006). The actual role of WAPL has not been properly tested, but it has been associated with Warsaw Breakage Syndrome (WABS)(Faramarz et al., 2019), which is a cohesinopathy that causes growth retardation, severe microcephaly, sensorineural hearing loss, cochlear anomalies, intellectual disability and abnormal skin pigmentation (Alkhunaizi et al., 2018; Faramarz et al., 2019).

In the common chaffinch, the demographic analysis estimate of divergence time between mainland and island populations was around 0.8-0.9 my, which coincides

with previous reconstructions of the species evolutionary history. A study of the entire common chaffinch radiation across the Atlantic archipelagos revealed that it started in Azores, then Madeira and finally the Canary Islands (Chapter II: Recuerda et al., 2021a). This sequential colonization of isolated archipelagos has left a genomic signature of recurrent selection along the genome, leading to regions with low absolute divergence due to selection in the ancestor, that were subsequently selected in the daughter populations, reducing genetic diversity and generating a F_{ST} peak (Irwin et al., 2016). This recurrent-selection model fits perfectly with the known colonization history, as the first selective episode probably occurred upon colonization of the Azores, and then at every subsequent colonization step between islands, there were successive selective events at the same genomic regions, leading to an increasing loss of genetic diversity. Among the outlier genes there were several involved in metabolism (i.e., *fabp2*, *kars1*, *lipa*, *nfrkb*, *pdha1*), five involved in pigmentation and six related to singing. Among the genes related to pigmentation, three were several related to avian plumage coloration, *ap3b1*(Ren et al., 2021), *hps6* (Domyan et al., 2019) and *ric1* (Bruders et al., 2020), one was related to sexual dichromatism in birds (Gazda, 2019), and the *atrn* gene was related to melanin production and has also been associated with coat coloration in macaques (Bradley et al., 2013). Regarding the genes related to song, we detected, *chrm2* and *chrm5*, which have shown differential expression associated with song learning and production in zebra finch (Osogwa, 2018), the mrps27 (Qi et al., 2012) and *upf3b* (Shi et al., 2021), which are involved in the song control system in the zebra finch, the *paip1* gene, which has been associated with song learning (Lovell et al., 2008), and the *ube2d3* gene, which was related to musical abilities using a convergent evidence method including data from humans, songbirds and other animals (Oikkonen et al., 2016). Interestingly, within the top-ten significant GO terms we detect "positive regulation of endosome organization" and endosomes play an important role in neural development (Yap & Winckler, 2012). We also find the term "regulation of focal adhesion assembly" and it has been shown that cell adhesion plays an important role in tissue morphogenesis (Harris & Tepass, 2010).

In the dark-eyed junco, the demographic inference revealed that the insular population diverged around 400,000 years, which is similar to previous estimates

(Aleixandre et al., 2013). The differentiated regions were mainly distributed at the ends of chromosomes, coinciding with telocentric centromeres, as previously found in Swainson's thrushes by Delmore et al., (2015). Consistent with this pattern, among the top-ten GO terms we identified several that were related to the centrosomes. An increasing number of studies recognize the centrosome as a signalling machine that is capable of regulating many cellular functions (Doxsey,McCollum & Theurkauf, 2005; Doxey, Zimmerman & Mikule, 2005).

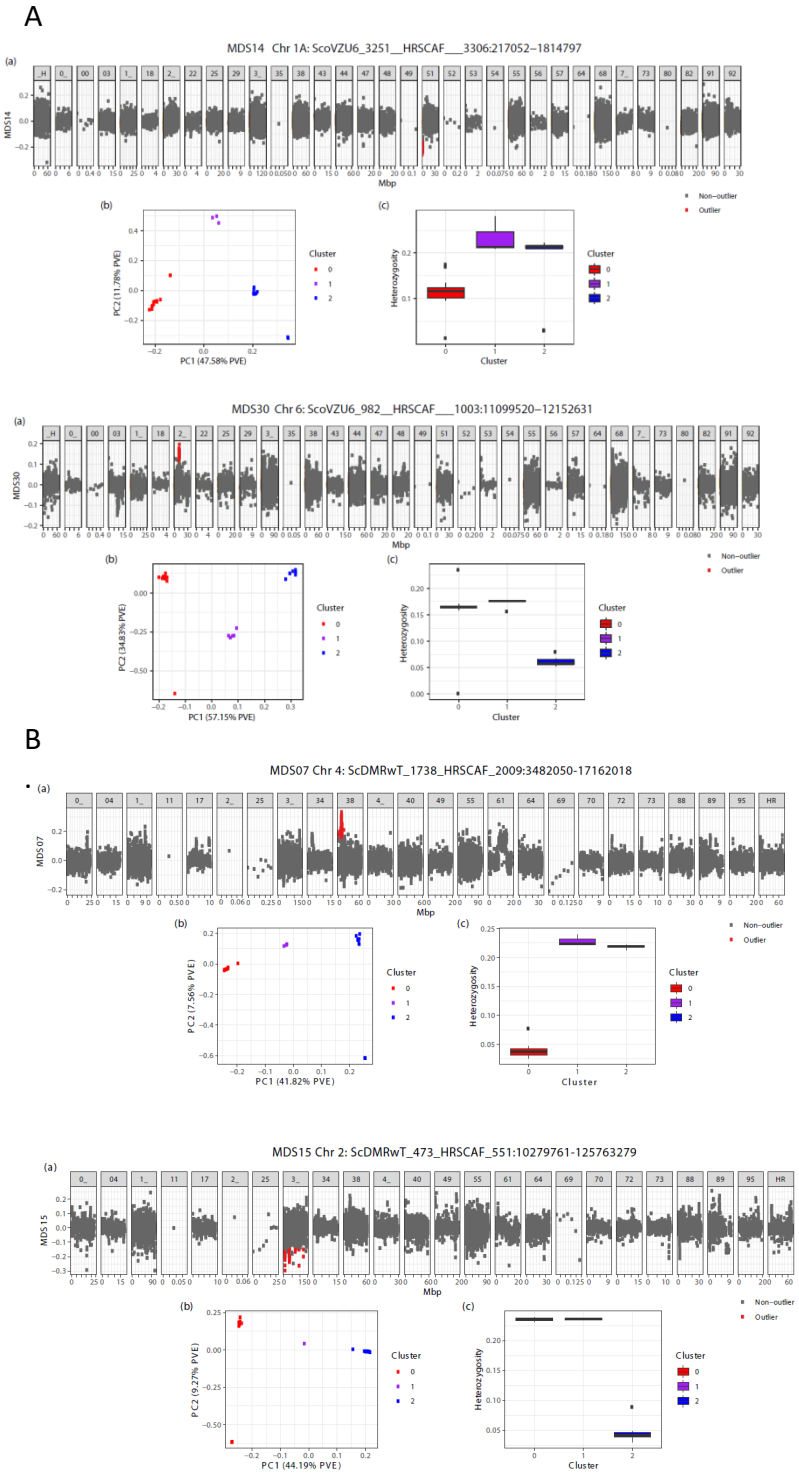
In the house finch, the genomic landscape showed signatures of different processes. Even though the demographic inference revealed a recent divergence time of about 40,000 years, we did no detect signatures of significant selective sweeps. The large region showing high differentiation and very low recombination in chromosome 3 represents a putative chromosomal inversion. Genomic islands of differentiation could be generated by chromosomal rearrangements that cluster highly differentiated loci together due to genomic hitchhiking (Yeaman, 2013; Huang et al., 2020). However, that could represent either a group of adaptive alleles or several neutral loci linked to a focal selected allele (Yeaman, 2013). Several studies have found regions highly diverged within chromosomal inversions (Hoffmann et al., 2004; Ayala et al., 2014; Christmas et al., 2019; Huang et al., 2020). In this case, 20 genes putatively under selection were found within the inversion. Two of those genes (*fam162b* and *fig4*) are related to facial morphology or disorders affecting it. Little is known about the function of the *fam162b* gene, but it is expressed in mouse facial prominences (Feng et al., 2009), and the *fig4* gene is associated with the Yunis-Varon syndrome, characterized by skeletal defects including cleidocranial dysplasia, digital anomalies and neurological impairment (Campeau et al., 2013). Another interesting candidate is the *Iyd* gene, which is also within an inversion in chromosome 2 in the white-throated sparrow (*Zonotrichia albicollis*) and has shown differences in expression between two morphs that differed in territorial aggression including song (Zinzow-Kramer et al., 2015). Within that inversion, they found mainly genes related to behavior and plumage color. Some genes within the inversion in the house finch are related to mental retardation including FMN2 (Law et al., 2014; Gorukmez et al., 2020), or to behavior in mice, like *pnisr* (Moloney et al., 2019). Interestingly, within the house finch inversion we also found the gene *gtf3c6*, which was found to be a candidate

involved in sexual selection in a comparison of 11 bird genomes (Jaiswal et al., 2021). Within the top-ten significant GO terms, we found "growth plate cartilage chondrocyte morphogenesis", which is involved in skeletal development and morphogenesis and regulated by multiple signalling pathways including, among others, the bone morphogenetic proteins (Bmp; De Luca et al., 2001), fibroblast growth factors (FGFs; Deng et al., 1996) and Wingless/int.1 molecules (Wnt; Yang et al., 2003). Among these pathways, the Bmp and Wnt signalling pathways are known to be involved in facial development in different organisms including beak morphology in birds (Abzhanov, 2004; Brugmann et al., 2010). We also found the term "zonula adherens maintenance" and it has been shown that the adherens junctions are also involved in tissue morphogenesis (Harris & Tepass, 2010).

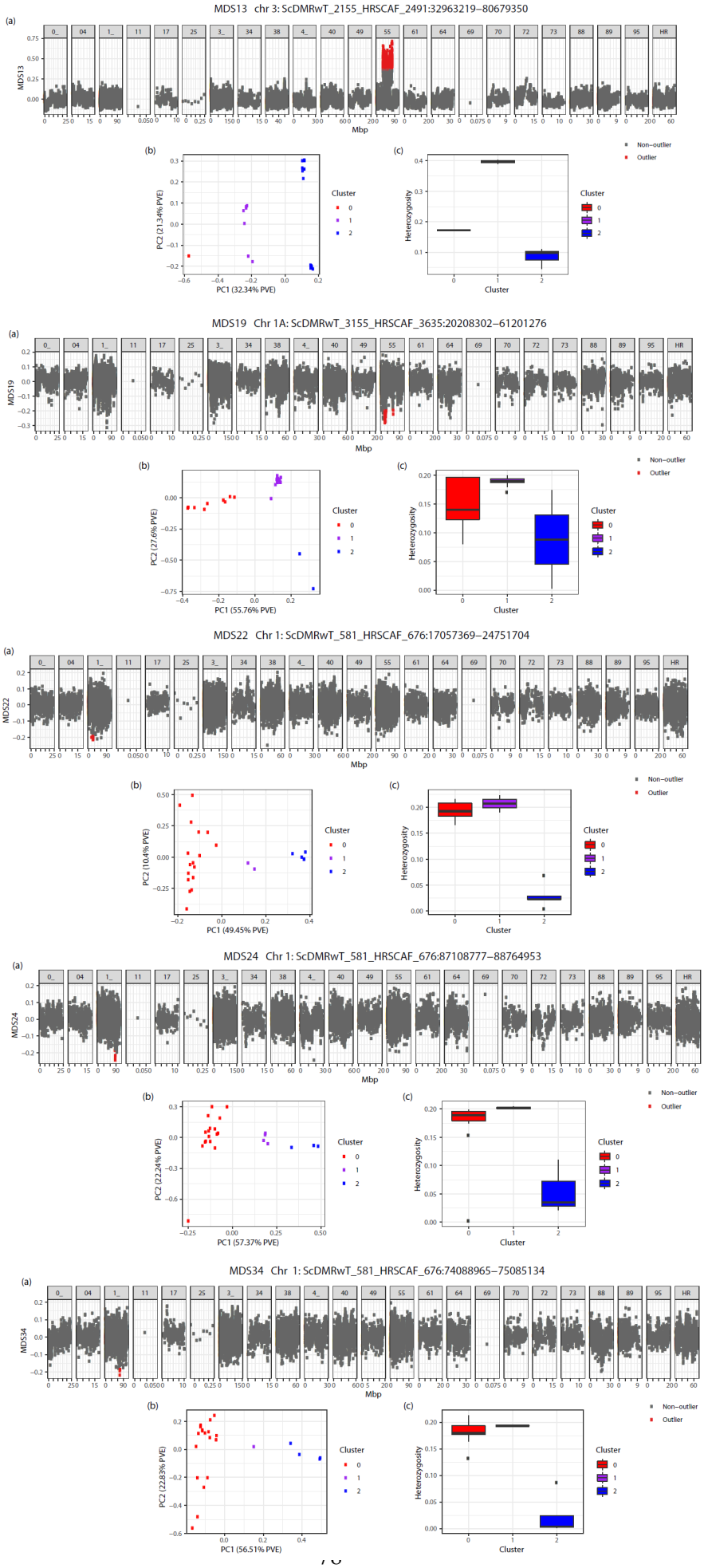
Different lineages that suffer similar selective pressures or have similar functional requirements can evolve similar morphological structures independently. We studied if four different species of birds that have colonized oceanic islands, and therefore were subjected to similar selective pressures, have undergone similar phenotypic changes and if the underlying genetic mechanisms were also shared. We detected that highly differentiated regions of the genome are lineage specific, which suggests that the genetic basis of phenotypic divergence is different in each species (Van Doren et al., 2017). Even if the same regions had been detected as putatively under selection or with shared genomic features involved in genomic differentiation, such as the stable recombination landscape in avian lineages (Singhal et al., 2015), it would be difficult to determine whether that pattern is generated by background and linked selection or by directional selection. Moreover, due to the polygenic nature of most adaptive traits, such as the bill, it is difficult to find the same genes involved because selection is likely to act on many loci of small effect (Pritchard & Di Rienzo, 2010; Bosse et al., 2017), so as that convergent phenotypes could be due to divergent genotypes. Several examples show that phenotypic change in a given trait can be driven by different sets of genes, such as mouth morphology in cichlid fishes (Elmer et al., 2014), or color pattern in mice (Hoekstra et al., 2006; Steiner et al., 2009) and flies (Wittkopp et al., 2003). Even though the outlier genes differ among species, there could be common significant GO terms because different genes share functions and pathways. Interestingly, between the common chaffinch and the house finch we found several similar GO terms related to tRNA aminoacylation, histone acetylations and cell adherens junctions. Remarkably,

we found that in all four species GO terms are mostly related to gene regulation, for instance by modifying histones or altering chromatin binding and chromosome condensation, which are essential for differentiation and development. Recently, Monroe et al., (2021) reported that mutations occur less often in functional regions of the genome, and that epigenomic and physical chromosomal features account for the position of the mutations. In our case, most of the terms related to outlier loci are involved in epigenetic modifications, suggesting that changes in gene regulation, instead of specific core genes, may be the main drivers of divergence. Currently, several models are being developed to understand the role of gene regulation in the evolution of complex traits (Boyle et al., 2017; Liu et al., 2019), implying that regulatory regions are disproportionately targeted by polygenic selection, highlighting the key role of gene regulatory networks in evolution (Fagny & Austerlitz, 2021).

Figure S1. Detection of chromosomal inversions using multidimensional scaling analysis (MDS). Information about the MDS outlier regions detected with LOSTRUCT for (A) the dark-eyed junco, (B) the common chaffinch and (C) the house finch. Within each species, (a) genome plot of MDS values for the chromosomes were inversions are detected, where each dot represents a window of 1,000 SNPs and outliers are in shown in red; (b) PCA performed with SNPs in the specific region. The three clusters correspond to the two homozygote groups (blue and red) and the heterozygote group (purple); and (c) heterozygosity values for each of the groups identified in the PCA.



C.


MDS13 chr 3: ScDMRwT_2155_HRSCAF_2491:32963219–80679350
MDS19 Chr 1A: ScDMRwT_3155_HRSCAF_3635:20208302–61201276
MDS22 Chr 1: ScDMRwT_581_HRSCAF_676:17057369–24751704
MDS24 Chr 1: ScDMRwT_581_HRSCAF_676:87108777–88764953
MDS34 Chr 1: ScDMRwT_581_HRSCAF_676:74088965–75085134
(a)
(b)
(c)
Mbp
Non−outlier
Outlier
Cluster
Heterozygosity
PC1 (32.34% PVE)
PC2 (21.34% PVE)
PC1 (55.76% PVE)
PC2 (27.6% PVE)
PC1 (49.45% PVE)
PC2 (10.4% PVE)
PC1 (57.37% PVE)
PC2 (22.24% PVE)
PC1 (56.51% PVE)
PC2 (22.83% PVE)

CHAPTER II: Sequential colonization of oceanic archipelagos led to a species-level radiation in the common chaffinch complex (Aves: *Fringilla coelebs*)

María Recuerda, Juan Carlos Illera, Guillermo Blanco, Rafael Zardoya, Borja Milá

Recuerda, M., Illera, J. C., Blanco, G., Zardoya, R., & Milá, B. (2021). Sequential colonization of oceanic archipelagos led to a species-level radiation in the common chaffinch complex (Aves: *Fringilla coelebs*). *Molecular Phylogenetics and Evolution*, *164*, 107291.

Pictures by:
F.c.coelebs coelebs : Miguel Montoro Peinado.
F.c.coelebs africana : José Ardaiz Ganuza
F.c.coelebs moreletti : Oleg Nomad/Flickr/ Permission.
F.c.coelebs maderensis : Victor Guimera O'Dogherty
F.c.coelebs palmae : Fred Hoorn. License: CC-BY-NC-ND.
All pictures are modified.

Abstract

Oceanic archipelagos are excellent systems for studying speciation, yet inference of evolutionary process requires that the colonization history of island organisms be known with accuracy. Here, we used phylogenomics and patterns of genetic diversity to infer the sequence and timing of colonization of Macaronesia by mainland common chaffinches (*Fringilla coelebs*), and assessed whether colonization of the different archipelagos has resulted in a species-level radiation. To reconstruct the evolutionary history of the complex we generated a molecular phylogeny based on genome-wide SNP loci obtained from genotyping-by-sequencing, we ran ancestral range biogeographic analyses, and assessed fine-scale genetic structure between and within archipelagos using admixture analysis. To test for a species-level radiation, we applied a probabilistic tree-based species delimitation method (mPTP) and an integrative taxonomy approach including phenotypic differences. Results revealed a circuitous colonization pathway in Macaronesia, from the mainland to the Azores, followed by Madeira, and finally the Canary Islands. The Azores showed surprisingly high genetic diversity, similar to that found on the mainland, and the other archipelagos showed the expected sequential loss of genetic diversity. Species delimitation methods supported the existence of several species within the complex. We conclude that the common chaffinch underwent a rapid radiation across Macaronesia that was driven by the sequential colonization of the different archipelagos, resulting in phenotypically and genetically distinct, independent evolutionary lineages. We recommend a taxonomic revision of the complex that takes into account its genetic and phenotypic diversity.

Keywords: islands, phylogenomics, phylogeography, speciation, species delimitation, systematics.

Introduction

Oceanic archipelagos are excellent model systems to study evolution and have been crucial in advancing our understanding of species diversification and ecosystem assembly processes (Emerson, 2002, Losos and Ricklefs, 2009, Warren et al., 2015, Patiño et al., 2017, Whittaker et al., 2017). According to island biogeography theory, the number of species that can colonize and thrive on an oceanic island is a dynamic process primarily determined by the size of the island and its distance from the mainland (MacArthur and Wilson, 1967, Valente et al., 2017, 2020). Upon arrival, the original colonizers would start diverging from their mainland ancestors through neutral and/or selective processes (Warren et al., 2015). In many cases, the colonization of an archipelago is accompanied by an acceleration of net diversification rates (e.g., Delmore et al., 2020). This leads to species radiations in which phenotypic diversification could be driven either by adaptation to vacant ecological niches and available resources in the different islands (Schluter, 2000, Grant & Grant, 2008, Blanco et al., 2014), or by genetic drift and sexual selection in geographic isolation (Rundell and Price, 2009), although both types of processes can be at work within a single radiation (Gillespie et al., 2020).

Although evolutionary history is often simplified in oceanic archipelagos relative to continents, island colonization can be a complex process that can include multiple colonization and extinction events, back colonizations, as well as the maintenance of gene flow within and between archipelagos, and even with the continent (Illera et al., 2012; Morinha et al., 2020). When inferring the colonization history of oceanic archipelagos, it has been usually assumed that the original settlers originated from the closest mainland area (Grant, 1979, Thornton, 2007), subsequently following a chronological sequence of colonization consistent with a "stepping-stone model" (Funk & Wegner, 1995, Juan et al., 2000, Beheregaray et al., 2004, VanderWerf et al., 2010). However, this basic model is one of many possible ones (Sanmartín, 2008), and molecular phylogenetic analyses using exhaustive regional sampling are increasingly reporting counterintuitive colonization routes, suggesting that long distance migration events could be disrupted by a diverse range of factors (Emerson et al., 1999, Nathan, 2006, Felicísimo et al., 2008, Sequeira et al., 2008, Illera et al., 2012, Stervander et al., 2015, Morinha et al., 2020). Hence, in order to understand the

evolutionary divergence of island biota, it is essential to set a robust phylogenetic framework to identify the closest living mainland relative, the phylogenetic relationships among insular species and populations, the timing and sequence of colonization (i. e., the order in which different islands were occupied), and the history of gene flow among insular populations within and between archipelagos (Whittaker & Fernández-Palacios, 2007, Losos & Ricklefs, 2009, Warren et al., 2015).

The common chaffinch (*Fringilla coelebs*) complex represents a sound system to study speciation processes on oceanic islands, as its broad geographic range includes Eurasia, Northern Africa, and the Atlantic Ocean archipelagos of Macaronesia, including Azores, Madeira and the Canary Islands, but not the Selvagens and Cabo Verde (Shirihai & Svensson, 2018). Common chaffinches on the mainland and the archipelagos differ genetically and in color pattern, morphology, and vocalizations (Grant, 1979, Lynch & Baker, 1994, Illera et al., 2018, Samarasin-dissanayake, 2010, Lachlan et al., 2013). Insular common chaffinches have characteristic dark blue-gray dorsal plumage, a larger body mass, shorter wings, as well as longer tarsi and bills compared to continental specimens (Grant, 1979). In addition, there are notable genetic and phenotypic differences among populations between and within the different archipelagos (see below). Although all common chaffinches are currently classified as a single species with several subspecific taxa, it has been suggested that mainland populations and the different archipelago radiations could be part of a multi-species complex (Illera et al., 2016).

Early proposals for the origin of Macaronesian chaffinches assumed the independent colonization of each archipelago from its nearest mainland, with phenotypic similarities among insular populations resulting from evolutionary convergence (Grant, 1979). In contrast, more recent studies based on mitochondrial DNA sequence data favored a single wave of colonization starting from Europe to Azores, Madeira, and finally the Canary Islands (Marshall & Baker 1999), though limited genetic sampling and weak phylogenetic signal provided only tentative support for this hypothesis. Here, we tested these alternative hypotheses on the timing and colonization route of the common chaffinch radiation by building a robust phylogeny based on thousands of genome-wide loci. Genome-wide datasets based on SNP (single nucleotide polymorphism) loci have proven useful in resolving phylogenetic

relationships at various evolutionary timescales, from deep nodes (Sackton et al., 2019) to very recent radiations (Stervander et al., 2015, Friis et al., 2016, Kozak et al., 2018, Meier et al., 2018). Based on a well-resolved phylogeny, we used biogeographical inference to estimate ancestral ranges using a dispersal-cladogenesis-extinction model that takes founder-event speciation into account and is thus particularly suited for oceanic island systems (A. De Queiroz, 2005, Gillespie et al., 2012, Matzke, 2013). Finally, in order to determine whether the colonization of oceanic archipelagos has resulted in a species-level radiation, we took an integrative taxonomy approach to determine the number of species in the complex according to different methods of species delimitation. This exercise has clear evolutionary and taxonomic implications, but also potentially major conservation impact for the taxa involved, most of which have restricted ranges and small population sizes (Whittaker et al., 2005).

Materials and Methods

Study system and sample collection

The common chaffinch is currently considered to be a polytypic species composed of about 16 subspecies (Clement, 2020) which can be divided into three main geographic groups: a Eurasian group that includes the nominate form (*coelebs*) and related subspecies; a North African group that includes forms *africana*, *spodiogenys* and *harterti* (Svensson, 2015); and a Macaronesian group that includes *moreletti* from the Azores, *maderensis* from Madeira, and four subspecies on the Canary Islands, *canariensis* on Tenerife and La Gomera, *palmae* on La Palma, *ombriosa* on El Hierro, and the recently described *bakeri* on Gran Canaria (Martín & Lorenzo, 2001, Suárez et al., 2009, Illera et al., 2018).

For the present study we obtained blood samples from wild populations in Europe (Segovia, Spain), North West Africa (Ceuta, Spain), the Azores, Madeira and the Canary Islands, so that subspecies included were *coelebs*, *africana*, *moreletti*, *maderensis*, *canariensis*, *palmae*, *ombriosa* and *bakeri* (Fig. 1, Table S1). Birds were captured in the field using mist nets, and each individual was marked with a uniquely numbered

Portuguese or Spanish aluminium band to avoid resampling. Birds were captured during the breeding season. Blood samples were obtained by venipuncture of the brachial vein and stored in absolute ethanol at -20°C in the laboratory until DNA extraction.

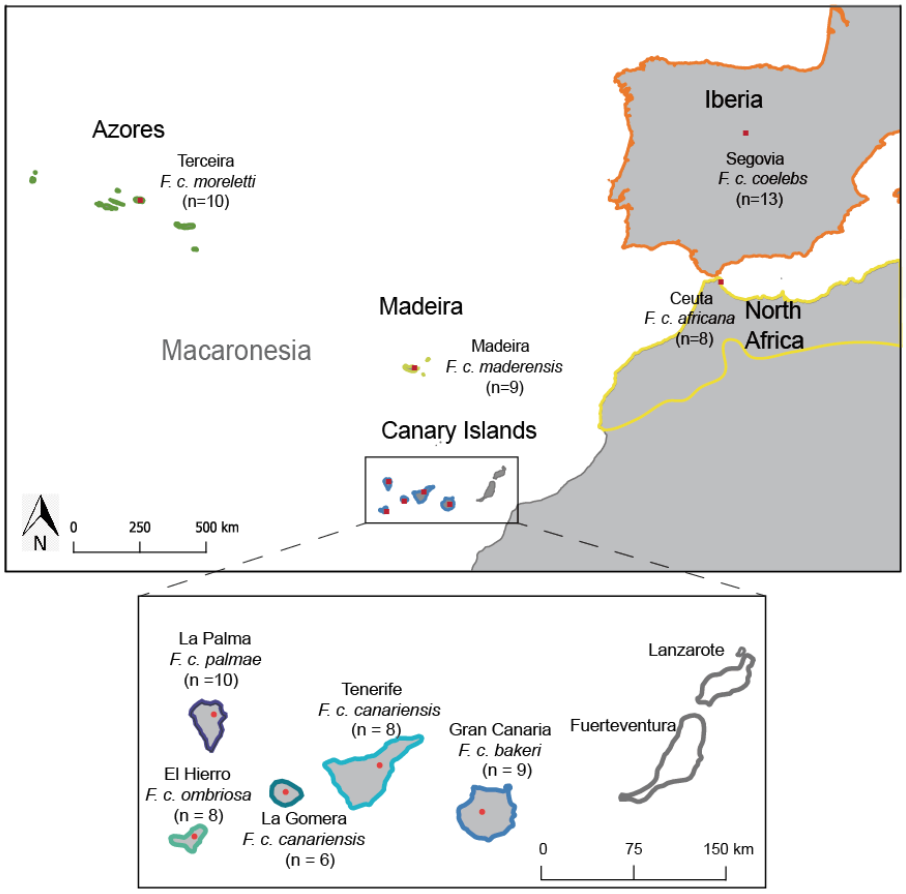


Figure 1. Distribution map of the common chaffinch in the study area. Note the species is absent in the eastern Canary islands of Fuerteventura and Lanzarote. Red dots correspond to sampling sites and sample sizes are indicated in parentheses.

SNP genotyping and analysis

High quality genomic DNA was extracted using a QIAGEN Blood and Tissue kit (Qiagen, Valencia, CA) following the manufacturer's protocol. SNP discovery was done using a genotyping-by-sequencing approach (Elshire et al., 2011) with restriction enzyme *PstI*, and sequencing was carried out on an Illumina Hiseq X Ten platform. Forward raw reads were trimmed to remove low quality ends using TrimGalore! V, 0.4.4 (http://www.bioinformatics. babraham.ac.uk/projects/trim_galore). We aligned the reads against the first version of the high-quality common chaffinch reference

genome (GCA_015532645.1, Chapter IV: Recuerda et al., 2021b) using BWA 0.7.16 (Li and Durbin, 2009), using the "–mem" algorithm and default parameters. The reference genome was mapped against the zebra finch (*Taeniopygia guttata*) genome v87 available in Ensembl (Yates et al., 2016). We used the Chromosembler tool available in Satsuma (Grabherr et al., 2010) obtaining a final assembly 906.9 Mb in length and an N50 of 69.09 Mb. Variant calling was performed with GATK 3.6 HaplotypeCaller and GenotypeGVCFs tools (McKenna et al., 2010), calling all samples together with a minimum base and mapping quality score of 30. The variant dataset obtained was filtered using VCFtools version 0.1.15 (Danecek et al., 2011) keeping biallelic sites with a depth ranging between 4 and 60, a phred quality score over 30, and a minor allele frequency over 0.018. Indels were also removed along with sites with over 75% missing data and showing significant deviation from Hardy-Weinberg equilibrium (p-value < 10^{-4}). To recover the chromosomal coordinates of the scaffolds obtained with HiRiseTM, we mapped and oriented them against the zebra finch (*Taeniopygia guttata*) genome v87 available in ensembl (Yates et al., 2016). We used the Chromosembler tool available in Satsuma (Grabherr et al., 2010) resulting in a final genome assembly 955.9 Mb length and a N50 of 71.46 Mb.

In order to separate neutral loci from loci under divergent selection we used BayeScan v2.1 (Foll & Gaggiotti, 2008) to detect outlier loci in an F_{ST} distribution. We ran the program on a dataset of 159,534 loci with the default sample size of 5,000, a thinning interval of 200, a total of 20 pilot runs of 10,000 iterations each, and a burn-in of 100,000. We checked for convergence and set the false discovery rate (FDR) parameter at 0.1, obtaining 157,366 neutral SNPs and 2168 outliers. We filtered the neutral dataset for linkage disequilibrium (LD) using the snpgdsLDpruning function from the {SNPRELATE} package (Zheng et al., 2012) in R version 3.4.3 (R Core Team, 2017), resulting in a final dataset of 100,166 neutral SNPs.

Genetic diversity

Our final SNP dataset was composed of 81 individuals of the common chaffinch divided into two mainland and seven insular populations (Table S1). Using the complete SNP dataset (159,534 loci), we calculated for each population: nucleotide

diversity (π), the expected and observed heterozygosities (H_e and H_o) and pairwise F_{ST} among populations. All statistics were calculated using STACKS v 1.47 (Catchen et al., 2013). A one-sample *t*-test was used to determine whether the mean F_{IS} score in each population was statistically different from zero using R version 3.4.3 (R Core Team, 2017).

For comparative purposes, we also estimated genetic diversity and demographic parameters using coding regions from the mitochondrial genome (900 bp of the *atp8* and *atp6* genes, and 835 bp of the *nad2* gene), both individually and as a concatenated dataset (1,735 bp). The mitochondrial genes were amplified using primers L5215 (5'-TATCGGGCCCATACCCCGAAAAT-3') (Hackett, 1996) and H6313 (5'-CTCTTATTTAAGGCTTTGAAGGC-3') (Sorenson et al., 1999) for *nad2* and L8929 (5'-GGACAATGCTCAGAAATCTCGCGG-3') (Eberhard & Bermingham, 2005) and H9855 (5'-ACGTAGGCTTGGATTATKGCTACWGC-3') (Sorenson et al., 1999) for *atp8* and *atp6*. PCR products were purified with an ethanol precipitation and sequenced by Sanger sequencing. Sequences were aligned using Sequencher 4.1.1 (Gene-codes Inc., Ann Arbor, MI, USA) and the accuracy of variable sites was checked visually on the chromatograms. We calculated haplotype (*h*) and nucleotide (π) diversity indices per population, pairwise genetic distances and performed Fu's neutrality test (designed to detect changes in population growth; Fu, 1997) using Arlequin v. 3.5 (Excoffier & Lischer, 2010).

Phylogenetic analysis and estimation of divergence times.

To infer the evolutionary history of common chaffinches in the Macaronesian region we reconstructed a phylogenetic tree based on the neutral SNP dataset (100,166 loci), including a Tenerife blue chaffinch (*Fringilla teydea*) as outgroup. We built a maximum-likelihood (ML) tree using RAxML v8.1.16 (Stamatakis, 2014), using a GTR+GAMMA substitution model with the Lewis ascertainment bias correction as recommended. We implemented the rapid bootstrap algorithm (Stamatakis et al., 2008) and evaluated node support with 1000 replicates.

To estimate the timing of island colonization, we used three mitochondrial genes (*nad2*, *atp8* and *atp6*, Table S1) to reconstruct a chronogram with Bayesian inference

in BEAST v 1.8.4 (Drummond et al., 2012), using the CIPRES Science Gateway (Miller et al., 2010) and excluding the outgroup to avoid long-branch effects (Drummond and Bouckaert, 2015). We concatenated all genes (1,735 bp) and selected the best-fitting substitution model with Partitionfinder 2.1.1 (Lanfear et al., 2016), using the Akaike Information Criterion (AIC). The model selected for all markers and all codon positions was GTR+I. Based on results from preliminary runs, we implemented a strict molecular clock with a lognormal distribution of the mutation rate, setting mean values of 0.029 and 0.019 substitutions/site/My for *nad2* and *atp8&6* genes, respectively (Lerner et al., 2011). The haplotype networks for *nad2* and *atp8&6* genes were generated using Hapview (Salzburger et al., 2011) with maximum likelihood trees constructed using Geneious 10.2.2 (https://www.geneious.com) with default parameters.

We also estimated divergence times from a Bayesian phylogenetic tree using SNAPP, a template within BEAST version 2.5.1 (Bouckaert et al., 2019) using the CIPRES Science Gateway (Miller et al., 2010,). SNAPP infers the species tree from biallelic SNPs integrating over all possible gene trees by the implementation of the multispecies coalescent model. We used the neutral SNP dataset restricted to two individuals per population and allowing 5% of missing data, which resulted in 15,836 SNP loci. We used the script "snapp_prp.rb" (Stange et al., 2018) to generate the XML input file keeping the original settings, except that the MCMC chain was set to 2,000,000 generations. We used the RAxML tree as starting tree and set four constraints: (1) The monophyly of North Africa; (2) the monophyly of Europe; (3) the monophyly of the clade including all the insular populations; and (4) given the lack of common chaffinch fossil records in Macaronesia, we used a secondary calibration point based on our dating of the common chaffinch colonization of Macaronesia with mtDNA. We set a lognormal distribution for the divergence time of the insular clade with mean at 0.83 Ma (offset=0, standard deviation = 0.1). A previous study based on a standard cyt-*b* calibration of 0.01 subs/site/lineage/ma obtained a similar date of 0.82 ma (Illera et al., 2018).

For both Bayesian analyses we checked for convergence using Tracer v 1.6 (Rambaut et al., 2014), ensuring that the estimated sample sizes (ESS) were over 200. Node ages and credible intervals (95% highest posterior density, HPD) were estimated, the best

tree was generated using TREEANNOTATOR v1.8.4. (Drummond et al., 2012) and was displayed using FigTree v.1.4.3 (Rambaut, 2017).

Ancestral range estimation

Ancestral range estimation for the common chaffinch across Macaronesia was performed with the SNAPP phylogeny using the BIOGEOBEARS package in R (Matzke, 2013). Among the dispersal-extinction-cladogenesis (DEC) models, we selected the DEC+J model. The "j" parameter allows for "founder-event speciation", which assumes that upon colonization of a remote locality, the founding population becomes instantly genetically distinct from the ancestral population (Matzke, 2014), a model that is appropriate for oceanic island systems, in which speciation takes place relatively quickly following colonization (A. De Queiroz, 2005, Cowie & Holland, 2006, Gillespie et al., 2012). Even though the DEC+J model may have a tendency to underestimate anagenetic events of dispersal and local extinction, which are probabilistic with respect to time, while inflating cladogenetic events of range expansion which are not time related (Ree & Sanmartin, 2018), we selected this model as the most biologically appropriate for our island scenario, where each taxon occupies a unique area that was likely sequentially colonized. We did not compare different models because according to Ree and Sanmartin (2018), their likelihoods are not statistically comparable, so that biological considerations are recommended for model selection instead. We set nine locations corresponding to the two continental areas (Europe and North Africa), which are also separated by sea, and the seven insular populations (Terceira in the Azores, Madeira, Gran Canaria, Tenerife, La Gomera, La Palma and El Hierro).

Genetic structure

To assess patterns of genetic structure and admixture between and within archipelagos, we used the program STRUCTURE (Pritchard et al., 2000) with the neutral SNP dataset, excluding the outgroup and filtered for missing data (5%), which resulted in a total of 16,416 loci. We used PGDSPIDER (Lischer & Excoffier, 2012) to convert the vcf file to the STRUCTURE format, ran preliminary analyses to infer the

lambda value, and then ran analyses five times per K value, each one including 100,000 iterations and a burn-in of 50,000 iterations. The first analysis included individuals from all localities, with K values ranging from 2 to 9. To improve resolution in specific areas, we also ran separate region-specific analyses of the two mainland populations (K = 2-5), and the Canary Islands (K = 2-5). The structure plots were generated using CLUMPAK (Kopelman et al., 2015). The optimal K value was determined by the natural logarithm of the probability of the data [ln(Pr(X|K)] as described in the STRUCTURE manual. In order to check the robustness of results, we performed the same three analyses of population structure with ADMIXTURE v1.3.0 (Alexander et al., 2009) using the complete dataset of neutral SNPs (100,166 loci) with 200 bootstrap replicates.

To estimate fine-scale population structure and quantify the ancestry sources of each common chaffinch population, we used fineRADstructure (Malinsky et al., 2018b), which uses information on haplotype linkage and common ancestry among individuals to produce a summary of nearest-neighbor haplotype relationships in the dataset in the form of a co-ancestry matrix. We converted the vcf file of neutral SNPs into fineRADstructure format using radiator (Gosselin, 2019) and we ran the pipeline using default parameters with 100,000 MCMC generations, sampling every 1,000 steps, and a burn-in of 100,000 steps. The tree was constructed with the fineSTRUCTURE algorithm (Lawson et al., 2012) with 10,000 iterations. The results obtained were plotted in R by adapting the scripts provided in http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html.

Species delimitation

To estimate the number of species in the common chaffinch radiation, we applied the multi-rate Poisson Tree Processes (mPTP) method for species delimitation (Kapli et al., 2017). The mPTP method is based on a rooted phylogenetic tree obtained by probabilistic methods, and it attempts to differentiate speciation from coalescence processes, allowing different intraspecific coalescent rates and a constant speciation rate, assuming that branching events within species are more frequent than between species. For input, we used the RAxML tree based on neutral SNPs, and we ran 10

independent MCMC chains of 10^8 steps, logged every one million generations, with a burn-in of two million steps. We used the "-multi" option to allow variance in coalescent rates among species, and the minimum branch length used was 0.001831, as calculated with the tool "minbr_auto". Average node support values (AVS) were generated for each clade by the MCMC method, with values close to one indicating a robust ML delimitation. We set a conservative threshold for support values over 75 to consider clusters as different candidate species (Kapli et al., 2017). We ensured chain convergence using the Average Standard Deviation of Support Values (ASDDSV), which quantifies the similarity among independent MCMC runs.

In addition to the mPTP analysis we applied an integrative taxonomic approach to species delimitation (Padial et al., 2010). In addition to the genetic data, we took into account differences in plumage coloration (Fig. 2) as well as previously published morphological data (Grant, 1979)(Fig. 3) and bioacoustic data (Lachlan et al., 2013). Finally, we applied a scoring system for avian species delimitation proposed by Tobias et al., (2010) which is based on phenotypic and geographic data, and has been adopted by some major avian taxonomic systems (del Hoyo et al., 2020; Handbook of the Birds of the World and BirdLife International, 2019). The avian species delimitation method proposed by Tobias et al., (2010) is based in a scoring system that takes into account phenotypic traits and geographic data. The method scores and combines the strongest differences in five types of avian traits: morphology, acoustics, plumage, ecology, behavior, and geographical relationship, and assigns species status if the total score reaches or exceeds an arbitrary threshold value of seven points. In our case as the common chaffinch populations considered are allopatric the geographic relationship is not adding any point to the comparisons. From the rest of the traits we just considered plumage coloration and morphology. We used the three most prominent differences in plumage coloration traits for each pairwise comparison across the study area. We generated a table summarizing the qualitative differences in color among forms obtained from digital photographs (Fig. 2) and categorized the plumage color differences following the criteria proposed by Tobias et al., (2010). We also included the morphology using biometric measurements obtained from Grant (1979) (Fig. 3).

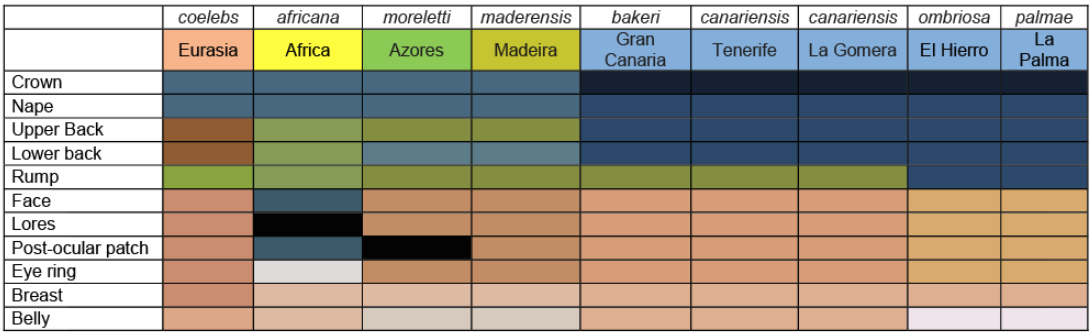


Figure 2. Summary of the main phenotypic differences among males of the different chaffinch taxa. Colours depicted for the different body parts are approximate estimates of real colours obtained from photographs (see Methods).

We scored common chaffinch taxa by means of pairwise comparisons using the two most prominently differentiated traits with opposing direction (i.e., strongest increase and strongest decrease) calculating the effect size for the morphological traits (Cohen's d) and adding the points according to the criteria described in Tobias et al., (2010). Because phenotypic differences among chaffinches in the different Canary Islands are relatively minor, we used inter-island means for the morphological traits, and for the plumage coloration we used patches shared among islands (i.e., nape, back, breast) or patches with similar coloration among islands but clearly differentiated from the counterpart (i.e., face). This points-based scoring system has received some criticism due to the subjectivity involved in the scoring itself, and because the quantitative criteria are based on fairly arbitrary magnitudes of difference that are broadly applied across taxa (Winker, 2010a). However, the method has demonstrated to be useful when used for taxonomical purposes (Winker, 2021) and its performance has been found to be high when tested against recently accepted splits (Tobias et al., 2021).

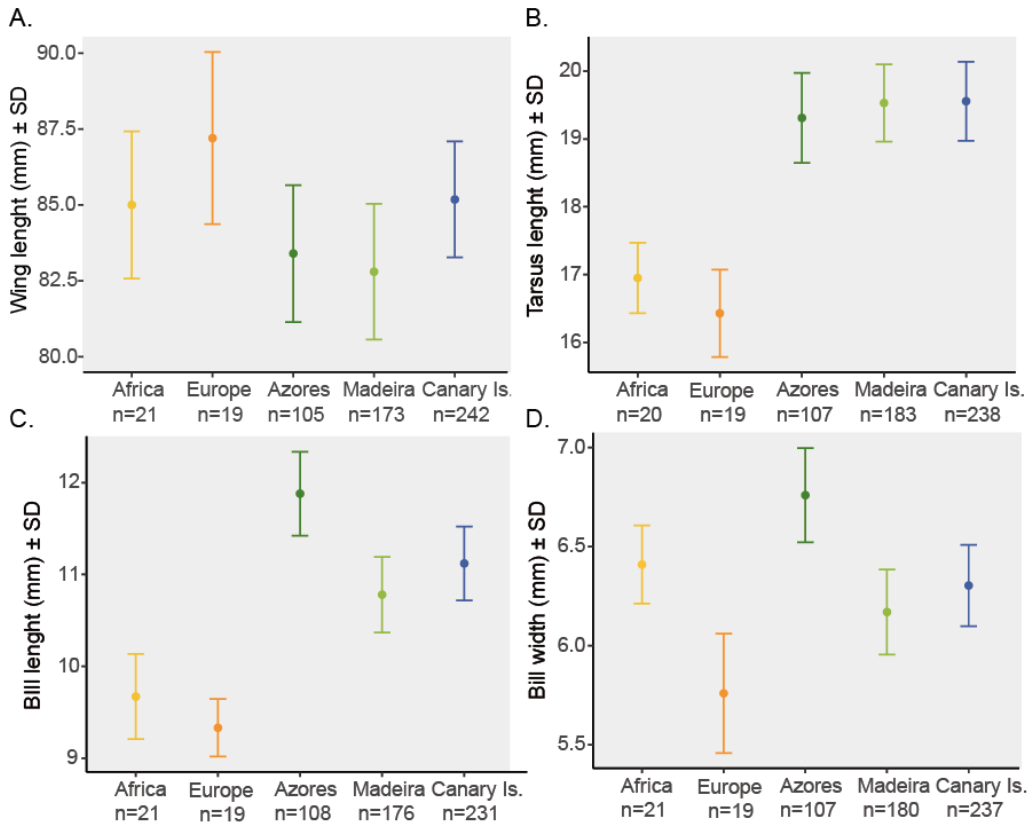


Figure 3. Mean and standard deviation of four morphological traits of the common chaffinch by population. Data from Grant (1979). (A) wing length, (B) tarsus length, (C) bill length and (D) bill width.

Results

SNP genotyping

We obtained 27,052,300 reads from GBS, which resulted in 15,506,115 reads after trimming. The mapping using BWA resulted in 207,339,592 primary aligned reads mapped to the common chaffinch reference genome. The variant calling with GATK generated 1,988,317 variants and after filtering with VCFTOOLS we obtained 159,534 variants with an average depth per site of 16.5.

Genetic diversity and differentiation

Genetic diversity indices were lower on islands than on the mainland (Table 1). Nucleotide diversity and heterozygosity were highest in mainland populations, followed by Azores and Madeira, with the Canary Islands showing the lowest values (Table 1). Pairwise F_{ST} values among populations ranged from 0.07 to 0.16, with an

average of 0.13 (Table 2). The lowest differentiation was found among the common chaffinches of Europe and North Africa, the latter being more differentiated from all insular populations than the former. The Azores population showed the lowest differentiation from mainland populations, and both Azores and Madeira showed similar values of differentiation with respect to the Canary Islands. Within the Canary Islands, F_{ST} values were generally consistent with geographic proximity among islands, with values ranging from 0.09 between Tenerife and La Gomera, and 0.14 between Gran Canaria and the other islands. Genetic distances calculated with the mtDNA dataset showed a similar pattern to that found for SNP markers (Table 3).

Table 1. Descriptive genetic statistics of the common chaffinch populations obtained with 159,534 SNPs: Locality, sample size (n), nucleotide diversity (π), observed heterozygosity (H_o) and expected heterozygosity (H_e).

Region/Locality	n	π	H_o	H_e
Mainland	21	0.193	0.160	0.187
Africa (Ceuta)	8	0.177	0.160	0.165
Europe (Segovia)	13	0.188	0.159	0.177
Macaronesia	60	0.075	0.049	0.074
Azores (Terceira)	10	0.140	0.116	0.130
Madeira	9	0.051	0.047	0.048
Canary Islands	41	0.045	0.034	0.045
Gran Canaria	9	0.035	0.033	0.033
Tenerife	8	0.032	0.030	0.030
La Gomera	6	0.041	0.039	0.038
La Palma	10	0.031	0.031	0.029
El Hierro	8	0.042	0.039	0.039

Phylogenetic analysis, colonization route and divergence times

The ML phylogenetic tree based on 100,166 neutral SNPs was highly resolved, with maximal node support for clades separating the different archipelagos and the different islands within the Canary archipelago (Fig. 4). The phylogenetic tree based on mitochondrial markers showed a similar topology to the genome-wide phylogeny, except for two relationships, which were not highly supported: (1) the two mainland populations (Europe and North Africa) formed a single clade (Fig. 5a); (2) the population of Gran Canaria was sister to a clade including two sister subclades: (a) the westernmost islands of La Palma and El Hierro, and (b) the geographically close islands of Tenerife and La Gomera. Individuals within the Tenerife-La Gomera clade showed an incomplete sorting of haplotypes despite a higher proportion of private haplotypes per island (Fig. 5a).

Table 2. Fixation index (F_{ST}) values among populations of the common chaffinch obtained with 159,534 SNPs. EUR (Iberian Peninsula), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma).

	EUR	AZO	MAD	GC	TEN	GOM	PAL	HIE
AFR	0.069	0.127	0.152	0.148	0.150	0.133	0.152	0.143
EUR		0.094	0.113	0.109	0.111	0.097	0.113	0.105
AZO			0.155	0.157	0.159	0.143	0.161	0.151
MAD				0.159	0.163	0.147	0.158	0.150
GCA					0.135	0.127	0.140	0.136
TEN						0.089	0.134	0.126
GOM							0.118	0.113
PAL								0.096

Haplotype networks revealed that the *nad2* gene showed higher diversity than the *atp8&6* genes except in Madeira and La Palma, and showed better sorting of haplotypes between the two mainland populations and the Tenerife/La Gomera clade, yet neither marker showed complete lineage sorting relative to the genome-wide phylogeny (Fig. 5b,c, Table 4), which provided higher phylogenetic resolution than the mtDNA data. Dating estimates indicated that insular populations diverged from the mainland around 0.83 million years ago (HPD: 0.38-1.48 Ma), Madeira diverged from the Canary Islands about 0.70 Ma ago (HPD: 0.34-1.28), and the Canary Islands differentiated from each other within the last half million years (Fig. 5a).

The SNAPP phylogenetic tree recovered the same topology as the mitochondrial phylogenetic tree but separated the insular populations of Tenerife and La Gomera (Fig. 6). The ancestral range estimation confirmed that colonization of the Atlantic Islands started in Azores, then Madeira, and finally the Canary Islands (Fig. 6). However, the mainland starting point was not clear, with both Europe and North Africa showing similar probabilities. Within the Canary Islands, the analysis suggested that the first island to be colonized was Gran Canaria, but the ancestral range of the remaining islands was not resolved.

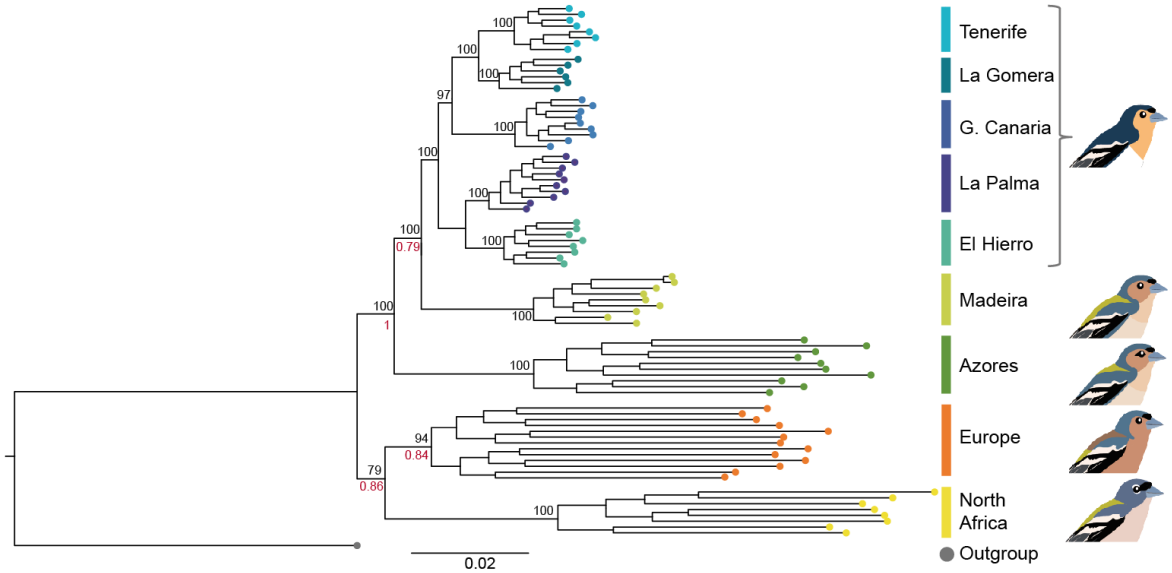


Figure 4. Maximum likelihood phylogenetic tree based on 100,166 genome-wide neutral SNP loci performed using RAxML with 1000 rapid bootstraps and using the blue chaffinch (*Fringilla teydea)* as the outgroup. Figures in black are node support values. Figures in red correspond to Average support values (AVS) from the mPTP species delimitation method. Sketches on the right depict the main phenotypic differences between forms, with chaffinches from the Canary Islands represented by subspecies *palmae*.

Table 3. Genetic distances between the different lineages of the common chaffinch using the *atp8* and *atp6* genes (900 bp) and *nad2* (835 bp) and both concatenated (1,735 bp). Above the diagonal: average number of pairwise differences between populations. Below the diagonal: corrected average pairwise differences. Along the diagonal (in italics): average number of pairwise differences within populations. EUR (Iberian Peninsula), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma).

	AFR	**EUR**	**AZO**	**MAD**	**GCA**	**TEN**	**GOM**	**PAL**	**HIE**
atp8&6									
+ nad2									
AFR	*7.46*	8.38	59.75	52.13	48.13	48.31	48.88	36.92	36.13
EUR	0.76	*7.76*	61.86	53.33	49.36	50.53	50.88	39.24	38.55
AZO	52.91	54.87	*6.22*	50.78	50.22	46.46	47.85	40.92	40.22
MAD	46.31	47.37	45.58	*4.17*	39.00	37.35	38.80	31.91	31.11
GCA	43.64	44.73	46.36	36.17	*1.50*	26.38	28.00	18.80	18.00
TEN	42.37	44.43	41.13	33.05	23.41	*4.43*	8.92	15.18	14.38
GOM	41.74	43.60	41.34	33.31	23.85	3.30	*6.80*	16.80	16.00
PAL	32.72	34.89	37.34	29.36	17.58	12.49	12.93	*0.93*	0.80
HIE	32.39	34.66	37.11	29.03	17.25	12.16	12.60	0.33	*0.00*
atp8&6									
AFR	*3.29*	3.82	23.6	21.58	22.20	20.50	22.33	16.80	16.00
EUR	0.021	*4.31*	24.15	21.89	22.38	20.86	22.70	17.05	16.36
AZO	20.96	20.99	*2.00*	19.93	23.80	17.95	19.53	16.30	15.60
MAD	18.52	18.32	17.52	*2.83*	18.09	15.91	15.91	12.69	11.89
GCA	20.36	20.03	22.60	16.47	*0.40*	18.53	18.53	13.00	12.20
TEN	18.36	18.21	16.45	14.13	16.00	*2.25*	2.25	7.30	6.50
GOM	19.82	19.68	17.67	13.62	17.47	1.73	*1.73*	9.13	8.33
PAL	14.69	14.43	14.83	10.81	12.33	7.80	7.80	*0.93*	0.80
HIE	14.35	14.21	14.60	10.47	12.00	7.47	7.47	0.33	*0.00*
nad2									
AFR	*4.18*	5.10	36.19	30.54	26.12	27.81	26.54	20.13	20.13
EUR	1.28	*3.45*	38.30	31.98	27.18	30.22	28.73	22.73	22.73
AZO	32.04	34.52	*4.11*	30.89	26.67	28.54	28.33	24.67	24.67
MAD	27.79	29.59	28.17	*1.33*	21.11	21.31	22.89	19.22	19.22
GCA	23.29	24.70	23.86	19.69	*1.50*	9.87	9.67	6.00	6.00
TEN	24.01	26.77	24.77	18.92	7.41	*3.43*	6.67	7.88	7.88
GOM	21.92	24.47	23.74	19.69	6.38	2.42	*5.07*	7.67	7.67
PAL	18.04	21.00	22.61	18.56	5.25	6.16	5.13	*0.00*	0.00
HIE	18.04	21.00	22.61	18.56	5.25	6.16	5.13	0.00	*0.00*

Table 4. Genetic diversity and population expansion indices of common chaffinch populations. MtDNA genes used include *atp8* and *atp6* genes (900 bp), *nad2* (835 bp) concatenated (1,735 bp) and individually. Included are DNA marker, geographic region, sample size (n), number of haplotypes (No. haps), haplotype diversity (*h*), nucleotide diversity (π), Fu's neutrality test (F_S). Statistical ignificance of F_s values is indicated by asterisks (* p = 0.05, ** p = 0.01 and ***, p = 0.001). EUR (Iberian Peninsula), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma).

DNA marker	Region	n	No. haps.	h ± SD	π ± SD	F_S
atp8&6 + nad2	AFR	8	6	0.93 ± 0.084	0.0043 ± 0.0405	0.33
	EUR	11	11	1.00 ± 0.039	0.0045 ± 0.0394	-5.10 **
	AZO	9	4	0.69 ± 0.15	0.0036 ± 0.0430	3.27
	MAD	9	7	0.94 ± 0.07	0.0024 ± 0.0261	-1.67
	GCA	8	6	0.90 ± 0.11	0.0009 ± 0.1912	-3.44 ***
	TEN	8	7	0.96 ± 0.08	0.0024 ± 0.0293	-2.32
	GOM	6	4	0.87 ± 0.13	0.0039 ± 0.0408	1.78
	PAL	10	4	0.89 ± 0.08	0.0005 ± 0.0145	-1.02
	HIE	10	1	0.00 ± 0.00	0.0000 ± 0.0000	0.00
atp8&6	AFR	8	5	0.86 ± 0.11	0.0037 ± 0.0368	-0.09
	EUR	11	10	0.98 ± 0.05	0.0048 ± 0.0404	-5.13 **
	AZO	10	3	0.60 ± 0.13	0.0022 ± 0.0321	1.98
	MAD	9	7	0.94 ± 0.07	0.0032 ± 0.0314	-2.71 *
	GCA	10	2	0.20 ± 0.15	0.0004 ± 0.0094	0.59
	TEN	8	2	0.25 ± 0.18	0.0011 ± 0.0166	1.94
	GOM	6	3	0.73 ± 0.16	0.0019 ± 0.0296	0.76
	PAL	10	4	0.71 ± 0.12	0.0010 ± 0.0201	-1.02
	HIE	10	1	0.00 ± 0.00	0.0000 ± 0.0000	0.00
nad2	AFR	8	5	0.89 ± 0.09	0.0050 ± 0.0441	0.41
	EUR	11	11	1.00 ± 0.04	0.0041 ± 0.0382	-8.95 ***
	AZO	9	3	0.64 ± 0.13	0.0049 ± 0.0512	3.80
	MAD	9	4	0.58 ± 0.18	0.0016 ± 0.0187	-0.45
	GCA	8	6	0.89 ± 0.11	0.0018 ± 0.0275	-3.44 ***
	TEN	8	7	0.96 ± 0.08	0.0042 ± 0.0363	-3.05 *
	GOM	6	4	0.87 ± 0.13	0.0061 ± 0.0500	1.20
	PAL	10	1	0.00 ± 0.00	0.0000 ± 0.0000	0.00
	HIE	10	1	0.00 ± 0.00	0.0000 ± 0.0000	0.00

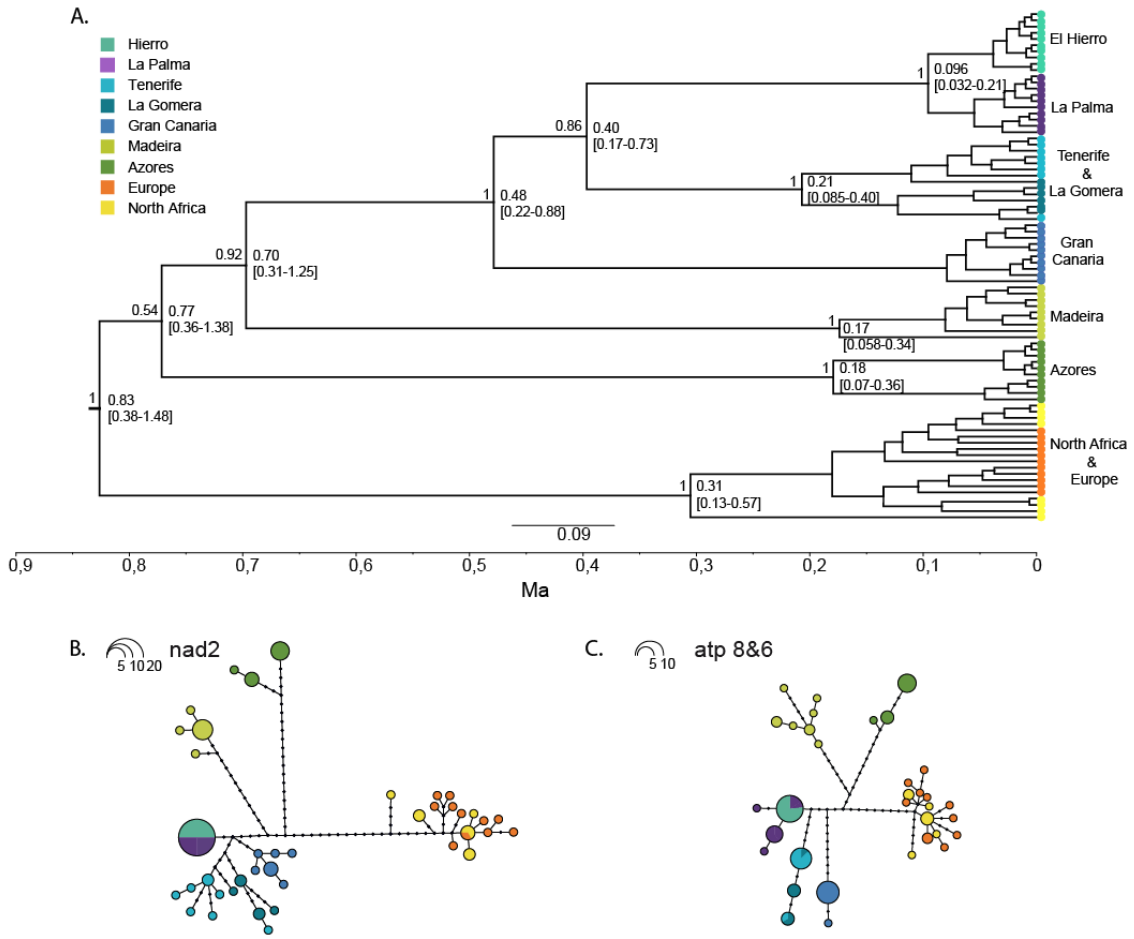


Figure 5. (A) Ultrametric Bayesian tree based on three mitochondrial genes (*atp8*, *atp6* and *nad2*,) obtained with BEAST. Values on the left of each node represent posterior probability of node support. Values on the right of each node represent node age in million years, with confidence intervals (95% HPD) in brackets. (B) Haplotype networks based on *nad2* and (C) *atp8&6* genes. Circles correspond to haplotypes, and their size is proportional to the frequency of each haplotype in the population. Black dots along branches correspond to unsampled or extinct haplotypes.

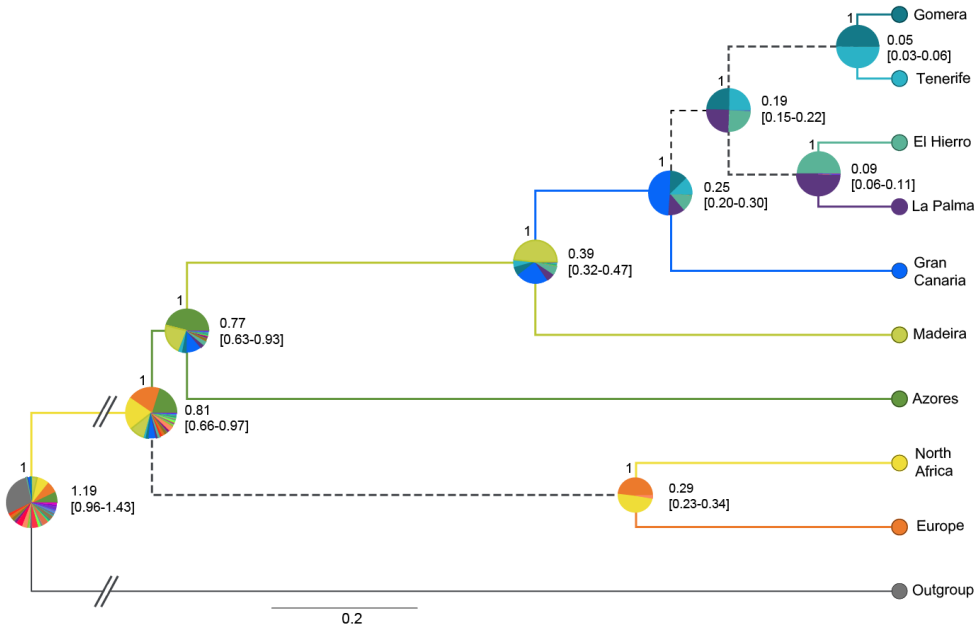


Figure 6. Ancestral range estimation of common chaffinch populations. Inference based on a dispersal-extinction-cladogenesis model with founder event (DEC+J), with the Bayesian phylogeny based on 15,836 neutral SNPs. Pie diagrams at each node represent the inferred geographical ranges for each ancestral taxon, with the probability of each area indicated by its respective color. Branch color represents the most likely state for each branch. Dashed branches indicate that multiple states were tied. Figures above pies represent posterior probabilities of node support, and figures to the right of each node correspond to age in Ma, with confidence intervals (95% HPD) in brackets.

Genetic structure and admixture analysis

The STRUCTURE analysis based on the genome-wide SNP dataset revealed marked genetic structure across the region that was consistent with the ML phylogeny. The optimal number of genetic clusters was K = 6, with clusters corresponding to North Africa, Europe, Azores, Madeira, Gran Canaria and the remaining Canary Islands, respectively (Fig. 7a). An analysis restricted to the mainland individuals confirmed the separation of both populations as the best clustering (Fig. 7b, 8, 9), and a separate analysis of the Canarian archipelago yielded five clusters with high posterior probability of assignment of all individuals to each of the five islands at K = 5 (Fig. 7c). In the latter analysis, K = 2 separated Gran Canaria from the rest, K = 3 additionally separated the western islands (La Palma and El Hierro) and the central islands (Tenerife and La Gomera), and K = 4 and K = 5 separated these two pairs of islands from each other, although La Gomera showed a small proportion of admixture with

Tenerife. The ADMIXTURE results were generally consistent with the STRUCTURE analysis, with the same optimal number of clusters but some differences in the sequence of population separation (Fig. 8, Table 5). In both analyses, the Azores shared some variance with the mainland at K = 2, and Gran Canaria shared some variance with Madeira, being the first island to separate from the rest within the Canary archipelago.

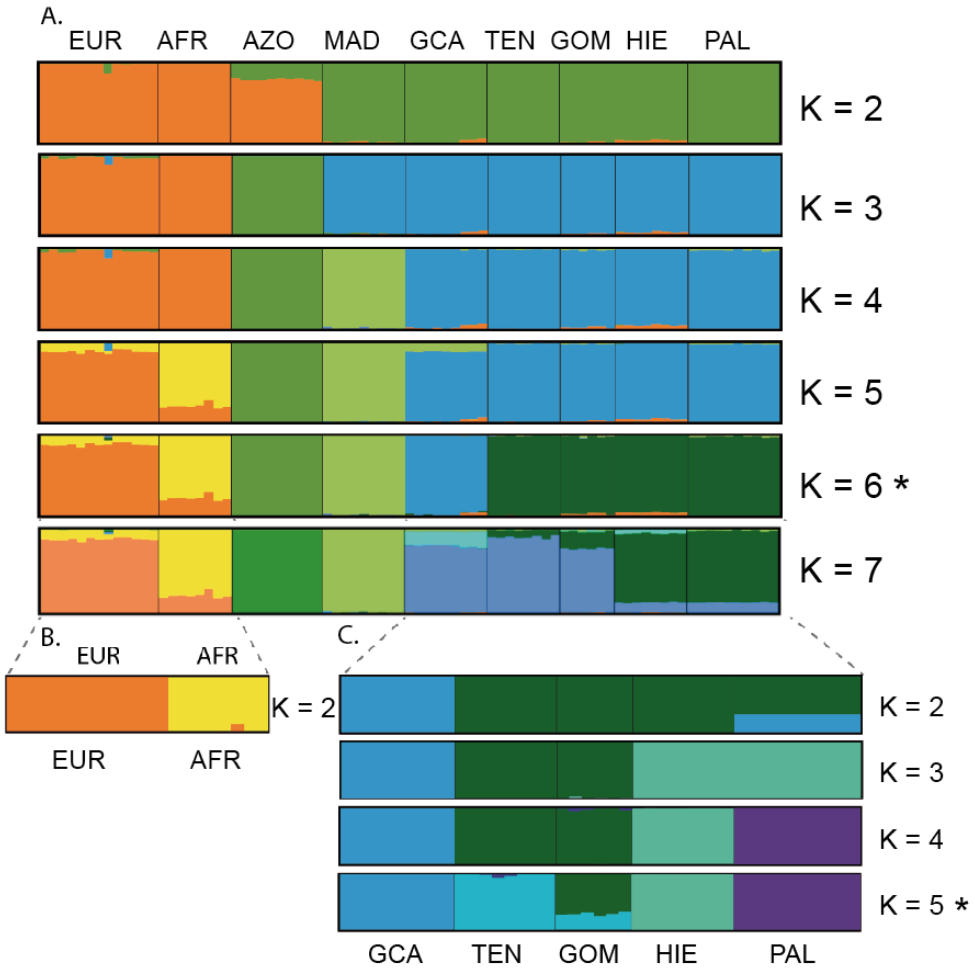


Figure 7. STRUCTURE analysis plots for (A) all chaffinch populations with K ranging from 2 to 7 (plots for K = 8 and 9 are not shown as they do not differ from K=7), (B) mainland populations only for K = 2, and (C) Canary Islands populations only with K ranging from 2 to 5. EUR (Iberia), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma). Asterisks (*) mark the optimal K value for each analysis.

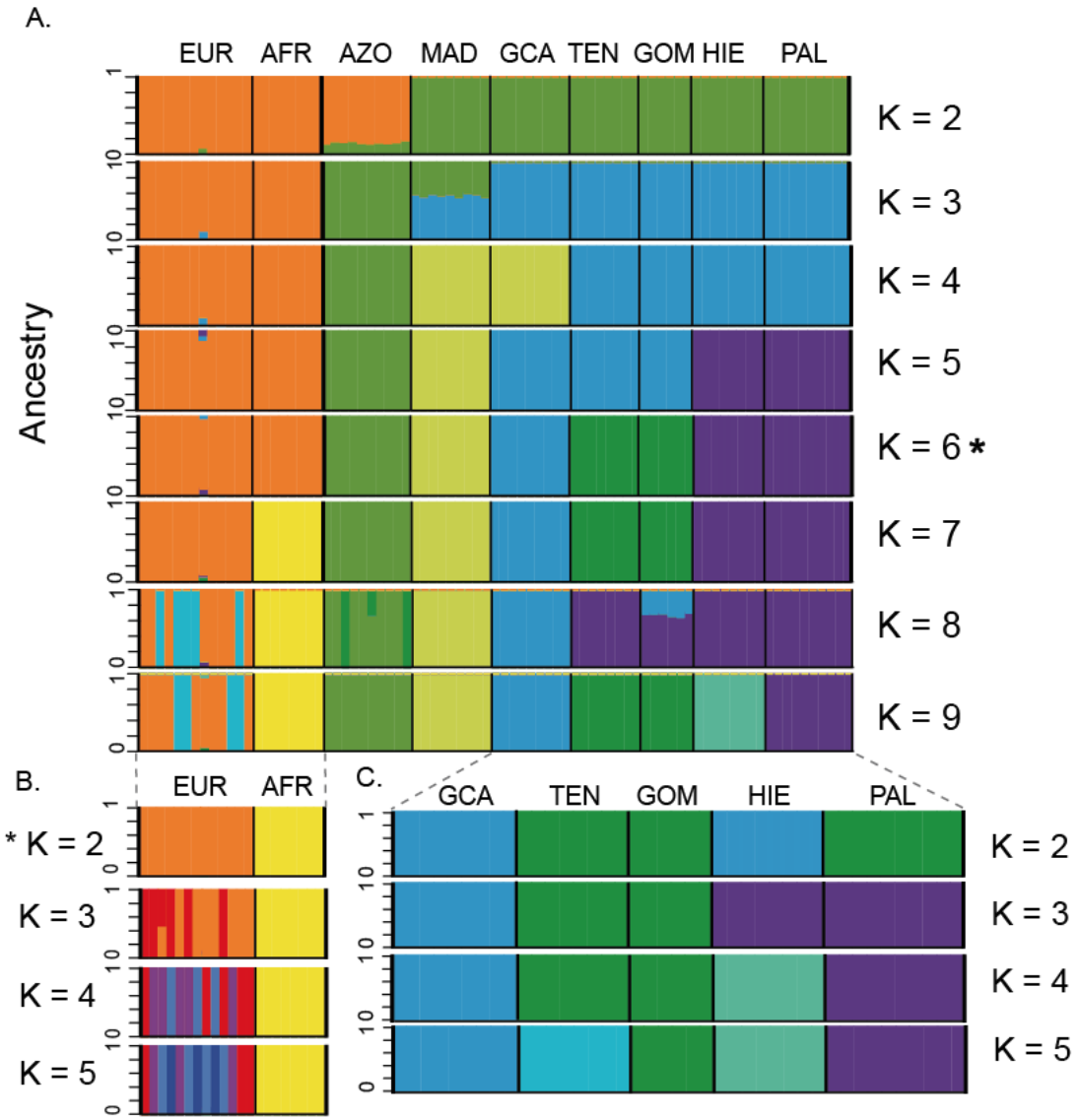


Figure 8. Genetic structure of common chaffinch populations using ADMIXTURE. (A) All populations with K ranging from 2 to 9. (B) Mainland populations only, with K ranging from 2 to 5. (C) Canary Islands only, with K ranging from 2 to 5. EUR (Iberia), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma). Asterisks (*) mark the optimal K value for each analysis.

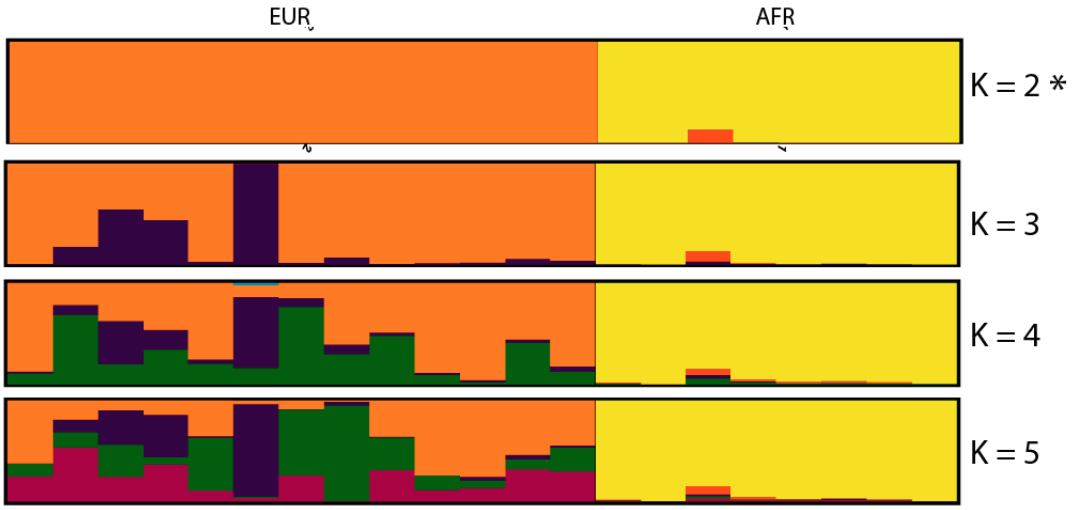


Figure 9. STRUCTURE analysis plots for mainland chaffinch populations with K ranging from 2 to 5. EUR (Iberia), AFR (North Africa). Asterisk (*) mark the optimal K value.

The FINERADSTRUCTURE analysis showed consistent results with previous analyses and divided individuals into the same nine populations (Fig. 10). The plot also showed clear regional structure among populations with two main clusters, one formed by the continental individuals along with Azores, and the other including the remaining insular populations. Coancestry relationships among populations revealed that the Azores shares more ancestry with Europe than with North Africa. Within the insular cluster, two pairs within the Canary Islands show high coancestry (Tenerife and La Gomera, and La Palma and El Hierro, respectively).

Table 5. Coefficients of variation for the different K values in the three ADMIXTURE Analyses.

	CV error							
	K=2	**K=3**	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**	**K=9**
All	0.31745	0.32174	0.30949	0.30230	0.29629	0.33647	0.39782	0.39518
Continent	0.61969	0.81042	1.06934	1.33184				
Canary Islands	0.18014	0.16307	0.16236	0.16635				

Species delimitation

The 10 independent MCMC runs of mPTP suggested species-level designation for the five main clades in the ML phylogeny, corresponding to Europe, North Africa, Azores, Madeira and the Canary Islands, with support values ranging from 0.79 to 1 (Fig. 4,

values in red). In addition, mPTP suggested one additional clade within Europe, with a support value of 0.84.

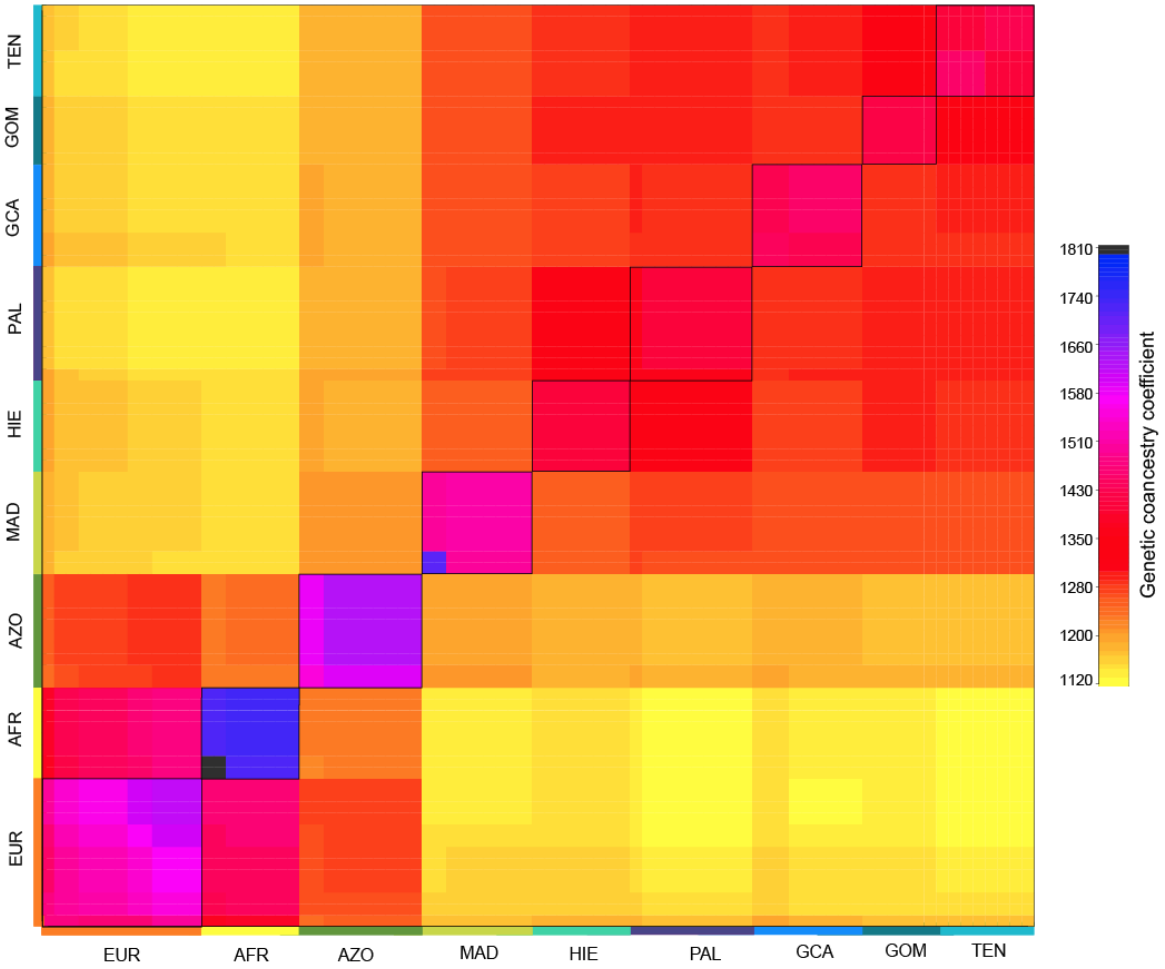


Figure 10. Matrix of pairwise genetic co-ancestry values among chaffinch populations. Averaged co-ancestry coefficients per population are color-coded from low (yellow) to high (black). Individuals clustering into populations are shown along the diagonal (squares framed in black). EUR (Iberia), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma).

Table 6. Pairwise comparisons among subspecies of *Fringilla coelebs* of the strongest traits (ST) and four biometric traits: Wing (Wg), Tarsus (T), Bill length (BL) and Bill width (BW). In the ST table, below the diagonal are summarized the traits showing the strongest increase (+) and the strongest decrease (-) per comparison and above the diagonal the total score obtained. In the biometric traits tables, below the diagonal are the effect sizes (Cohen's d) calculated as described in Tobias et al., (2010) and above the diagonal the corresponding scores.

ST	Africa	Europe	Azores	Madeira	Canarias
Africa	-	3	3	3	3
Europe	BW/Wg	-	4	4	4
Azores	Wg/BL	Wg/BL	-	3	3
Madeira	BW/BL	Wg/T	BW/T	-	1
Canary Is.	BW/T	Wg/T	BW/Wg	0/Wg	-
W					
Africa	-	1	1	1	1
Europe	-0.86	-	1	1	1
Azores	0.71	1,63	-	1	1
Madeira	0.98	1.92	0.27	-	1
Canary Is.	-0.09	1.02	-0.88	-1.16	-
T					
Africa	-	1	2	2	2
Europe	0.91	-	2	3	3
Azores	-3.70	-4.40	-	1	1
Madeira	-4.60	-5.41	-0.37	-	1
Canary Is.	-4.53	-5.35	-0.41	-0.05	-
BL					
Africa	-	1	2	2	2
Europe	0.87	-	3	2	2
Azores	-4.86	-5.85	-	2	1
Madeira	-2.68	-3.62	2.57	-	1
Canary Is.	-3.57	-4.53	1.81	-0.84	-
BW					
Africa	-	2	1	1	1
Europe	2.65	-	2	1	2
Azores	-1.52	-4.06	-	2	2
Madeira	1.13	-1.84	2.65	-	1
Canary Is.	0.52	-2.56	2.12	-0.64	-

We integrated the molecular data from the mPTP analysis with phenotypic data and all five clades identified by mPTP showed congruent differentiation in phenotypic traits, mainly in terms of plumage color but also morphology and bioacoustics. When scoring differences in plumage coloration (Fig. 2) and morphology (Table 6) among pairs of subspecies using the five most prominent traits (Tobias et al., 2010), all comparisons reached the minimum threshold for species designation (Table 7).

Table 7. Pairwise comparison among subspecies of *Fringilla coelebs* using the scoring system provided by Tobias et al., (2010) using the strongest increase and the strongest decrease for biometric traits obtained from Grant (1979) and the three most prominent characters of plumage coloration.

Subspecies comparison	Morphology	Plumage	Total Score
coelebs vs *africana*	Bill width (2) Wing (1)	Back: brown/green (3) Face: redish brown/gray (3) Eyering: Redish brown/white (3)	12
coelebs vs *moreletti*	Wing (1) Bill length (3)	Upper back: brown/green (3) Lower back: brown/blue-gray (3) Face: redish brown/pale orange (2)	12
coelebs vs *maderensis*	Wing (1) Tarsus (3)	Upper back: brown/green (3) Lower back: brown/blue-gray (3) Face: brown redish/pale orange (2)	12
coelebs vs *canariensis*	Wing (1) Tarsus (3)	Back: brown/ blue (3) Face: redish brown/pale orange (2) Breast: redish brown/pale orange (2)	11
africana vs *moreletti*	Bill length (2) Wing (1)	Lower back: green/blue-gray (3) Face: gray/ pale orange (3) Eyering: white/ pale orange (2)	11
africana vs *maderensis*	Bill width (1) Bill length (2)	Lower back: green/blue-gray (3) Face: gray/ pale orange (3) Eyering: white/ pale orange (2)	11
africana vs *canariensis*	Bill width (1) Bill length (2)	Back: green / blue-gray (3) Face: gray/ pale orange (3) Eyering: white/ light orange (2)	11
moreletti vs *maderensis*	Bill width (2) Tarsus (1)	Eye patch: black / light orange (3)	6
moreletti vs *canariensis*	Bill width (2) Wing (1)	Eye patch: black/orange (3) Upper back: green/blue-gray (3) Lower back: blueish gray/blue-gray (2)	11
maderensis vs *canariensis*	Wing (1)	Upper back: green/blue-gray (3) Lower back: gray/blue-gray (2) Nape : gray /blue-gray (2)	8

Discussion

Colonization history in the common chaffinch radiation

Our results from molecular phylogenies, ancestral range estimation, and coancestry analyses, provide strong and consistent support for a colonization of Macaronesia by the common chaffinch that took place from the mainland, via the Azores and Madeira to the Canary Islands, and resulted in a rapid species-level radiation. This circuitous colonization route seems counterintuitive from a biogeographic perspective, given the large distance separating the Azores from the mainland (*ca*. 1300 km) compared to the other archipelagos, and suggests that factors other than mere geographic distance were at play in the common chaffinch radiation. Although the topologies of the phylogenetic trees do not allow determining whether the original colonizers of the Azores came from Europe or North Africa, the coancestry analysis with fineRADstructure, along with the genetic distances based on both datasets, suggests that a European origin is more likely. The estimation of the colonization time of these Atlantic islands by the common chaffinch obtained with BEAST coincides with previous estimates of about one million years before present (Illera et al., 2018), which is relatively recent compared to the age of most of the islands (Illera et al., 2012). The estimated colonization time falls within the last 3 million years, a period found to include most colonization events by Macaronesian bird taxa (Valente et al., 2017). This period coincides with the establishment of most Macaronesian laurel forests in the Plio-Pleistocene (2.6 Ma), and with the movement of the trade wind zone over the islands during the Pleistocene (2.6-0.01 Ma), which provided sufficient precipitation and moisture (Kondraskov et al., 2015). The phylogenomic tree obtained with ~100,000 neutral SNPs provided enough resolution to reveal independently evolving, monophyletic lineages of the common chaffinch on each archipelago. Results also suggest shared ancestry of all the Macaronesian islands, followed by divergence with restricted gene flow among islands. This single-wave colonization history is supported by shared phenotypic characters among insular populations. Macaronesian chaffinches show plumage patterns with blue-gray dorsal coloration and reduced green and red patches (Grant, 1980); longer tarsi and shorter wings than their mainland counterparts (Grant, 1979; Dennison & Baker, 1991), as documented for other passerines (Wright et al., 2016); and decreasing song complexity after each

colonization event (Lynch & Baker, 1994, Lachlan et al., 2013). Overall, this pattern of shared traits among all insular populations is more consistent with common ancestry than convergence following independent colonizations from the nearest mainland (Marshall & Baker, 1999). Given the phylogenetic relationships among all insular populations, common ancestry is more parsimonious than the alternative hypothesis of repeated, independent evolution of these traits on each island under common selective pressures.

Unlike the Azores, where gene flow appears to have prevented the differentiation of common chaffinch populations among islands (Baker et al., 1990; Rodrigues et al., 2014), those in the Canary Islands have diverged markedly from each other, giving rise to a range of phenotypes currently grouped into four different subspecific taxa (Illera et al., 2018). Partly because of this recent inter-island differentiation, inferring the specific order in which the Canary Islands were colonized is challenging (Marshall & Baker, 1999). The absence of the common chaffinch in the eastern-most islands of Lanzarote and Fuerteventura may be due to the current lack of suitable habitat, which is known to have varied widely over time due to the frequent extinction-recolonization events of their flora (García-Verdugo et al., 2019), but whether or not the common chaffinch was present there in the past cannot be determined from available data. For the islands where the common chaffinch is present, we obtained conflicting results and found evidence consistent with both an east-to-west and a west-to-east pattern of colonization. On one hand, our results support the eastward colonization because La Palma is closest to Madeira in the haplotype networks and shows lower genetic distance with Madeira than Gran Canaria. This route may have been favoured by the wind patterns that blow south-eastwards from the Azores in winter (Grant, 1980), as previously proposed (i.e., Grant, 1980, Marshall & Baker, 1999, Suárez et al., 2009, Lachlan et al., 2013). On the other hand, the mitochondrial DNA tree, the ancestral range estimation and the population structure analysis are more consistent with a westward colonization starting from Gran Canaria. More research will be needed to disentangle the specific common chaffinch colonization within the Canary Islands, an archipelago with a diverse range of avian colonization histories given its proximity to neighboring archipelagos and mainland (Illera et al 2012, Morinha et al., 2020).

The progressive reduction of genetic diversity from the Azores to the Canary Islands is also consistent with the colonization route, and expected when islands are sequentially colonized from other islands by small groups of individuals from source populations of progressively smaller effective population size (Clegg, Degnan, Kikkawa et al., 2002). Genetic diversity in the Azores was similar to that found on mainland populations and an order of magnitude higher than that found on other archipelagos. This suggests that a relatively large group of original colonizers (or multiple colonization events in a short period of time), arrived to Azores, avoiding a major founder event (James et al., 2016), but also that effective population size was maintained relatively large over time. Indeed, in addition to the magnitude of potential founder events, the surface area of suitable common chaffinch habitat in the different islands and the presence of gene flow among them are also likely to have influenced present levels of genetic diversity. Except for La Palma, where common chaffinches have stable breeding populations in dry pine forests, Macaronesian common chaffinches are largely restricted to *monteverde* humid habitats, from cloud forest to moist heaths, and the geographic area of these habitat types varies widely among islands (Martín & Lorenzo, 2001). While most of the Azores are humid enough to sustain common chaffinch populations, suitable habitat decreases markedly with latitude, becoming less abundant in Madeira, and restricted to small "islands within islands" in the Canaries, where humid habitats are more restricted than in the other archipelagos (Fernández-Palacios, 2009). In turn, gene flow among the Azores, which has prevented genetic differentiation among islands (Rodrigues et al., 2014), has favored the maintenance of high population sizes and genetic diversity, in contrast to the Canary Islands, where populations have become isolated from each other due to highly restricted gene flow.

The chaffinch taxa produced in the Macaronesian archipelagos differ from each other mostly in plumage coloration, and to a much lesser degree in morphological characters. This is similar to what has been observed in non-adaptive avian radiations, such as those in South American capuchino seedeaters (Campagna et al., 2012), North American juncos (Friis & Milá 2020), or European wagtails (Ödeen & Björklund 2003), where taxa differ in color traits with a simple genetic basis (Campagna et al., 2017, Abolins-Abols et al., 2018), yet are relatively uniform in morphology. This suggests

that drift and sexual selection have been the main drivers of the phenotypic diversification, with morphological adaptation to local ecological conditions playing a relatively minor role (Rundell & Price 2009), likely due to the ecological similarity between Macaronesia and its mainland. This is in contrast to well-studied adaptive radiations such as that of the Darwin's finches in the Galapagos Islands (Grant & Grant 2008, Lamichhaney et al., 2015), or the honeycreepers in Hawaii (Lerner et al., 2011). Within a similar time frame to that of the chaffinch diversification, these two radiations gave rise to markedly diverse beak morphologies as populations of the original colonizers adapted through strong directional selection to the food resources available in the different islands. In the case of the honeycreepers, which belong to the same family Fringillidae as chaffinches, morphological divergence was accompanied by a stunning diversification in color patterns and other ornamental traits (Freed et al., 1987), suggesting the combined action of natural and sexual selection (Gillespie et al., 2020). Even though the common chaffinches have not diversified bill morphology to that extent, natural selection has likely played a role in modifying their morphology, especially the size and shape of their beaks (Grant, 1979).

Systematics and taxonomy of the chaffinch radiation

Our species delimitation analyses suggest that the common chaffinch radiation has resulted in several species-level taxa. The genome-wide analysis of genetic variation revealed the existence of several distinct evolutionary lineages evolving independently from each other, and species delimitation analyses provided support for the existence of at least five different species within the complex. The mPTP method provided support for the five nodes corresponding to North Africa, Europe, Azores, Madeira and Canary Islands, respectively. The additional supported clade within Europe could be due to high genetic diversity of the European population, and does not seem to be associated with phenotypic differences or geographical limits. The STRUCTURE and ADMIXTURE analyses for the continental clades showed that for K > 2, some individuals of the Iberian population show some divergence, but do not correspond to the clades in the phylogenomic tree (Fig. S3). Marked phenotypic divergence among major lineages was confirmed by Tobias' et al., (2010) delimitation method, which was also consistent with the five-species hypothesis. Even though,

plumage coloration and morphological differences among *F. c. moreletti* and *F. c. maderensis* were less prominent than between other members of the complex, they are known to differ in other characters relevant to reproductive isolation like territorial male song (Lachlan et al., 2013), that were not included in our analysis.

We concur with previous studies on this system (Marshall & Baker, 1999; Suárez et al., 2009, Rodrigues et al., 2014, Illera et al., 2016, Perktaş et al., 2017, Clement, 2018), on the need for a taxonomic revision of this group, and based on their and our results, we propose that the common chaffinch be divided into five different species, corresponding to Eurasia (*Fringilla coelebs*), North Africa (*Fringilla spodiogenys/africana*), Azores (*Fringilla moreletti*), Madeira (*Fringilla maderensis*) and the Canary Islands (*Fringilla canariensis*). *F. coelebs* would include all subspecies closely related and phenotypically similar to *F. c. coelebs* found across continental Eurasia. Although populations on the different Canary Islands are genetically distinct, their phenotypic differentiation is relatively minor, and we propose to maintain their current subspecific status within *F. canariensis*. Such a subspecific classification would be as follows: *F. canariensis canariensis* on Tenerife and La Gomera, *F. canariensis palmae* on La Palma, *F. canariensis ombriosa* on El Hierro, and *F. canariensis bakeri* on Gran Canaria.

North African subspecies *spodiogenys* and *harterti* were not included in this study, yet they are phenotypically similar to *africana* (Svensson, 2015, Perktas et al., 2017). The early molecular study by Marshall and Baker (1999) reported *spodiogenys* as a divergent lineage that was basal to the *Fringilla coelebs* complex in a mtDNA phylogeny, yet more recent molecular analyses using nuclear DNA markers indicate that the two North African subspecies are indeed closely related sister taxa (Samarasin-Dissanayake, 2010). This result is consistent with both phenotype and geography, and suggests that mtDNA may not be suitable to recover the evolutionary history of these taxa. Based on this evidence, and since *spodiogenys* Bonaparte 1841 was described before *africana* Levaillant 1850 and *harterti* Svensson 2015, we recommend recognizing species *Fringilla spodiogenys* with three subspecies (*F. spodiogenys spodiogenys*, *F. spodiogenys africana*, and *F. spodiogenys harterti*).

Recognizing the new proposed species should be consistent with most species concepts that take into account evidence for independent evolving lineages and phenotypic differentiation (K. De Queiroz, 2007; Sangster, 2013; Gill, 2014). The taxonomic upgrade from subspecies to species is likely to have important conservation implications, as species tend to receive more conservation attention than subspecies (Winker, 2010b; Sangster et al., 2016). Specifically, species status would guarantee that the conservation status of each chaffinch taxon is evaluated by the International Union for Conservation of Nature (IUCN), taking into account their distribution area and population size independently, making the difference especially for the more restricted insular populations (Martín, 2009). Hence, conservation biogeography (Whittaker et al., 2005), which includes the distribution of taxa in the conservation criteria by applying biogeographical analysis is important for the improvement of biodiversity conservation. This may in turn help preserve the genetic diversity of the species complex, which is crucial for the resilience to environmental change in the current scenario of climate change, especially given the reduced genetic variability found across the region.

Conclusions

The colonization of Macaronesia by the common chaffinch has resulted in an evolutionary radiation as populations differentiated phenotypically and genetically in the different archipelagos, and even between islands within the Canary archipelago. The molecular phylogeny was instrumental in revealing a circuitous colonization route from the mainland to the faraway Azores, and then south to Madeira and the Canary Islands. Relatively minor differences in morphology between insular and mainland chaffinches compared to differences in coloration, suggest that drift due to founder events, along with sexual selection acting on plumage coloration and song, are likely the major factors driving the common chaffinch radiation in Macaronesia. The sequential colonization of three Atlantic archipelagos and Northern Africa has led to the formation of at least four new species-level taxa in the genus *Fringilla*, and our results should help further our understanding of the evolutionary processes involved.

Data availability

Mitochondrial markers sequences are deposited at Gen Bank, the accession numbers are (MW460715- MW460796) for *atp8&6* and (MW460797-MW460875) for *nad2* (see Table S1 for details). SNP raw data is deposited at NCBI under the SRA data project PRJNA692563 with accession numbers (SAMN17349018 - SAMN17349100), see Table S1 for details) and the vcf datasets are deposited in *Figshare* (https://doi.org/10.6084/m9.figshare.13562582). The *Fringilla coelebs* reference genome is deposited at NCBI (Accession number: JADKPM000000000.1).

Supplementary Materials

Table S1. List of the samples of *F. coelebs* used in the study for mtDNA, including specimen NCBI Accesion numbers for atp, nd2 and GBS data, Field and Specimen ID, locality and Region information, and haplotype designation for *atp* and *nad2* markers. The specimens with a "Y" in the field GBS were used sequenced by genotyping-by-sequencing to generate the SNP data. The individual TEN4 marked with an asterisk (*) is an individual of *F. Teydea* used as an outgroup for SNP data.

Accession atp	Accesion nd2	Accesion GBS	MNCN	Field ID	Specimen ID	Locality	Region	*atp*	*nad2*	*GBS*
MW460725	MW460806	SAMN17349018	112224	18-642	CEU1	Ceuta	Africa	1	1	Y
MW460726	MW460807	SAMN17349019	112240	18-658	CEU2	Ceuta	Africa	1	1	Y
MW460727	MW460808	SAMN17349020	112244	18-662	CEU3	Ceuta	Africa	2	2	Y
MW460728	MW460809	SAMN17349021	112251	18-669	CEU4	Ceuta	Africa	3	3	Y
MW460729	MW460810	SAMN17349022	112253	18-671	CEU5	Ceuta	Africa	2	2	Y
		SAMN17349023	112266	18-684	CEU6	Ceuta	Africa			Y
		SAMN17349024	112267	18-685	CEU7	Ceuta	Africa			Y
		SAMN17349025	112268	18-686	CEU8	Ceuta	Africa			Y
MW460730	MW460811		112262	18-680	CEU9	Ceuta	Africa	1	4	
MW460731	MW460812		112263	18-681	CEU10	Ceuta	Africa	4	4	
MW460732	MW460813		112264	18-682	CEU11	Ceuta	Africa	5	5	
MW460775	MW460852	SAMN17349026	112273	16-100	SEG1	Segovia	Iberia	6	6	Y
MW460776	MW460853	SAMN17349027	112275	16-102	SEG2	Segovia	Iberia	7	7	Y
MW460777	MW460854	SAMN17349028	112276	16-103	SEG3	Segovia	Iberia	8	8	Y
MW460778	MW460855	SAMN17349029	112280	16-107	SEG4	Segovia	Iberia	9	4	Y
MW460779	MW460856	SAMN17349030	112282	16-109	SEG5	Segovia	Iberia	10	9	Y
MW460780	MW460857	SAMN17349031	112286	16-113	SEG6	Segovia	Iberia	11	10	Y
MW460781	MW460858	SAMN17349032	112293	16-120	SEG7	Segovia	Iberia	12	11	Y
MW460782	MW460859	SAMN17349033	112296	16-123	SEG8	Segovia	Iberia	13	12	Y
MW460783	MW460860	SAMN17349034	112302	16-129	SEG9	Segovia	Iberia	14	13	Y
		SAMN17349035	112303	16-130	SEG10	Segovia	Iberia			Y
MW460773	MW460861	SAMN17349036	112305	16-132	SEG11	Segovia	Iberia	10	14	Y

MW460774	MW460862	SAMN17349037	112308	16-135	SEG12	Segovia	Iberia	15	15	Y
		SAMN17349038	112309	16-136	SEG13	Segovia	Iberia			Y
MW460716	MW460797	SAMN17349039		A407593	AZ1	Terceira	Azores	16	16	Y
MW460717		SAMN17349040		A407594	AZ2	Terceira	Azores	16		Y
MW460718	MW460798	SAMN17349041		A407595	AZ3	Terceira	Azores	16	16	Y
MW460719	MW460799	SAMN17349042		A407596	AZ4	Terceira	Azores	17	17	Y
MW460720	MW460800	SAMN17349043		A407600	AZ5	Terceira	Azores	17	17	Y
MW460721	MW460801	SAMN17349044		AZ407UN1	AZ6	Terceira	Azores	16	16	Y
MW460722	MW460802	SAMN17349045		AZ407UN2	AZ7	Terceira	Azores	17	18	Y
MW460723	MW460803	SAMN17349046		AZ407UN3	AZ8	Terceira	Azores	16	16	Y
MW460724	MW460804	SAMN17349047		AZ407UN4	AZ9	Terceira	Azores	18	17	Y
MW460715	MW460805	SAMN17349048		AZ407UN6	AZ10	Terceira	Azores	16	16	Y
MW460764	MW460843	SAMN17349049		A407538	MAD1	Madeira	Madeira	19	19	Y
MW460765	MW460844	SAMN17349050		A407539	MAD2	Madeira	Madeira	19	19	Y
MW460766	MW460845	SAMN17349051		A407541	MAD3	Madeira	Madeira	20	19	Y
MW460767	MW460846	SAMN17349052		A407542	MAD4	Madeira	Madeira	21	20	Y
MW460768	MW460847	SAMN17349053		A407543	MAD5	Madeira	Madeira	22	21	Y
MW460769	MW460848	SAMN17349054		A407544	MAD6	Madeira	Madeira	23	19	Y
MW460770	MW460849	SAMN17349055		A407546	MAD7	Madeira	Madeira	20	19	Y
MW460771	MW460850	SAMN17349056		A407548	MAD8	Madeira	Madeira	24	19	Y
MW460772	MW460851	SAMN17349057		MA407UN1	MAD9	Madeira	Madeira	25	19	Y
MW460739	MW460819	SAMN17349058		4L88806	GC1	G. Canaria	Canary Is.	26	22	Y
MW460740	MW460820	SAMN17349059		4L88807	GC2	G. Canaria	Canary Is.	26	23	Y
MW460741	MW460821	SAMN17349060		4L88811	GC3	G. Canaria	Canary Is.	26	23	Y
MW460742	MW460822	SAMN17349061		4L88812	GC4	G. Canaria	Canary Is.	26	23	Y
MW460743	MW460823	SAMN17349062		4L88813	GC5	G. Canaria	Canary Is.	26	24	Y
MW460744	MW460824	SAMN17349063		4L88814	GC6	G. Canaria	Canary Is.	26	25	Y
MW460745		SAMN17349064		4L88815	GC7	G. Canaria	Canary Is.	27		Y
MW460746	MW460825	SAMN17349065		4L88816	GC8	G. Canaria	Canary Is.	26	26	Y
MW460747		SAMN17349066		4L88817	GC9	G. Canaria	Canary Is.	26		Y
MW460738	MW460826			4L88819	GC10	G. Canaria	Canary Is.	26	24	

MW460784	MW460863	SAMN17349067		2L55180	TEN1	Tenerife	Canary Is.	28	27	Y
MW460785	MW460864	SAMN17349068		N636250	TEN2	Tenerife	Canary Is.	29	27	Y
MW460786	MW460865	SAMN17349069		2N96525	TEN3	Tenerife	Canary Is.	29	28	Y
		SAMN17349070		2N96527	TEN4*	Tenerife	Canary Is.			Y*
MW460787	MW460866	SAMN17349071	112582	18-609	TEN5	Tenerife	Canary Is.	29	28	Y
MW460788	MW460867	SAMN17349072		1N23582	TEN6	Tenerife	Canary Is.	29	29	Y
MW460789	MW460868	SAMN17349073		2L55177	TEN7	Tenerife	Canary Is.	29	30	Y
MW460790	MW460869	SAMN17349074	112584	18-611	TEN8	Tenerife	Canary Is.	29	31	Y
MW460791	MW460870	SAMN17349075		2L55179	TEN9	Tenerife	Canary Is.	29	32	Y
MW460748	MW460827	SAMN17349076		1N23571	GOM1	La Gomera	Canary Is.	30	33	Y
MW460749	MW460828	SAMN17349077		1N23572	GOM2	La Gomera	Canary Is.	29	34	Y
MW460750	MW460829	SAMN17349078		4L88825	GOM3	La Gomera	Canary Is.	28	35	Y
MW460751	MW460830	SAMN17349079		4L88826	GOM4	La Gomera	Canary Is.	28	35	Y
MW460752	MW460831	SAMN17349080		4L88827	GOM5	La Gomera	Canary Is.	30	33	Y
MW460753	MW460832	SAMN17349081		4L88828	GOM6	La Gomera	Canary Is.	30	36	Y
MW460733	MW460814		112404	16-046	PAL1	La Palma	Canary Is.	31	37	
MW460734	MW460815		112405	16-047	PAL2	La Palma	Canary Is.	31	37	
MW460735	MW460816		112406	16-048	PAL3	La Palma	Canary Is.	31	37	
MW460736	MW460817		112407	16-049	PAL4	La Palma	Canary Is.	32	37	
MW460737	MW460818		112411	16-053	PAL5	La Palma	Canary Is.	33	37	
MW460792	MW460871		112433	17-005	PAL6	La Palma	Canary Is.	31	37	
MW460793	MW460872		112434	17-006	PAL7	La Palma	Canary Is.	31	37	
MW460794	MW460873		112438	17-010	PAL8	La Palma	Canary Is.	33	37	
MW460795	MW460874		112444	17-016	PAL9	La Palma	Canary Is.	33	37	
MW460796	MW460875		112453	17-025	PAL10	La Palma	Canary Is.	34	37	
		SAMN17349082	112360	16-002	PAL11	La Palma	Canary Is.			Y
		SAMN17349083	112361	16-003	PAL12	La Palma	Canary Is.			Y
		SAMN17349084	112435	17-007	PAL13	La Palma	Canary Is.			Y
		SAMN17349085	112437	17-009	PAL14	La Palma	Canary Is.			Y
		SAMN17349086	112440	17-012	PAL15	La Palma	Canary Is.			Y
		SAMN17349087	112442	17-014	PAL16	La Palma	Canary Is.			Y

		SAMN17349088	112443	17-015	PAL17	La Palma	Canary Is.			Y
		SAMN17349089	112445	17-017	PAL18	La Palma	Canary Is.			Y
		SAMN17349090	112483	18-508	PAL19	La Palma	Canary Is.			Y
		SAMN17349091	112484	18-509	PAL20	La Palma	Canary Is.			Y
MW460755	MW460833	SAMN17349092	112566	18-593	HI1	El Hierro	Canary Is.	33	37	Y
MW460756	MW460834	SAMN17349093	112567	18-594	HI2	El Hierro	Canary Is.	33	37	Y
MW460757	MW460835	SAMN17349094	112568	18-595	HI3	El Hierro	Canary Is.	33	37	
MW460758	MW460836	SAMN17349095	112569	18-596	HI4	El Hierro	Canary Is.	33	37	Y
MW460759	MW460837	SAMN17349096	112570	18-597	HI5	El Hierro	Canary Is.	33	37	Y
MW460760	MW460838	SAMN17349097	112571	18-598	HI6	El Hierro	Canary Is.	33	37	Y
MW460761	MW460839	SAMN17349098	112573	18-600	HI7	El Hierro	Canary Is.	33	37	Y
MW460762	MW460840	SAMN17349099	112578	18-605	HI8	El Hierro	Canary Is.	33	37	Y
MW460763	MW460841	SAMN17349100	112579	18-606	HI9	El Hierro	Canary Is.	33	37	Y
MW460754	MW460842	SAMN17349101	112580	18-607	HI10	El Hierro	Canary Is.	33	37	Y

CHAPTER III: Habitat-driven adaptive divergence in the common chaffinch within a small oceanic island

María Recuerda, Mercè Palacios, Guillermo Blanco, Benoit Nabholz, Keith Hobson, Rafael Zardoya, Borja Milá

Pictures *F.c.palmae* by Borja Milá (modified).

Abstract

Populations adapted to contrasting habitat types can become genetically differentiated given strong enough environment-driven selection to counteract the homogenizing effect of gene flow. We test this hypothesis in the common chaffinch (*Fringilla coelebs*), a passerine on the small island of La Palma (Canary Islands), where it occupies two markedly different habitats, humid cloud forest and dry pine forest. Isotopic analysis indicated that birds in the two habitats have different diets, and analysis of phenotypic traits revealed significant differences in morphology and plumage coloration between habitats that are consistent with ecomorphological and ecogeographical predictions, respectively. A genome-wide survey of genetic variation using single-nucleotide polymorphism (SNP) loci revealed that neutral structure between localities was generally consistent with geography and isolation by distance, and reveals that dispersal is highly restricted in the species despite its capacity for flight. In contrast, Genome-Wide Association (GWA) and Genotype/Phenotype-Environment Association (GEA) analyses revealed marked adaptive divergence between birds in the two habitats. Loci associated with phenotypic and environmental differences among habitats were distributed across the genome, as expected for polygenic traits involved in local adaptation. Whole-genome sequencing (WGS) data from a subset of individuals from each habitat revealed numerous selective sweeps that were more prominent in the recently occupied pine forest. Our results suggest a strong role for habitat-driven local adaptation in population divergence in the chaffinches of La Palma, an excellent system for studying the evolutionary mechanisms of adaptive phenotypic divergence.

Keywords: *Fringilla coelebs*, GEA, GWAS, speciation, local adaptation.

Introduction

Patterns of phenotypic and genomic variation across heterogeneous landscapes provide the opportunity to investigate the processes and mechanisms that lead to divergence and eventually the formation of independent evolutionary lineages (Schluter, 2000; Schluter, 2001; Coyne & Orr, 2004). In the presence of environmental heterogeneity, selective pressures vary across space and this can drive phenotypic and genetic divergence among populations in order to increase their fitness in each environment, resulting in local adaptation (Kawecki & Ebert, 2004; Wang & Bradburd 2014). According to models of ecological speciation, adaptation to local environmental conditions can drive evolutionary divergence between populations even in the presence of gene flow, given that the magnitude of directional selection is strong and dispersal is low or non-random (Rundle & Nosil, 2005; Aeschbacher et al., 2017; Yeaman & Whitlock, 2011; Wang & Bradburd, 2014). In turn, the magnitude of gene flow is determined by the dispersal capacity of the species and the degree of geographic isolation (Lenormand, 2002). In organisms with high dispersal capacity, like birds, gene flow is considered to prevent divergence at small spatial scales and major geographic barriers to dispersal are thought to be necessary for genetic differentiation (Price, 2008). During the divergence process, geographic isolation can vary across time, including periods of sympatry and periods with restricted gene flow due to geographic barriers or other factors (Rundle & Schluter, 2004). Barriers to gene flow may arise between populations due to their respective interactions with the environment, which generates phenotype-environment correlations, resulting in ecological divergent selection (Rundle & Nosil, 2005; Nosil et al., 2008). Thus, in order to unravel the effects of divergent selection and gene flow, when phenotypic differences are found among populations at small spatial scales, is important to find the link between them, genetic variation and environmental heterogeneity. Recently evolved systems showing phenotypic variation across different environments provide the opportunity to study the process of local adaptation and reveal the genetic basis of fitness traits (Jones et al., 2012; Chaves et al., 2016; Szulkin et al., 2016; Wolf & Ellegren, 2016; Byers et al., 2017; Campagna et al., 2017; Meier et al., 2018; Mikles et al., 2020; Salmón et al., 2021).

Most fitness traits are quantitative and polygenic, thus selection is likely to act simultaneously on many loci of small effect, making more difficult their detection (Pritchard & Di Rienzo, 2010). Although less common in nature, there are also examples of divergence driven by selection acting on few genes of large effect (i.e., Barson et al., 2015, Knief et al., 2019). In order to find genetic evidence for divergence, molecular signatures of selection can be found across the genome (Barrett & Hoekstra, 2011). Advances in high-throughput sequencing methods have allowed the recognition of genomic footprints of recent adaptation by the identification of selective sweeps in non-model organisms (i.e., Hejase et al., 2020; Moest et al., 2020; Ravi Kumar et al., 2020; De La Torre et al., 2021). Selective sweeps leave diagnostic patterns of reduced genomic diversity around the selected locus, which leads to extended haplotypes due to genetic hitchhiking of the surrounding neutral loci (Berry et al., 1991; Nielsen, 2005). During hard selective sweeps, the selected variant appears as a novel mutation and goes to fixation rapidly, whereas soft selective sweeps arise either due to the selection of an allele that was already segregating in the population and thus increases its frequency, or by selection favoring several independent mutations at a specific locus, causing a concerted increase in frequency (Hermisson & Pennings, 2005; Pennings & Hermisson, 2006; Pritchard et al., 2010). Hard sweeps are easier to detect because the effects of soft sweeps on linked sites are weaker and could be indistinguishable from the genetic background (Pritchard et al., 2010; Pennings & Hermisson, 2006).

The common chaffinch (*Fringilla coelebs*) colonized the cloud forests on the island of La Palma in the Canaries Archipelago within the last 100K years (Chapter II, Recuerda et al., 2021a), and subsequently occupied the extensive dry pine forests covering most of the island, likely due to the absence there of its congeners the blue chaffinches, pine specialists in the larger islands of Tenerife and Gran Canaria (*Fringilla teydea* and *F. polatzeki*, respectively). On La Palma, the two adjacent habitats are markedly different in terms of plant species composition, vegetation structure, food resources, and climatic conditions (Irl & Beierkuhnlein, 2011), yet they are separated by narrow ecotones.

Here, we use a multidisciplinary approach to detect genomic divergence between chaffinch populations in the two habitats in the face of gene flow, and correlate the loci under selection with the divergent environmental variables and phenotypic traits. A wide range of methods for detecting signatures of selection is available, and it is the combination of several approaches, the best strategy to improve confidence at identifying candidate genes. In addition, it is important to control for population structure to avoid false positives (Bourgeois & Warren, 2021). Genetic-Environment Association (GEA; Hedrick et al., 1976) methods are designed to detect genetic changes explained by differences in environmental variables independently of geographic distance. Several methods, such as Redundancy Analysis (RDA, Legendre & Legendre. 1988) and BayPass (Gautier, 2015) can control for population structure in order to account for geographic distance and neutral genetic structure. Genome-Wide Association Studies (GWAS) are powerful methods to detect the correlations among SNPs and the phenotypic variation of the traits of interest, and at the same time take into account population structure (Bush & Moore, 2012; Forester et al., 2018). The advances in sequencing methods have facilitated the application of GWAS to several wild species (Ellegren, 2014; Schielzeth & Hysby, 2014; Santure & Garant, 2018). Both GEA and GWAS methods are useful for detecting weak signatures of selection distributed across the whole genome, usually generated by polygenic traits (Hoban et al., 2016; Forester et al., 2018; Santure & Garant, 2018).

We characterized phenotypic traits of the common chaffinch in both habitats by measuring morphological traits and plumage coloration, and also searched for differences in diet by analysing the feather isotopic composition as a diet proxy. We then used genotyping-by-sequencing (GBS, Elshire et al., 2011) and whole-genome resequencing to obtain SNP loci across the genome of common chaffinches from different populations of La Palma distributed in cloud and pine forests, respectively. Our goal was to (a) assess neutral population genomic structure, (b) detect the adaptive genomic variation explained by the differences in phenotype and environmental conditions performing GWAS and GEA, (c) search for signatures of recent selection by detecting hard and soft selective sweeps by combining complementary methods using whole-genome data, and (d) identify

candidate genes under selection associated with selective sweeps and explore their relationship with environmental variables and phenotypic traits.

Methods

Study area and fieldwork

We sampled common chaffinch populations across La Palma, the most northwestern island of the Canary Islands archipelago. The island is relatively small (708 km^2, 47 km long and 29 km wide), and reaches 2,426 m in elevation. The vegetation communities are shaped by the trade winds that blow in a NE to SW direction, which combined with the island's rugged topography leads to a wide range of local climatic conditions. Humid cloud forests occupy mid-elevations along the northern and eastern slopes and are dominated by evergreen trees from the Lauraceae family, tree heather (*Erica arborea*) and other fruiting shrubs (Fernández-Palacios, 2009). Extensive pine forests consist of open dry woodland dominated by the endemic Canary pine (*Pinus canariensis*) and extend between 500 m a.s.l. on the leeward side, and 1,100-2,000 m a.s.l. on the windward side (Fig. 1A).

Individual chaffinches were captured in the field using mist nets at five cloud forest and five pine forest localities between January and June of 2016-2020 under the corresponding sampling and handling permits (Fig. 1A). Individuals on the island are sedentary and remain on territory year round (BM Pers. obs.). Each individual was marked with a uniquely numbered aluminum band, sexed, aged and measured. A wing ruler was used to measure unflattened wing cord to the nearest 0.5 mm, and dial callipers of 0.1-mm precision were used to measure tail length, tarsus length, bill culmen (from the anterior end of the narine to the tip of the upper mandible), bill length (from the base of the beak to the tip of the upper mandible) and bill width and depth measured at the anterior end of the narine, following (Milá et al., 2008). All measurements were taken by a single observer (BM). A blood sample was obtained by venipuncture of the sub-brachial vein and stored in absolute ethanol at -20°C in the laboratory for DNA extraction. A few

body feathers (5-10 per body area) were collected from four different body areas (nape, back, rump and breast) per individual to measure plumage color in the laboratory, and one tail feather was collected to study the diet by isotope analysis. After processing, birds were released unharmed at the site of capture.

Phenotypic characterization and analysis

Ecomorphology

We compared the morphology of adult males in the cloud forest (n = 151) and pine forest (n = 113) habitats by comparing the means of morphological variables using univariate ANOVA or Kruskal-Wallis tests, adjusting p values for multiple tests with Holm's correction (Holm, 1979). Because of apparent differences in body size between birds in the two habitats, and in order to control for allometric effects, we corrected each variable for structural body size (Rising & Somers, 1989). We obtained marginal means from an analysis of covariance (ANCOVA) for each morphological variable previously log transformed, using habitat type as factor, and adding as a covariate the principal component 1 values from a principal component analysis (PCA) of all variables except the variable being tested (Jolicoeur, 1963, McCoy *et al.,* 2006). We performed this analysis in SPSS (v.22, IBM Corp, 2013). We also calculated the mean bill dimensions (width, length and depth) per locality and in order to test if the beak area was playing a role in thermoregulation, we approximated the total bill surface area using bill dimensions to calculate the lateral surface of a circular elliptical cone (Greenberg & Danner, 2012) and tested if it was correlated with local temperature.

Plumage coloration

To characterize differences in plumage color between habitats, we used a spectrophotometer to measure reflectance of body feathers from adult males collected in cloud forest (n = 108) and pine forest (n = 89). We used a JAZ-EL200 spectrophotometer with a Xenon light source via a bifurcate optical fibre probe (Ocean Insight™, Orlando, FL, USA). The reflectance captor probe was mounted in a black rubber holder that excluded all external light and we maintained the probe at a fixed distance of 3 mm from the feather surface at a 90° angle (Friis & Mila,

2020). For each individual and body patch, five feathers were piled on top of each other and three replicate measurements of three different readings per replicate were taken, randomly changing the order of the feathers each time. The spectrum of each measurement ranged from 300 to 700 nm and replicates were averaged before analysis. All reflectance data are expressed as the percentage of reflectance from a white standard (WS-1, Ocean Insight™). The white standard was measured after each specimen, and the spectrophotometer was recalibrated regularly. All measurements were taken by a single observer (MP). Negative values were transformed to zero and the noise in the reflectance curves was smoothed (0.2 value). We obtained colorimetric variables by applying the avian visual model by Stoddard and Prum (2008), based on Goldsmith (1990) tetrahedral colour space for spectral data. We used the R-package {PAVO} (Maia et al., 2019) to calculate the relative quantum catch for each cone using the function "vismodel" and obtained the spherical coordinates describing the hue (Θ and φ) and the achieved chroma (r) using the function "tcs". We included the normalized brilliance as a fourth variable, computed as described in Stoddard and Prum (2008). To analyze the colorimetric variables, we applied a non-parametric ANOVA using Euclidean distance matrices of standardized data (mean=0, SD=1) using the "adonis" function in the vegan R-package (Oksanen et al., 2019). Analyses were carried out with R version 4.1.0 (R Development Core Team 2021).

Environmental variables

To characterize the habitats within La Palma, we selected ten remotely-sensed environmental variables related to temperature, precipitation and vegetation cover based on previous knowledge of common chaffinch ecology. Variables were downloaded from Worldclim (Hijmans et al., 2005), averaged for a 30-year period (1970-2000), and included annual mean temperature (Bio1), isothermality (Bio3), temperature seasonality (Bio4), mean temperature of the warmest quarter (Bio8), annual precipitation (Bio12), precipitation seasonality (Bio15), precipitation of the driest quarter (Bio17) and precipitation of the warmest quarter (Bio18). For vegetation variables, we downloaded Normalized Difference Vegetation Index (NDVI, Fig. 1B) data from the MODIS satellite from NASA, which are calculated

every 16 days with a 250-m resolution (available at https://modis.gsfc.nasa.gov/data/), and the tree cover data for the year 2000 from (Hansen et al., 2013).

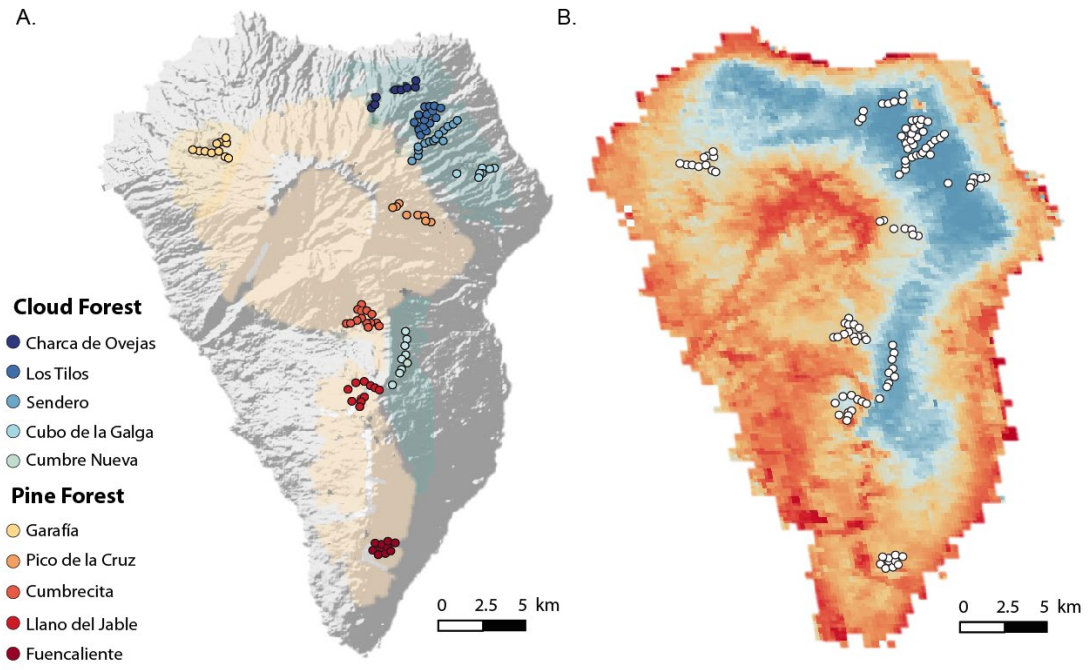


Figure 1. Sampling localities and distribution of habitat types on the island of La Palma. (A) Sampling sites, with cold and warm-colored markers corresponding to cloud and pine forest localities, respectively. Blue and light orange areas correspond to cloud and pine forest, respectively. (B) Map of La Palma showing the Normalized Difference Vegetation Index (NDVI) showing in colder and warmer colors regions with higher and lower greenness, respectively.

Diet characterization and analysis

To determine whether chaffinches in the two habitats had different diets, we obtained carbon and nitrogen isotope ratio data from tail feather fragments from a cloud forest locality (Los Tilos, n = 12) and a pine forest locality (Fuencaliente, n = 12), all of them from birds captured in 2016, following the methods in Hobson & Clark (1992). The ratio of stable isotopes of carbon ($^{13}C/^{12}C$) and nitrogen ($^{15}N/^{14}N$) in consumer tissues can be related to the isotopic composition of diet, and therefore are useful to understand dietary patterns in wild populations (Peterson & Fry, 1987). The results are "delta" (δ) values expressed in units of

permil (‰), which are presented relative to universal international standards to make these measures comparable across studies. The $\delta^{13}C$ allows determining the relative contributions of C_3 and C_4 plants to avian diets in areas where these two plant types coexist (Von Schirnding et al., 1982). There is a difference in the diffusion rate of ^{13}C from the atmosphere into the plant tissues between C_3, C_4 and Crassulacean acid metabolism (CAM) photosynthetic pathways, with C_3 plants being associated with cooler and wetter habitats and showing more depleted $\delta^{13}C$ values, in contrast to C_4 and CAM plants which are usually associated with hotter and drier environments and enriched $\delta^{13}C$ values (Michener & Lajtha, 2008). Stable isotopes of nitrogen have been used to determine an individual's position in the food web, but high levels of $\delta^{15}N$ can also indicate protein catabolism caused by nutritional stress (Hobson et al., 1993). Enriched levels of $\delta^{15}N$ are also usually associated with hotter and drier environments (Hobson, 1999). The ratios of carbon ($\delta^{13}C$) and nitrogen ($\delta^{15}N$) isotopes were obtained for each habitat and compared using univariate ANOVA.

Genotyping of genome-wide SNP loci

Genomic libraries from 200 individual chaffinches (Table S1) were constructed using genotyping-by-sequencing (GBS) (Elshire et al., 2011) from genomic DNA digested with the enzyme *PstI*. Sequencing was performed on a HiSeq2000 (Illumina) in a single lane as a multiplex using custom GBS barcodes. Reads were demultiplexed using the command axe-demux from the software AXE (Murray & Borevitz, 2018) and their quality was evaluated using FASTQC. Trimming and filtering was performed with TrimGalore v.0.3.7 (Krueger, 2015) to a minimum length of 40 and a mean genotyping phred quality score of 30, with no positions below 20. Reads were then mapped against the assembled chaffinch genome (GCA_015532645.2, Chapter IV: Recuerda et al., 2021b), using the mem algorithm in the Burrows-Wheeler Aligner (BWA, Li & Durbin, 2009). Variant calling was done with the Genome Analysis Toolkit version 3.6-0 (GATK, (Mckenna et al., 2010). First, we used the HaplotypeCaller tool, to call the individual and then the GenotypeGVCFs tool to join all the GVCFs files in vcf format. We used VCFTOOLS to filter the vcf file, keeping biallelic SNPs with depth coverage between 5 and 60 and

a genotyping phred quality over 30. Positions with a minor allele frequency (MAF) below 0.01, fewer than 75% of individuals genotyped, and those out of Hardy-Weinberg equilibrium (HWE) with a *p*-value below 10^{-4}, were removed from the dataset. We obtained 1,384,105,602 reads from the GBS sequencing, which resulted in 1,008,888,912 reads after trimming. BWA mapped 1,099,976,024 reads and 2,130,850 variants were called with GATK, which resulted in 9,960 SNPs after filtering with VCFTOOLS v.0.1.15 (Danecek et al., 2011).

Genome-wide population structure from SNP data

To explore genome-wide population structure we ran a PCA using a neutral SNP dataset. We filtered out the loci under selection detected with BayPass v.2.2 (Gautier, 2015) under the Standard model (see below), and then filtered out loci under linkage disequilibrium (LD) using the function snpgdsLDpruning from the {SNPRELATE} package (Zheng et al., 2012) in R version 3.2.2 (R Core Team 2015). We applied the correlation coefficient method with a threshold of 0.75 (method = 'corr', ld.threshold=0.75) to obtain 6,881 independent SNPs. We then performed the PCA using the function snpgdsPCA, also available in {SNPRELATE}. In addition, we conducted a maximum-likelihood estimation of individual ancestries using ADMIXTURE (Alexander et al., 2009) to infer population structure. We performed 200 runs for values of K from 1-10. The K value with the lowest cross-validation error was selected as optimal. We computed the weighted pairwise F_{ST} among populations and habitats using VCFTOOLS v.0.1.15 (Danecek et al., 2011). We also estimated the Expected and Observed Heterozygosities (H_e, H_o), genetic diversity (π) and the inbreeding coefficient (F_{IS}) per locality and per habitat using STACKS v.2.54 (Catchen et al., 2013).

We used the mean coordinates for each population to compute the geographic distance matrix using "distHarvesine" function from the R package geosphere v.1.5 (Hijmans, 2019). Pairwise F_{ST} values among populations were computed using VCFTOOLS (Danecek et al., 2011). We transformed the vcf file to genlight object using the Radiator package in R (Gosselin, 2019). To test for isolation-by-distance

we performed a Mantel test between the F_{ST} (from all, neutral and selected loci) and geographic distance matrices using the R package vegan (Oksanen et al., 2019) implemented in the "gl.ibd" function from the DARTR package in R (Gruber et al., 2018). We performed 10,000 permutations to test for significance.

Adaptive divergence across habitats

To assess the role of selection in driving genome-wide differentiation across habitats, we used redundancy analysis (RDA) (Legendre & Legendre, 1988), an ordination approach which allows estimating the variance in a response variable that is accounted for by a set of explanatory variables such as those related to environmental conditions. Moreover, a so-called conditioned RDA allows correcting for the effects of a set of covariables. First, we performed a regular, non-conditioned RDA to determine which environmental variables explained the most morphological differentiation among habitats, using the climatic conditions as explanatory variables. We selected NDVI as the main predictor variable and in order to select the rest of predictors, we evaluated the collinearity of the environmental variables, keeping those that showed the smallest Pearson correlation with NDVI ($|r| < 0.7$), as recommended by Dormann et al., (2013). All precipitation variables and tree cover were highly correlated with NDVI, whereas temperature variables showed low correlation with NDVI but high correlation among them, so we only selected the temperature variable showing the least collinearity with NDVI (i.e., temperature seasonality). We ensured that the variance inflation factor (VIF) of the retained variables was below 10 (Borcard et al., 2018) and a permutation test was performed on the final RDA, as recommended by Borcard et al., (2018). We also conducted a conditioned or partial RDA (pRDA) controlling for neutral population structure by including as a covariable the PC1 values from the genomic PCA performed with the neutral SNP dataset, using only the individuals with both morphological and genomic data (n = 151).

For plotting RDA results we used the correlation biplot (i.e., scaling=2) and the site scores. The statistical significances of the complete and per-axis models were

tested using an ANOVA-like permutation test setting $\alpha = 0.01$ and 10,000 permutations. The analyses were conducted in R version 3.2.2 using the R-package vegan (Oksanen et al., 2019).

Detection of outlier loci on the GBS dataset

To identify genetic markers under selection associated with specific covariates (i.e., environmental or phenotypic variables), we used two methods, RDA and BayPass (Gautier, 2015). The RDA approach to outlier detection is powerful for detecting adaptive loci even when selection is weak and polygenic (Forester et al., 2018). We performed a regular RDA and a pRDA (partial RDA controlling for neutral population structure as above) using the unlinked and neutral SNP dataset to detect outlier loci and their correlation with the selected environmental variables (i.e., NDVI and temperature seasonality) (Forester et al., 2018). As the RDA requires datasets without missing data, we imputed missing genotypes (17%) using the most common genotype across all individuals (Forester et al., 2018). We then performed an outlier analysis by setting the threshold for considering outlier loci at ±3 SD from the mean of the loading distribution of each axis. Finally, we identified the environmental variable that explained the most variance for each outlier SNP by correlating the observed allele frequencies across populations with each predictor (Forester et al., 2018).

BayPass accounts for shared demographic history by including the population genetic structure as a covariance matrix Ω (Gautier, 2015). First, we ran BayPass v.2.2 under the standard covariate model using the default importance sampling approach, scaling the environmental/phenotypic variables using the "-scalecov" option. This core model calculates the covariance matrix that is used to control for population structure when testing for associations between SNPs and environmental or phenotypic variables. We tested for associations between SNP frequencies and the two previously selected environmental variables (i.e., NDVI and temperature seasonality) and the three morphological bill traits that showed significant differences among habitats (length, width, and depth). We performed three independent runs with different initial seeds, setting 20 pilot runs of 5,000

iterations, a burn-in of 2,500 iterations, a sampling of 1,000 MCMC samples and a thinning interval of 15 iterations. We checked that all runs converged to similar estimates of the covariance matrix by calculating the pairwise Förstner and Moonen distance (FMD<0.2, Förstner & Moonen, 2003) among runs using the "fmd.dist" R function. In order to detect significant associations, we simulated pseudo-observed data (POD) of 10,000 SNPs for each environmental variable including the covariance matrix using the "simulate.baypass" R function. Then, we obtained the significance threshold to define outliers by calculating the top 5% quantile of the Bayes factor (BF) parameter of the simulated data for each climatic and phenotypic variable (BF_{NDVI} > -2.19, $BF_{temp\_seasonality}$ > -2.64, BF_{bill_width} > -2.75, BF_{bill_depth} > -1.81, BF_{bill_length} > -2.55). We also ran the auxiliary model ("covaux" option) with the covariance matrix from the standard model, scaling the covariables and the same MCMC settings. This model accounts explicitly for multiple testing and computes a BF that supports the association between each SNP and the selected covariables. Loci with BF > 10 were considered to be associated with the corresponding covariables. Finally, in order to assess the effect of selected SNPs on population structure, we also performed a PCA and an ADMIXTURE analysis with K ranging from 2 to 5 for every set of candidate loci detected by the pRDA and Baypass under the standard model using PLINK v.1.9 (Purcell et al., 2007).

Whole-genome resequencing

Resequencing of 12 individuals per habitat at 18X coverage was conducted using an Illumina SE50 platform at Novogene commercial services. Reads were trimmed with Trim galore (Krueger, 2015) and mapped against the common chaffinch reference genome (GCA_015532645.2, Chapter IV: Recuerda et al., 2021b) using BWA (Burrows-Wheeler Aligner). Variant calling was performed using BCFTOOLS v.1.3.1 including invariant sites. The variant calling format file (vcf) was filtered using VCFTOOLS v. 0.1.15 (Danecek et al., 2011) separately for variant and invariant sites. For both variant and invariant sites, we kept those with read depth between 10 and 40, and genotype quality over 30. Also, for variant sites we removed indels, sites with more than two alleles, sites with minor allele frequency

below 0.01, missing data over 25%, and with Hardy-Weinberg equilibrium threshold below $1 \cdot 10^{-4}$. Finally, variant and invariant sites were merged using BCFTOOLS concat.

Selection scans and candidate gene detection from whole genome sequences

We combined different methods in order to detect genomic signatures of selection among chaffinches in both habitats within the island of La Palma from whole-genome sequences. First, we calculated the fixation index (F_{ST}, Weir & Cockerham, 1984), d_{xy} and π using non-overlapping windows of 10 kb using Pixy v. 1.0.4 (Korunes & Samuk, 2021). Most programs that use VCF files to calculate d_{xy} and π do not differentiate between invariant and missing sites, resulting in deflated estimates, but Pixy takes into account both invariant and missing sites for π and d_{xy} calculations, thus overcoming the problem (Korunes & Samuk, 2021). We also calculated Tajima's D and the number of SNPs for each population in 10Kb windows with VCFTOOLS v. 0.1.15 (Danecek et al., 2011). The averaged values of each variable were transformed to z- scores using the "scale" command in R. We corrected for multiple testing setting the FDR to 0.05 (Benjamini & Hochberg, 1995). Then, we used two different methods to detect selective sweeps, XP-EHH and SweepFinder2.

To use XP-EHH, we first phased the data using SHAPEIT (Delaneau et al., 2013) using a 0.5 Mb window size, then we used the phased genotypes to calculate the extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2007), which compares at each position the integrated extended haplotype homozygosity (EHH) among two populations using the R package rehh (Gautier et al., 2017). A Z-test was applied to detect significant SNPs and the genomic regions showing a –log10 (*p*-value) ≥ 4 were considered to be under selection. We also computed the iHS statistic, which detects genomic signatures of selection within populations using the R package rehh with default options. This test is based on the genetic hitchhiking that leads to extended haplotypes around a focal allele under selection. The test calculates the standardized ratio of integrals of the EHH decay for ancestral and derived alleles of the focal SNP. A high iHS score is related with

longer haplotypes and reduced genetic diversity in the neighboring loci. As there was no available outgroup, the ancestral allele was inferred as the most common allele in the dataset. As with XP-EHH, a Z-test was applied to identify outlier SNPs and loci showing –log10 (*p*-value) ≥ 4 were considered to be under selection.

We also applied the composite likelihood ratio CLR test (Nielsen et al., 2005) using SweepFinder2 (SF2, Degiorgio et al., 2016), a method that compares the neutral model of evolution to a selective-sweep model while accounting for background selection, which can leave a similar genomic signal to that of selective sweeps. We used a precomputed background Site Frequency Spectrum (SFS) using the option "-lu" and the grid –points were defined at every variant except for scaffold "ScDMRwT_473_HRSCAF_551" that were defined every five variants to reduce computational time. To detect outliers for each population, we set the threshold at the 99.5th percentile of the CLR score distribution. Manhattan plots for all statistics were generated using the R package qqman v.0.1.8 in R v. 3.6 (R Core Team, 2018), and we examined the plots to detect overlapping outlier-rich regions shared by all methods.

To identify candidate genes potentially associated with the observed adaptive divergence, we extracted the annotation of the SNPs under selection detected by each method and each variable applied on the GBS and WGS datasets from the common chaffinch genome annotation (gff file) using bedtools intersect (Quinlan & Hall, 2010). We then obtained descriptions of putative functions and gene ontologies from www.genecards.org (Stelzer et al., 2016) and performed a bibliographic search. GO term enrichment analysis was performed using the R package TopGO (Alexa & Rahnenfuhrer, 2016), with 132 genes obtained by combining the genes putatively under selection from table 7 (see below) including only the genes related to NDVI and bill width, the genes shared between the GBS and WGS datasets, and also genes from table 10 detected by several methods in the WGS. The GO terms were obtained using the zebra finch dataset in biomaRt in R. We implemented the weight01 method with the Fisher statistic to assess significance. We used a gene universe containing 16,563 genes and 99 significant genes, out of which the software identified 9,065 from the gene universe and 85

significant genes to be used in the analysis. Following TopGO manual recommendations, we did not perform a correction for multiple testing and present the raw *p*-values for the top-ten GO terms associated with biological processes.

Demographic inference

Finally, we inferred the demographic history of the La Palma chaffinch populations using the whole genomes and implementing the Pairwise Sequentially Markovian Coalescent (PSMC) analysis (Li & Durbin, 2011). The PSMC model infers the demographic history based on genome-wide heterozygous sequence data. We obtained diploid consensus sequences from the BAM files generated with BWA-mem (Li & Durbin, 2009) using SAMTOOLS (Li et al., 2009). We kept sites with sequencing depth between 10X and 35X and only from autosomes because sex chromosomes can show different rates and patterns of evolution (reviewed by Irwin, 2018; Wright & Mank, 2013). We used the fq2psmcfa tool included in the PSMC software to convert the diploid consensus sequence to PSMC input files (psmcfa). Then the program PSMC was used to infer the population history with the options '-N25 –t5 –r1 –p "4+30*2+4+6+10'. The atomic time ("-p") interval was set following Nadachowska-Brzyska et al., (2016). Generation time was set to two years (Baker & Marshall, 1999) and we used a mutation rate of 4.6 e-9 mutations/site/generation (Smeds et al., 2016), which has been used in other avian systems for PSMC analysis (i.e., Ericson et al., 2017; Hanna et al., 2017; Sato et al., 2020; Campana et al., 2020). The plots were generated using the function "plot_psmc" within the PSMC software.

Results

Ecomorphology and plumage coloration

Birds in the cloud forest were generally larger and heavier than those in the pine forest (Fig. 2A). Controlling for structural body size to avoid possible allometric effects revealed that birds in the pine forest have longer, wider and shallower bills than those in the cloud forest (bill length $F_{1,261} = 24.54$, $p < .001$; bill width $F_{1,261} =$

7.34, $p < .01$; bill depth $F_{1,261} = 5.85$, $p < .05$) (Fig. 2), although without size correction, variation among localities was high (Fig. 3A-C).

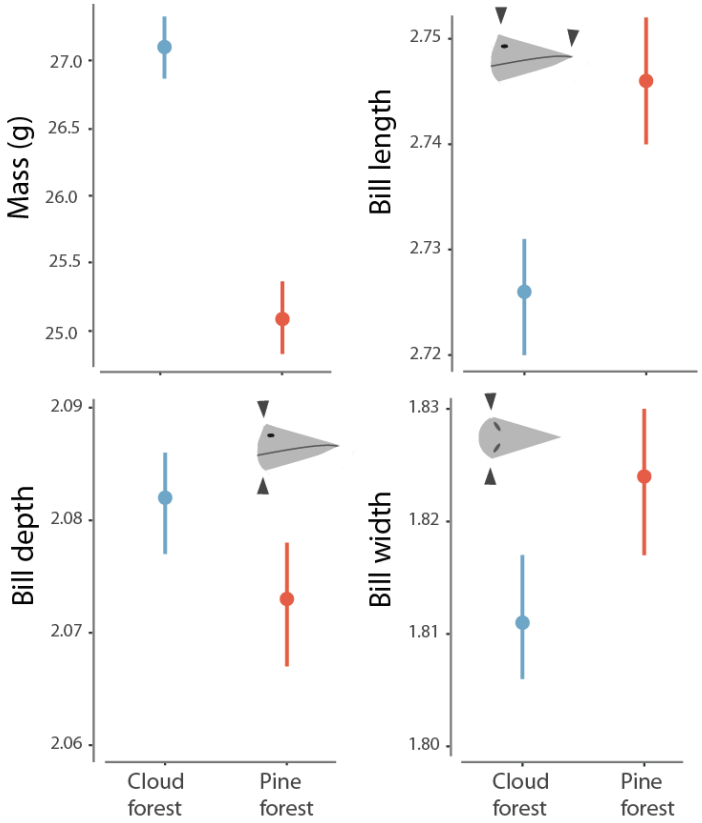


Figure 2. Divergence in morphological traits associated with habitat type. Differences between cloud and pine forest in mean mass (top left), and marginal means for size-corrected bill traits (bill width, bill depth and bill length). Vertical bars represent 95% confidence intervals.

We found no correlation between bill surface area and mean annual temperature ($F_{1,219} = 0.29$, $p = .59$), and bill surface area showed the highest values at Pico de la Cruz and Cumbrecita, two localities with very different temperature regimes (Fig. 3D and E). With respect to plumage coloration, we found significant differences among habitats in the hue (θ) of the four patches analyzed (nape $F_{1,197} = 7.13$, $p < .01$, back $F_{1,197} = 15.96$, $p < .001$, rump $F_{1,195} = 8.36$, $p < .005$ and breast $F_{1,197} = 26.75$, $p < .001$), and all breast variables were also significant (hue φ $F_{1, 197} = 33.82$, $p < .001$; chroma $F_{1, 197} = 18.44$, $p < .001$; brilliance $F_{1, 197} = 8.60$, $p < .005$) (Tables 1, 2).

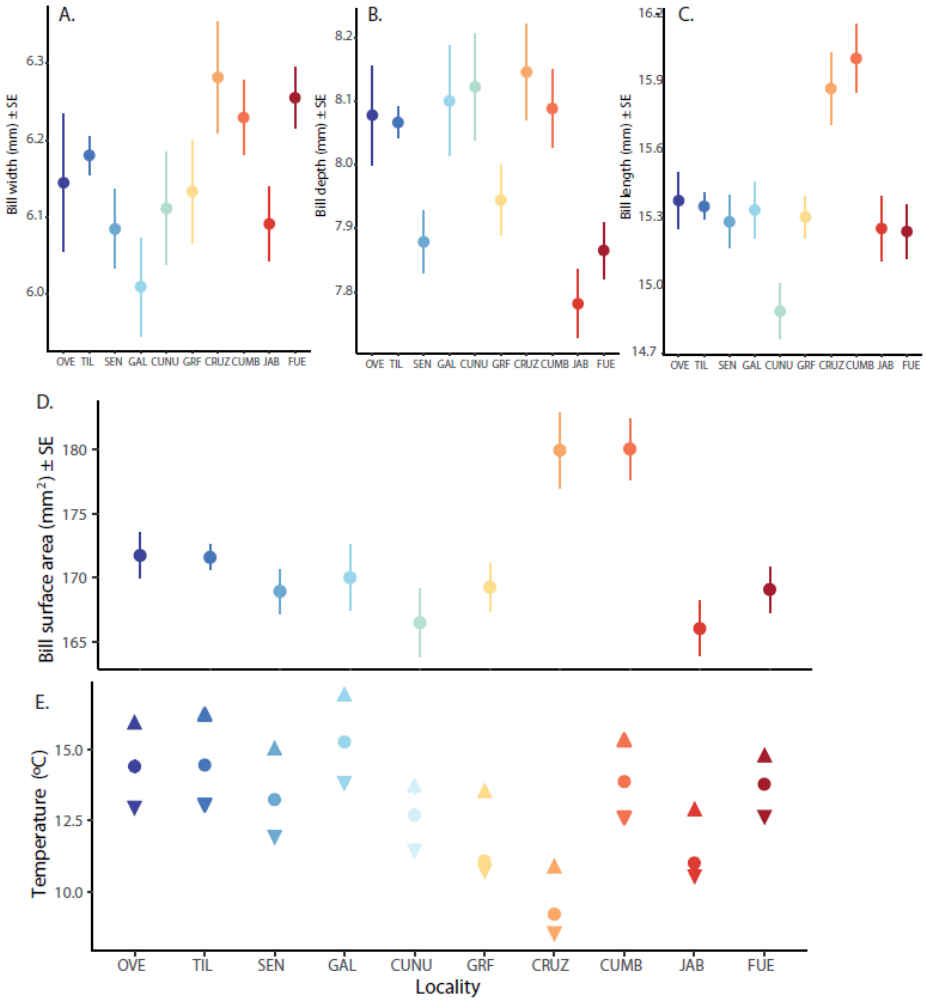


Figure 3. Bill dimensions, surface area and annual temperature regimes by locality. (A, B and C) Mean bill width, depth and length per locality in mm, without size correction, the error bars correspond to the Standard Error (SE). (D) Mean bill surface per locality in mm^2 and the error bars correspond to the Standard Error (SE). (E) Mean annual temperature (circles) by locality, including mean temperature of warmest quarter (above triangles) and mean temperature of coldest quarter (below triangles). Blue and red localities correspond to cloud and pine forest respectively. OVE (Charca de Ovejas), TIL (Los Tilos), SEN (Sendero), GAL (Cubo de la Galga), CUNU (Cumbre Nueva), GRF (Garafía), CRUZ (Pico de la Cruz), CUMB (Cumbrecita), JAB (Llano del Jable) and FUE (Fuencaliente).

Table 1. Comparison of plumage coloration variables among cloud and pine forest individuals for four different patches (nape, back, rump and breast). *F* statistic (degrees of freedom and sample size) and *p*-values adjusted for multiple testing using Holm's correction. Statistically significant comparisons are highlighted in bold.

	NAPE		BACK		RUMP		BREAST	
	F (1, 197)	adj *p*-value	*F* (1, 197)	adj *p*-value	*F* (1, 195)	adj *p*-value	*F*(1, 197)	adj *p*-value
Hue (θ)	7.133	**0.008**	15.961	**< 0.001**	8.367	**0.004**	26.754	**< 0.001**
Hue (φ)	2.344	0.127	2.657	0.104	1.226	0.270	33.820	**< 0.001**
Chroma	2.344	0.129	1.308	0.252	0.026	0.877	18.435	**< 0.001**
Norm Brilliance	0.117	0.734	1.368	0.274	1.027	0.313	8.608	**0.003**

Diet characterization

The isotopic analysis of the tail feather samples showed a significant difference among habitats in the $\delta^{13}C$ isotope (F = 48.47, $p = 5.48e^{-7}$), being more depleted in the wet cloud forest (Fig. 4). No difference in the $\delta^{15}N$ isotope was found between habitats (F = 0.52, $p = 0.48$).

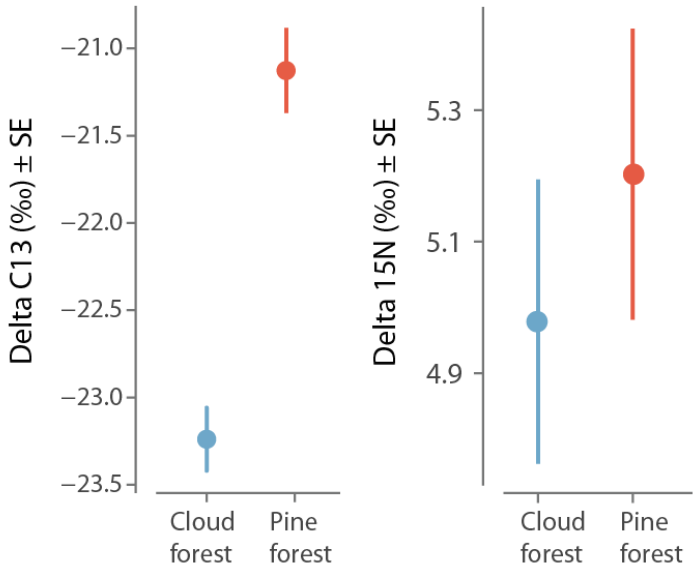


Figure 4. Differences in diet proxies among habitats. Mean feather isotopes $\delta^{13}C$ (‰) and for $\delta^{15}N$ (‰) for cloud forest (Los Tilos) and pine forest (Fuencaliente).Error bars represent Standard Error (SE).

Genome-wide population structure and isolation by distance

The PCA analysis showed no clearly differentiated clusters within the island, with individuals structured according to geography along the PC1, which explained 4.6% of the variance (Fig. 5).

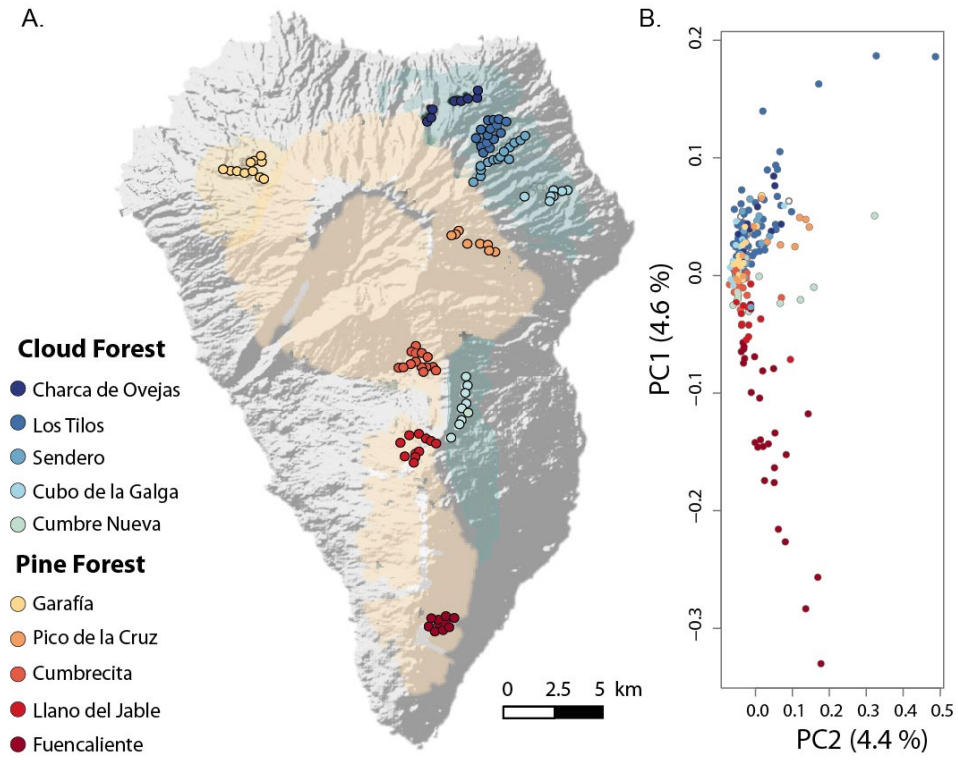


Figure 5. Sampling localities and patterns of neutral genetic variation of La Palma common chaffinches. (A) Distribution range of the common chaffinch, with blue and light orange areas corresponding to cloud and pine forest, respectively. Sampling sites, with cold and warm-colored markers corresponding to cloud and pine forest localities, respectively. (B) Principal components analysis plot, based on independent neutral SNPs (6,876 loci). The percent variance explained by each component is shown on the axes.

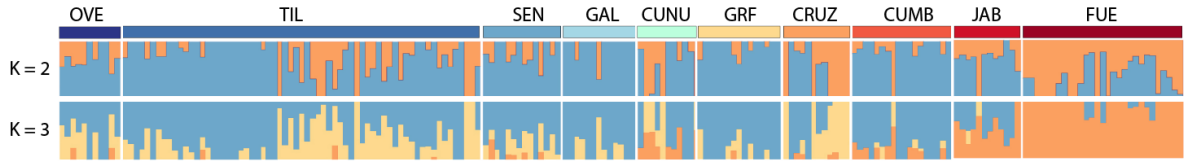


Figure 6. Admixture analysis of neutral genome-wide variation for the common chaffinch populations within the island of La Palma. Plots for K = 2 and 3 using a SNP dataset of 6,881 unlinked neutral loci.

Table 2. Mean values of color variables Hue (theta and phi), chroma and brilliance for the four color patches measured (nape, back, rump and breast).

		Nape	Back	Rump	Breast
Hue (theta)	Cloud forest	-0.90	-0.92	-0.81	-0.091
	Pine forest	-0.93	-0.93	-0.046	-0.080
Hue (phi)	Cloud forest	0.94	0.94	0.70	-0.88
	Pine forest	0.99	0.99	0.57	-0.89
Chroma	Cloud forest	0.26	0.26	0.20	0.41
	Pine forest	0.25	0.26	0.20	0.42
Brilliance	Cloud forest	$4.84 \cdot 10^{-3}$	$3.51 \cdot 10^{-3}$	$5.01 \cdot 10^{-3}$	0.24
	Pine forest	$4.79 \cdot 10^{-3}$	$3.47 \cdot 10^{-3}$	$4.57 \cdot 10^{-3}$	0.25

Regarding the Admixture analysis, and according to the smallest CV error, the structure is best explained by either two or three evolutionary clusters (Table 3), separating mainly Fuencaliente from the rest, and some individuals from the closer populations to Fuencaliente showing some admixture (Fig. 6). Values of genetic diversity, heterozygosity and the inbreeding coefficient showed very low values for all populations (Table 4). Genetic differentiation among populations was also very low, and increased only when using the outlier loci, even though the values remained low (Table 5).

Table 3. Coefficients of variation for the different K values in the ADMIXTURE Analyses using the dataset of neutral unlinked SNPs (6,876 loci, K ranging from 2 to 10), the outliers using NDVI as predictor from BayPass using the Standard model (50 loci, K ranging from 2 to 5) and the partial RDA (114 loci, K ranging from 2 to 5).

	CV error								
	K=2	**K=3**	**K=4**	**K=5**	**K=6**	**K=7**	**K=8**	**K=9**	**K=10**
Neutral	0.23	0.24	0.26	0.28	0.30	0.32	0.34	0.35	0.37
Outliers BayPass NDVI	0.35	0.47	0.53	0.61					
Outliers pRDA	0.42	0.44	0.47	0.52					
Outliers BayPass Bill width	0.28	0.30	0.32	0.33					

Table 4. Sample sizes, locality information and genetic diversity indices. Information per habitat and per population: sample size of SNP dataset, morphological data and individuals with both genetic and morphological data. Mean latitude, longitude and elevation per locality. For habitat, mean elevation for cloud and pine forest. Summary statistics calculated using STACKS v 2.54 from the vcf file containing 12,930 variant sites and including invariant sites from the GBS dataset: total number of sites (Sites), polymorphic sites (Polym), observed and expected heterozygosity (H_o, H_e, respectively), genetic diversity (π) and inbreeding coefficient (F_{IS}).

POP	N_SNPs	N_morph	N_both	Latitude	Longitude	Elev. (m)	Sites	Polym.	H_o	H_e	π	*Fis*
CLOUD	114	134	82	-	-	798	580,529	11,132	0.00074	0.0009	0.00092	0.00396
OVE	11	9	8	28.80901	-17.82527	809	417,723	3,798	0.00119	0.00136	0.00144	0.00113
TIL	65	86	42	28.78731	-17.80396	554	512,240	8,840	0.00086	0.00103	0.00104	0.00288
SEN	14	19	15	28.77965	-17.80051	1005	412,875	3,774	0.00099	0.0012	0.00125	0.00135
GAL	13	11	10	28.76108	-17.77891	530	380,359	3,256	0.00097	0.00116	0.00122	0.00122
CUNU	11	9	7	28.64116	-17.81913	1093	427,528	3,214	0.00103	0.00122	0.00129	0.00107
PINE	86	86	69	-	-	1245	558,098	9,857	0.00073	0.0009	0.00091	0.00343
GRF	15	18	13	28.77420	-17.93944	1218	393,005	3,322	0.00087	0.0011	0.00115	0.00143
CRUZ	12	11	9	28.73183	-17.81207	1682	494,077	4,019	0.00118	0.00132	0.00141	0.00095
CUMB	18	17	15	28.66633	-17.84358	910	375,763	3,768	0.0009	0.00107	0.0011	0.00137
JAB	12	11	10	28.61880	-17.84561	1303	370,788	3,032	0.00101	0.00115	0.00121	0.00098
FUE	29	29	22	28.51912	-17.83315	1110	433,533	5,361	0.00101	0.00121	0.00124	0.00181

Habitat-associated adaptive divergence

The RDA with the complete SNP dataset as the response variable showed a clear structure separating cloud and pine forest individuals along the first RDA axis, which correlates with NDVI, and explained 0.8 % of the variation (Fig. 7A). Cloud forest individuals were positively correlated with NDVI. The pine forest localities showed some structure along the second RDA axis that was related to temperature seasonality, which explained 0.5% of the variance. Overall, 114 loci showed a significant correlation with the environmental predictors, of which 70 were related to temperature seasonality and 44 to NDVI. These loci were mapped onto 22 and 8 genes, respectively (Table S2).

Table 5. Mean pairwise F_{ST} values per locality. Above the diagonal: F_{ST} values computed with neutral SNPs (9,882 loci, excluding the outliers detected by BayPass under the Standard method using NDVI, bill width and bill length). Below the diagonal: F_{ST} values based on the outliers (78 loci). Blue and red correspond to pairwise F_{ST} values among localities within cloud and pine forest respectively, and green F_{ST} values correspond to comparisons of localities between the two habitats.

	OVE	TIL	SEN	GAL	CUNU	GRF	CRUZ	CUMB	JAB	FUE
OVE		0.005	0.005	0.004	0.008	0.003	0.001	0.011	0.008	0.018
TIL	0.018		0.004	0.000	0.011	0.001	0.007	0.001	0.004	0.017
SEN	0.025	0.041		0.003	0.010	0.001	0.004	0.003	0.004	0.013
GAL	0.042	-0.005	0.014		0.009	0.004	0.004	0.003	0.006	0.017
CUNU	0.068	0.032	0.069	0.035		0.014	0.005	0.009	0.005	0.014
GRF	0.022	0.033	0.074	0.054	0.131		0.007	0.004	0.006	0.014
CRUZ	0.077	0.078	0.120	0.063	0.049	0.138		0.013	0.007	0.016
CUMB	0.064	0.062	0.094	0.079	0.095	0.088	0.075		0.003	0.012
JAB	0.085	0.064	0.089	0.065	0.050	0.134	0.098	0.075		0.003
FUE	0.098	0.112	0.119	0.126	0.085	0.132	0.134	0.062	0.040	

The complete model was statistically significant, yet only NDVI was significant when testing by axis. The partial RDA analysis controlling for neutral structure also showed a separation between individuals in cloud and pine forest (Fig. 7B), yet the model was no longer significant. The partial RDA found 230 outliers, 114 of them related to NDVI and 116 to temperature seasonality, which mapped to 28 and 38 genes, respectively (Table S2). Both methods showed 11 genes in common, of which only one (*col8a2*) was related to NDVI.

Detection of loci under selection on the GBS dataset

The BayPass analysis under both the standard and the auxiliary models revealed five shared genes for the association with NDVI, six for temperature seasonality, three for bill depth and five for bill width (Table S2). The standard model identified 120 genes associated with the environmental variables (64 with NDVI and 54 with temperature seasonality), and 153 genes were associated with phenotypic traits (52, 49, and 52 with bill length, depth and width, respectively; Table S2). The auxiliary model detected a strong association of six genes related to NDVI, 14 to temperature seasonality, 13 to bill length, 7 to bill depth and 14 to bill width, all showing some overlap with the standard model (Table S2). The population

structure revealed by the PCA performed with the SNPs identified by the pRDA and BayPass using the NDVI, and the outliers identified by BayPass using bill width corresponded with habitat type, the latter showing higher overlap among habitats (Fig. 8), but no clear structure was apparent using the loci associated with temperature seasonality and bill length and depth.

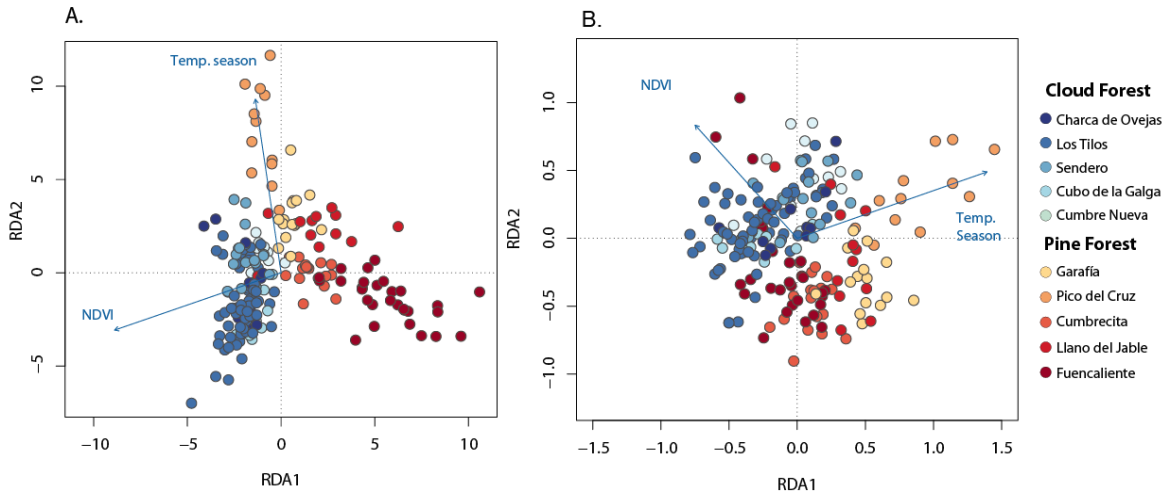


Figure 7. Habitat-associated adaptive genomic divergence. Plots showing results from redundancy analysis (RDA) performed with the complete SNP dataset: (A) regular RDA, and (B) Partial RDA controlling for neutral genetic structure. Vectors indicate the environmental predictors: temperature seasonality and NDVI. Each point represents an individual colored by locality. Blue and red markers correspond to cloud and pine forest localities, respectively.

The Admixture analysis performed with outliers showed structure by habitat, with the NDVI outliers detected by the partial RDA (K=3, Fig. 9A, Table 3) and BayPass under the Standard model (K=2, Fig. 9B). The outliers from Baypass under the Standard model using bill width as predictor showed a similar pattern except for Llano del Jable, which showed a higher assignment probability to the cloud forest cluster (Fig. 9C, Table 3). For the rest of the variables, as in the PCA, we did not detect habitat-related structure. The pattern of separation by habitat is found by different loci found by both methods using both phenotypic and environmental variables but the NDVI is the variable that better performs detecting habitat related loci (Fig. 8).

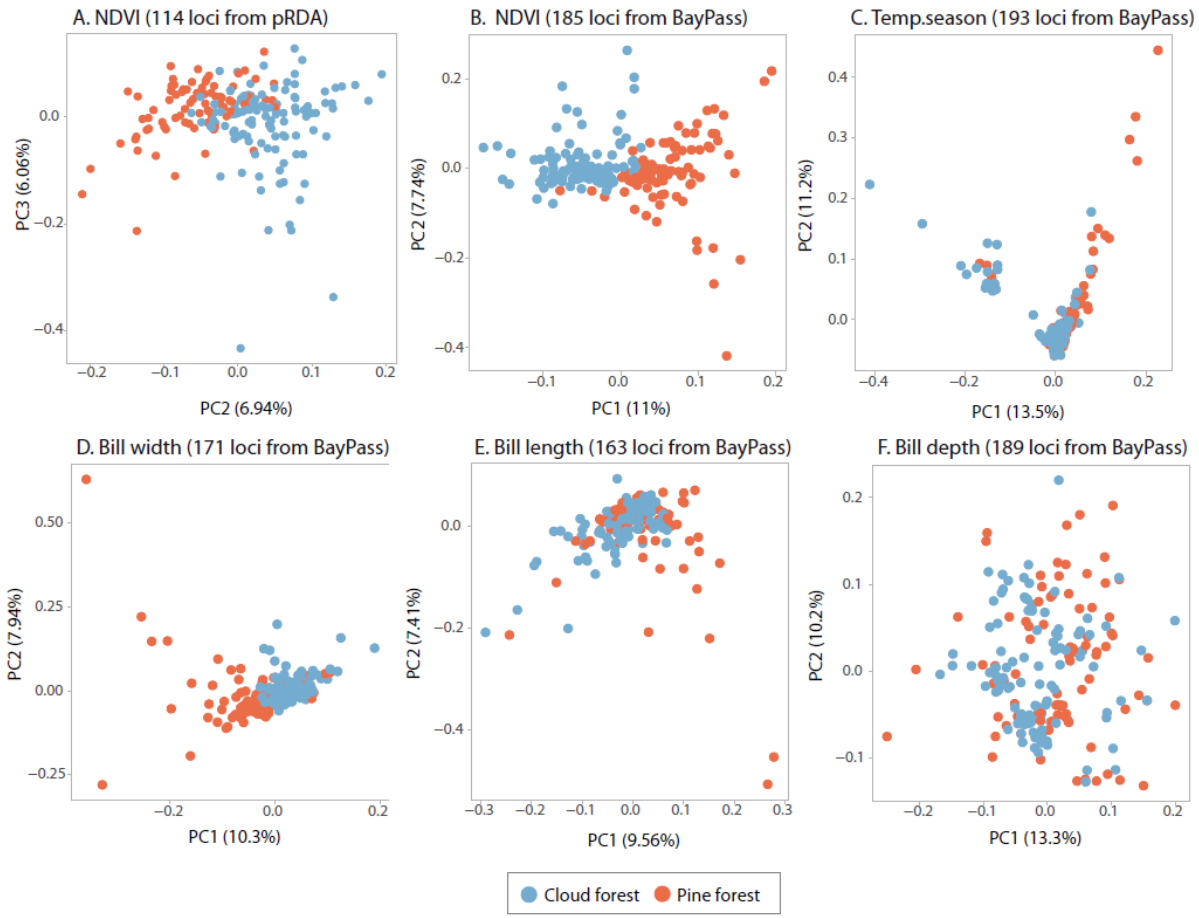


Figure 8. Adaptive divergence across habitat types. PCA plots based on outlier loci detected through partial RDA and BayPass, for loci associated with NDVI (A, and B), and for BayPass related with temperature seasonality (C), bill width (D), bill length (E) and bill depth (F). All the outliers from BayPass are under the Standard model. Blue and red dots correspond to individuals from cloud and pine forest, respectively.

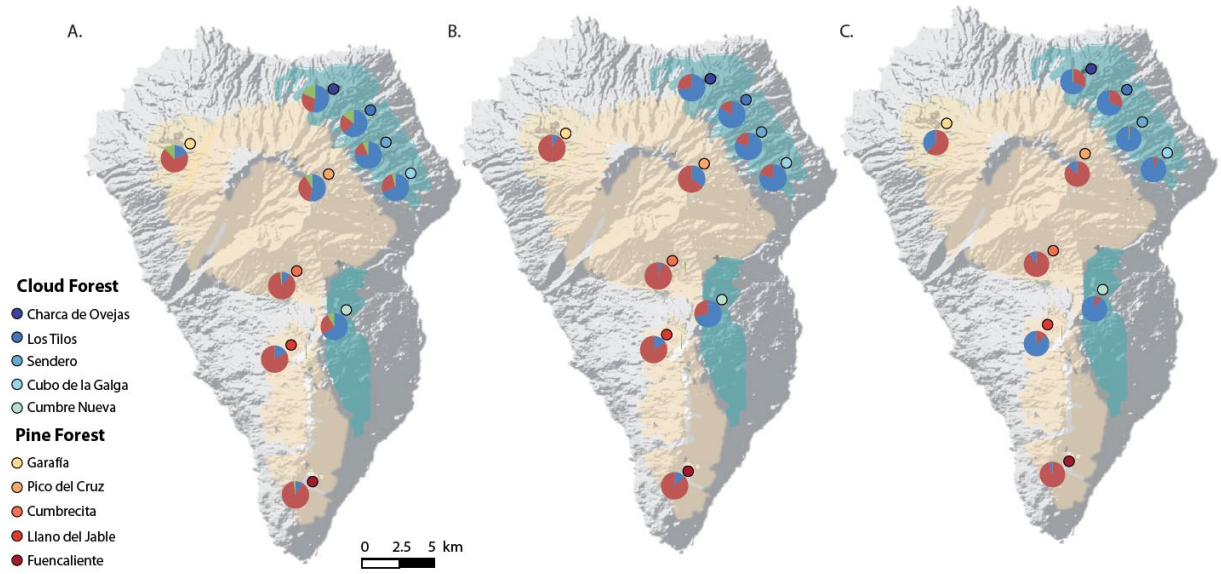


Figure 9. ADMIXTURE analysis performed with outlier loci. Pie charts per locality show the proportion of ancestry from the outliers detected by (A) the partial RDA using NDVI as predictor (K=3); (B) BayPass under the standard model using NDVI as predictor (K=2); and (C) BayPass under the standard model using bill width as predictor (K=2). The blue and light orange areas correspond to cloud and pine forest, respectively.

Candidate genes from the GBS dataset

There were common genes identified by several outlier-detection methods (Table S3). Both the pRDA and BayPass detected more genes in common using the same environmental variable, but there were a few genes that were also common among methods related to the other environmental variable (Table S3, Fig. 10A). Among the genes detected by BayPass related to bill dimensions, there was only one gene in common among the three (*tox2*), but there were 5 shared genes between length-width and width-depth and 8 among depth-length (Fig. 10B). A total of 17 candidate genes related to NDVI were detected by both the pRDA and BayPass, 10 of which were also detected by BayPass related to different bill dimensions (Table S3, Fig. 10C). Among the shared genes detected by pRDA and Baypass there were 10 related to temperature seasonality, of which only one was also related to bill (Table S3, Fig. 10D). Several candidate genes detected are related to known functions such as growth (*unc80*, *igf1r*, *cadps*), craniofacial development (*unc80*, *igf1r*, *med15*, *ndst1*, *fras1*, *dst*), melanogenesis, (*mitf*, *pold2*), vision (*cacna2d4*,

col8a2, *mttl24*), limb development (*cux1*) and metabolic pathways (*acss1*, *gmps*, *slc34a2*), which are consistent with the differences found in phenotype (i.e., morphology, plumage coloration and diet).

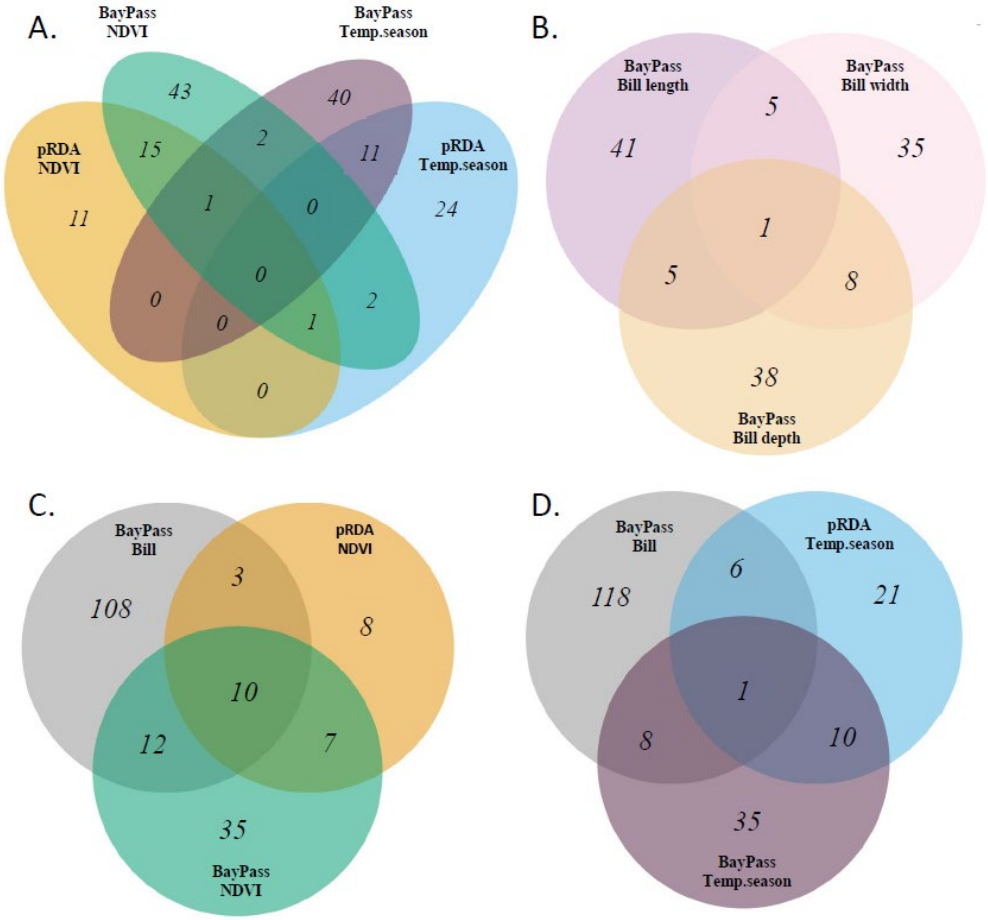


Figure 10. Venn diagrams showing the number of shared genes among outlier-detection methods and variables. (A) Common genes between the pRDA and BayPass related to NDVI and temperature seasonality; (B) Common genes detected by BayPass between the three bill dimensions (width, length and depth); (C) Common genes between all genes detected by all bill dimensions together, and genes related to NDVI detected by pRDA and BayPass; and (D) Common genes between all genes detected by bill and genes related to temperature seasonality detected by pRDA and BayPass.

The GO enrichment analysis performed with 132 genes including genes detected by several methods related to NDVI and bill width from the GBS dataset, the common genes between the GBS and WGS dataset and the genes detected by several methods from the WGS dataset, revealed that among the ten most

significant GO terms there were several related to the regulation of development, gene expression and cellular proliferation. These included negative regulation of osteoclast development (GO:2001205), small GTPase mediated signal transduction (GO:0007264), cellular response to amyloid-beta (GO:1904646) and regulation of catalytic activity (GO:0050790)(Table 6). We also found several GO terms related to neural development such as the roundabout signalling pathway (GO:0035385), axon extension involved in axon guidance (GO:0048846), and gamma-aminobutyric acid signalling pathway (GO:0007214)(Table 6).

Table 6. Top-ten GO terms obtained using topGO with 132 genes including from table S3 only the genes related to NDVI or bill width, and the common genes with the WGS dataset and genes detected by several methods in the WGS dataset from table 7.

GO.ID	Term	Annot.	Sign.	Exp.	Fisher
GO:2001205	Negative regulation of osteoclast development	2	2	0.02	7.70E-05
GO:0051414	Response to cortisol	3	2	0.03	0.00023
GO:0070100	Negative regulation of chemokine-mediated signaling pathway	3	2	0.03	0.00023
GO:0061364	Apoptotic process involved in luteolysis	3	2	0.03	0.00023
GO:0035385	Roundabout signaling pathway	4	2	0.04	0.00046
GO:0048846	Axon extension involved in axon guidance	10	2	0.09	0.00331
GO:0050790	Regulation of catalytic activity	910	14	8.53	0.00425
GO:0007264	Small GTPase mediated signal transduction	204	7	1.91	0.0054
GO:1904646	Cellular response to amyloid-beta	13	2	0.12	0.00563

Whole-genome resequencing and demographic inference

Variant calling identified 10,486,574 variants and 968,325,845 invariant sites, which after filtering resulted in 7,498,475 and 967,278,628 variant and invariant sites, respectively. The demographic inference using PSMC revealed that N_e showed a sharp decline about one million years ago that reduced the effective population size from 30 x 10^4 to 5 x 10^4 individuals, and remained stable after that (Fig. 11). Populations in both habitats showed identical patterns.

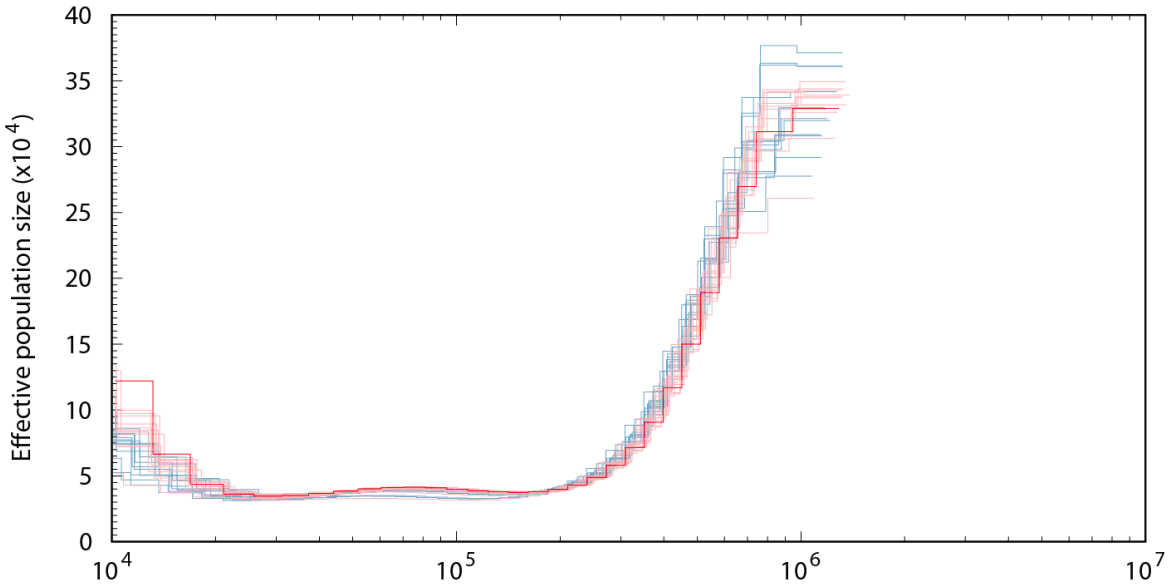


Figure 11. Demographic inference using Pairwise Sequentially Markovian Coalescent (PSMC) model. All the genomes from both habitats are represented. Blue and red lines correspond to cloud and pine forest, respectively. The mutation rate used was 4.6 e^{-9} mutations/site/generation and the generation time (g) was set to two years.

Genome wide scans of whole-genome sequences

F_{ST} outliers were distributed across the genome, with a low average F_{ST} between habitats (0.044) and the threshold for outliers was set at the top 1% (F_{ST} > 0.225) representing 149 genes (Table S4). Genomic divergence (d_{xy}) and genetic diversity of both populations (π) showed very similar genomic landscapes (Fig. 12). Some F_{ST} peaks coincided with drops in d_{xy} and π (Fig.12). These also coincided with the same regions found in the mainland-island comparison, identified as being consistent with the recurrent selection model (see Fig. 3 in Chapter 1). In contrast, Tajima's D showed clear drops in diversity that coincided with significant F_{ST} peaks, strong evidence for selective sweeps as detected by the SF2. However, an unexpected result is that the drops in Tajima's D coincided with CLR peaks in the population belonging to the opposite habitat type (Fig. 12). This pattern of low Tajima's D matched regions of higher SNP density within the same population, a pattern that is not expected in selective sweeps, suggesting other processes like introgression may be at work (Fig. 12). The iHS method detected 468 and 275 genes within the selected sweeps in the cloud and pine forest, respectively, and the SF2 detected 30 and 139 genes, respectively (Table S4).

Candidate genes from whole genome sequencing

There were few genes detected by the two methods used to detect selective sweeps within populations (iHS and SF2). Four genes were detected by both methods within the same populations, and eight in the opposite population (Table 7). Among the genes detected by both iHS and SF2, three were related to the "developmental biology" pathway (*slit2*, *slit3* and *myo10*), two to the ERK signalling pathway (*fbn1* and *cd8*), and *cd8* is also involved in the Wnt signalling pathway. Also, there was an overlap of 17 genes among the F_{ST} and the XP-EHH outliers. Among those 17 genes there were genes involved in different signalling pathways (RET, Notch and Akt) suggesting a major role in gene regulation. Moreover, there were 49 coincidences between the genes detected by all methods in the GBS dataset and the WGS dataset (Table S3).

Table 7. Genes putatively under selection detected by at least two methods using the whole-genome data. The methods include: F_{ST} scan (F_{ST}), Cross Population Extended Haplotype Homozygosity (XP), Integrated Haplotype Score (iHS) and SweepFinder 2 (SF2). The genes in common between iHS and SF2 are specified if they were detected in different (*) or the same population (=) and the population, "c" and "p" for cloud and pine forest, respectively.

Gene	Method	Name	Pathways	GO terms
ACER3	F_{ST} /iHSp	Alkaline Ceramidase 3	Metabolism and Sphingolipid metabolism	*Hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in linear amides* and *phytoceramidase activity*
ACTN2	F_{ST} /SF2c	Actinin Alpha 2	RET signaling and Post NMDA receptor activation events	*Calcium ion binding* and *protein dimerization activity*
ADAMTS12	F_{ST} /XP	ADAM Metallopeptidase With Thrombospondin Type 1 Motif 12	Diseases of glycosylation and Metabolism of proteins.	*Peptidaseactivity and metalloendopeptidase activity*
ANKHD1	F_{ST} /iHSc	Ankyrin Repeat And KH Domain Containing 1	X	*nucleic acid binding* and *RNA binding*
CEP152	SF2/iHS *p/c	Centrosomal Protein 152	Regulation of PLK1 Activity at G2/M Transition and Organelle biogenesis and maintenance.	*Protein kinase binding*
CFAP47	F_{ST} /iHSc	Cilia And Flagella Associated Protein 47	X	X
CORIN	F_{ST} /iHSc	Corin, Serine Peptidase	Myometrial Relaxation and Contraction Pathways and Cardiac conduction	*Serine-type endopeptidase activity* and *serine-type exopeptidase activity*
CDH8	SF2/iHS =c	Cadherin 8	WNT Signaling and Ectoderm Differentiation, Cell junction organization and ERK signaling	*Calcium ion binding*

DPF3	SF2/iHS *c/p	Double PHD Fingers 3		
FBN1	SF2/iHS =*p/cp	Fibrillin 1	ERK signaling, Elastic fibre formation, Integrin Pathway and Hypothesized Pathways in Pathogenesis of Cardiovascular Disease	*Calcium ion binding* and *extracellular matrix structural constituent*.
FBXW7	F_{ST} /XP	F-Box And WD Repeat Domain Containing 7	Notch Signaling Pathway,Cell cycle Role of SCF complex in cell cycle regulation and Chaperonin-mediated protein folding.	
FHOD3	F_{ST} /XP/iHScp	Formin Homology 2 Domain Containing 3	X	*Binding* and *actin binding*
FLI1	F_{ST} /XP/iHS	Fli-1 Proto-Oncogene, ETS Transcription Factor	Hematopoietic Stem Cell Differentiation and NF-kappaB Signaling.	*DNA-binding transcription factor activity* and *chromatin binding*
GABRA1	F_{ST} /XP	Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha1	Akt Signaling and Apoptosis Pathway.	*Drug binding* and *extracellular ligand-gated ion channel activity*
GABRA4	F_{ST} /XP	Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha4	Akt Signaling and Apoptosis Pathway.	*Chloride channel activity* and *GABA-A receptor activity*
GRAMD1C	F_{ST} /iHSp	GRAM Domain Containing 1C	X	Lipid binding
MAPK11	F_{ST} /XP	Mitogen-Activated Protein Kinase 11	RET signaling, Mitochondrial Gene Expression and Activated TLR4 signalling	*Transferase activity, transferring phosphorus-containing groups* and *protein tyrosine kinase activity*
MAPK14	F_{ST} /XP	Mitogen-Activated Protein Kinase 14	RET signaling, Mitochondrial Gene Expression and Activated TLR4 signalling	*Transferase activity, transferring phosphorus-containing groups* and *protein tyrosine kinase activity*.
MTMR8	F_{ST} /XP	Myotubularin Related Protein 8		*Phosphatase activity* and *phosphatidylinositol-3-phosphatase activity*.
MYO10	SF2/iHS *c/p	Myosin X	Developmental biology, PAK pathway, Actin Nucleation by ARP-WASP Complex and Regulation of actin dynamics for phagocytic cup formation	*Calmodulin binding* and *cytoskeletal motor activity*.
NCOR2	F_{ST} /XP	Nuclear Receptor Corepressor 2	fMLP Pathway and Signaling by NOTCH1	*Sequence-specific DNA binding* and *protein-containing complex binding*.
PDE4B	F_{ST} /XP	Phosphodiesterase 4B	cAMP signaling, Sweet Taste Signaling and Myometrial Relaxation and Contraction Pathways.	*Transmembrane transporter binding* and *cAMP binding*.
PDE4C	F_{ST} /XP	Phosphodiesterase 4C	Sweet Taste Signaling and G alpha (s) signalling events	*3',5'-cyclic-nucleotide phosphodiesterase activity* and *3',5'-cyclic-AMP phosphodiesterase activity*
RHOBTB3	F_{ST} /iHSc	Rho Related BTB Domain Containing 3	Vesicle-mediated transport and Golgi-to-ER retrograde transport	*GTP binding* and *small GTPase binding*
SCN1A	F_{ST} /iHSc	Sodium Voltage-Gated Channel Alpha Subunit 1	Developmental biology, Neuroscience and Activation of cAMP-Dependent PKA	*Ion channel activity* and *voltage-gated sodium channel activity*
SCN2A	F_{ST} /iHSc	Sodium Voltage-Gated Channel Alpha Subunit 2	Developmental biology, Neuroscience and Activation of cAMP-Dependent PKA	*Ion channel activity* and *voltage-gated sodium channel activity*
SCN8A	F_{ST} /iHSc	Sodium Voltage-Gated Channel Alpha Subunit 8	Developmental biology, Neuroscience and Activation of cAMP-Dependent PKA	*Ion channel activity* and *voltage-gated sodium channel activity*
SIPA1L1	SF2/iHS *c/p	Signal Induced Proliferation Associated 1 Like 1	Protein-protein interactions at synapses and Transmission across Chemical Synapses.	*GTPase activator activity* and *ephrin receptor binding*.
SLIT2	SF2/iHS =p	Slit Guidance Ligand 2	Signaling by slit, Developmental Biology, Guidance Cues and Growth	*Calcium ion binding* and *identical protein*

			Cone Motility and Signaling by Robo receptor.	*binding*
SLIT3	SF2/iHS =p	Slit Guidance Ligand 3	Signaling by slit, Developmental Biology, Guidance Cues and Growth Cone Motility and Signaling by Robo receptor.	*Calcium ion binding* and *Roundabout binding*.
TLCD4	F_{ST} /XP	TLC Domain Containing 4	X	*X*
TMEM56	F_{ST} /XP	Transmembrane Protein 56	X	*X*
TRIM8	SF2/iHS *p/c	Tripartite Motif Containing 8	Interferon gamma signaling and Cytokine Signaling in Immune system.	*Protein homodimerization activity* and *ligase activity*
LOC100220956	F_{ST} /XP	X	X	*X*
LOC100221786	F_{ST} /XP	X	X	*X*
LOC101232962	F_{ST} /XP	X	X	*X*
LOC107049564	SF2/iHS* p/c	X	X	*X*
LOC100229858	SF2/iHS *c/p	X	X	*X*

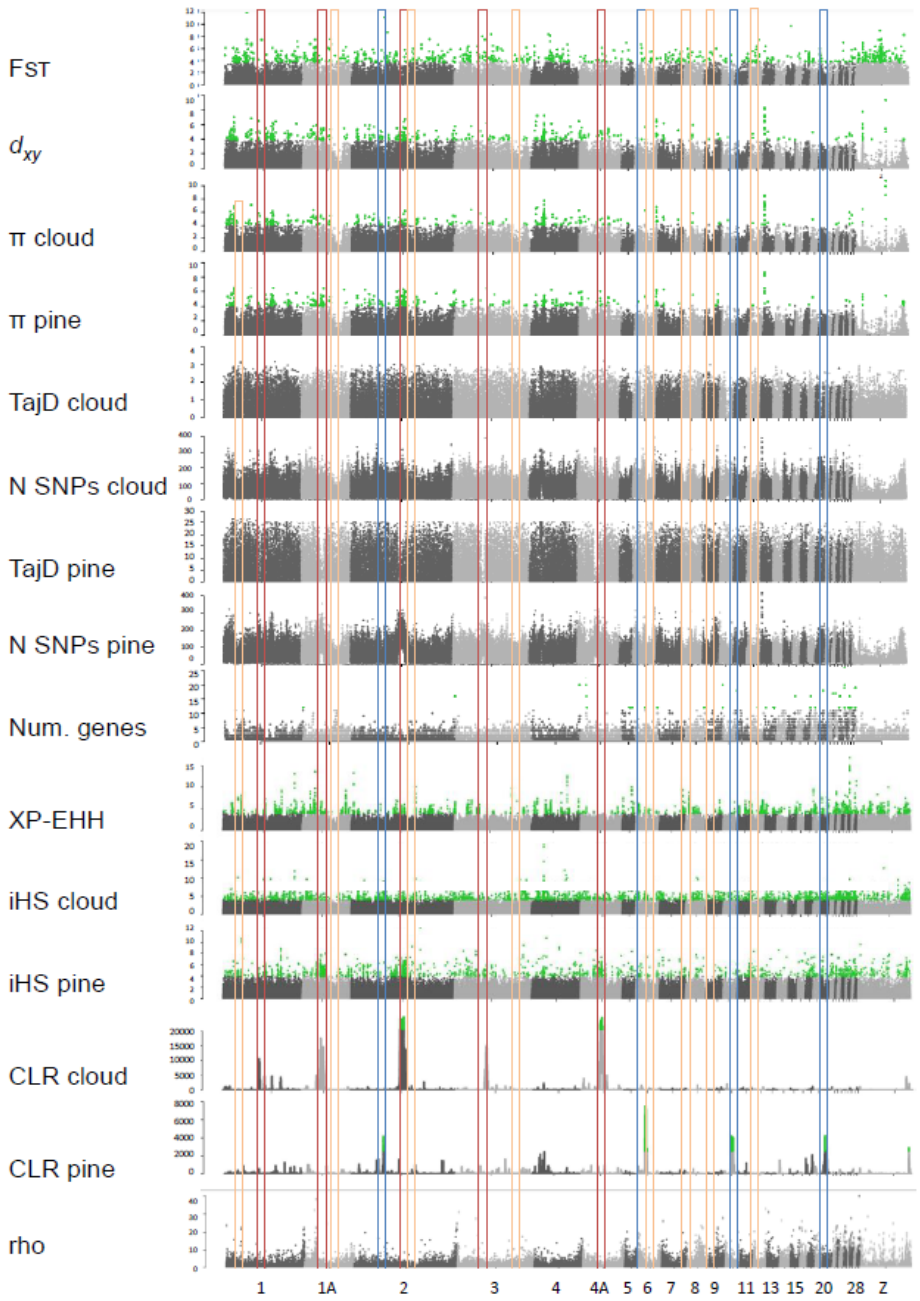


Figure 12. Genomic scans of chaffinches in cloud and pine forest. Pairwise genomic differentiation (F_{ST}) and genomic divergence (d_{xy}), population genetic diversity (π), Tajima's D, and number of SNPs (N SNPs) per habitat, number of genes, Cross population Extended Haplotype Heterozygosity (XP-EHH), Integrated Haplotype Score (iHS) per habitat, Composite Likelihood Ratio (CLR) and recombination rate (rho). The red and blue rectangles mark the selective sweeps detected by SF2 for cloud and pine forest, respectively, which correspond to a drop in Tajima's D value in the opposite habitat. The yellow rectangles show areas with high F_{ST}, but low d_{xy} and π that are also detected in the island-mainland comparison (see Chapter I). Chromosomes are shown alternating colors and are numbered along the X-axis according to the Zebra finch alignment.

Discussion

Overall, our phenotypic and genomic results indicate that chaffinch populations sampled in contrasting cloud and pine forest habitats on the island of La Palma have become locally adapted despite the presence of gene flow at very small geographic scales. Recent studies are showing examples of bird divergence at small scale (Bardwell et al., 2001; Ryan et al., 2007; McCormack & Smith, 2008; Bertrand et al., 2014; Langin et al., 2015; Szulkin et al., 2016; Dubuc-Messier et al., 2018; Bourgeois et al., 2020; Gabrielli et al., 2020; Cheek et al., 2022), indicating that marked reductions in dispersal upon colonization of oceanic islands may be general in small birds. As the common chaffinch arrived in a single colonization wave (Chapter II: Recuerda et al., 2021a), the differentiation within the island cannot be attributed to two separate colonizations. The shared demographic history between habitats revealed by the PSMC analysis confirms that divergence was *in situ*. Organisms such as birds are likely affected by gene flow at small spatial scales due to their high mobility. Within La Palma there are no major geographical barriers to dispersal, and the restriction in gene flow observed may be explained by the "insular syndrome" which can contribute to local adaptation by increasing the genetic differentiation at small spatial scale (Blondel, 2000; Whittaker & Fernández-Palacios 2007; Losos & Ricklefs 2009; Whittaker et al., 2017). Another important factor is the absence of competitors and therefore the possibility to expand to different habitats; for instance, in Tenerife and Gran Canaria, the blue chaffinches (*F. teydea* and *F. polatzeki,* respectively) occupy pine forest and the common chaffinch are found in more humid habitats, so that no overlap occurs (Lynch & Baker, 1991). The evidence of local adaptation related to habitat is striking given the low levels of genetic diversity and differentiation found within the island.

The ADMIXTURE analysis of independent neutral SNPs did not show a clear pattern, but clustering methods are less effective when characterizing patterns of continuous genetic variation than when populations are divided into relatively distinct groups (Petkova et al., 2016). The PCA performed with unlinked neutral loci showed slight structure that corresponded to geographical distribution across the island. Geography can influence population genetic structure, generating the

pattern of isolation by distance (IBD; Wright 1943). Since environmental variables are usually correlated with geographic distance (Shafer & Wolf, 2013; Wang, 2013), it is difficult to determine whether genetic distances are generated by environmental differences independent from geographic distance (IBE; Wang & Summers, 2010). Since we found support for IBD within the island of La Palma, we took into account neutral genetic structure when exploring the relationship between genotype and the environment, using outlier-detection methods that allow controlling for population structure in order to reduce the number of false positives (Rellstab et al., 2015; Forester et al., 2016).

The environmental variable that best represents the clear difference between the cloud and pine forests within La Palma is NDVI, a measure of "greenness" which distinguishes among vegetation types (i.e., Lauraceae vs *Pinus canariensis*). This also implies the availability of completely different resources which are important drivers of local adaptation and potentially ecological speciation (Rundle & Nosil, 2005). The differences found in diet proxies among the cloud and pine forest are found in the isotope ^{13}C and are probably related to the differences in vegetation between habitats. Chaffinches in cloud and pine forest showed δC^{13} values closer to C_3 and C_4 plants, respectively (Peterson & Fry, 1987; Hobson & Clark, 1992). This result is consistent with previous studies where wetter environments show more depleted $\delta^{13}C$ than drier environments (e.g., Marra et al., 1998). However, a more detailed diet analysis using meta-barcoding will be necessary to confirm these differences and further explore dietary changes across habitats.

We find differences between habitats in the three bill dimensions measured (i.e., width, depth and length), although variation among localities is high and larger sample sizes will be necessary to obtain definitive results. In any case, differences in beak shape are likely due to adaptation to the different available resources in both habitats, although performance experiments will be needed to confirm the link between beak shape and fitness. Our finding that bills are wider in the pine forest is consistent with results from Grant (1976), who suggested a relatively broader bill is useful to extract and crack pine seeds.

Identifying the genetic basis of complex polygenic traits remains a challenge, even for model species (Pritchard & Di Rienzo, 2010; Rockman, 2012). If the trait that favors the adaptation to new conditions is a continuous quantitative trait, adaptation would probably occur by small allele frequency shifts spread across many loci along the genome (Pritchard & Di Rienzo, 2010). As pointed out by Santure and Garant (2018), the GWAS approach has shown that most of the traits in nature are polygenic and therefore controlled by many loci of small effect; as shown by several studies in a wide range of organisms (i.e., *Arabidopsis thaliana* (Zan & Carlborg, 2019), *Mus* mice (Valdar et al., 2006), stickleback fish (Laurentino et al., 2020) and *Drosophila* (Ayroles et al., 2009). In our case, the adaptive phenotypes studied, such as beak shape, are complex polygenic traits (Abzhanov et al., 2006; Mallarino et al., 2011; Lamichhaney et al., 2015; Bosse et al., 2017; Lundregan et al., 2018; Cheek et al., 2022). The high evolvability of the beak allows birds to rapidly adapt to environmental changes (Grant & Grant, 2008) and several studies have shown that small differences in any of the three dimensions of the beak (depth, width and length) can have a major impact on fitness (Boag & Grant, 1981; Price et al., 1984; Grant, 1986; Gibbs & Grant, 1987).

Another possible explanation for changes in bill dimensions is thermoregulation, as bill surface has been shown to increase in hot, dry environments to facilitate heat dissipation (Greenberg & Danner, 2012; Tattersall et al., 2017). We found no clear evidence for this hypothesis in La Palma chaffinches, and found no correlation between bill surface and temperature, in contrast with other studies where birds found in colder environments had smaller beaks (Gamboa et al., 2022). However, birds at Pico de la Cruz, a high elevation locality with the lowest mean temperature, showed relatively large bill surface areas. Moreover, in the RDA this locality separates from the rest along the temperature axis, and some of the genes involved in bill morphology overlap with genes related to temperature, which may indicate that temperature has an effect in high-elevation populations. Further sampling will be necessary to properly test this hypothesis. Differences between habitats were also found in body mass, with cloud forest individuals being larger than those in pine forest. This is consistent with ecomorphological predictions as this trait could be related with habitat productivity (Boyer & Jetz,

2010), which is expected to be positively correlated with precipitation and temperature (Gustafson et al., 2017).

As expected for polygenic traits, outlier detection methods in the GBS dataset revealed multiple loci related to environmental and phenotypic variables. In cases where selection is spread across the genome, signatures of selection may be difficult to detect by changes in allele frequencies (Pritchard & Di Rienzo, 2010). The combination of different methodologies is recommended for the detection of SNPs under selection (Bourgeois & Warren, 2021), and in our case, the implementation of two different methods for detecting outlier loci showed limited overlap among them, but in cases of recent and weak selection it is unlikely that different methods will detect the same signals limiting the ability to find strong candidate genes (Forester et al., 2018). Despite the limited overlap, the population structure revealed by the PCAs performed with the outlier loci detected by pRDA and BayPass, showed structure among cloud and pine forest suggesting that all of them are strong candidates involved in local adaptation to the different habitats. This analysis also confirms that bill width is the best bill dimension explaining the habitat-related adaptation. Even though the pRDA model was not significant, we considered the outliers because the candidates associated with NDVI showed clear structure by habitat and there was some overlap with RDA and BayPass candidates. Previous studies have identified several genes involved in beak morphology in different avian species (Abzhanov, 2004; Wu et al., 2004; Lamichhaney et al., 2016; Chaves et al., 2016; Bosse et al., 2017; Bai et al., 2018). For instance, the BMP and Calmodulin pathways are known to be related to beak morphological variation (Abzhanov, 2004; Wu et al., 2004; Abzhanov et al., 2006; Wu et al., 2006). Moreover, Brugmann et al., (2010) showed experimentally that the Wnt signalling pathway is able to induce Bmp expression and therefore may function together with the BMP and Calmodulin pathways towards shaping facial morphology. Among the candidate genes related to NDVI identified by BayPass, we found two genes related to Wnt signalling (*cux1*, *med15*), other three associated to the BMP pathway (*cacna2d3*, *cacna2d4* and *col8A2*), and one related to the Calmodulin pathway (*myo3a*). Moreover, many of the candidate genes found are related to abnormalities of the head, face, neck and/or mouth in humans (i.e., *igf1r*,

dst, *ndst1*, *unc80*) and as pointed out by Brugmann et al., (2010) avian and mammalian facial development seem to share the same genetic bases. For instance, Bmp4 is also known to originate cleft lip in mice (Liu et al., 2005) and humans (Suzuki et al., 2009) and even in cichlids it is associated with variations in jaw morphology (Terai et al., 2002). Among our candidates, the *med15* gene interacts with the *srebf1* gene (found by BayPass and pRDA as an outlier with temperature seasonality) that has been related to beak morphology and also to craniofacial abnormalities in humans (Brugmann et al., 2010). We also found as a candidate gene the *igf1r* (Insulin-like growth factor 1 receptor, also detected as an outlier by the whole genome analysis), which has also been identified as an outlier related to bill morphology in the island scrub-jay (Cheek et al., 2022). The Igf1r mediates actions of the protein encoded by the *igf1* gene, which was found to be associated with bill size in the black-bellied seedcracker (*Pyrenestes ostrinus*) by vonHoldt et al., (2018). Among cloud and pine forest common chaffinches, we also detected differences in wing length and we find as a candidate the *cux1* gene which has been shown to be involved in the regulation of joint formation in the limb development in chicks (Lizarraga et al., 2002) and is related to the small wings causing flightlessness in the Galapagos cormorant (Burga et al., 2017). We also detected two genes related with temperature (*fat1* and *epha4*), the first of which is also associated with the flight loss in the cormorant (Burga et al., 2017) and the latter is related with limb development in quail (Arisawa et al., 2005).

Birds in the humid cloud forest had darker breasts than those in the drier pine forest, a result that is consistent with Gloger's rule, which states that birds and mammals in warmer and more humid habitats tend to be darker (Gloger, 1833; Delhey, 2017; Delhey, 2019). We find support for Gloger's rule but only in the breast, which is the only patch to show statistically significant lower brilliance in cloud forest individuals. The rest of the patches, showed the opposite pattern, but the differences are not significant. Several studies have shown that plumage coloration is associated with the light environment (McNaught & Owens, 2002; Gomez & Théry, 2004; Shultz & Burns, 2013; Ribot et al., 2019; Fang et al., 2022) and habitat type or other environmental variables (Friedman & Remeš, 2017; Ribot et al., 2019). Regarding the light environment, we found that in the brighter

pine forest habitat birds showed higher UV reflectance (i.e., higher h.phi) than in the darker cloud forest, but the difference was not significant. Previous studies (i.e., Fang et al., 2022) have found that plumages with better UV reflectance may be selected for in brighter habitats.

Sexual selection could also drive differences in plumage coloration due to its importance as a signalling and communication trait (Hill et al., 2006), which could potentially lead to premating reproductive barriers. In both the GBS and WGS analysis we detected the *cntn1* gene that has been detected within differentially methylated regions with respect to breast brightness and stress resilience in swallows (Taff et al., 2019). Moreover, using NDVI and bill width as predictors, BayPass identified the MITF gene (microphthalmia-associated transcription factor) which is another interesting candidate due to its function as a regulator of melanocyte development and its influence in the expression of other pigmentation genes (Tachibana et al., 1996). This gene has been found to be related to pigmentation in several mammals (Schmutz et al., 2009; Philipp et al., 2011; Yusnizar et al., 2015), fish (Li et al., 2013; Chen et al., 2020) and birds (Minvielle et al., 2010; Wang et al., 2014; Sultana et al., 2018). Poelstra et al., (2014) found the MIFT gene in a divergent genomic region differentiating carrion and hooded crows. Moreover, a transcriptomic analysis by Poelstra et al., (2015) found that even though MITF did not show significant differences in expression between carrion and hooded crows, it is involved in the downregulation of several genes that are related to gray feather melanogenesis in hooded crows. Among those genes they found *corin*, which was detected as an outlier in our WGS analysis and has been found to be associated also with plumage coloration in *Zosterops borbonicus* (Bourgeois et al., 2016) and the capuchino seedeater radiation (Campagna et al., 2017). We also detected, with both the GBS and the WGS data, the *camk2d* gene, which is also involved in cell communication during melanogenesis and it has been found in a highly differentiated region in the capuchino seedeater radiation (Campagna et al., 2017). Interestingly, we have also found several genes related to vision and ocular abnormalities (i.e., *cacna2d4*, *crybb3*, *col8a2*, *mettl24*) which can be associated with the need to adapt to the differences in light availability among habitats since the pine forest has an open canopy and is considerably brighter than

the cloud forest. For instance, the *cacna2d4* gene has been identified to be positively selected in the early history of owls, which are adapted to nocturnal conditions (Espındola-Hernandez et al., 2020).

We used whole-genome resequencing data to explore the genomic landscapes of individuals in the two habitats. Genomic scans showed regions with high relative divergence in F_{ST} but low absolute divergence (d_{xy}) and genetic diversity (π), a pattern consistent with recurrent selection (Nachman & Payseur, 2012; Cruickshank & Hahn, 2014; Irwin et al., 2016). As expected, those regions coincided regions showing the same pattern in the insular-mainland comparison (see Chapter I). In this within-island comparison we also found regions were relative and absolute divergences along with genetic diversity were high, a pattern consistent with divergence-with-gene-flow, except that genetic diversity is supposed to be low (Irwin et al., 2016; Irwin et al., 2018). The genomic scans also identified recent signatures of selection in the form of selective sweeps. As expected, several sweeps detected by iHS were coincident with regions of negative values of Tajima's D, indicating a frequency spectrum with an excess of rare variants (Braverman et al., 1995; Jensen et al., 2005). However, the pattern of SF2 was unexpectedly reversed between habitats, showing CLR peaks that were perfectly coincident with drops in Tajima's D in the opposite habitat. Those regions of low Tajima's D correspond to regions with high SNP density witin the same population. After checking individual heterozygosity values in order to test for possible contamination, we found that there were different individuals generating the pattern in each region, and that allelic depth was similar among all individuals, suggesting that introgression might be the cause of that high number of rare variants. Further analysis will be necessary to confirm this hypothesis.

Methods SF2 and iHS detected sweeps in the same regions where XP-EHH and F_{ST} scans detected common outliers, so that those regions are considered as strong candidates as targets of selection. iHS and XP-EHH are methods based on detecting perturbations of haplotype structure and can detect relatively soft sweeps (Pennings & Hermisson, 2006), while CLR (SF2) and Tajima's D are better at detecting hard sweeps (Williamson et al., 2007), so that both approaches can be

considered as complementary. The differences between methods could explain the low number of sweeps and genes detected by SF2 and suggest that the common sweeps detected by both methods are hard sweeps and that most of the sweeps identified by iHS and XP-EHH are relatively soft sweeps. Messer and Petrov (2013) argue that soft sweeps might be the main method of adaptation in many species but are usually difficult to detect. The sweeps detected by both iHS and XP-EHH are spread throughout the genome, as expected for polygenic traits. Accordingly, the outliers identified by the GEA and GWAS methods implemented in the GBS dataset also detected signatures of polygenic selection along the genome and showed some overlap with the outliers detected in the WGS dataset. We expected to find sweeps mainly in the pine forest, considering that is the newly colonized habitat because in the previous steps of the colonization the common chaffinch inhabits mainly in cloud forest, however with the iHS method, we found otherwise. We do not detect sweeps in common between habitats with the SF2 method but we do with the iHS, again suggesting that the SF2 detects hard sweeps. We have to consider also the possibility of global adaptation and therefore some of the sweeps and highly differentiated regions detected in each habitat may be just ephemeral genetic differentiation at the moment of the sampling that would disappear when mutations spread to all the populations (Booker et al., 2021).

The genes within selective sweeps detected by both iHS and SF2 are involved in pathways related to development (*slit2*, *slit3* and *myo10*), ERK (*fbn1*) and Wnt signalling pathways (*fbn1* and *cd8*). The Wnt signalling pathway is related to the Bmp pathway which is related to avian craniofacial morphology (Brugmann et al., 2010) and among the F_{ST} outliers we detected the *bmp5* gene, which has been shown to regulate skeletal anatomy, specifically in the nasal cartilages and ribs (Guenther et al., 2008) and it has also been associated with the development and growth of the goose knob (Ji et al., 2021). The *myo10* gene has also been related to craniofacial development in zebra fish (Yancey, 2015) and the *slit2* and *slit3* genes are involved in brain development (Andrews et al., 2007). Interestingly, the *slit1* gene, which was found as an outlier with the iHS method, has been found to be related to vocal learning in birds and is also associated with autism, dyslexia and speech sound language disorders in humans (Pfenning et al., 2014; Wang et al.,

2015). Other genes detected by both F_{ST} and XP-EHH are also involved in brain processes. For instance, there is evidence that the *mapk11* gene is related to vocal learning in the zebra finch (Lovell et al., 2008; Burkett et al., 2018). The *pde4b* gene has been suggested to play a role in learning processes in birds (Thompson et al., 2000) and the *mapk14*, *ncor2* and *pde4c* genes have been identified as neural targets of the *foxp2* gene, which affects speech and language disorders (Vernes et al., 2007). We also detected three genes (*scn1A*, *scnd2A* and *scn8A*) which are differentially expressed in the song system nuclei in zebra finches (Friedrich et al., 2019) . All these genes involved in vocal learning suggest that there might be differences in song between habitats which could be promoting the divergence or preventing gene flow across habitats. The *mapk14* gene has also been related to responses to external stress factors, and has been identified as a candidate for thermal adaptation in the chicken (Valero et al., 2014). Another interesting candidate gene is the Erythroblast transformation-specific (ETS) family transcription factor (*fli1*), which was detected by several methods in the WGS dataset. This gene is known to repress fibrillar collagen genes (Asano et al., 2009) and differences in collagen fibril composition could be responsible for structural coloration in birds (Saranathan & Finet, 2021). We find differences in the breast plumage coloration among habitats and male common chaffinches breast is an example of pigment structural color, it shows a pinkish-brown color, more orange in La Palma, that has been related to separated aggrupations of phaeomelanin granules that reflect the light (Frank, 1939; Ralph, 1969; Jeon et al., 2021). We also found the *ncor2* gene, which is within a genomic region that has been related to the orange pigmentation in the canary (Toomey et al., 2017), even though is related with carotenoid pigmentation which is not the common chaffinch case.

Overall, we identified genes associated with several pathways related to regulation of development, gene expression and cell proliferation as the most significant GO terms, suggesting that the phenotypic differences among habitats are probably due to changes in the genes involved in regulation. Interestingly, there were several genes related to neural development that could suggest differences in behavior among habitats (Nomura & Izawa, 2017). Moreover, it would also be interesting to study differences in vocalizations between cloud and pine forest individuals

because one of those terms (Roundabout signalling pathway) is related to bird vocal learning (Wang et al., 2015).

Conclusions

Our study highlights that when highly mobile organisms as birds occupy differentiated habitats, adaptive population structure can appear at small spatial scales without obvious barriers for dispersal, leading to phenotypic and genomic divergence due to local adaptation. The striking reduction in the common chaffinch dispersal showed by the neutral genomic structure is consistent with the insular syndrome in birds. Candidate genes related to habitat type known to be involved in bill morphology and plumage coloration in other bird species, which are consistent with the phenotypic differences detected among habitats, point toward local adaptation of the common chaffinch to cloud and pine forest. The genomic signatures of selection are spread throughout the whole genome and mainly affect genes involved in regulation as previously observed for polygenic traits. Future research will focus on understanding the importance of fitness of ecological and signalling traits, such as color and song, which might be acting as premating barriers and therefore contributing to non-random dispersal promoting local adaptation. The recent eruption of the volcano Cumbre Vieja may have had an impact on the common chaffinch and it will be necessary to continue monitoring the populations in both habitats to ensure their conservation.

Supplementary Materials

Table S1. List of the samples of *F. coelebs* used in the study, including Field and Specimen ID, latitude and longitude (Lat, long), Habitat and Locality information. The specimens marked with a "Y" in the field Morph had also morphological data.

Field ID	Specimen ID	LAT	LONG	Habitat	Locality	Morph
20-059	OVE1	28.80187	-17.83719	Cloud Forest	Charca de Ovejas	Y
20-060	OVE2	28.80623	-17.83656	Cloud Forest	Charca de Ovejas	Y
20-061	OVE3	28.80056	-17.83972	Cloud Forest	Charca de Ovejas	Y
20-063	OVE5	28.81152	-17.82067	Cloud Forest	Charca de Ovejas	
20-074	OVE6	28.80931	-17.82528	Cloud Forest	Charca de Ovejas	Y
20-075	OVE7	28.81064	-17.82278	Cloud Forest	Charca de Ovejas	Y
20-076	OVE8	28.81152	-17.82067	Cloud Forest	Charca de Ovejas	
20-077	OVE9	28.81152	-17.82067	Cloud Forest	Charca de Ovejas	Y
20-123	OVE10	28.81110	-17.81709	Cloud Forest	Charca de Ovejas	Y
20-124	OVE11	28.81332	-17.81665	Cloud Forest	Charca de Ovejas	Y
20-126	OVE12	28.81152	-17.82067	Cloud Forest	Charca de Ovejas	
19-001	GAL1	28.76107	-17.77943	Cloud Forest	Cubo de la Galga	
19-002	GAL2	28.76107	-17.77943	Cloud Forest	Cubo de la Galga	Y
19-003	GAL3	28.76007	-17.78059	Cloud Forest	Cubo de la Galga	
19-004	GAL4	28.76007	-17.78059	Cloud Forest	Cubo de la Galga	Y
19-005	GAL5	28.76007	-17.78059	Cloud Forest	Cubo de la Galga	Y
19-006	GAL6	28.76007	-17.78059	Cloud Forest	Cubo de la Galga	Y
19-007	GAL7	28.75890	-17.78034	Cloud Forest	Cubo de la Galga	Y
19-008	GAL8	28.76271	-17.77407	Cloud Forest	Cubo de la Galga	Y
19-009	GAL9	28.76263	-17.77590	Cloud Forest	Cubo de la Galga	Y
19-010	GAL10	28.76270	-17.77852	Cloud Forest	Cubo de la Galga	Y
19-011	GAL11	28.76242	-17.77870	Cloud Forest	Cubo de la Galga	Y
19-012	GAL12	28.76105	-17.77866	Cloud Forest	Cubo de la Galga	
19-014	GAL14	28.76127	-17.77844	Cloud Forest	Cubo de la Galga	Y
18-529	CUNU8	28.65284	-17.81758	Cloud Forest	Cumbre Nueva	Y
18-531	CUNU10	28.64954	-17.81824	Cloud Forest	Cumbre Nueva	Y
18-538	CUNU11	28.64195	-17.81773	Cloud Forest	Cumbre Nueva	Y
18-539	CUNU4	28.64195	-17.81773	Cloud Forest	Cumbre Nueva	
18-540	CUNU12	28.63964	-17.81936	Cloud Forest	Cumbre Nueva	Y
18-541	CUNU13	28.63759	-17.81637	Cloud Forest	Cumbre Nueva	Y
18-542	CUNU5	28.63371	-17.81968	Cloud Forest	Cumbre Nueva	
18-545	CUNU7	28.63035	-17.82126	Cloud Forest	Cumbre Nueva	
20-035	CUNU1	28.65273	-17.81708	Cloud Forest	Cumbre Nueva	Y
20-036	CUNU2	28.64844	-17.81864	Cloud Forest	Cumbre Nueva	Y
20-037	CUNU3	28.62408	-17.82676	Cloud Forest	Cumbre Nueva	
17-034	CUMB11	28.66779	-17.83996	Pine Forest	Cumbrecita	Y
17-035	CUMB12	28.66688	-17.84039	Pine Forest	Cumbrecita	Y
17-036	CUMB1	28.66214	-17.84242	Pine Forest	Cumbrecita	Y
17-038	CUMB15	28.66569	-17.84504	Pine Forest	Cumbrecita	Y
17-039	CUMB16	28.66216	-17.83706	Pine Forest	Cumbrecita	Y
17-040	CUMB17	28.66175	-17.83537	Pine Forest	Cumbrecita	Y

17-041	CUMB18	28.66166	-17.83961	Pine Forest	Cumbrecita	Y
17-042	CUMB19	28.66146	-17.84190	Pine Forest	Cumbrecita	Y
17-044	CUMB20	28.66457	-17.84638	Pine Forest	Cumbrecita	Y
18-570	CUMB10	28.66337	-17.84999	Pine Forest	Cumbrecita	
18-571	CUMB2	28.66305	-17.85091	Pine Forest	Cumbrecita	Y
18-572	CUMB3	28.66979	-17.84617	Pine Forest	Cumbrecita	Y
18-573	CUMB4	28.67009	-17.84540	Pine Forest	Cumbrecita	
18-574	CUMB5	28.67009	-17.84540	Pine Forest	Cumbrecita	Y
18-575	CUMB6	28.67011	-17.84394	Pine Forest	Cumbrecita	Y
18-576	CUMB7	28.67159	-17.84510	Pine Forest	Cumbrecita	Y
18-577	CUMB8	28.67089	-17.84469	Pine Forest	Cumbrecita	Y
18-578	CUMB9	28.67089	-17.84469	Pine Forest	Cumbrecita	
16-024	FUE1	28.51912	-17.83315	Pine Forest	Fuencaliente	
16-027	FUE2	28.51912	-17.83315	Pine Forest	Fuencaliente	
16-028	FUE3	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-030	FUE5	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-031	FUE6	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-032	FUE7	28.51912	-17.83315	Pine Forest	Fuencaliente	
16-033	FUE8	28.51912	-17.83315	Pine Forest	Fuencaliente	
16-035	FUE10	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-038	FUE11	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-039	FUE12	28.51912	-17.83315	Pine Forest	Fuencaliente	
16-040	FUE13	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-041	FUE14	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-042	FUE15	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-043	FUE16	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-044	FUE17	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-046	FUE18	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-047	FUE19	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-050	FUE22	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-051	FUE23	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-052	FUE24	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
16-053	FUE25	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
20-085	FUE26	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
20-090	FUE27	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
20-095	FUE28	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
20-096	FUE29	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
20-100	FUE30	28.51912	-17.83315	Pine Forest	Fuencaliente	
20-101	FUE31	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
20-102	FUE32	28.51912	-17.83315	Pine Forest	Fuencaliente	Y
20-103	FUE33	28.51912	-17.83315	Pine Forest	Fuencaliente	
19-051	GRF4	28.77759	-17.93611	Pine Forest	Garafía	
19-052	GRF5	28.77759	-17.93611	Pine Forest	Garafía	Y
19-053	GRF6	28.77596	-17.93766	Pine Forest	Garafía	Y
19-054	GRF7	28.77679	-17.93916	Pine Forest	Garafía	Y
19-055	GRF8	28.77586	-17.93555	Pine Forest	Garafía	Y
19-056	GRF9	28.77619	-17.93781	Pine Forest	Garafía	Y
19-057	GRF10	28.77478	-17.93848	Pine Forest	Garafía	Y
19-058	GRF11	28.77478	-17.93848	Pine Forest	Garafía	

19-086	GRF13	28.77254	-17.93839	Pine Forest	Garafía	Y
19-087	GRF14	28.76952	-17.93506	Pine Forest	Garafía	Y
19-088	GRF15	28.77065	-17.93611	Pine Forest	Garafía	Y
19-089	GRF16	28.77178	-17.94148	Pine Forest	Garafía	Y
19-090	GRF17	28.77248	-17.94444	Pine Forest	Garafía	Y
19-092	GRF19	28.77274	-17.94699	Pine Forest	Garafía	Y
19-093	GRF20	28.77371	-17.94985	Pine Forest	Garafía	Y
18-586	JAB9	28.62190	-17.84071	Pine Forest	Llano del Jable	Y
18-587	JAB10	28.62370	-17.84244	Pine Forest	Llano del Jable	Y
18-589	JAB12	28.62127	-17.83866	Pine Forest	Llano del Jable	Y
18-590	JAB13	28.62090	-17.83882	Pine Forest	Llano del Jable	Y
20-078	JAB1	28.62434	-17.85027	Pine Forest	Llano del Jable	Y
20-079	JAB2	28.62013	-17.85454	Pine Forest	Llano del Jable	Y
20-080	JAB3	28.61277	-17.85153	Pine Forest	Llano del Jable	Y
20-081	JAB4	28.62495	-17.84495	Pine Forest	Llano del Jable	Y
20-082	JAB5	28.61506	-17.84485	Pine Forest	Llano del Jable	Y
20-083	JAB6	28.61470	-17.84697	Pine Forest	Llano del Jable	Y
20-107	JAB7	28.61360	-17.84676	Pine Forest	Llano del Jable	
20-108	JAB8	28.61225	-17.84678	Pine Forest	Llano del Jable	Y
16-002	TIL7	28.79035	-17.80171	Cloud Forest	Los Tilos	
16-003	TIL8	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-004	TIL9	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-006	TIL11	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-007	TIL12	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-008	TIL13	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-009	TIL14	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-011	TIL16	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-012	TIL17	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-013	TIL18	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-014	TIL19	28.79035	-17.80171	Cloud Forest	Los Tilos	
16-015	TIL20	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-016	TIL21	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-017	TIL22	28.79035	-17.80171	Cloud Forest	Los Tilos	
16-018	TIL23	28.79035	-17.80171	Cloud Forest	Los Tilos	
16-019	TIL24	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-020	TIL25	28.79035	-17.80171	Cloud Forest	Los Tilos	
16-055	TIL27	28.79035	-17.80171	Cloud Forest	Los Tilos	
16-057	TIL29	28.79035	-17.80171	Cloud Forest	Los Tilos	
16-058	TIL30	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-059	TIL31	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-060	TIL32	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-061	TIL33	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-062	TIL34	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-063	TIL35	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-065	TIL36	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-066	TIL37	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-067	TIL38	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-068	TIL39	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
16-069	TIL40	28.79035	-17.80171	Cloud Forest	Los Tilos	

17-007	TIL1	28.79035	-17.80171	Cloud Forest	Los Tilos	
17-009	TIL2	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
17-012	TIL3	28.79035	-17.80171	Cloud Forest	Los Tilos	
17-014	TIL4	28.79035	-17.80171	Cloud Forest	Los Tilos	
17-015	TIL5	28.79035	-17.80171	Cloud Forest	Los Tilos	
18-508	TIL81	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
18-509	TIL82	28.79035	-17.80171	Cloud Forest	Los Tilos	Y
19-060	ESP1	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-061	ESP2	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-062	ESP3	28.78218	-17.80796	Cloud Forest	Los Tilos	
19-063	ESP4	28.78218	-17.80796	Cloud Forest	Los Tilos	
19-064	ESP5	28.78218	-17.80796	Cloud Forest	Los Tilos	
19-066	ESP7	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-067	ESP8	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-068	ESP9	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-069	ESP10	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-075	CAR1	28.78478	-17.80558	Cloud Forest	Los Tilos	Y
19-076	CAR2	28.78478	-17.80558	Cloud Forest	Los Tilos	
19-077	CAR3	28.78478	-17.80558	Cloud Forest	Los Tilos	
19-078	CAR4	28.78478	-17.80558	Cloud Forest	Los Tilos	
19-079	CAR5	28.78478	-17.80558	Cloud Forest	Los Tilos	Y
19-081	CAR7	28.78478	-17.80558	Cloud Forest	Los Tilos	Y
19-082	CAR8	28.78478	-17.80558	Cloud Forest	Los Tilos	Y
19-083	CAR9	28.78478	-17.80558	Cloud Forest	Los Tilos	Y
19-084	CAR10	28.78478	-17.80558	Cloud Forest	Los Tilos	Y
19-085	CAR11	28.78478	-17.80558	Cloud Forest	Los Tilos	
19-094	ESP11	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-095	ESP12	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-096	ESP13	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-097	ESP14	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-098	ESP15	28.78218	-17.80796	Cloud Forest	Los Tilos	
19-099	ESP16	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-100	ESP17	28.78218	-17.80796	Cloud Forest	Los Tilos	Y
19-101	CAR12	28.78478	-17.80558	Cloud Forest	Los Tilos	Y
19-102	CAR13	28.78478	-17.80558	Cloud Forest	Los Tilos	
18-525	CRUZ11	28.73108	-17.81683	Pine Forest	Pico de la Cruz	
18-526	CRUZ12	28.73108	-17.81683	Pine Forest	Pico de la Cruz	Y
18-527	CRUZ2	28.73646	-17.82387	Pine Forest	Pico de la Cruz	Y
18-528	CRUZ3	28.73692	-17.82279	Pine Forest	Pico de la Cruz	Y
18-546	CRUZ4	28.73156	-17.80580	Pine Forest	Pico de la Cruz	Y
18-547	CRUZ5	28.73122	-17.80643	Pine Forest	Pico de la Cruz	Y
18-549	CRUZ6	28.73131	-17.80970	Pine Forest	Pico de la Cruz	Y
18-551	CRUZ7	28.73814	-17.82295	Pine Forest	Pico de la Cruz	Y
18-553	CRUZ13	28.72844	-17.80407	Pine Forest	Pico de la Cruz	
18-554	CRUZ9	28.72844	-17.80407	Pine Forest	Pico de la Cruz	Y
18-555	CRUZ14	28.72869	-17.80537	Pine Forest	Pico de la Cruz	
18-556	CRUZ10	28.72866	-17.80610	Pine Forest	Pico de la Cruz	Y
19-030	SEN1	28.77710	-17.80651	Cloud Forest	Sendero	Y
19-031	SEN2	28.77872	-17.80384	Cloud Forest	Sendero	Y

19-032	SEN3	28.76990	-17.81003	Cloud Forest	Sendero	Y
19-033	SEN4	28.77229	-17.80930	Cloud Forest	Sendero	
19-034	SEN5	28.77489	-17.80838	Cloud Forest	Sendero	Y
19-035	SEN6	28.77987	-17.80173	Cloud Forest	Sendero	Y
19-037	SEN8	28.78093	-17.79928	Cloud Forest	Sendero	Y
19-038	SEN9	28.77975	-17.80023	Cloud Forest	Sendero	Y
19-039	SEN10	28.77973	-17.79900	Cloud Forest	Sendero	Y
19-041	SEN12	28.77902	-17.79564	Cloud Forest	Sendero	Y
19-070	SEN13	28.78814	-17.79032	Cloud Forest	Sendero	Y
19-071	SEN14	28.78658	-17.79208	Cloud Forest	Sendero	Y
19-072	SEN15	28.78516	-17.79345	Cloud Forest	Sendero	Y
19-074	SEN16	28.78303	-17.79739	Cloud Forest	Sendero	Y

Table S2. Candidate genes identified in the GBS dataset by the three separate methods: RDA, partial RDA and BayPass including the standard model (St) and the auxiliary model (AUX). The variables used were NDVI (N) and Temperature seasonality (T) for all models, and for BayPass we also used bill width (W), bill depth (D) and bill length (L). Genes highlighted in bold are shared between the RDA and the partial RDA, and between the Standard and auxiliary models using the same variable.

RDA (N)	pRDA (N)	RDA (T)	pRDA (T)	St (N)	AUX (N)	St (T)	AUX (T)	St (W)	AUX (W)	St (L)	AUX (L)	St (D)	AUX (D)
ANKRD6	ACSS1	**ACOX1**	**ACOX1**	AAK1	**BUB1**	ACOX1	ADAM15	AACS	**ADGRD1**	ACCS	**CDK1**	ACBD6	**CAST**
CADPS	ANKRD11	**ARHGAP19**	ALPI	ADGRD1	**GAS7**	AKIP1	**CDK1**	ABCG1	ARL3	ARHGAP18	**CDK3**	ATP6V0A1	GAS7
COL8A2	ATP8B4	DBR1	ALPL	AJAP1	NRF1	AMPH	**CDK3**	**ADGRD1**	C1H2orf49	ASZ1	**CDK5**	ATP6V0A4	**GMPS**
FRAS1	CACNA2D3	EIF4G3	**ARHGAP19**	AMFR	**POLD2**	ASZ1	**CDK5**	ASTN1	DHX30	ATP6V0A1	**CDK6**	**CAST**	**RRM2B**
LOC115491069	CACNA2D4	**EPHA4**	CTHRC1	ASB2	**RXRG**	BBS4	**CDK6**	ATP11A	**FBXO31**	ATP6V0A4	CFAP50	CCDC191	SRD5A2
METTL24	**COL8A2**	**FAT1**	DENND6A	ATF1	**UNC80**	BCO1	CENPV	ATP11C	**GMPS**	ATP8B4	FRMD4A	CCDC27	SRGAP1
NDST1	EP400	**FAT2**	DENND6B	ATP8B4		CAPZB	DENND1B	CACNA2D3	MAP2K5	B4GALT5	GAS7	CEP350	SRGAP3
NDST2	IGF1R	**FAT3**	DUSP3	BEGAIN		**CDK1**	DST	CACNA2D4	**MYO3A**	CAMK2A	GTF2H3	CSMD3	
	JMJD1C	FER	EDEM1	**BUB1**		CDK2	EPRS1	COL2A1L	PSD2	CAMK2D	LMINA	EBF2	
	LOC100221408	FER1L5	**EPHA4**	CACNA2D4		**CDK3**	**FAT2**	COMMD7	RGS3	CAMK2G	LMNA	EP400	
	LOC115491069	FES	**FAT1**	CDKN1A		**CDK5**	**FAT3**	CRYBB3	**SH3KBP1**	**CDK1**	LMNB2	ERBB4	
	LOC116807726	G3BP1	**FAT2**	CREB1		**CDK6**	LIN54	CTBP1	SPIRE2	CDK2	**SPSB3**	EZH2	
	MYO3A	G3BP2	**FAT3**	CRISPLD2		CELF4	NIN	DENND10	TEDC1	**CDK3**	TOX2	GIPC2	
	PCDH1	**IGDCC4**	FLT1	CUX1		CNTRL	WDR31	DHDDS	USP3	**CDK5**		**GMPS**	
	PDP1	**LOC417372**	FN1	DDHD2		DAZL		EHBP1		**CDK6**		HSD11B1a	
	PDP2	LRP1B	HSD17B12	DNAH10		DHX30		ELN		CNTN1		IGF1R	
	PDZRN4	PLOD2	**IGDCC4**	DOCK9		DNAH9		EP400		COMMD7		INTS8	
	POLD2	RNF216	KALRN	DPP10		DST		FAM45A		CSMD2		JADE2	
	PPTC7	SORL1	LATS1	DPP6		DUSP3		FAM65C		DENND6A		LHX4-AS1	
	RASGEF1A	**SREBF1**	LOC100218887	DST		EIF4G3		FBXO25		DENND6B		LOC100225021	
	RASGEF1B	**TNIK**	LOC100220302	EFHD1		ELP1		**FBXO31**		DNAI1		LOC100232465	
	RASGEF1C	WDR31	LOC115491069	EP400		ENGASE		FBXO32		FLII		LOC115491050	
	RIMBP2		LOC115497819	EPFIP1L		EPHA4		**GMPS**		FRMD4B		LOC115491069	
	RNH1		LOC116806803	FBXO31		FAT1		HCN3		FRMD5		LOC115491081	
	TSPAN6		LOC116807107	**GAS7**		**FAT2**		ISLR		HCN3		LOC115491097	
	UBR3		**LOC417372**	GTF2H3		**FAT3**		LOC100222482		HSD11B1a		MGAT5B	
	VEGFA		LRBA	IGF1R		G3BP1		LOC115491069		KDF1		MMP2	
	ZBTB49		MCM9	LAMA5		G3BP2		LOC116807107		LOC100220956		MOS	
			PAX4	LHFPL3		GAS7		LOC116807726		LOC100221408		MSH3	
			PRKCH	LOC100217869		GMPR		LRP1B		LOC100223101		NAA60	
			PRR5L	LOC100220728		GPRIN3		MITF		LOC100225021		P4HA1	
			PTPRD	LOC100221408		HSF4		**MYO3A**		LOC115492412		P4HA2	
			RHBDF1	LOC115490544		IFFO2		NDST1		LOC116809274		PATJ	
			SLC34A2	LOC115491069		IGDCC4		NDST2		MOV10L1		PRKAR1A	
			SREBF1	LOC116807107		IKBKAP		NIN		MYO1D		PRKAR1B	
			TBCEL	LOC116807726		ITIH5		PCDH1		NAA60		PRKCA	

	TNIK	LRP1B	LOC100218152	PDZRN4	PASK	PRKCB
	UTP6	MDGA1	LOC100220302	PNPT1	PDE4D	PRR5L
		MED15	LOC100220956	POLD2	PDZRN4	**RRM2B**
		MIPEP	LOC100225836	PRR5L	PIK3C2G	SLC12A5
		MITF	LOC417372	PTPRD	PRKAR1A	SLC12A7
		MROH7L2	MYLIP	RIC8B	PRKAR1B	SLC13A2
		MYO3A	PDE4D	RIPOR3	RRM2B	SORL1
		PDZRN4	PGM1	**SH3KBP1**	RXRG	SYTL4
		POLD2	PLOD2	SLC34A2	SAMD11	THADA
		PPTC7	POGLUT1	SLC6A1	SLC34A2	TMEM132C
		RALGAPA2	PON2	SLC6A2	SPG11	TOX2
		RASGEF1A	PPTC7	SLC6A6	TGFBI	WDR7
		RASGEF1B	PRR5L	SLC9A9	TOPAZ1	XRCC2
		RASGEF1C	RBM33	TOX2	**TOX2**	
		RXRG	SPSB3	UNC80	UNC80	
		SLC34A2	SREBF1	XRCC2	VEGFA	
		SLC6A1	TAB3			
		SLC6A2	USP36			
		SLC6A6				
		SMKR1				
		SSH1				
		SYTL4				
		TSPAN6				
		UBR3				
		UNC80				
		USP3				
		ZBTB49				
		ZNF804B				

Table S3. Candidate genes identified in the GBS dataset by the three separate methods (RDA, partial RDA and BayPass) using NDVI (N) and Temperature seasonality (T), and for BayPass using also bill width (W), bill depth (D) and bill length (L). Genes identified by the RDA are just marked by "N/T", those identified by the partial RDA are marked with (p), and those identified by both are marked with an (X). The model used for BayPass is specified in brackets: Standard model (S), the Auxiliary model (*), or both (S*). Gene names preceded by an asterisk (*) have also been detected in a high F_{ST} region or within a selective sweep in the whole-genome dataset.

GENE	RDA/ pRDA	BAYPASS	Name	Pathways	GO
*AKIP1		T (S)	A-Kinase Interacting Protein 1		
*AMPH		T (S)	Amphiphysin	Vesicle-mediated transport and Clathrin -mediated endocytosis.	*Phospholipid binding*
*ARHGAP19	T(X)		Rho GTPase Activating Protein 19	Signaling by GPCR and Signaling by Rho GTPases.	*GTPase activator activity*
*CACNA2D3	N (p)	W(S)	Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 3	BMP Pathway and ITK and TCR Signaling	*Ion and calcium ion transport*
*CACNA2D4	N(p)	NW (S)	Calcium Voltage-Gated Channel Auxiliary Subunit Alpha2delta 4	Arrhythmogenic right ventricular cardiomyopathy and T CR Signaling	*Voltage-gated calcium channel activity and calcium channel regulator activity*
*CAMK2D		L(S)	Calcium/Calmodulin Dependent Protein Kinase II Delta	RET signaling and Development Angiotensin activation of ERK	*Protein homodimerization activity and protein kinase activity*
*CAST		D (S*)	Calpastatin	Degradation of the extracellular matrix and Neurodegenerative Diseases	*RNA binding and cysteine-type endopeptidase inhibitor activity*
*CEP350		D (S)	Centrosomal Protein 350		*Microtubule binding*
*CNTN1		L(S)	Contactin 1	Signaling by NOTCH1 and Developmental biology	*Carbohydrate binding and protein binding*
*COL2A1L		W (S)	X	X	*X*
*CSMD2		L(S)	CUB And Sushi Multiple Domains 2	X	*X*
*DENND6A	T(p)	L(S)	DENN Domain Containing 6A	Vesicle-mediated transport and RAB GEFs exchange GTP for GDP on RABs	*Guanyl-nucleotide exchange factor activity*
*DENND6B	T(p)	L(S)	DENN Domain Containing 6B	Vesicle-mediated transport and RAB GEFs exchange GTP for GDP on RABs	*Guanyl-nucleotide exchange factor activity*
*DNAH9		T (S)	Dynein Axonemal Heavy Chain 9	X	*ATP hydrolysis activity and cytoskel etal motor activity*
*DPP10		N (S)	Dipeptidyl Peptidase Like 10	X	*Serine-type peptidase activity and dipeptid yl-peptidase activity*

*DPP6		N (S)	Dipeptidyl Peptidase Like 6	X	*Serine-type peptidase activity and dipeptid yl-peptidase activity*
*DST		N (S)	Dystonin	Cell junction organization and Cytoskeleton remodeling Neurofilaments.	*Cytoskeleton organization and cell adhesion*
*EHBP1		W (S)	EH Domain Binding Protein 1	X	*Actin cytoskeleton organization and protein transport*
*EPHA4	T (p)	T (S)	EPH Receptor A4	GPCR Pathway, ERK signaling and EPHA forward signaling	*Identical protein binding and protein kinase activity*
*FBXO25		W (S)	F-Box Protein 25	X	*Actin binding and ubiquiti n-protein transferase activity*
*FBXO32		W (S)	F-Box Protein 32	Factors and pathways affecting insulin-like growth factor (IGF1)-Akt signaling and innate inmune system	*X*
*FLT1	T (p)		Fms Related Receptor Tyrosine Kinase 1	GPCR Pathway and ERK signaling	*identical protein binding and protein kinase activity*
*FRMD4A		L (*)	FERM Domain Containing 4A	X	*Protein binding, bridging and establishment of epithelial cell polarity*
*FRMD5		L(S)	FERM Domain Containing 5		*Cytoskeletal protein binding*
*HCN3		LW(S)	Hyperpolarization Activated Cyclic Nucleotide Gated Potassium Channel 3	Potassium Channels adn Transmission across Chemical Synapses	*Oon channel activity and cAMP binding.*
*IFFO2		T (S)	Intermediate Filament Family Orphan 2		*Structural molecule activity*
*ITIH5		T (S)	Inter-Alpha-Trypsin Inhibitor Heavy Chain 5		*Serine-type endopeptidase inhibitor activity*
*JADE2		D (S)	Jade Family PHD Finger 2	Chromatin organization	*Ubiquitin protein ligase activity*
*KALRN	T (p)		Kalirin RhoGEF Kinase	EPH-Ephrin signaling and p75 NTR receptor-mediated signalling	*Transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity*
*LIN54		T (*)	Lin-54 DREAM MuvB Core Complex Component	Regulation of PLK1 Activity at G2/M Transition and Cell Cycle, Mitotic	*DNA binding*
*LOC100220 728		N (S)	X	X	*X*
*LOC100220 956		LT(S)	X	X	*X*
*LOC100225 836		T (S)	X	X	*X*
*LOC115490 544		N (S)	X	X	*X*
*LOC115491 069	N(X)T(p)	DNW(S)	X	X	*X*
*LRP1B	T	NW (S)	LDL Receptor Related Protein 1B	X	*Calcium ion binding and low-density lipoprotein particle receptor activity*

*MAP2K5		W (*)	Mitogen-Activated Protein Kinase Kinase 5	IL-2 Pathway, CNTF Signalingand TGF-Beta Pathway	*MAPK cascade and protein phosphorylation*
*METTL24	N		Methyltransferase Like 24	X	*Methyltransferase activity*
*MGAT5B		D (S)	Alpha-1,6-Mannosylglycoprotein 6-Beta-N-Acetylglucosaminyltransferase B		*Alpha-1,6-mannosylglycoprotein 6-beta-N-acetylglucosaminyltransferase activity.*
*MROH7L2		N (S)	X	X	*X*
*MSH3		D (S)	MutS Homolog 3	Mismatch Repair and Homology Directed repair	*Protein homodimerization activity and single-stranded DNA binding*
*NRF1		N (*)	Nuclear Respiratory Factor 1	Glucose / Energy Metabolism and Apelin signaling pathway.	*RNA polymerase II proximal promoter sequence-specific DNA binding and proximal promoter DNA-binding transcription activator activity, RNA polymerase II-specific*
*PASK		L(S)	PAS Domain Containing Serine/Threonine Kinase	Glucose / Energy Metabolism	*Transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity*
*PATJ		D (S)	PATJ Crumbs Cell Polarity Complex Component	Cell junction organization and Sertoli-Sertoli Cell Junction Dynamics	*X*
*PIK3C2G		L(S)	Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Gamma	MAPK Pathway and Apoptosis pathway	*transferase activity, transferring phosphorus-containing groups and kinase activity*
*PTPRD	T (p)	W(S)	Protein Tyrosine Phosphatase Receptor Type D	Commom Cytokine Receptor G Chain Siganaling Pathway and PAK pathway	*signaling receptor binding and protein tyrosine phosphatase activity*
*RGS3		W (*)	Regulator Of G Protein Signaling 3	Signaling by GPCR and Ephrin B reverse signaling.	*Signal transduction and inactivation of MAPK activity*
*SPG11		L(S)	SPG11 Vesicle Trafficking Associated, Spatacsin		
*SRGAP1		D (*)	SLIT-ROBO Rho GTPase Activating Protein 1	Developmental Biology, Signaling by Robo receptor and Signaling by Slit	*GTPase activator activity and small GTPase binding*
*SRGAP3		D (*)	SLIT-ROBO Rho GTPase Activating Protein 3	Developmental Biology, Signaling by Robo receptor and Signaling by Slit	*GTPase activator activity and small GTPase binding*
ACOX1	T (p)	T (S)	Acyl-CoA Oxidase 1	Metabolism	*Signaling receptor binding and protein N-terminus binding*
ADGRD1		N(S) W(*)	Adhesion G Protein-Coupled Receptor D1	GPCRs, Other	*Signal transduction and transmembrane signaling receptor activity*

ARL3		W (*)	ADP Ribosylation Factor Like GTPase 3	Cargo trafficking to the periciliary membrane and Organ elle biogenesis and maintenance.	*Mitotic cytokinesis and obsolete signal transducer activity*
ASZ1		LT(S)	Ankyrin Repeat, SAM And Basic Leucine Zipper Domain Containing 1	Gene Expression and Mitotic Prophase	*Obsolete signal transducer activity*
ATP6V0A1		DL(S)	ATPase H+ Transporting V0 Subunit A1	Signaling by GPCR,RET signaling and Innate Immune system	*ATPase binding and proton-transporting ATPase activity, rotational mechanism.*
ATP6V0A4		DL(S)	TPase H+ Transporting V0 Subunit A4	Signaling by GPCR,RET signaling and Innate Immune system	*ATPase binding and proton-transporting ATPase activity, rotational mechanism*
ATP8B4	N (p)	LN (S)	ATPase Phospholipid Transporting 8B4	Innate immune system, ion channel transport and Cardiac conduction	*Nucleotide binding and ATPase-coupled cation transmembrane transporter activity*
BUB1		N (*)	BUB1 Mitotic Checkpoint Serine/Threonine Kinase	MAPK-Erk Pathway and Cell cycle	*Protein phosphorilation and cell cycle*
C1H2orf49		W (*)	Chromosome 2 Open Reading Frame 49	Gene Expression and tRNA processing.	*Embryonic morphogenesis and tRNA splicing, via endonucleolytic cleavage and ligation*
CDK1		LT (*)	Cyclin Dependent Kinase 1	Cell Cycle and ATM Signaling Pathway	*Microtubule cytoskeleton organization and mitotic cell cycle*
CDK2		LT (S)	Cyclin Dependent Kinase 2	Cell Cycle and Homology Directed Repair	*Transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity*
CDK3		LT (*)	Cyclin Dependent Kinase 3	Cell Cycle Control of Chromosomal Replication and GADD 45 Pathway.	*Transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity*
CDK5		LT (*)	Cyclin Dependent Kinase 5	Regulation of TP53 Activity and Developmental Biology	
CDK6		LT (*)	Cyclin Dependent Kinase 6	Aldosterone synthesis and secretion and Cell cycle	*Positive regulation of cell-matrix adhesion and cell-cycle*
CFAP50		L (*)	X	X	*X*
COL8A2	N (X)		Collagen Type VIII Alpha 2 Chain	Collagen chain trimerization and ERK Signaling.	*Extracellular matrix structural constituent and prot ein binding, bridging*
COMMD7		LW (S)	COMM Domain Containing 7		
DHX30		W (*)	DExH-Box Helicase 30	X	*Central nervous system development and chromatin binding*
DUSP3	T (p)	T (S)	Dual Specificity Phosphatase 3	MAP Kinase Signaling, Ret signaling and Signal transduction_Erk Interactions-Inhibition of Erk.	*Protein kinase binding and protein tyrosine phosphatase activity*

EP400	N (p)	DNW (S)	E1A Binding Protein P400	Chromatin organization and Cellular Senescence (REACTOME)	*Chromatin binding and helicase activity*
FAT1	T (p)	T (S)	FAT Atypical Cadherin 1	Primary Focal Segmental Glomerulosclerosis FSGS	*X*
FAT2	T (p)	T (S)	FAT Atypical Cadherin 2	X	*X*
FAT3	T (p)	T (S)	FAT Atypical Cadherin 3	X	*X*
FBX031		N(S) W(*)	F-Box Protein 31	Class I MHC mediated antigen processing and presentation and Innate Immune System.	*Cyclin binding and cell cycle*
GAS7		N (S*)T (S)LD(*)	Growth Arrest Specific 7	X	*DNA-binding transcription factor activity, actin filament binding and cell differentiation*
GMPS		DW (S*)	Guanine Monophosphate Synthase	Drug metabolism - cytochrome P450 and Glucose / Energy Metabolism.	*Enzyme binding and transferase activity.*
GTF2H3		L (*)	General Transcription Factor IIH Subunit 3	Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3 and Apoptotic Pathways in Synovial Fibroblasts.	*DNA-binding transcription factor activity and protein N-terminus binding*
HSD11B1a		DL(S)	X	X	*X*
IGDCC4	T (p)	T (S)	Immunoglobulin Superfamily DCC Subclass Member 4		*Protein binding*
IGF1R	N(p)	DN (S)	Insulin Like Growth Factor 1 Receptor	ERK Signaling and Apoptotic Pathways in Synovial Fibroblasts	*Identical protein binding and protein kinase activity*
JMJD1C	N (p)		Jumonji Domain Containing 1C	Factors involved in megakaryocyte development and platelet production	*Thyroid hormone receptor binding and dioxygenase activity*
LMINA		L (*)	X	X	*X*
LMNA		L (*)	Lamin A/C	Cell Cycle, Mitotic and DNA Damage.	*Nucleus organization and muscle organ development*
LMNB2		L (*)	Lamin B2	Apoptosis and survival Caspase cascade and Cytoskeletal Signaling	*Structural molecule activity*
LOC100220302	T (p)	T (S)	X	X	*X*
LOC100221408	N (p)	NL(S)	X	X	*X*
LOC100225021		DL(S)	X	X	*X*
LOC116807107	T (p)	NW(S)	X	X	*X*
LOC116807726	N (p)	NW(S)	X	X	*X*
LOC417372	T (p)	T (S)	X	X	*X*
MITF		NW (S)	Melanocyte Inducing Transcription Factor	MAPK-Erk Pathway, RANK Signaling in Osteoclasts and Melanogenesis	*DNA-binding transcription factor activity and RNA polymerase II proximal promoter sequence-specific DNA binding*

MYO3A	N (p)	N(S) W (S*)	Myosin IIIA	PAK Pathway and Sweet Taste Signaling	*Transferase activity, transferring phosphorus-containing groups and protein tyrosine kinase activity*
NAA60		DL (S)	N-Alpha-Acetyltransferase 60, NatF Catalytic Subunit		*N-acetyltransferase activity*
NDST1	N	W (S)	N-Deacetylase And N-Sulfotransferase 1	heparan sulfate biosynthesis and Glyco saminoglycan metabolism	*Hydrolase activity and deacetyl ase activity*
PCDH1	N (p)	W (S)	Protocadherin 1	X	*Calcium ion binding, cell adhesion and cell-cell signaling*
PDE4D		LT(S)	Phosphodiesterase 4D	Signaling by GPCR and DAG and IP3 signaling	*Enzyme binding and protein domain specific binding*
PDZRN4	N (p)	LNW (S)	PDZ Domain Containing Ring Finger 4	X	*Ubiquitin-protein transferase activity and ubiquiti n protein ligase activity*
POLD2	N (p)	N(S*) W(S)	DNA Polymerase Delta 2, Accessory Subunit	Transcription-Coupled Nucleotide Excision Repair (TC-NER) and E2F mediated regulation of DNA replication.	*DNA-directed DNA polymerase activity*
PPTC7	N (p)	N(S)	Protein Phosphatase Targeting COQ7	Development Dopamine D2 receptor transactivation of EGFR	*Phosphoprotein phosphatase activity*
PRKAR1A		DL(S)	Protein Kinase CAMP-Dependent Type I Regulatory Subunit Alpha	Activation of cAMP-Dependent PKA, DAG and IP3 signaling and RET signaling.	*Ubiquitin protein ligase binding and protein kinase A catalytic subunit binding*
PRKAR1B		DL(S)	Protein Kinase CAMP-Dependent Type I Regulatory Subunit Beta	Activation of cAMP-Dependent PKA, DAG and IP3 signaling and RET signaling.	*cAMP binding and cAMP-dependent protein kinase regulator activity*
PRRL5	T (p)	DTW(S)	X	X	*X*
PSD2		W (*)	Pleckstrin And Sec7 Domain Containing 2	Endocytosis	*Regulation of catalytic activity*
RASGEF1A	N (p)	N(S)	RasGEF Domain Family Member 1A	Developmental Biology and RET signaling.	*MAPK cascade and cell migration*
RASGEF1B	N (p)	N(S)	RasGEF Domain Family Member 1B	X	*X*
RASGEF1C	N (p)	N(S)	RasGEF Domain Family Member 1C	X	*X*
RRM2B		DL(S)	Ribonucleotide Reductase Regulatory TP53 Inducible Subunit M2B	Gene Expression and superpathway of pyrimidine deoxyribonucleotides de novo biosynthesis	*Ribonucleoside-diphosphate reductase activity, thioredoxin disulfide as acceptor*
RXRG		N (*) L(S)	Retinoid X Receptor Gamma	Apoptotic Pathways in Synovial Fibroblasts and Gene Expression	*Regulation of transcription, DNA-templated and cell differentiation*
SH3KBP1		W (S*)	SH3 Domain Containing Kinase Binding Protein 1	Endocytosis and EGF/ EGFR Signaling Pathway and Lipoprotein metabolism	*SH3 domain binding*

SLC34A2	T (p)	LNW (S)	Solute Carrier Family 34 Member 2	Glucose / Energy Metabolism and Mineral absorption	*Protein domain specific binding and phosphate ion binding*
SLC6A1		NW (S)	Solute Carrier Family 6 Member 1	GABAergic synapse and Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds.	*Learning, memory and neurotransmitter transport*
SLC6A2		NW (S)	Solute Carrier Family 6 Member 2	Monoamine Transport and Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds.	*Beta-tubulin binding and neurotransmitter:sodium symporter activity*
SLC6A6		NW (S)	Solute Carrier Family 6 Member 6	Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds and NRF2 pathway	*Neurotransmitter and amino acid transport*
SPIRE2		W (*)	Spire Type Actin Nucleation Factor 2	TGF-Beta Pathway.	*Actin binding and actin cytoskeleton organization*
SPSB3		L (*)	SplA/Ryanodine Receptor Domain And SOCS Box Containing 3	X	*Protein ubiquitination*
SREBF1	T (p)	T (S)	Sterol Regulatory Element Binding Transcription Factor 1	AMPK Signaling Pathway, Regulation of cholesterol biosynthesis by SREBP (SREBF) and Metabolism	*DNA-binding transcription factor activity and chromatin binding*
SYTL4		ND(S)	Synaptotagmin Like 4	Response to elevated platelet cytosolic Ca2+ and Deregulation of Rab and Rab Effector Genes in Bladder Cancer	*Calcium ion binding and phospholipid binding.*
TEDC1		W (*)	Tubulin Epsilon And Delta Complex 1	X	*X*
TOX2		L (*)	TOX High Mobility Group Box Family Member 2	X	*Regulation of transcription by RNA polymerase II*
TSPAN6	N (p)	N(S)	Tetraspanin 6	X	*Obsolete signal transducer activity*
UBR3	N (p)	N (S)	Ubiquitin Protein Ligase E3 Component N-Recognin 3	Protein modification; protein ubiquitination	*Ligase activity and ubiquitin-protein transferase activity*
UNC80		N (*)LW(S)	Unc-80 Homolog, NALCN Channel Complex Subunit	Ion channel transport and Transport of glucose and other sugars, bile salts and organic acids, metal ions and amine compounds.	*Ion transmembrane transport and cation homeostasis*
USP3		W (*)	Ubiquitin Specific Peptidase 3	Deubiquitination and Metabolism of proteins	*Negative regulation of transcription by RNA polymerase II and mitotic cell cycle*
VEGFA	N (p)	L (S)	Vascular Endothelial Growth Factor A	VEGF Signaling Pathway	*Protein homodimerization activity and protein heterodimerization activity*

XRCC2		DW(S)	X-Ray Repair Cross Complementing 2	Homologous DNA Pairing and Strand Exchange and Resolution of D-Loop Structures	*ATP-dependent activity, acting on DNA and four-way junction DNA binding.*
ZBTB49	N (p)	N(S)	Zinc Finger And BTB Domain Containing 49	X	*Negative regulation of transcription by RNA polymerase II and cell cycle*

Table S4. List of candidate genes detected to be putatively under selection in the whole genome dataset by the different methods including: F_{ST} scan (F_{ST}), Cross Population Extended Haplotype Homozygosity (XP-EHH), Integrated Haplotype Score (iHS) for cloud and pine forest and SweepFinder 2 (SF2) for cloud and pine forest. The genes highlighted in bold are the common genes detected by the iHS method in both cloud and pine forest.

F_{ST}	XP-EHH	iHS cloud				iHS pine		SF2 cloud	SF2 Pine
ABTB1	ACAD8	ACACA	FOXP4L	NIBAN1	USP22	ACER3	MAP2K4	ACTN2	ABHD3
ACER3	ADAMTS12	ACAD8	FRMD4A	NIPSNAP3L	USP31	ACOX3	MARCH11	ARG2	ACTR1A
ACTN2	ADAMTS20	ACSF3	**GATA5**	NKAIN4	**USP40**	ADAMTS17	MARCHF11	ATP6V1D	ADA
ADAMTS12	AGAP3	ADAM28	GATC	NOA1	USP51	ADAMTS20	MBNL1	CDH8	ANKRD2
ADGRA3	ANKRD28	ADCY2	GATD1	NOL4	**UVSSA**	ADCY5	MBNL2	DPF3	ANKRD29
ADGRA3L	ASS1	**ADCY7**	GBE1	NOS1	VAC14	**ADCY7**	MBNL3	EIF2S1	ARHGAP19
AFG1L	BCR	ADGRA1	GBX1	NOTCH1	**VPS13B**	ADORA2B	MCAM	GPHN	ARMH3
AGTPBP1	CAMK2D	ADGRB1	**GGA3**	NOTCH2	VWF	ADORA3	MEF2A	LOC100220017	AVPI1
ANKHD1	CCDC88C	ADGRB2	GLP2R	NOX4	WDFY4	AFAP1L2	MEF2C	LOC100226769	BORCS7
ANKRD27	CDH12	ADGRF4	GNA11	NRF1	WDR36	AKAP6	MEF2D	LOC100229858	BPGM
ANKRD50	CEPT1	AFAP1	GNB5	NRG1	WDR78	AMPH	MGAT5B	LOC112530475	BTAF1
AOPEP	CEPT1	AGAP1	GPD1L	**NRP1**	WHRN	**ANKFY1**	MICAL2	LOC423277	C6H10orf76
AQR	CHN2	AGBL2	GRM5	NRXN1	YARS2	ANO6	MIDEAS	MAP3K9	C6H10orf95
ARFGEF3	CHPT1	AHCYL2	GSTCD	NUDT9	ZC3H18	APOB	MKL2	MPP5	CEP152
ARHGAP10	CNTN1	AKAP11	GTF3C5	OPRD1	ZNF106	ARHGAP11A	**MKLN1**	MYO10	CGNL1
ARMC9	CNTN6	AKIP1	GUK1	OPRK1	ZNF236	ARHGEF18	MLLT6	PCNX1	COPS2
ARVCF	CSF1R	ALKBH8	H3F3B	OPRM1	ZNF330	ARID1A	MON2	PIGH	CRTAC1
BCAR3	CSMD1	AMMECR1	H3F3C	OSBP2	ZRANB1	ARL8A	**MROH7L2**	PLEK2	CTXN2
BMP5	CSMD2	AMMECR1L	HAPLN1	OXR1	ACER3	ARL8B	MRTFB	PLEKHH1	CTXN2
CBLB	CTSD	AMPD3	HCN3	PARP11	ACOX3	ARL8BL	MRVI1	PLEKHH2	CYP17A1
CCDC126	CTSE	ANK1	HERC3	PARP8	ADAMTS17	ASB7	MSH3	RAD51B	DPCD
CCDC83	CUBN	ANKDD1B	HERC4	PATJ	ADAMTS20	ASXL3	**MTMR8**	RDH12	DUSP11RC1I
CEP78	DENND6A	**ANKFY1**	HLF	PCDH19	ADCY5	ATG5	MYH15	RGS6	DUT
CERCAM	DENND6B	ANKHD1	HMCN1	PCLO	ADCY7	**ATP6AP1L**	MYO10	SIPA1L1	ESCO1
CERCAM	DEPDC1	ARAP2	HMGXB3	PDCD6IP	ADORA2B	BAZ1A	NCAM2	SRSF5	EXOSC1
CFAP100	DEPDC1B	ARHGEF38	HOMER1	PDDC1	ADORA3	BCAN	**NEK10**	SRSF5A	FBN1
CFAP47	DHX32	**ATP6AP1L**	HOMER3	PDGFRB	AFAP1L2	**BECN1**	NELL2	TMEM229B	FBXW11
CFTR	DST	BCAS1	IDE	PDS5B	AKAP6	BEND4	NID2	VTI1B	FBXW4
CHPF	EOGT	**BECN1**	IGFBP7	**PEBP4**	AMPH	BSN	NLRC5	ZFP36L1	FGF8
CHPF2	EXOC4	BGLAP	ILKAP	PEX7	ANKFY1	BVES	NPAS3	ZFYVE26	FITM2
CHRDL2	FARP2	BRD7	INTS12	PI4KB	ANO6	C5H14orf180	NRIP3		GALK2

CORIN	FAT4	BSX	**IQGAP3**	PIK3C2A	APOB	C8B	**NRP1**		GCOM1
CYFIP1	FBXW7	C18H17orf62	**ITGA9**	PIK3C2G	ARHGAP11A	CACNA1S	NSF		GDAP1
CYFIP2	FGFR2	C2H8orf76	ITIH5	**PIK3C3**	ARHGEF18	CACNA2D2	NUDCD1		GDAP1L1
DNAH12	FHOD3	C4H20ORF194	ITPKB	PKHD1	ARID1A	CAD	OC90		GOLGA7B
DOCK2	FLI1	CACNA1E	JADE2	PKP2	ARL8A	CADPS2	PASK		HNF4A
DOT1L	FOXP1	CACNA2D1	KALRN	PLA2G4E	ARL8B	**CAMK4L**	PDE10A		HOGA1
DYNC1LI1	FOXP2	CACNA2D3	KANK4	PLEKHA5	ARL8BL	CAST	**PEBP4**		HPS1
DYNC1LI2	GABRA1	CACNA2D4	KAT2A	PLEKHA7	ASB7	CCDC172	PHF20L1		HPS6
DYNLT3	GABRA4	**CAMK4L**	KBP	PLEKHB1	ASXL3	CDCP1	PHTF2		HPSE2
EED	GASTL	CAPN1	KCNK10	PLEKHB2	ATG5	CDH11	**PIK3C3**		JPH1
EIPR1	GATB	CAPN9	KCNK13	PLEKHD1	ATP6AP1L	CDH5	PLG		JPH2
EPB41L2	GGA3	CCDC138	KCNK2	PLXNA1	BAZ1A	**CEPT1**	POLA1		JPH3
ERC1	GLP2R	CCNI	KCNMA1	PLXNA2	BCAN	CHD7	POLR3B		KAZALD1
EXD3	GNB1	CD164L2	KIAA0319L	PNPLA7	BECN1	**CHPT1**	POU2F3		KCNIP1
FAM237A	GNB4	CD300LG	KIAA0430	POLR2B	BEND4	CNGA3	POU4F1		KCNIP2
FANCC	GNS	CDH12	KIF16B	POLR2C	BSN	CNTN1	POU4F3-2		KCNIP4
FBP1	GPM6B	CDH13	KIF21B	PPA2	BVES	CNTN6	PRDM16		LAMA3
FBP2	GSDMA	CDH18	KIF25	PPFIA2	C5H14orf180	**COL19A1**	**PREP**		LBX1
FBXW7	GTF2IRD1	CDH20	KIF26A	PPM1L	C8B	**COL5A1**	PSMA6		LCOR
FCHSD2	HADHA	CDH8	KNDC1	**PREP**	CACNA1S	CPLANE1	PSMD14		LDB1
FHOD3	HDAC4	CEL	**KRT9L**	PRIM2	CACNA2D2	CPXM2	**PUM1**		LOC100218933
FLI1	IDO2	CENPE	LASP1	PRKAG1	CAD	CSMD2	RAB18		LOC100218935
FNBP1L	IFA3L	CEP152	LGR5	PRPF19	CADPS2	CYTH1	RAB3IP		LOC100219398
FOCAD	ISY1	CEP350	LIMCH1	PSAP	CAMK4L	CYTH3	RALGAPA1		LOC100220010
FRMD5	ITGA9	CEP43	LIMS1	PSMG2	CAST	DCDC2	RAPGEF4		LOC100227831
GABRA1	ITPR2	CEP89	LIMS2	PTPN13	CCDC172	DESI2	RASGRP1		LOC100232500
GABRA4	KCNK5	**CEPT1**	LIPML4	PTPRD	CDCP1	DISP3	RETREG1		LOC101233387
GINM1	KIF16B	CFAP47	LMBR1L	PTPRF	CDH11	DLD	REV1		LOC101234170
GLDC	KLHL1	CHD9	LOC100190135	**PUM1**	CDH5	**DNAH9**	RGS3		LOC107049564
GRAMD1C	KLHL4	**CHPT1**	LOC100218816	PXDNL	CEPT1	DPF3	RIN3		LOC107049953
HBS1L	LDLRAP1	CHRNA3	LOC100219000	RAB37	CHD7	EDMPN2	RNF38		LOC107049998
HHIPL1	LOC100220956	CHRNA4	LOC100219113	RALGPS1	CHPT1	ELMSAN1	RNF44		LOC107053690
ITGB6	LOC100221786	CHRNA5	LOC100219353	RAVER2	CNGA3	EXTL3	RPTOR		LOC107057318
KATNA1	LOC100222224	CHRNB2	**LOC100220416**	RB1CC1	CNTN1	FAM124A	RSRC1		LOC101234170
KATNAL1	LOC100225462	CHRNB3	LOC100220650	RBM17	CNTN6	FAM134B	RYBP		LZTS2
LOC100218384	LOC100226213	CHRNB4	LOC100220938	RECK	COL19A1	**FAM149A**	SBF1		MACIR
LOC100219012	LOC100227703	CLGN	LOC100222254	REEP3	COL5A1	FAM185A	**SCO2**		MFSD13A
LOC100219846	LOC100228347	CLVS1	LOC100224002	RHBDL3	CPLANE1	**FBN1**	SEC31A		MGEA5
LOC100220956	LOC100228379	CLVS2	LOC100224543	RHOBTB3	CPXM2	**FHAD1**	**SEPT9**		MIB1
LOC100221786	LOC101232962	CNNM2	LOC100224572	RHOT1	CSMD2	**FHOD3**	**SEPTIN9**		MMS19
LOC100222802	LOC101234162	CNTN5	LOC100225408	RHOT2	CYTH1	**FLI1**	**SETDB1**		MORN4
LOC100224567	LOC115491068	COG6	LOC100225462	RHPN1	CYTH3	FOXP2	SIPA1L1		MRPL43
LOC100226935	LOC115492592	**COL19A1**	LOC100225953	RIMS1	DCDC2	FSD1	SLA		MYEF2

LOC100858068	LOC115493448	COL20A1	LOC100228347	RNF17	DESI2	FSD1L	SLC13A1		MYZAP
LOC101232962	LOC115495938	COL22A1	LOC100230245	RNF2	DISP3	GALNT1	SLC25A16		NPC1
LOC101233767	LOC116808964	COL26A1	**LOC100230285**	ROBO2	DLD	GALNT13	SLC35F2		NPM3
LOC105758686	LOC116809290	COL2A1L	LOC100231218	RPH3AL	DNAH9	**GATA5**	SLC38A1		OCSTAMP
LOC115490732	LOC428335	COL4A6	LOC100231240	SAG	DPF3	GATB	SLC38A4		OGA
LOC115491052	LRP6	**COL5A1**	LOC100231703	SAP130	EDMPN2	GDF6	SLC4A10		OSER1
LOC429098	LRRC39	CORIN	LOC101233514	SCN1A	ELMSAN1	GDF7	SLIT2		PABPC1
LPIN2	MAPK11	CP	LOC101234197	SCN2A	EXTL3	GFRAL	SLIT3		PDZD7
LRRC41	MAPK12	CPEB2	LOC101747868	SCN8A	FAM124A	**GGA3**	SPAM1		PGAM1
LRRIQ1	MAPK14	CPEB3	LOC105758623	**SCO2**	FAM134B	GNB1	SRGAP1		PI15
MAPK11	MFSD14A	CPEB4	LOC105760451	SDK1	FAM149A	GNB4	SRGAP3		PI4K2A
MAPK14	MTMR8	CPNE3	LOC105760826	SEC14L2	FAM185A	GPR26	STARD9		PI4K2B
METTL24	MYO10	CRACR2A	LOC107049564	SEC14L3	FBN1	GPR78	SUPV3L1		PKIG
MTMR8	MYO5C	CRPPA	LOC107051278	**SEPT9**	FHAD1	GRAMD1C	SYT16		POLL
MXRA7	NCOR2	CRY4	LOC107055605	**SEPTIN9**	FHOD3	GRIN2A	TACC2		PPRC1
NCOR2	NIBAN1	CSMD1	LOC112530493	SETD7	FLI1	GRIN2B	**TECRL**		PYROXD2
NFIB	NPHP3	CST3	**LOC112530773**	**SETDB1**	FOXP2	HADHA	TET1		R3HCC1L
NPR3	NUP98	CSTF1	LOC112532368	SH3BP1	FSD1	HELB	TG		R3HDML
NSUN4	OAF	CTR9	LOC115490544	SH3GL1	FSD1L	HELZ	TMEM131		RBBP8
NVL	OSBP2	CYP2AB1	LOC115491038	SH3GL2	GALNT1	HERC1	TMEM242		RIMS2
OVST	PCCA	DCAF13	LOC115491069	SH3GL3	GALNT13	IDO2	TMEM38A		RIOK3
PDE4B	PCDHAC2	DCBLD1	LOC115491137	SIK2	GATA5	IFFO2	TMEM38B		RRP12
PDE4C	PDE4B	DCLK1	**LOC115492592**	SKIV2L2	GATB	IGSF10	TP53BP2		SECISBP2L
PGM5	PDE4C	DCPS	LOC115492903	SLC40A1	GDF6	ILDR2	TPM1		SEMA4C
PHC1	PEBP4	DDN	LOC115493789	SLC5A7	GDF7	IQCK	TPM4		SEMA4G
PLA2R1	PGA3	DDX54	LOC115493890	SLC6A11	GFRAL	**IQGAP3**	TTC25		SEMA6A
PLXDC2	PHTF2	DGCR2	LOC115494024	SLC9B2	GGA3	IQUB	UBASH3A		SEMA6D
POLD3	PINLYP	DHX32	LOC115495264	SMYD3	GNB1	IRAG1	**UBE2Q1**		SERINC1
PRDM15	PLCH2	DHX8	LOC115495663	SOX1	GNB4	IRX2	**UBE2Q2**		SERINC3
PRPSAP1	PLP1	DIAPH3	LOC115498023	SPAG5	GPR26	ISY1	UCK1		SFN
PTCH1	PLPPR5	DIS3L	LOC116806597	SPG11	GPR78	ITFG1	UCK2		SFRP1
PTPRG	POLR1E	DLGAP1L	LOC116806773	SSH2	GRAMD1C	ITGA6	ULK4		SFRP5
RASEF	POU2F1	DLGAP2	LOC116806777	STAM	GRIN2A	**ITGA9**	UNC79		SFXN1
RBPJ	RAB3IP	DLGAP3	LOC116807647	STAM2	GRIN2B	KATNBL1	**USP40**		SFXN3
RFT1	RBBP4	DLGAP4	LOC116808964	STARD4	HADHA	KBTBD12	**UVSSA**		SHC4
RHOBTB3	RBBP7	**DNAH9**	LOC751995	STOML3	HELB	KDM4C	VCAN		SLC12A1
RNF144A	RBM19	DPP10	LPIN1	SUZ12	HELZ	**KRT9L**	**VPS13B**		SLC13A3
RNF144B	RBM33	DPP6	**LRP1B**	SYNJ2	HERC1	LEMD3	VWA3B		SLC24A5
ROS1	RECK	DPYS	LRTM1	TACR2	IDO2	LIN54	WASF1		SLC2A10
SCHIP1	RELN	DPYSL5	LSAMP	TACR3		LMBRD1	WDFY3		SLC2A11L5
SCN1A	RHBDD2	DPYSL5L	LY6CLEL	TAOK1		LOC100190084	WDR77		SLF2
SCN2A	RIN3	DRAXIN	MACROD2	TAOK3		LOC100217483	WNK2		SLIT1
SCN8A	SCO2	DSCAM	MAP1LC3A	TBC1D22A		**LOC100220416**	XKR4		SLIT2

SEMA5B	SH3GL1	DSTYK	MAP2K5	TBC1D7	LOC100220728	XKR7	SLIT3
SIN3B	SLC37A2	DTD1	MAP3K5	TBC1D9	LOC100220760	YAF2	SOX17
SLC25A14	SLC44A3	EDAR	MARF1	TBC1D9B	LOC100222224	ZCRB1	SOX18
SLC25A30	SMS	EDRF1	MATN1	**TECRL**	LOC100225836	ZDHHC18	STK4
SLC35A3	SNIP1	EFL1	MBD3	TENM4	LOC100226561	ZMAT4	SUFU
SNRPB2	SNX7	EHBP1	MFSD2B	TFAP2A	LOC100226810	ZNF385C	TCEA1
SORCS1	SYPL1	ELF1	MGST2	TFAP2B	LOC100227703	ZSWIM7	TCEA2
SPIRE1	SYPL1	ENOX1	**MKLN1**	TFAP2C	LOC100229858	ZZEF1	TCF12
ST6GALNAC1	SYT11	ENOX2	MMD	TFAP2E	LOC100229949		TLX3
ST6GALNAC2	TENM1	EPC2	**MROH7L2**	TMCO4	LOC100230122		TOMM34
STAU2	TENM2	EPHA10	MRPS15	TMOD1	**LOC100230285**		TP53RK
SYK	TJP2	EPHA4	**MTMR8**	TMOD2	LOC100231371		TRIM8
SYT4	TLCD4	EPHA5	MTREX	TMOD3	LOC101748650		TTPAL
TAAR1	TMC1	EPHA7	MUC4	TNFRSF19	LOC107051263		TWNK
TARS	TMEM56	EPHB2	MYBPH	TNS3	LOC107052803		UBTD1
THOC7	TMOD1	ERH	MYH10	TRAIL-LIKE	LOC107053715		WBP1L
TLCD4	TMOD2	ESRP1	MYH9	TRIM2	LOC107055438		YWHAB
TM2D1	TMOD4	ESRP2	MYO10L	TRIM33	LOC112530482		YWHAE
TMEM56	TPO	ESYT2	MYO16	TRIM8	**LOC112530773**		YWHAG
TRA2A	TPR	FAM107B	MYO1A	TTC12	LOC115490710		YWHAH
TRAPPC12	TTC7A	FAM13A	MYO5A	TUBGCP2	LOC115490962		ZDHHC16
TSSC1	TTC7B	**FAM149A**	MYOCD	TXLNA	**LOC115492592**		ZFYVE27
TSSK6L1	UBA5	FAM173A	MYOM1	TXNDC16	LOC115494277		
TTC26	UNC5A	FANCE	N4BP2	UBE2D1	LOC115494435		
UBE2O	UNC5D	**FBN1**	NAA15	UBE2D2	LOC116807077		
UQCRH	UROS	FBXO25	NAALAD2	UBE2D3	LOC116808208		
VPS37A	USP15	FBXO32	NALCN	**UBE2Q1**	LOC418421		
	USP6NL	FGFR1OP	NANOS3	**UBE2Q2**	LRP1		
	WDR77	**FHAD1**	NAPG	UMODL1	**LRP1B**		
	ZBTB16	**FHOD3**	NCDN	UNC119	LRP6		
	ZBTB32	**FLI1**	NDE1	UNC119B	LRP8		
	ZNF592	FLT1	**NEK10**	UNC5A	LRRFIP1		
	ZNF592	FMN2	NFIA	UNC5D	MAGI1		

CHAPTER IV: Chromosome-level genome assembly of the common chaffinch (Aves: *Fringilla coelebs*): a valuable resource for evolutionary biology

María Recuerda, Joel Vizueta, Cristian Cuevas-Caballé, Guillermo Blanco, Julio Rozas, Borja Milá

Recuerda, M., Vizueta, J., Cuevas-Caballé, C., Blanco, G., Rozas, J., & Milá, B. (2021). Chromosome-level genome assembly of the common chaffinch (Aves: *Fringilla coelebs*): a valuable resource for evolutionary biology. *Genome biology and evolution*, *13*(4), evab034.

Abstract

The common chaffinch, *Fringilla coelebs*, is one of the most common, widespread and well-studied passerines in Europe, with a broad distribution encompassing Western Europe and parts of Asia, North Africa and the Macaronesian archipelagos. We present a high-quality genome assembly of the common chaffinch generated using Illumina shotgun sequencing in combination with Chicago and Hi-C libraries. The final genome is a 994.87 Mb chromosome-level assembly, with 98% of the sequence data located in chromosome scaffolds and a N50 statistic of 69.73 Mb. Our genome assembly shows high completeness, with a complete BUSCO score of 93.9% using the avian dataset. Around 7.8 % of the genome contains interspersed repetitive elements. The structural annotation yielded 17,703 genes, 86.5% of which have a functional annotation, including 7,827 complete universal single-copy orthologs out of 8,338 genes represented in the BUSCO avian data set. This new annotated genome assembly will be a valuable resource as a reference for comparative and population genomic analyses of passerine, avian and vertebrate evolution.

Keywords: *Common chaffinch,* Fringilla coelebs*, reference genome, whole genome assembly.*

Introduction

The decreasing costs of DNA sequencing, along with advances in computational genomics, are promoting a rapid increase in the availability of high-quality reference genomes of non-model species, which greatly improves our capacity to address a range of biological questions from a genomic perspective. Among them, the correct annotation of protein-coding genes in whole genomes allows to identify new genes involved in the process of evolutionary adaptation and provides a better understanding of the evolutionary mechanisms involved in the speciation process. Avian genomes are particularly suited for studying the molecular basis of speciation as they have a relatively simple architecture and are among the smallest within amniotes, ranging from 0.91 to 1.3 Gb (Gregory, 2002). In the last decade, the number of bird reference genomes has increased dramatically (e.g., Dalloul et al., 2010; Warren et al., 2010; Zhang et al., 2012; Jarvis et al., 2014; Poelstra et al., 2014; Frankl-Vilches et al., 2015; Friis et al., 2018; Louha et al., 2020; Peñalba et al., 2020; Ducrest et al., 2020, W. Wang et al., 2020), providing major scientific breakthroughs in phylogenetics (i.e., Alström et al., 2018; Braun et al., 2019; Jarvis et al., 2015), comparative genomics (Zhang, 2014, Feng et al., 2020), adaptation genomics (Wirthlin et al., 2014; Lawson & Petren 2017), and genomic architecture (Poelstra et al., 2014; Vijay et al., 2016), among others. Moreover, the Ten-Thousand Bird Genomes (B10K) consortium has generated and analyzed over 300 avian genomes from 92.4% of bird families, providing an unprecedent genomic resource for avian comparative studies (Zhang, 2015, Feng et al., 2020).
The common chaffinch (Aves, Passeriformes, Fringillidae, *Fringilla coelebs*) is a widely distributed species, ranging from across Eurasia to the north of Africa, and has colonized three Macaronesian archipelagos in the Atlantic Ocean (Azores, Madeira and the Canary Islands) (Collar et al., 2020). With about 15 currently recognized subspecies, the common chaffinch is an ideal system for testing hypotheses on the evolutionary process given its distribution across the continent and the colonization of several oceanic islands, recognized as excellent natural laboratories for studying evolution (Brown et al., 2013). Island systems have inspired the development of biogeographical theories (MacArthur & Wilson, 1967) and are of central importance for understanding the role of area and isolation in colonization, extinction and speciation rates (Valente et al., 2020), which are

processes influencing global patterns of species richness (Losos & Schluter, 2000). Species that have colonized insular environments, like the common chaffinch, are also excellent systems for the study of demographic events, such as bottlenecks leading to small effective population size (N_e) (Leroy et al., 2020), or the roles of drift and selection in the divergence process (Barton, 1996). The common chaffinch has been intensively studied using molecular tools, so that the availability of a reference genome represents a valuable resource to improve our understanding of avian evolution, biogeography and demography (Illera et al., 2018).

Materials and Methods

Sequencing and genome assembly pipeline

Sample collection and genome assembly

A blood sample was extracted from a common chaffinch female captured in Torreiglesias, Segovia, Spain, in 2017 and frozen immediately in liquid nitrogen. The sample was handled by Dovetail Genomics for DNA extraction, sequencing and genome assembly using the HiRise pipeline (Putnam et al., 2016).

Chicago Library preparation and sequencing

The Chicago library was prepared as described previously (Putnam et al., 2016). Briefly, about 500 ng of high molecular weight genomic DNA (mean fragment length = 80 kb) was used to reconstruct chromatin *in vitro* and fixed with formaldehyde. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq X

platform to produce over 100 million 2x151 bp paired-end reads, which provided 197.67 x physical coverage of the genome (10-100 kb pairs).

Dovetail Hi-C library preparation and sequencing

A Dovetail Hi-C library was prepared in a similar manner as described previously (Lieberman-Aiden et al., 2009). For each library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed and the DNA purified from protein. Purified DNA was treated to remove biotin that was not internal to ligated fragments. The DNA was then sheared to ~350 bp mean fragment size and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina HiSeq X platform to produce 202 million 2x151 bp paired-end reads, which provided 27,400.40 x physical coverage of the genome (10-10,000 kb pairs).

De novo shotgun assembly of the common chaffinch genome

A *de novo* assembly was constructed using a combination of paired end reads (mean insert size ~350 bp). De novo assembly was performed using Meraculous (v. 2.2.2.5 diploid_mode1; Chapman et al., 2011) with a *k*-mer size of 73. The input data consisted in 397.3 million read pairs sequenced from paired-end libraries (totaling 1.204 Gb). Reads were trimmed for quality, sequencing adapters, and mate pair adapters using Trimmomatic (Bolger et al., 2014).

Scaffolding the assembly with HiRise

The input *de novo* assembly, shotgun reads, Chicago library reads, and Dovetail Hi-C library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al., 2016), conducting an iterative analysis. First, shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper (Zaharia et al., 2011; http://snap.cs.berkeley.edu). The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to

produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold. After aligning and scaffolding Chicago data, Dovetail Hi-C library sequences were aligned and scaffolded following the same method. After scaffolding, shotgun sequences were used to close gaps between contigs. After Hi-C assembly, microchromosomes 18 and 20-28 were found to be located in a single superscaffold which was subsequently split into the different microchromosomes based on the alignment with the zebra finch chromosome sequences.

Z chromosome assembly

Given the total absence of the Z chromosome in the resulting HiRise chromosome level assembly, or fragmented scaffolds assigned to this chromosome in zebra finch genome, a *de novo* assembly was conducted using the using MaSuRCA (v. 3.3.1) genome assembler with default parameters (Zimin et al., 2013). Sex-linked sequences were identified by aligning the assembled scaffolds with the Zebra finch (bTaeGut2.pat.W.v2) Z chromosome and studying the coverage distribution (expected to be half that of the autosomes), as described in (Xu et al., 2019). Briefly, nucmer from the MUMmer package (v.4.0, '-b 400' and filtering with 'delta-filter -1'; Marçais et al., 2018) was used to align the MaSuRCa resulting scaffolds with the Zebra finch genome. In addition, raw reads (from a female sample) were mapped against the assembled scaffolds with BWA 'mem' v0.7.17 (Li and Durbin 2009), and the coverage per scaffold was calculated using 'samtools depth -q 20 -Q 10' and 'bedtools map'. To assemble the Z chromosome, we retained any scaffolds longer than 1 Kb with a 50% or better alignment with the reference Z chromosome, and showing half the coverage of autosomes. These Z-linked scaffolds were then input into the HiRise pipeline described above to obtain the common chaffinch Z chrosomome.

Assembly quality control

The final assembly was screened for potential contaminants using BlobTools v1.0.1 (Laetsch & Blaxter 2017). Briefly, taxonomic annotations for all scaffolds were retrieved by comparing contigs against the NCBI nucleotide database (nt) using BLASTN (E-value ≤ 10^{-25}) (Altschul et al., 1990; Altschul et al., 1997). Next, raw reads were mapped against the assembly with BWA 'mem' v0.7.17 (Li and Durbin 2009) in order to estimate read coverage per scaffold. Finally, GC coverage plots were obtained using BlobTools, integrating all information to identify putative contaminant sequences based on taxonomic classification, read coverage and GC content.

Gene completeness in the chaffinch genome assembly (and in the annotated gene set) was assessed through BUSCO (Benchmarking Universal Single-Copy Orthologs) v4.0.5 (Seppey et al., 2019) by using the 8,338 single-copy orthologous genes in the class Aves (Aves_odb10; https://busco.ezlab.org/), using the chicken as the Augustus reference species. The 8,338 orthologous genes are present in at least 90% of the 40 species included within the Aves lineage group, and thus are likely to be found in the genome of related species. In order to check for consistency and synteny of the common chaffinch assembly, we aligned it with the zebra finch assembly (bTaeGut2.pat.W.v2, downloaded from the RefSeq FTP site: https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/008/822/105/GCF_008822105.2_bTaeGut2.pat.W.v2/) using nucmer from the MUMmer package (v.4.0; Marçais et al., 2018), and retaining only the best one-to-one alignments (deltafilter -1). The alignment coordinates among genomes were represented in a circos plot (Krzywinski et al., 2009).

Identification of repetitive regions

Repetitive regions were identified and masked prior to gene prediction. First, repeats were modelled *ab initio* using Repeat Modeler 1.0.11 (Smit & Hubley, 2019) in scaffolds longer than 100 Kb with default options. The repeats obtained were merged with known bird repeat libraries from the RepBase database (RepBase-20181026) (Bao et al., 2015), Dfam_Consensus-20181026 and repeats from the zebra finch (obtained from B10K). The resulting repeat library was

compared against the complete assembly with Repeat Masker 4.0.7 (Smit et al., 2013) and the identified regions were soft-masked. For the identification and description of microsatellites in the common chaffinch genome assembly we used GMATA v.2.01 (Wang & Wang, 2016), with sequence motif length between 2 and 20 bp.

Gene annotation and function prediction

Gene prediction was conducted using BRAKER v2.1.5 (Hoff et al., 2016) and GeMoMa v1.7.1 (Keilwagen et al., 2016, 2018). In order to assess and obtain the more complete protein coding gene set, different pipelines were used, and the resulting assemblies were compared using BUSCO. In summary, the most complete gene set was obtained using the conserved orthologous genes from BUSCO Aves_odb10 as proteins from short evolutionary distance to train Augustus (Gremme et al., 2005; Stanke et al., 2006; see Figure 3B from Hoff et al., 2019). The predicted proteins were combined with homology-based annotations using the the zebra finch (GCF_008822105.2; Warren et al., 2010) and chicken (GCF_000002315.6; Hillier et al., 2005) annotated genes with GeMoMa pipeline, obtaining the final reported gene models. We applied a similarity-based search approach to assist the functional annotation of the chaffinch predicted proteins. We first used BLASTP against the UniProt SwissProt database and the annotated proteins from the zebra finch genome (Warren et al., 2010; UniProt Consortium 2015) (*E*-value 10^{-5}). We only considered as positives those hits covering at least 2/3 of the query sequence length or 80% of the total subject sequence. We also used InterProScan v5.31 (Jones et al., 2014) in order to identify specific protein-domain signatures in the predicted genes. The functional annotation, including Gene Ontology terms, was integrated from all searches providing a curated set of chaffinch coding genes. We used GenomeTools (Gremme et al., 2013) to calculate the number and mean length of genes, exons, introns and CDS (Coding Sequence) from the annotation file in general feature format (GFF).

Non-coding RNA prediction and identification

In order to predict Transfer RNAs (tRNAs) in the common chaffinch genome we used tRNAscan-SE v2.0 (Lowe & Chan, 2016). The tRNA search across the genome was conducted using the eukaryotic type as training set for the covariance models implemented in the software package Infernal v1.1.1 (Nawrocki, 2014). The functional classification of the tRNAs was also conducted using tRNAscan-SE by analyzing the gene prediction with a set of isotype-specific covariance models dividing the set into subgroups of the universal 20 aminoacids. In order to check if the predicted tRNAs were consistent, we contrasted a random sample of 20 against the databases GtRNAdb (Chan & Lowe, 2016) and tRNAdb (Jühling et al., 2009). In addition, the identification of ncRNA (non-coding RNA) homologues on the common chaffinch genome was performed using Infernal "cmscan" program in conjunction with the Rfam library of covariance models (Gardner et al., 2010). The overlapping hits and the hits with *E*-value higher than 10^{-5} were removed from the analysis. Hits in the final set were classified in different categories according to Rfam database (Kalvari et al., 2018): CREs (Cis regulatory element), Ribozymes, Gene, miRNAs (microRNAs), snoRNAs (small nucleolar RNAs), snRNAs (small nuclear RNAs), rRNAs (ribosomal RNAs), and lncRNAs (long non-coding RNAs). We added our results to the comparison made by Louha et al., (2020) among different genome assemblies of avian species.

Results and Discussion

Assembly and quality control

The total length obtained by the HiRise software for the common chaffinch assembly was 994.87 Mb. Nevertheless, the estimate from *k*-mer metrics is 1.2 Gb. The discrepancy between these estimates could be caused by the presence of repetitive elements given the assembly strategy used, which could have been improved including long-read sequencing technologies. This final assembly consists of 3,255 scaffolds, 3,239 over 1kb and an N50 of 69.73 Mb (Table 1) with a sequence coverage of 249x. The use of Chicago and Hi-C libraries provided a clear improvement in quality by increasing 917 times the scaffold N50, reducing the

number of scaffolds from 38,666 to 3,255 (Table 1). In fact, 98 % of the total genome sequence maps in the 30 described chromosomes.

Table 1. Quality statistics of the different stages of the common chaffinch (*Fringilla coelebs*) assembly.

	Shotgun	Chicago HiRise	Dovetail Hi-C
Total length	903.83 Mb	907.32 Mb	994.87 Mb
Scaffold N50	0.076 Mb	8.033 Mb	69.727 Mb
Scaffold N90	0.011Mb	1.812 Mb	13.238 Mb
Scaffold L50	3,104	31	5
Scaffold L90	14,974	125	19
Longest scaffold	1,012,806 bp	43,054,799 bp	147,056,713 bp
Number of scaffolds	38,666	3,767	3,255
Number of scaffolds > 1kb	38,666	3,767	3,255
Contig N50	67.15 kb	67.08 kb	66.81 kb
Number of gaps	8,240	43,147	49,330
Percent of genome in gaps	0.11 %	0.50 %	0.49 %

The chaffinch genome showed high synteny with the zebra finch genome (Figure 1), evidencing the completeness of the assembly, with all micro-chromosomes and the Z chromosome present in the assembly. In addition, the alignment between these genomes suggests the presence of several inversions in chromosomes 1, 1A, 2, 3, 5, 7, 8 and 9. Several studies have documented that inversions are very common in birds (Aslam et al., 2010; Völker et al., 2010; Skinner & Griffin 2011; Zang et al., 2014). For instance, Hooper and Price (2017) identified 319 inversions on the 9 largest autosomes combined in 81 independent clades. No putative contaminations were detected and 89.6% of the reads were mapped in the genome assembly (Figure 2). The mean GC content of the assembly was 41.86% (±11 SD). The common chaffinch genome assembly included 7,832 complete copies (93.9%) out of the 8,338 BUSCO dataset from avian genomes, among which 7,816 were single-copy orthologs and 16 were duplicated. Only 1.8% of the gene models were fragmented, and 4.3% were missing in the genome. These few missing gene models could represent divergent or lost genes in our species, but also could be related with putative errors during the assembly process or missing data.

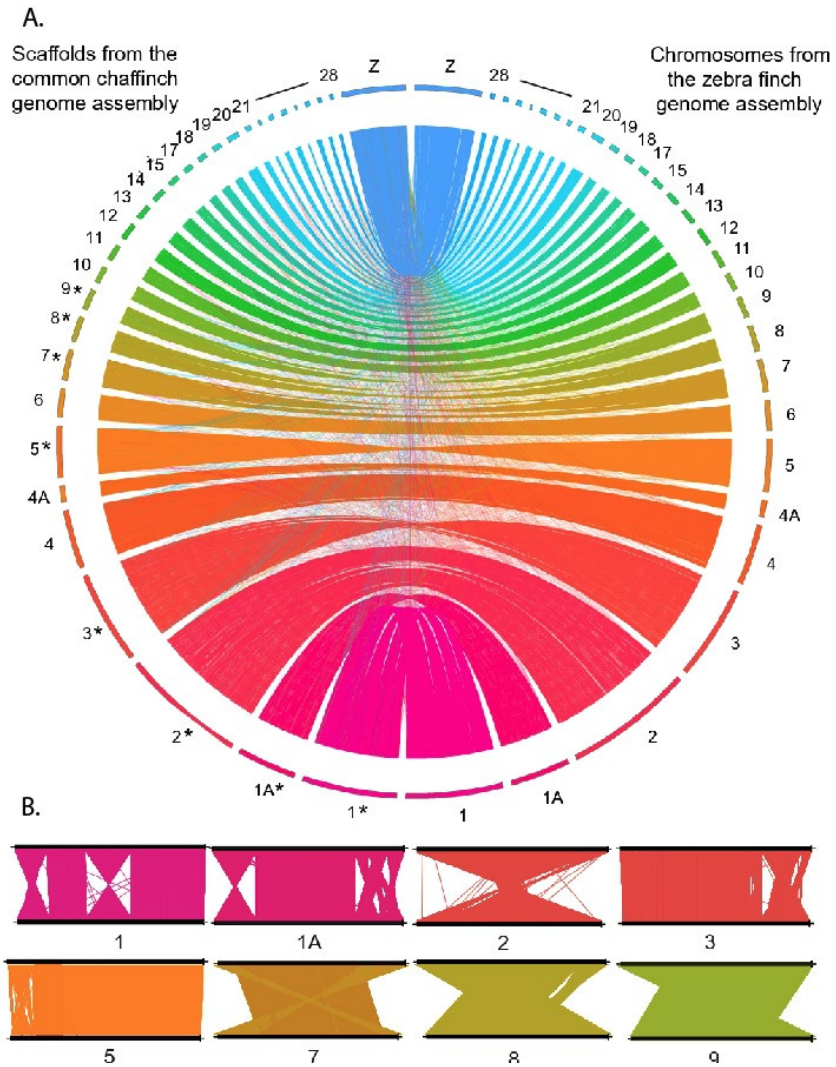


Figure 1. (A) Circos plot comparing the zebra finch (right hemisphere) and the common chaffinch (left hemisphere) genome assemblies. The common chaffinch chromosomes marked with an asterisk (*) show inversions with respect to the zebra finch assembly. (B) Linear synteny plots of the common chaffinch chromosomes showing inversions relative to the zebra finch generated with the R package genoPlotR (Guy et al., 2010). The zebra finch assembly (top) is compared to the common chaffinch assembly (bottom), and numbers designate specific chromosomes.

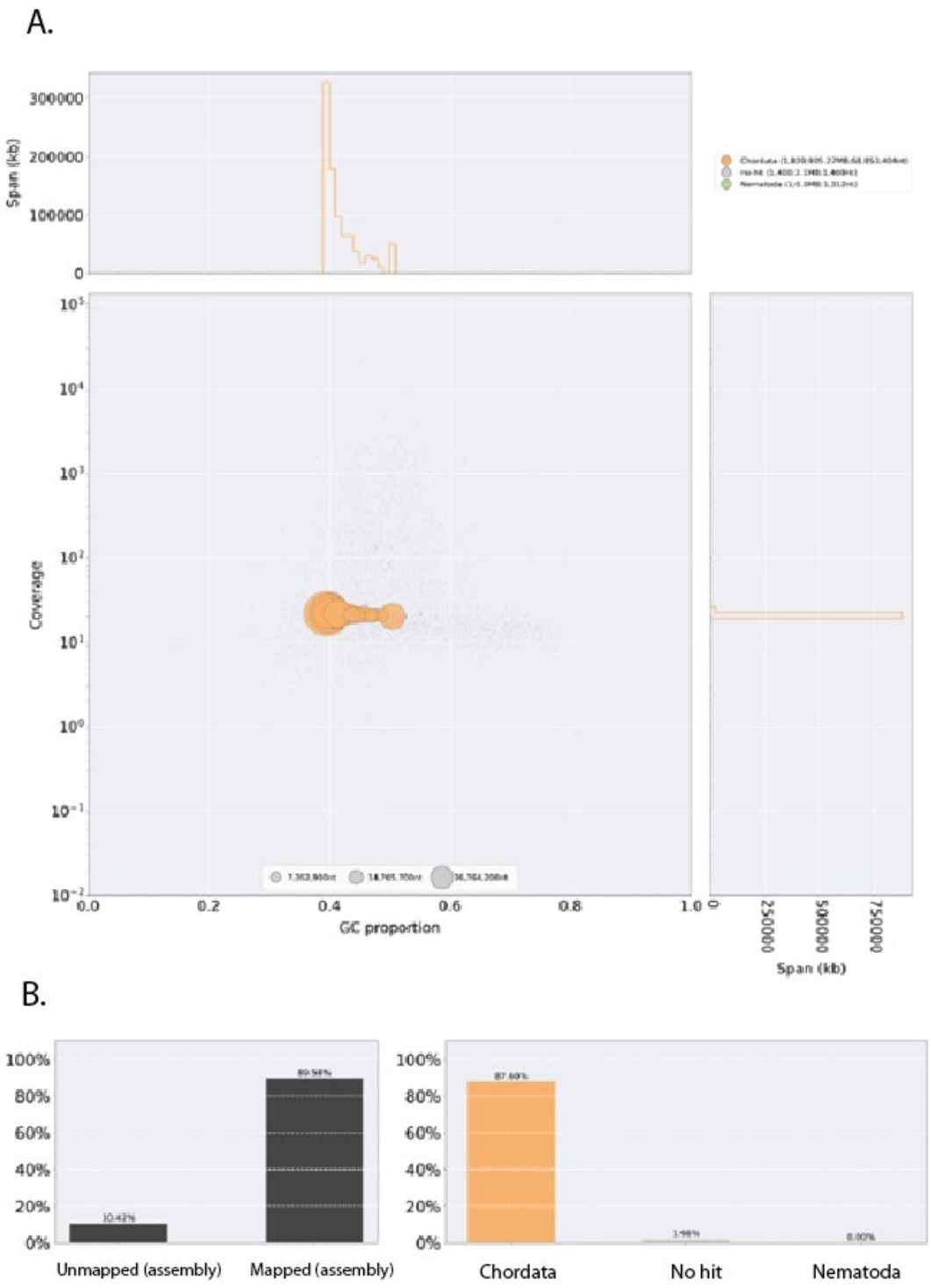


Figure 2. Screen for possible contamination using BlobTools of the common chaffinch genome assembly. (A) The taxon-annotated GCcoverage plot. (B) A summary of reads mapped to taxonomic groups as putative contaminants.

Repetitive regions

Overall, 7.82% of the genome assembly are repeats (~78 Mb), of which 85.4 % are transposable elements (TEs). The most abundant TEs are LINEs (53.5%) followed by LTR (29.4%), DNA elements (4.1 %) and SINEs (1.4 %), with the remaining 11.6% unclassified. The rest of repeats (14.6 %) contained simple repeats (75.4%), low complexity repeats (18.5 %), satellites (4.2%) and small RNA (1.9 %) (Table 2). The number of repetitive regions is within the expected range in birds, which is at 4 to 10% of the genome (Zhang et al., 2014).

Table 2. Classification of repeats: number, length in base pairs, and percentage of the common chaffinch assembly.

Classification	Number of elements	Length (bp)	% of assembly
SINEs	7950	963,528	0.10 %
LINEs	128,953	35,523,880	3.57 %
LTR elements	24,252	10,742,558	1.08 %
DNA elements	18,883	2,702,481	0.27 %
Unclassified	29,993	7,675,094	0.77 %
Total interspersed repeats		66,431,968	6.68 %
Small RNA	1,572	217,307	0.02 %
Satellites	1,794	488,121	0.05 %
Simple repeats	198,014	8,707,329	0.88 %
Low complexity	42,685	2,137,109	0.21 %

A total of 111,076 microsatellites, with motif length ranging between 2 and 20 bp, were identified in the common chaffinch genome (Fig. S3; their genomic locations are shown in File S1 in the Figshare repository). The most common *k*-mer sizes conforming the microsatellites were 2 (68.2%), 3 (15.9%) and 4 (8.2%) (File S1). The most common length of the microsatellites was 10 bp (40.4%), followed by 12 bp (13%) and 15 bp (8.8%) (see File S1 for the length distribution of microsatellites). In addition, the number of microsatellites was positively correlated with the sequence length (Fig 3; see File S1 for the frequency of occurrence in every scaffold).

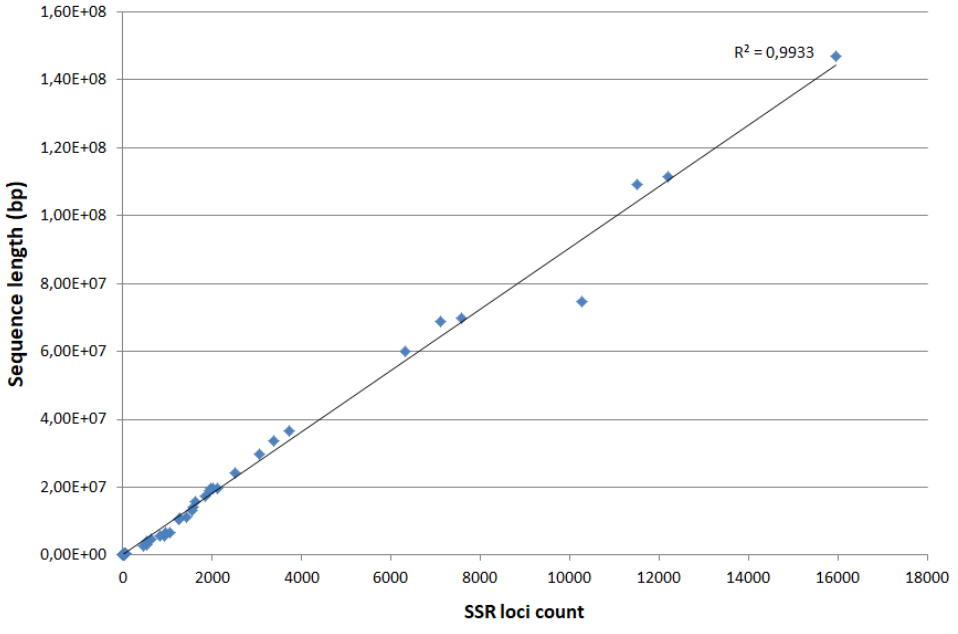


Figure 3. Microsatellite count and sequence length correlation.

Gene annotation and function prediction

Our annotation pipeline combining both *de novo* and homology-based predictions inferred 21,831 proteins encoded by 17,703 genes in the common chaffinch genome with a mean length of 15,818 bp (Table 3). The common chaffinch genome annotation (File S2 in Figshare) included 7,850 complete copies (94.2%) out of the 8,338 of BUSCO avian dataset used, retrieving all expected copies with a slight increase from that estimated in the un-annotated genome (see above). Among the complete BUSCO genes, 7,827 were single-copy orthologs (99.7%) and 23 were duplicated (0.3%). Around 1.9% (162) of the gene models were fragmented and 3.9% showed no significant matches (326).

Table 3. Genome statistics and predicted ncRNAs of the *F. coelebs* genome compared to other similarly sized avian species (*Melospiza melodia, Taeniopygia guttata, Ficedula albicollis, Manacus vitellinus,* and *Geospiza fortis*), modified from Louha et al., (2020).

	F. coelebs	*M. melodia*	*T. guttata*	*F. albicollis*	*M. vitellinus*	*G. fortis*
Number of genes	17,703	15,086	17,561	16,763	18,976	14,399
Mean gene length (bp)	15,818	14,457	26,458	31,394	27,847	30,164
Number of CDSs	17,703	15,086	17,561	16,763	18,976	14,399
Mean CDs length (bp)	1,679	1,325	1,677	1,942	1,929	1,766
Number of exons	221,872	131,940	171,767	189,043	190,390	164,721
Mean exon length (bp)	165	153	255	253	264	195
Mean number of exons/gene	10.16	8.67	10.25	12.22	11.51	11.41
Number of introns	200,041	116,724	153,909	171,236	171,089	149,563
Mean intron length (bp)	1,902	1,695	2,930	3,257	3,294	2,813
Total proteins	21,831					
ncRNA						
tRNA	325	267	184	179		
miRNA	140	166	302	510		
snRNA	18	16	44	32		
snoRNA	126	154	241	199		
rRNA	5	8	100	22		
lncRNA	17	20	908	1473		

Over all predicted proteins, 19,458 (89.1%) provided positive BLASTP hits against the Uniprot SwissProt database, and 19,617 (89.9%) against the annotated proteins from the zebra finch genome. In addition, InterproScan identified 18,551 (85%) specific protein-domain signatures in the predicted peptides. The combination of the annotation from these databases allowed assigning a functional annotation with GO terms to 19,425 proteins (89%) assigned to 15,309 genes (86.5%; File S3 in Figshare).

tRNAs and other non-coding RNA prediction

The search by tRNAscan-SE (File S4 in Figshare) identified 325 tRNAs in the common chaffinch genome, of which 167 decode for the standard twenty amino acids. Among all the tRNAs detected, 131 presented low scores and therefore were categorized as pseudogenes (i.e., lacking tRNA-like secondary structures). There were no suppressor tRNAs, 1 had undetermined isotopes, 25 were chimeric and 15 included introns within their sequences. One of the tRNAs was predicted to code for selenocysteine (sequences and structures of the predicted tRNAS are available in File S5 in Figshare). In addition, the search against both tRNA databases (GtRNAdb and tRNAdb) yielded positive results in many other species, suggesting that tRNA prediction in our assembly was correct. Moreover, our searches using Infernal identified 354 ncRNAs, which were classified as follows: 39 CREs, 2 Ribozymes, 7 Gene, 140 miRNAs, 126 snoRNAs, 18 snRNAs, 5 rRNAs, and 17 lncRNAs (File S6 in Figshare). The number of tRNAs predicted in the common chaffinch genome is the highest when compared to other passerine species (i.e., *M. melodia*, *T. guttata* and *F. albicollis*), but the other types of ncRNAs present similar values to the *M. melodia* genome and lower than the other two species (Table 3), probably because we applied a strict threshold to avoid an excess of false positives.

Conclusions

We provide here a high-quality assembly for the common chaffinch, a valuable resource as a reference genome to address a range of biological questions from a genomic perspective. Moreover, our annotation provides useful information to detect candidate genes involved in adaptation and divergence processes. The combination of the Chicago and shotgun sequencing with the HiRise assembly approach lead to a highly contiguous chromosome-level genome assembly. The genome assembly size was 994.87 Mb, with the 30 chromosomes accounting for 98% of it. Although the expected length of the genome was 1.2 Gb, closer to those obtained in other avian species by flow cytometry (Gregory 2002), the BUSCO analyses showed that both the assembly and structural annotation encode 93.9% and 94.2% complete copies out of the 8,338 orthologous conserved genes in avian species, respectively. This discrepancy of the genome size could be caused by the

absence of large repetitive elements in the assembly. The structural annotation predicted 17,703 coding genes, with most of them (86.5%) assigned to functional annotation and GO terms.

Data availability

The chaffinch genome assembly and the raw data have been deposited at NCBI under BioProject PRJNA674347 with accession number JADKPM000000000, and the annotation and all described datasets are publicly accessible in *Figshare* (https://doi.org/10.6084/m9.figshare.13296122.v3).

GENERAL RESULTS

In Chapter I, we detected that the exposure of four different species of passerines to the process of island colonization resulted in congruent phenotypic changes that were consistent with the island rule. We also detected differences in diet. However, the genomic signatures of selection were lineage-specific and caused by different factors. The main processes generating the heterogeneous landscapes of differentiation were selective sweeps for the red-billed chough, recurrent selection for the common chaffinch, a large chromosomal inversion in the house finch, and differentiation in centromeric regions for the dark-eyed junco. Even though we did not find parallel evolution in the strict sense, we did find that gene regulation was a crucial factor in all four species.

The results obtained in Chapter II revealed a species-level radiation of the common chaffinch in the Macaronesian region and revealed a circuitous colonization route that started in the mainland, through Azores, then Madeira and finally the Canary Islands. The sequential colonization of these archipelagos led to a progressive loss of genetic diversity that was consistent with the colonization order, with Azores showing highest diversity values and the Canary islands the lowest. The phylogenomic analysis resulted in a robust phylogeny which showed that all populations corresponded to a monophyletic group, including each island within the Canary Islands. Phenotypic data and species delimitation methods were consistent with the genomic conclusions, and we proposed that four current subspecies should be considered separate species. These new species would be *F. spodiogenys* from North Africa, *F. moreletti* from Azores, *F. maderensis* from Madeira and *F. canariensis* from the Canary Islands, with the latter maintaining the currently observed subspecies.

In Chapter III we document phenotypic and genomic evidence for local adaptation of the common chaffinch to two contrasting habitats within the island of La Palma. Individuals from the cloud forest tend to be larger and show relatively shorter, deeper, and narrower bills than the pine forest individuals. Moreover the isotopic composition of feathers revealed differences in the ^{13}C isotope among habitats, indicating that individuals from the cloud and the pine forest use different

resources for feeding. The neutral genomic structure revealed patterns of isolation by distance consistent with the geographic distribution within the island. However, Genome-Environment Association methods and Genome Wide Association Studies revealed loci that were structuring individuals by habitat type, which were associated with environmental variables (NDVI and temperature seasonality), and phenotypic variables (bill width), independent of geographic distance.

Finally, in Chapter IV we provide the genome assembly for the common chaffinch. We obtained a high-quality chromosome-level assembly using a combination of shotgun, Chicago and Hi-C libraries. The assembly also contains the sex chromosomes Z and W because the sample was a female (the heterogametic sex in birds), although the W chromosome was very fragmented and could not be properly assembled. The assembly showed no contamination and the annotation was very complete, as indicated by a BUSCO score of 93.9% using the avian dataset.

GENERAL DISCUSSION AND CONCLUSIONS

Combination of pictures from the covers of Chapters I, II and III

GENERAL DISCUSSION

Biological systems with populations and taxa at different stages of the speciation continuum provide an excellent opportunity to study the mechanisms involved in the divergence process (Shaw & Mullen, 2014; Supple et al., 2015; Henderson & Brelsford, 2020). Moreover, birds on oceanic islands are particularly suitable for the study of the divergence process and over the years have contributed to the development of ecological and evolutionary theories (Barton, 1996; Blondel, 2000; Grant, 2001; Price, 2008; Losos & Ricklefs, 2009, Warren et al., 2015). Studies using birds on islands have improved our understanding of the divergence process by testing the role of micro evolutionary processes such as drift and selection (Clegg, Degnan, Mortitz, Estoup, Kikkawa & Owens, 2002; Clegg, 2009), biogeographic processes such as colonization, speciation and extinction (Valente et al., 2020), niche shifts (Lapiedra et al., 2021), adaptive radiations (Grant & Grant, 2002; Ricklefs & Bermingham, 2007; Campana et al., 2019), convergent evolution (Cooper & Uy, 2017), the relationship between effective population size and adaptive potential (Leroy et al., 2021), and even changes in brain size (Sayol et al., 2018) and song complexity (Robert et al., 2021). In this PhD thesis, using insular bird species and populations as a model system, we have combined different approaches to understand evolutionary processes at three different scales, and addressing several of the topics mentioned above. The three scales cover different stages along the speciation continuum, ranging from a broad comparison among four species that have colonized oceanic islands to study evolution repeatability, to a species-level radiation from the colonization of several oceanic archipelagos by a single species, down to the small local scale within a small oceanic island, where different populations have adapted to different habitats.

The genetic basis of evolutionary divergence on oceanic islands: Inferring parallel or non-parallel evolution on species subjected to similar selective pressures

Understanding the repeatability of evolution and the genomic basis for the emergence of similar adaptive responses upon exposure to similar selective pressures is a long-standing goal of evolutionary biology (Stern & Orgogozo, 2008; Gompel & Prud'homme, 2009; Morris, 2010; Losos, 2011; Conte et al., 2012; Martin & Orgogozo, 2013; Stern, 2013; Rosenblum et al., 2014; Blount et al., 2018; Lamichhaney et al., 2019; Vizueta et al., 2019). In the first chapter of this PhD thesis, we focus on evaluating whether being subjected to insular selective pressures upon island colonization results in parallel phenotypic and genomic evolution. Parallel phenotypic evolution through similar selective pressures has been documented in several systems, such as the lateral plate number and body shape of sticklebacks upon independent colonization of freshwater environments (Schluter et al., 2004), eco-morphological adaptations to different crater lakes by cichlids (Elmer et al., 2014), eco-morphs of *Anolis* lizards on different Caribbean islands (Losos et al., 1998) and mice adapting their coat color to the lighter sand of dune environments (Steiner et al., 2009). However, these parallel phenotypic changes could arise through the same mutations and genes as selection acts on standing genetic variation, as is the case in sticklebacks (Schluter et al., 2004), or alternatively through completely different genes, as is the case with the cichlid eco-morphs (Elmer et al., 2014) and the changes in color coat in mice (Steiner et al., 2009).

In islands, both drift and selection could promote phenotypic divergence, but if the same phenotypic changes are detected, is likely that are the product of selection because the stochastic nature of drift is improbable to produce modifications in the same direction many times independently (Harvey & Page,l 1991; Clegg, 2009; Rosenblum et al., 2014). Following island colonization, species are exposed to similar pressures leading to the same adaptations in many organisms (i.e., island rule, Benítez-López et al., 2021), which are promoted by the reduced inter-specific competition and increased intra-specific competition, and by the different

resources available in insular environments (Blondel, 2000; Lomolino, 2005; Losos & Ricklefs, 2009; Lomolino et al., 2012). We identified parallel changes in body and bill size that were consistent with the island rule, with small birds becoming larger and with longer bills (Clegg & Owens, 2002; Benítez-López et al., 2021), changes that are likely the result of selection. Moreover, the differences in available resources usually promote dietary shifts, and using feather isotopic composition as a diet proxy we find changes between insular and mainland conspecifics. Because these phenotypic changes are due to the effect of natural selection through similar selective pressures, and is acting on avian species that share similar genomic features (Singhal et al., 2015), we asked whether shared phenotypic traits had a similar genetic basis (Rosenblum et al., 2014). Our results indicate a lack of evidence for parallel molecular evolution, and genomic changes were found in different regions and genes, indicating that they were not caused by the same mutations. We did however find that all species showed changes in pathways involved in gene regulation, suggesting that regulatory differences rather than variants in coding regions played a major role in divergence. In our system, the parallel phenotypic changes detected correspond to complex quantitative traits like morphological characters (Bosse et al., 2017), and these polygenic traits can be modified through many different possible pathways to obtain similar phenotypes (i.e., low genotype-phenotype degeneracy), so that the probability of parallel molecular evolution is reduced (Boyle et al., 2017; Szukala et al., 2022).

Other factors like demographic history also influence the probability of molecular parallelism (Rosenblum et al., 2014). The four species in our study share a pattern of reduced effective population size upon island colonization (Woolfit & Bromham, 2005; Leroy et al., 2021), which implies that all the species suffered a decrease in genetic diversity due to drift resulting from founder effects. This may limit the adaptive variability of certain genetic pathways on each specific lineage, thus reducing the probability of parallel molecular evolution. Therefore, among distantly related species, polygenic selection and species-specific demographic history can override the effect of shared selective pressures and common genomic features. Thus, when comparing the genomic landscapes among the four species we found heterogeneous differentiation patterns across the genome that were

shaped by very different processes during the divergence of each species, such as inversions, selective sweeps and recurrent selection. This resulted in lineage-specific outlier regions, suggesting that each species has its own important genes involved in divergence (Delmore et al., 2018). However, at a broader scale, we did find parallelism in the form of similar pathways between the common chaffinch and the house finch, including cell adhesion and histone modifications. And in general, as mentioned before, all species showed genetic changes related to gene regulation instead of directly related to specific functional genes. Recent studies are focusing on the major role of gene regulation in the divergence of polygenic traits (Fagny & Austerlitz, 2021; Mathieson 2021; Verta & Jacobs, 2021).

Results in Chapter II, which show the sequential colonization of the Atlantic archipelagos of Azores, Madeira and the Canary Islands by the common chaffinch, do not allow to test for parallel evolution upon island colonization because the colonization events are not independent (Losos, 2011). The common phenotypic changes among archipelagos compared to the mainland populations are likely due to common ancestry. Selection acted in the first step of the colonization in Azores and then the phenotypic changes spread through the subsequent colonization steps to the rest of the archipelagos.

In Chapter III we focus on studying the process of local adaptation to two different habitats within La Palma, the cloud and the pine forests. It would have been interesting to have a replicate island with the same scenario of both habitats available without competitors to look for patterns of parallel evolution to the same environment. However, in El Hierro, which is the only island that provides a similar ecological scenario (i.e., the presence of pine forest that is unoccupied by the blue chaffinch), the common chaffinch seems to be largely restricted to the cloud forest, and the pine forest is too small in area to host sufficiently large populations, which prevents this hypothesis from being tested.

Drift versus selection on islands

The roles of drift and selection in the divergence process are commonly studied in insular systems because they offer the ideal settings of small founding populations and novel environments to promote the action of both evolutionary forces (Barton, 1996; Kolbe et al., 2012). Even though random drift is an important factor shaping the genomic divergence upon island colonization due to founder events, it usually acts along with selection (Funk et al., 2016; Prentice et al., 2017; Sendell-Price et al., 2021). We have detected evidence of selection being crucial in the divergence of insular systems in Chapters I, II and III.

The role of selection as the main force generating the pattern of convergent phenotypes upon island colonization detected in Chapter I has been already discussed in the previous section. In Chapter II, first we observed that the insular populations of the common chaffinch follow the island rule (Clegg & Owens, 2002; Benítez-López et al., 2021), being larger than their mainland counterparts, and therefore, as mentioned before, that changes are likely to be promoted by selection to the insular selective pressures in the first colonization step, rather than drift (Harvey & Pagel, 1991). For instance, Sendell-Price et al., (2021) detected the same pattern of larger insular individuals of *Zosterops lateralis* (Passeriformes: Zosteropidae) compared to the mainland, and invoked selection to explain it, even though drift also played an important role in divergence. We also detected the expected dramatic effects of founder events in sequential colonizations by a successive reduction in genetic diversity along the colonization route due to drift (Clegg, Degnan, Kikkawa et al., 2002). However, in Chapter III we found evidence for local adaptation within the island of La Palma, which is one of the last Macaronesian islands to be colonized, suggesting that in addition to drift, selection is also promoting divergence. We observed that selection can still act on populations with impoverished genetic diversity, allowing the common chaffinch to adapt to different habitats and providing support for the strong role of selection as an evolutionary force on islands. The effect that drift and reduced effective population sizes on islands have on the evolutionary potential of species has been previously studied (Woolfit & Bromham, 2005; Funk et al., 2016; Leroy et al.,

2021). The fact that most adaptive traits in nature are polygenic may help avoid the effects of drift that diminish the genetic variability for selection to act on, by having many alternative genomic routes to obtain the same phenotype.

Evolutionary and colonization history of the common chaffinch radiation in the Macaronesian region

Across the broad geographic range of the common chaffinch, the colonization and diversification on the Atlantic archipelagos have been widely studied, proving again the key role of islands as model systems for the study of evolution (Grant, 1980; Dennison & Baker, 1991; Baker & Marshall, 1999; Marshall & Baker, 1999; Suarez, Betancor, Klassert, Almeida, Hernández & Pestano, 2009; Samarasin-dissanayake 2010; Lachlan et al., 2013; Rodrigues et al., 2014; Illera et al., 2018). A main objective of this PhD thesis was to reconstruct the evolutionary history and colonization route of the common chaffinch in the Macaronesian region using genome-wide data. Our results confirm the previous hypothesis on the colonization route by Marshall and Baker (1999), arriving first to Azores from the continent, then Madeira and finally the Canary islands. The genomic data provide high phylogenetic resolution, detecting that even within the Canary Islands, each island is an independent evolutionary lineage. Mitochondrial data could not resolve structure and relationships to that extent, probably due to incomplete lineage sorting (Maddison & Knowles, 2006; Cariou et al., 2013).

The colonization route within the Canary Islands remains an unresolved issue, probably because it happened within a short period of time, and further research is needed to clarify it, probably using whole-genome data in order to increase resolution (Szarmach et al., 2021). The phenotypic differentiation across the common chaffinch populations is congruent with the clades obtained in the phylogenomic analysis. Differences within the Canary Islands are less obvious but there are already four subspecies recognised based on morphological and genetic evidence that corresponded to each island occupied by the common chaffinch, except La Gomera and Tenerife that shared the same subspecies (Martín &

Lorenzo, 2001; Suarez, Betancor, Klassert, et al., 2009; Suarez, Betancor & Pestano, 2009; Illera et al., 2018).

We implemented an integrative taxonomy approach that included the combination of several lines of evidence, including phenotypic and genomic data, and results provided support for the taxonomic hypothesis that the current subspecies of North Africa, Azores, Madeira and the Canary Islands represent different species (Winker, 2021). Therefore, we proposed that the respective subspecies *spodiogenys*, *moreletti*, *maderensis* and *canariensis* be treated as separate biological species, maintaining the subspecies within the Canary Islands. This taxonomic proposal has relevant implications for conservation, especially for insular populations that have small population sizes and reduced genetic diversity (Martín, 2009; Sangster et al., 2016; Kearns et al., 2022). Moreover, volcanic activity, such as the eriuption of Cumbre Vieja in La Palma in 2021, could be an additional potential risk for insular populations in volcanic islands.

Moreover, knowing the colonization route of the common chaffinch allows a better interpretation of the genomic landscape of mainland (Iberia) versus insular (La Palma) counterparts of the common chaffinch in Chapter I, which is consistent with the recurrent selection model (Irwin et al., 2016). The strongest selection episode took place in the first step (i.e., the colonization of the Azores) and then those regions were subjected to selection again in the following steps (colonization of Madeira and Canaries), reducing genetic diversity, increasing F_{ST} and showing low d_{xy} due to past events of selection in the ancestor.

Local adaptation at a small spatial scale: Selection versus gene flow

Genomic divergence depends, among other factors, on the interaction between directional selection and the opposing effect of gene flow (Lenormand, 2002; Nosil & Feder, 2012). In turn, the magnitude of gene flow is determined by the dispersal ability of the species and their degree of geographic isolation (Lenormand, 2002); therefore, gene flow is expected to be intense for highly mobile organisms at small

spatial scales (Price, 2008). However, in Chapter III, we found evidence of local adaptation to different habitats, supported by phenotypic and genomic differentiation at very small spatial scale in the common chaffinch, which is an organism with high dispersal capacity and no apparent geographical barriers within La Palma. An increasing number of studies are finding this striking pattern in highly mobile organisms like birds (i.e., Porlier et al., 2012; Bertrand et al., 2014; Szulkin et al., 2016; Gabrielli et al., 2020). Therefore, the antagonist role of gene flow might be reduced by non-random dispersal that would, on the contrary, promote divergence (Garant et al., 2005; Senar et al., 2006; Edelaar et al., 2008; Edelaar & Bolnick, 2012), a process that may eventually lead to speciation (Schluter, 2000; Rundle & Nosil, 2005). Non-random dispersal between habitats could be promoted by different processes, such as strong natural selection, spatial autocorrelation, habitat choice, homing behvaior, and natural or sexual selective barriers against migrants (Richardson et al., 2014). Also, the insular syndrome is consistent with our results from Chapters II and III, explaining the reduced dispersal capacity, behavioural changes, changes in body size and niche expansion associated with island colonization by birds (Blondel, 2000).

High quality chromosome-level genome assembly of the common chaffinch

The availability of reference genomes in the last decade has increased dramatically for non-model organisms, especially for birds (Bravo et al., 2021). The access to these resources has allowed the improvement of genomic studies (Prasad et al., 2021) and biodiversity conservation (Brandies et al., 2019). For instance, in order to perform the comparative analysis in Chapter I, we have used three different reference genomes, the New Caledonian crow genome and two genomes generated by our group, the common chaffinch (Chapter IV) and the dark-eyed junco (Friis et al., 2022), both available in NCBI. Of course, we could have performed the analysis by aligning all the species to, for instance, the zebra finch genome, but since they are more distantly related to this species than to the references that we used, fewer data would have been available and the results would have shown lower resolution (Prasad et al., 2021). In our case we used the common chaffinch

reference genome also for the house finch, because the house finch reference genome currently available has low quality compared with our assembly, leading to poorer results than using as a reference a high-quality genome from a closely related species from the same family. The ideal situation would be to have high-quality reference genomes for all the species because it would improve the quality of comparative genomic studies. For instance, the challenging goal of sequencing the genomes of all eukaryotes in ten years by the Earth BioGenome Project (EBP) would have important implications for the advance of comparative genomics and biodiversity preservation (Lewin et al., 2018; Feng et al., 2020). The chromosome-level assembly of the common chaffinch that we present in Chapter IV, has proven already to be a valuable resource as a reference genome in the previous three chapters, and of course, will continue to be in future research. It will be included in the Ten-Thousand Bird Genomes consortium (B10K, Zhang 2015) becoming part of a project that would be a major pillar for any comparative genomic study involving avian species.

CONCLUSIONS

- Parallel molecular evolution in the red-billed chough, common chaffinch, house finch, and dark-eyed junco is dependent on the demographic history and the polygenic nature of adaptive traits, which outweigh the effects of being exposed to similar selective pressures and shared genomic features, especially for more distantly related taxa.

- Polygenic traits can be modified through several alternative genetic routes diminishing the probability of parallel molecular evolution but facilitating the adaptive potential of organisms, for instance when they have reduced genetic diversity. Modifying gene regulation rather than specific functional genes seems to play a key role in polygenic selection.

- Heterogeneous landscapes of differentiation among red-billed chough, common chaffinch, house finch, and dark-eyed junco reveal that loci promoting divergence are taxon-specific, and can be shaped by several processes, like selective sweeps, inversions and recurrent selection.

- The circuitous colonization route of the Atlantic archipelagos by the common chaffinch started from the continent to Azores, then Madeira and finally the Canary Islands resulting in phenotypic and genomic divergence among populations leading to an evolutionary radiation through the formation of four new species in the genus *Fringilla*.

- The sequential colonization route of the common chaffinch in the Macaronesian region revealed the dramatic effects of several successive founder events showing a progressive reduction of genetic diversity at each step.

- The common chaffinch within the Canary Islands shows genomic divergence among islands that is consistent with the established subspecies, but further phenotypic characterization of the populations from Tenerife and La Gomera is needed to separate them as subspecies.

- On islands, as demonstrated for the common chaffinch in cloud and pine forest populations of La Palma, local adaptation at small spatial scales can occur even for highly mobile organisms like birds, due to strong selective pressures and markedly reduced or non-random dispersal.

- The environmental variable that best explains the habitat-related local adaptation in the common chaffinch within the island of La Palma is the Normalized Difference Vegetation Index (NDVI), which reflects differences in vegetation, and implies different resources and conditions.

- The chromosome-level assembly of the common chaffinch is a valuable resource as a high-quality reference genome to implement in genomic research due to its completeness and the availability of a functional and structural annotation that allows the detection of candidate genes involved in evolutionary processes.

CONCLUSIONES

- La evolución molecular paralela en la chova piquirroja, el pinzón común, el camachuelo mexicano y el junco ojioscuro depende de la historia demográfica y de la naturaleza poligénica de los rasgos adaptativos, que superan los efectos de estar expuestos a presiones selectivas similares y compartir características genómicas, especialmente para taxones relacionados más lejanamente.

- Los rasgos poligénicos puede ser modificados a través de varias rutas genéticas alternativas, lo cual disminuye la probabilidad de la evolución molecular paralela pero al mismo tiempo facilita el potencial adaptativo de los organismos, por ejemplo cuando muestran una diversidad genética reducida. La modificación de la regulación génica en lugar de genes funcionales específicos parece jugar un papel clave en la selección poligénica.

- La heterogeneidad de los paisajes genómicos de diferenciación entre la chova piquirroja, el pinzón común, el camachuelo mexicano y el junco ojioscuro muestra que los loci encargados de promover la divergencia son específicos de cada taxón, y que pueden ser generados por procesos variados, como barridos selectivos, inversiones y selección recurrente.

- La colonización poco intuitiva de los archipiélagos Atlánticos por el pinzón común empezó en el continente hasta Azores, después Madeira y finalmente las Islas Canarias resultando en la divergencia fenotípica y genómica entre poblaciones dando lugar a una radiación a través de la formación de cuatro especies nuevas del género *Fringilla.*

- La ruta de colonización secuencial del pinzón común en la región Macaronésica reveló los dramáticos efectos de varios eventos fundador sucesivos mostrando un reducción progresiva de la diversidad genética en cada paso.

- El pinzón común dentro de las Islas Canarias muestra divergencia genómica entre islas que es consistente con las subespecies establecidas, pero es necesaria una mayor caracterización fenotípica de las poblaciones de Tenerife y La Gomera para separarlas como subespecies.

- En islas, la adaptación local a pequeña escala puede ocurrir incluso para organismos altamente móviles como las aves, debido a las fuertes presiones selectivas y una dispersión marcadamente reducida o no aleatoria.

- La variable ambiental que mejor explica la adaptación local relacionada con el tipo de hábitat en el pinzón común dentro de la isla de La Palma es el Índice de Vegetación de Diferencia Normalizada (NDVI, en inglés), que refleja las diferencias de vegetación lo cual implica diferentes recursos y condiciones.

- El ensamblaje a nivel de cromosoma del pinzón común es un recurso muy valioso como genoma de referencia de alta calidad para utilizar en estudios a nivel genómico, debido a su completitud y la disponibilidad de una anotación funcional y estructural que permite la detección de genes candidatos involucrados en los procesos evolutivos.

REFERENCES

Abolins-Abols, M., Kornobis, E., Ribeca, P., Wakamatsu, K.,Peterson, M.P., Ketterson, E., & Milá, B. (2018). A role for differential gene regulation in the rapid diversification of melanic plumage coloration in the dark-eyed junco (*Junco hyemalis*). *Molecular Ecology, 27*(22), 4501-4515. https://doi.org/10.1111/mec.14878

Abzhanov, A. (2004). Bmp4 and Morphological variation of beaks in Darwin's Finches. *Science, 305*(5689), 1462–1465. https://doi.org/10.1126/science.1098095

Abzhanov, A., Kuo, W. P., Hartmann, C., Grant, B. R., Grant, P. R., & Tabin, C. J. (2006). The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*, *442*(7102), 563-567. https://doi.org/10.1038/nature04843

Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences*, *114*(27), 7061–7066. https://doi.org/10.1073/pnas.1616755114

Albaladejo, R. G., Martín-Hernanz, S., Reyes-Betancort, J. A., Santos-Guerra, A., Olangua-Corral, M., & Aparicio, A. (2021). Reconstruction of the spatio-temporal diversification and ecological niche evolution of *Helianthemum* (Cistaceae) in the Canary Islands using genotyping-by-sequencing data. *Annals of Botany, 127*(5), 597–611. https://doi.org/10.1093/aob/mcaa090

Alexa, A., & Rahnenfuhrer, J. (2016). topGO: Enrichment analysis for Gene Ontology. R package version 2.28. 0. Cranio.

Aleixandre, P., Hernández Montoya, J., & Milá, B. (2013). Speciation on oceanic islands: rapid adaptive divergence vs. cryptic speciation in a Guadalupe Island songbird (Aves: Junco). *PLoS ONE, 8*(5) , e63242. https://doi.org/10.1371/journal.pone.0063242

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research, 19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109.vidual

Alkhunaizi, E., Shaheen, R., Bharti, S. K., Joseph-George, A. M., Chong, K., Abdel-Salam, G. M., ... & Chitayat, D. (2018). Warsaw breakage syndrome: Further clinical and genetic delineation. *American Journal of Medical Genetics Part A, 176*(11), 2404-2418. https://doi.org/10.1002/ajmg.a.40482

Alström, P., Rheindt, F. E., Zhang, R., Zhao, M., Wang, J., Zhu, X., ... & Olsson, U. (2018). Complete species-level phylogeny of the leaf warbler (Aves: Phylloscopidae) radiation. *Molecular Phylogenetics and Evolution, 126*, 141-152. https://doi.org/10.1016/j.ympev.2018.03.031

Alter, S., Munshi-South, J., & Stiassny, M. L. J. (2017). Genome-wide SNP data reveal cryptic phylogeographic structure and microallopatric divergence in a rapids-adapted clade of cichlids from the Congo River Accepted. *ARPN Journal of Engineering and Applied Sciences, 12*(10), 3218–3221. https://doi.org/10.1111/ijlh.12426

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*(3), 403-410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research, 25*(17), 3389-3402. https://doi.org/10.1093/nar/25.17.3389.

Andrews, W. D., Barber, M., & Parnavelas, J. G. (2007). Slit-Robo interactions during cortical development. *Journal of Anatomy, 211*(2), 188–198. https://doi.org/10.1111/j.1469-7580.2007.00750.x

Appelhans, M. S., Paetzold, C., Wood, K. R., & Wagner, W. L. (2020). RADseq resolves the phylogeny of Hawaiian Myrsine (Primulaceae) and provides evidence for hybridization. *Journal of Systematics and Evolution, 58*(6), 823–840. https://doi.org/10.1111/jse.12668

Arisawa, K., Yazawa, S., Eguchi, J. I., Kagami, H., & Ono, T. (2005). Spatiotemporal pattern of EphA4 gene expression in developing quail limb buds. *Animal Science Journal, 76*(1), 65–71. https://doi.org/10.1111/j.1740-0929.2005.00239.x

Armstrong, C., Richardson, D. S., Hipperson, H., Horsburgh, G. J., Küpper, C., Percival-Alwyn, L., … & Spurgin, L. G. (2018). Genomic associations with bill

length and disease reveal drift and selection across island bird populations. *Evolution Letters, 2*(1), 22–36. https://doi.org/10.1002/evl3.38

Árnason, Ú., Lammers, F., Kumar, V., Nilsson, M. A., & Janke, A. (2018). Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. *Science Advances, 4*(4): eaap9873. https://doi.org/10.1126/sciadv.aap9873

Asano, Y., Markiewicz, M., Kubo, M., Szalai, G., Watson, D. K., & Trojanowska, M. (2009). Transcription factor Fli1 regulates collagen fibrillogenesis in mouse skin. *Molecular and Cellular Biology, 29*(2), 425–434. https://doi.org/10.1128/mcb.01278-08

Aslam, M. L., Bastiaansen, J. W., Crooijmans, R. P., Vereijken, A., Megens, H. J., & Groenen, M. A. (2010). A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes. *BMC Genomics, 11*(1), 1-11. https://doi.org/10.1186/1471-2164-11-647.

Ayala, D., Ullastres, A., & González, J. (2014). Adaptation through chromosomal inversions in Anopheles. *Frontiers in Genetics, 5*, 129. https://doi.org/10.3389/fgene.2014.00129

Ayroles, J. F., Carbone, M. A., Stone, E. A., Jordan, K. W., Lyman, R. F., Magwire, M. M., … & Mackay, T. F. C. (2009). Systems genetics of complex traits in *Drosophila melanogaster*. *Nature Genetics, 41*(3), 299–307. https://doi.org/10.1038/ng.332

Bai, H., Sun, Y., Liu, N., Xue, F., Li, Y., Xu, S., … & Chen, J. (2018). Single SNP- and pathway-based genome-wide association studies for beak deformity in chickens using high-density 600K SNP arrays. *BMC Genomics, 19*(1), 1–10. https://doi.org/10.1186/s12864-018-4882-8

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., … & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE, 3*(10), 1–7. https://doi.org/10.1371/journal.pone.0003376

Baker, A. J., Dennison, M. D., Lynch, A., & Le Grand, G. (1990). Genetic divergence in peripherally isolated populations of chaffinches in the Atlantic

islands. *Evolution, 44*(4), 981-999. https://doi.org/10.1111/j.1558-5646.1990.tb03819.x

Baker, A. J., & Marshall, H. D. (1999). Population divergence in Chaffinches *Fringilla coelebs* assessed with control-region sequences. Proceedings XXII International Ornithological Congress (NJ Adams and RH Slotow, Eds.). BirdLife South Africa, Durban, 1899–1913.

Bao, W., Kojima, K. K., & Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mobile Dna, 6*(1), 1-6. https://doi.org/10.1186/s13100-015-0041-9.

Bardwell, E., Benkman, C. W., & Gould, W. R. (2001). Adaptive geographic variation in Western Scrub-Jays. *Ecology, 82*(9), 2617–2627. https://doi.org/10.1890/0012-9658(2001)082[2617:AGVIWS]2.0.CO;2

Barraclough, T. G., & Vogler, A. P. (2000). Detecting the geographical pattern of speciation from species-level phylogenies. *The American Naturalist, 155*(4), 419–434. https://doi.org/10.1086/303332

Barrett, R. D. H., & Hoekstra, H. E. (2011). Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics, 12*(11), 767–780. https://doi.org/10.1038/nrg3015

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., … & Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature, 528*(7582), 405–408. https://doi.org/10.1038/nature16062

Barton, N. H. (1996). Natural selection and random genetic drift as causes of evolution on islands. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 351*(1341), 785-795. https://doi.org/10.1098/rstb.1996.0073.

Beatty, J. (2006). Replaying Life ' s Tape. *Journal of Philosophy, 103*(7), 336–362.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences, 263*(1377), 1619–1626. https://doi.org/10.1098/rspb.1996.0237

Beheregaray, L. B., Gibbs, J. P., Havill, N., Fritts, T. H., Powell, J. R., & Caccone, A. (2004). Giant tortoises are not so slow: rapid diversification and

biogeographic consensus in the Galápagos. *Proceedings of the National Academy of Sciences, 101*(101), 6514-6519. https://doi.org/10.1073/pnas.0400393101

Benítez-López, A., Santini, L., Gallego-Zamorano, J., Milá, B., Walkden, P., Huijbregts, M. A. J., & Tobias, J. A. (2021). The island rule explains consistent patterns of body size evolution in terrestrial vertebrates. *Nature Ecology and Evolution, 5*(6), 768–786. https://doi.org/10.1038/s41559-021-01426-y

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological), 57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bernhardt, N., Brassac, J., Dong, X., Willing, E. M., Poskar, C. H., Kilian, B., & Blattner, F. R. (2020). Genome-wide sequence information reveals recurrent hybridization among diploid wheat wild relatives. *Plant Journal, 102*(3), 493–506. https://doi.org/10.1111/tpj.14641

Berry, A. J., Ajioka, J. W., & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics, 129*(4), 1111–1117. https://doi.org/10.1093/genetics/129.4.1111

Bertrand, J. A. M., Bourgeois, Y. X. C., Delahaie, B., Duval, T., García-Jiménez, R., Cornuault, J., … & Thébaud, C. (2014). Extremely reduced dispersal and gene flow in an island bird. *Heredity, 112*(2), 190–196. https://doi.org/10.1038/hdy.2013.91

Bhuiyan, Z. A., Klein, M., Hammond, P., Van Haeringen, A., Mannens, M. M. A. M., Van Berckelaer-Onnes, I., & Hennekam, R. C. M. (2006). Genotype-phenotype correlations of 39 patients with Cornelia de Lange syndrome: The Dutch experience. *Journal of Medical Genetics, 43*(7), 568–575. https://doi.org/10.1136/jmg.2005.038240

Bijlsma, R., & Loeschcke, V. (2012). Genetic erosion impedes adaptive responses to stressful environments. *Evolutionary Applications, 5*(2), 117–129. https://doi.org/10.1111/j.1752-4571.2011.00214.x

Blanco, G., Laiolo, P., & Fargallo, J. A. (2014). Linking environmental stress, feeding-shifts and the “island syndrome”: A nutritional challenge hypothesis.

Population Ecology, 56(1), 203–216. https://doi.org/10.1007/s10144-013-0404-3

Blanquart, F., Gandon, S., & Nuismer, S. L. (2012). The effects of migration and drift on local adaptation to a heterogeneous environment. *Journal of Evolutionary Biology, 25*(7), 1351–1363. https://doi.org/10.1111/j.1420-9101.2012.02524.x

Blanquart, F., Kaltz, O., Nuismer, S. L., & Gandon, S. (2013). A practical guide to measuring local adaptation. *Ecology Letters, 16*(9), 1195–1205. https://doi.org/10.1111/ele.12150

Bleidorn, C. (2017). *Phylogenomics: an introduction*. Springer. https://doi.org/10.1007/978-3-319-54064-1

Blondel, J. (2000). Evolution and ecology of birds on islands: trends and prospects. *Vie et Milieu/Life & Environment*, 205:220.

Blount, Z. D., Lenski, R. E., & Losos, J. B. (2018). Contingency and determinism in evolution: Replaying life's tape. *Science, 362*(6415), eaam5979. https://doi.org/10.1126/science.aam5979

Boag, P. T., & Grant, P. R. (1981). Intense natural selection in a population of Darwin's finches (Geospizinae) in the Galapagos. *Science, 214*(4516), 82–85. https://doi.org/10.1126/science.214.4516.82

Bolger, A. M., Lohse, M., & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15):2114-2120. https://doi.org/10.1093/bioinformatics/btu170.

Bolnick, D. I., & Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics, 38*, 459–487. https://doi.org/10.1146/annurev.ecolsys.38.091206.095804

Bolnick, D. I., Barrett, R. D. H., Oke, K. B., Rennison, D. J., & Stuart, Y. E. (2018). (Non) parallel evolution. *Annual Review of Ecology, Evolution, and Systematics, 49*, 303–330. https://doi.org/10.1146/annurev-ecolsys-110617-062240

Booker, T. R., Yeaman, S., & Whitlock, M. C. (2021). Global adaptation complicates the interpretation of genome scans for local adaptation. *Evolution Letters, 5*(1), 4–15. https://doi.org/10.1002/evl3.208

Borcard, D., Gillet, F., & Legendre, P. (2018). *Numerical Ecology with R*. Springer.

Bosse, M., Spurgin, L. G., Laine, V. N., Cole, E. F., Firth, J. A., Gienapp, P., … & Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science, 358*(6361), 365–368. https://doi.org/10.1126/science.aal3298

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kuhnert, D., De Maio, N. (2019). BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology, 15*(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650

Bourgeois, Y. X. C., Bertrand, J. A. M., Delahaie, B., Cornuault, J., Duval, T., Milá, B., & Thébaud, C. (2016). Candidate gene analysis suggests untapped genetic complexity in melanin-based pigmentation in birds. *Journal of Heredity, 107*(4), 327–335. https://doi.org/10.1093/jhered/esw017

Bourgeois, Y. X. C., Bertrand, J. A. M., Delahaie, B., Holota, H., Thébaud, C., & Milá, B. (2020). Differential divergence in autosomes and sex chromosomes is associated with intra-island diversification at a very small spatial scale in a songbird lineage. *Molecular Ecology, 29*(6), 1137–1153. https://doi.org/10.1111/mec.15396

Bourgeois, Y. X. C., & Warren, B. H. (2021). An overview of current population genomics methods for the analysis of whole- genome resequencing data in eukaryotes. *Molecular Ecology, 30*(23), 6036-6071. https://doi.org/10.1111/mec.15989

Boyer, A. G., & Jetz, W. (2010). Biogeography of body size in Pacific Island birds. *Ecography, 33*(2), 369–379. https://doi.org/10.1111/j.1600-0587.2010.06315.x

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell, 169*(7), 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Bradley, B. J., Gerald, M. S., Widdig, A., & Mundy, N. I. (2013). Coat Color Variation and Pigmentation Gene Expression in Rhesus Macaques (*Macaca mulatta*). *Journal of Mammalian Evolution, 20*(3), 263–270. https://doi.org/10.1007/s10914-012-9212-3

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes, 10*(11), 846. https://doi.org/10.3390/genes10110846

Braun, E. L., Cracraft, J., & Houde, P. (2019). Resolving the avian tree of life from top to bottom: the promise and potential boundaries of the phylogenomic era. *Avian Genomics in Ecology and Evolution*, 151-210. Springer, Cham. https://doi.org/10.1007/978-3-030-16477-5_6.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H., & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics, 140*(2), 783–796. https://doi.org/10.1093/genetics/140.2.783

Bravo, G. A., Schmitt, C. J., & Edwards, S. V. (2021). What have we learned from the first 500 avian genomes?. *Annual Review of Ecology, Evolution, and Systematics, 52*, 611-639. https://doi.org/10.1146/annuerev-ecolsys-012121-085928

Brookfield, J. (2002) Review of genes, categories, and species by Jody Hey. *Genetic Research, 79*(1):107–108. https://doi.org/10.1017/S0016672302215608

Brown, R. M., Siler, C. D., Oliveros, C. H., Esselstyn, J. A., Diesmos, A. C., Hosner, P. A., ... & Alcala, A. C. (2013). Evolutionary processes of diversification in a model island archipelago. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 411-435. https://doi.org/10.1146/annurev-ecolsys-110411-160323

Bruders, R., van Hollebeke, H., Osborne, E. J., Kronenberg, Z., Maclary, E., Yandell, M., & Shapiro, M. D. (2020). A copy number variant is associated with a spectrum of pigmentation patterns in the rock pigeon (*Columba livia*). *PLoS Genetics, 16*(5), 1–25. https://doi.org/10.1371/journal.pgen.1008274

Brugmann, S. A., Powder, K. E., Young, N. M., Goodnough, L. H., Hahn, S. M., James, A. W., Helms, J. A., & Lovett, M. (2010). Comparative gene expression analysis of avian embryonic facial structures reveals new candidates for human craniofacial disorders. *Human Molecular Genetics, 19*(5), 920–930. https://doi.org/10.1093/hmg/ddp559

Burga, A., Wang, W., Ben-David, E., Wolf, P. C., Ramey, A. M., Verdugo, C., … & Kruglyak, L. (2017). A genetic signature of the evolution of loss of flight in

the Galapagos cormorant. *Science, 356*(6341), eaal3345. https://doi.org/10.1126/science.aal3345

Burkett, Z. D., Day, N. F., Kimball, T. H., Aamodt, C. M., Heston, J. B., Hilliard, A. T., … & White, S. A. (2018). FoxP2 isoforms delineate spatiotemporal transcriptional networks for vocal learning in the zebra finch. *ELife, 7*, e30649. https://doi.org/10.7554/eLife.30649

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters, 1*(3), 118–131. https://doi.org/10.1002/evl3.14

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., … & 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*(11), 1656-1665.https://doi.org/10.1101/gr.196485.115

Burt, D. W. (2002). Origin and evolution of avian microchromosomes. *Cytogenetic and Genome Research, 96*(1–4), 97–112. https://doi.org/10.1159/000063018

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLoS Computational Biology, 8*(12), e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Butlin, R. K., Galindo, J., & Grahame, J. W. (2008). Review. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1506), 2997–3007. https://doi.org/10.1098/rstb.2008.0076

Butlin, R. K., & Stankowski, S. (2020). Is it time to abandon the biological species concept? No. *National Science Review, 7*(8), 1400–1401. https://doi.org/10.1093/nsr/nwaa108

Byers, K. J. R. P., Xu, S., & Schlüter, P. M. (2017). Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? *Molecular Ecology, 26*(1), 277-290. https://doi.org/10.1111/mec.13678

Campagna, L., Benites, P., Lougheed, S.C., Lijtmaer, D.A., Di Giacomo, A.S., Eaton, M.D., & Tubaro, P.L. (2012). Rapid phenotypic evolution during incipient speciation in a continental avian radiation. *Proceedings of the Royal Society*

B: Biological Sciences, 279(1734), 1847-1856. https://doi.org/10.1098/rspb.2011.2170

Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances, 3*(5), e1602404. https://doi.org/10.1126/sciadv.1602404

Campana, M. G., Corvelo, A., Shelton, J., Callicrate, T. E., Bunting, K. L., Riley-Gillis, B., ... & Fleischer, R. C. (2020). Adaptive radiation genomics of two ecologically divergent Hawai ‘ian honeycreepers: the ‘akiapōlā ‘au and the Hawai ‘i ‘amakihi. *Journal of Heredity, 111*(1), 21-32. https://doi.org/10.1093/jhered/esz057

Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is Speciation Genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society, 124*(4), 561-583. https://doi.org/10.1093/biolinnean/bly063

Campeau, P. M., Lenk, G. M., Lu, J. T., Bae, Y., Burrage, L., Turnpenny, P., ... & Lee, B. H. (2013). Yunis-Varon syndrome is caused by mutations in FIG4, encoding a phosphoinositide phosphatase. *The American Journal of Human Genetics*, *92*(5), 781-791.https://doi.org/10.1016/j.ajhg.2013.03.020

Cariou, M., Duret, L., & Charlat, S. (2013). Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution, 3*(4), 846–852. https://doi.org/10.1002/ece3.512

Casillas, S., & Barbadilla, A. (2017). Molecular population genetics. *Genetics, 205*(3), 1003–1035. https://doi.org/10.1534/genetics.116.196493

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

Cezard, T., Rüber, L., Gharbi, K., Dasmahapatra, K. K., Ford, A. G. P., & Day, J. J. (2015). High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology, 24*(13), 3421–3440. https://doi.org/10.1111/mec.13247

Chan, P. P., & Lowe, T. M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleid Acids Research, 44*(D1), D184-D189. https://doi.org/10.1093/nar/gkv1309.

Chapman, J. A., Ho, I., Sunkara, S., Luo, S., Schroth, G. P., & Rokhsar, D. S. (2011). Meraculous: de novo genome assembly with short paired-end reads. *PloS One, 6*(8), e23501. https://doi.org/10.1371/journal.pone.0023501

Charlesworth, B. (1998). Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution, 15*(5), 538–543. https://doi.org/10.1093/oxfordjournals.molbev.a025953

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics, 70*(2), 155–174. https://doi.org/10.1534/genetics.115.178558

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics 18*(1), 237-268.. https://doi.org/10.1146/annurev.es.18.110187.001321

Chase, M. A., Ellegren, H., & Mugal, C. F. (2021). Positive selection plays a major role in shaping signatures of differentiation across the genomic landscape of two independent *Ficedula* flycatcher species pairs. *Evolution, 75*(9), 2179-2196.https://doi.org/10.1111/evo.14234

Chaves, J. A., Cooper, E. A., Hendry, A. P., Podos, J., De León, L. F., Raeymaekers, J. A. M., … & Uy, J. A. C. (2016). Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology, 25*(21), 5282-5295. https://doi.org/10.1111/mec.13743

Cheek, R. G., Forester, B. R., Salerno, P. E., Trumbo, D. R., Chen, N., Sillett, S., … & Funk, W. C. (2022). Habitat-linked genetic variation supports microgeographic adaptive divergence in an island- endemic bird species. *Molecular Ecology*. https://doi.org/10.1111/mec.16438

Chen, Y., Gong, Q., Lai, J., Song, M., Liu, Y., Wu, Y., … & Long, Z. (2020). Transcriptome analysis identifies candidate genes associated with skin color variation in *Triplophysa siluroides*. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics, 35*, 100682. https://doi.org/10.1016/j.cbd.2020.100682

Christmas, M. J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O., & Webster, M. T. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology, 28*(6), 1358–1374. https://doi.org/10.1111/mec.14944

Clegg, S. (2009). Evolutionary changes following island colonization in birds. *The Theory of Island Biogeography Revisited*, (1), 293–325. https://doi.org/10.1515/9781400831920.293

Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. F. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 8127–8132. https://doi.org/10.1073/pnas.102583399

Clegg, S. M., Degnan, S. M., Moritz, C., Estoup, A., Kikkawa, J., & Owens, I. P. F. (2002). Microevolution in island forms: The roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution, 56*(10), 2090–2099. https://doi.org/10.1111/j.0014-3820.2002.tb00134.x

Clegg, S. M., & Owens, I. P. F. (2002). The "island rule" in birds: Medium body size and its ecological explanation. *Proceedings of the Royal Society B: Biological Sciences, 269*(1498), 1359–1365. https://doi.org/10.1098/rspb.2002.2024

Clement, P. (2018). Common chaffinch (*Fringilla coelebs*). *Handbook of the birds of the world alive*. Lynx Edicions, Barcelona (retrieved from http://www. hbw. com/node/61286 on 2 Jan 2018).

Clement, P., (2020). Common Chaffinch (*Fringilla coelebs*), version 1.0. In Birds of the World (J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie, and E. de Juana, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.comcha.01

Collar, N., Newton, I., & Bonan, A. (2020). Finches (Fringillidae). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A. & de Juana, E. (eds.). *Handbook of the Birds of the World Alive*. Lynx Edicions, Barcelona. (retrieved from https://www.hbw.com/node/52376 on 15 January 2020).

Conte, G. L., Arnegard, M. E., Peichel, C. L., & Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of*

the Royal Society B: Biological Sciences, 279(1749), 5039–5047. https://doi.org/10.1098/rspb.2012.2146

Cooper, E. A., & Uy, J. A. C. (2017). Genomic evidence for convergent evolution of a key trait underlying divergence in island birds. *Molecular Ecology, 26*(14), 3760–3774. https://doi.org/10.1111/mec.14116

Cornetti, L., Valente, L. M., Dunning, L. T., Quan, X., Black, R. A., Hébert, O. H., & Savolainen, V. (2015). The genome of the "great speciator" provides insights into bird diversification. *Genome Biology and Evolution, 7*(9), 2680-2691. https://doi.org/10.1093/gbe/evv168

Cowie, R. H., & Holland, B. S. (2006). Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. *Journal of Biogeography, 33*(2), 193–198. https://doi.org/10.1111/j.1365-2699.2005.01383.x

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.

Cronquist, A. (1978). Once again, what is a species? Biosystematics in agriculture. Beltsville Symposia in Agr. Res., 2, 3–20.

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology, 23*(13), 3133-3157. https://doi.org/10.1111/mec.12796

Cunha, R. L., Castilho, R., Rüber, L., & Zardoya, R. (2005). Patterns of cladogenesis in the venomous marine gastropod genus *Conus* from the Cape Verde islands. *Systematic Biology, 54*(4), 634-650. https://doi.org/10.1080/106351591007471

Czekanski-Moir, J. E., & Rundell, R. J. (2019). The ecology of nonecological speciation and nonadaptive radiations. *Trends in Ecology and Evolution, 34*(5), 400–415. https://doi.org/10.1016/j.tree.2019.01.012

Dalloul, R. A., Long, J. A., Zimin, A. V., Aslam, L., Beal, K., Ann Blomberg, L., ... & Reed, K. M. (2010). Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biology, 8*(9), e1000475. https://doi.org/10.1371/journal.pbio.1000475.

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

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., …& Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience, 10*(2), 1–4. https://doi.org/10.1093/gigascience/giab008

Darwin, C. (1859). *On the origins of species by means of natural selection*. London: Murray, 247. https://doi.org/10.1126/science.146.3640.51-b

De La Torre, A. R., Sekhwal, M. K., & Neale, D. B. (2021). Selective sweeps and polygenic adaptation drive local adaptation along moisture and temperature gradients in natural populations of coast redwood and giant sequoia. *Genes, 12*(11), 1826. https://doi.org/10.3390/genes12111826

De Luca, F., Barnes, K. M., Uyeda, J. A., De-Levi, S., Abad, V., Palese, T., Mericq, V., & Baron, J. (2001). Regulation of growth plate chondrogenesis by bone morphogenetic protein-2. *Endocrinology, 142*(1), 430–436. https://doi.org/10.1210/endo.142.1.7901

De Queiroz, A. (2005). The resurrection of oceanic dispersal in historical biogeography. *Trends in Ecology & Evolution, 20*(2), 68-73. https://doi.org/10.1016/j.tree.2004.11.006

De Queiroz, K. (1998). The general lineage concept of species , species criteria , and the process of speciation and terminological recommendations. *Endless Forms: Species and Speciation*, 57–75.

De Queiroz, K. (2005). Ernst Mayr and the modern concept of species. *Proceedings of the National Academy of Sciences, 102*(S1), 6600-6607. https://doi.org/10.1073/pnas.0502030102

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology, 56*(6), 879–886. https://doi.org/10.1080/10635150701701083

De Queiroz, K., & Donoghue, M. J. (1988). Phylogenetic systematics and the species problem. *Cladistics, 4*(4), 317–338. https://doi.org/10.1111/j.1096-0031.1988.tb00518.x

Degiorgio, M., Huber, C. D., Hubisz, M. J., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics, 32*(12), 1895–1897. https://doi.org/10.1093/bioinformatics/btw051

del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A., & De Juana, E. (2018). *Handbook of the Birds of the World Alive*. Barcelona: Lynx Edicions; 2017.

del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A., & Kirwan, G. (2020). *Handbook of the Birds of the World Alive*. Barcelona: Lynx Edicions; 2020.

Delaneau, O., Zagury, J. F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature Methods, 10*(1), 5–6. https://doi.org/10.1038/nmeth.2307

Delhey, K. (2017). Gloger's rule. *Current Biology, 27*(14), R689–R691. https://doi.org/10.1016/j.cub.2017.04.031

Delhey, K. (2019). A review of Gloger's rule, an ecogeographical rule of colour: definitions, interpretations and evidence. *Biological Reviews, 94*(4), 1294–1316. https://doi.org/10.1111/brv.12503

Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., Guigõ, R., & Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology, 24*(8), 1873–1888. https://doi.org/10.1111/mec.13150

Delmore, K. E., Lugo Ramos, J. S., Van Doren, B. M., Lundberg, M., Bensch, S., Irwin, D. E., & Liedvogel, M. (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters, 2*(2), 76-87. https://doi.org/10.1002/evl3.46

Delmore, K. E., Illera, J. C., Pérez-Tris, J., Segelbacher, G., Lugo Ramos, J. S., Durieux, G., Ishigohoka, J., & Liedvogel, M. (2020). The evolutionary history and genomics of European blackcap migration. *Elife, 9*, e54462. http://dx.doi.org/10.7554/eLife.54462

Deng, C., Wynshaw-Boris, A., Zhou, F., Kuo, A., & Leder, P. (1996). Fibroblast growth factor receptor 3 is a negative regulator of bone growth. *Cell, 84*(6), 911–921. https://doi.org/10.1016/S0092-8674(00)81069-7

Dennenmoser, S., Vamosi, S. M., Nolte, A. W., & Rogers, S. M. (2017). Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (*Cottus asper*) revealed by Pool-Seq. *Molecular Ecology, 26*(1), 25–42. https://doi.org/10.1111/mec.13805

Dennison, M. D., & Baker, A. J. (1991). Morphometric variability in continental and Atlantic island populations of chaffinches (*Fringilla coelebs*). *Evolution, 45*(1), 29-39. https://doi.org/10.2307/2409479

DeSilva, R., & Dodd, R. S. (2020). Association of genetic and climatic variability in giant sequoia, Sequoiadendron giganteum, reveals signatures of local adaptation along moisture-related gradients. *Ecology and Evolution, 10*(19), 10619-10632. https://doi.org/10.1002/ece3.6716

Domyan, E. T., Hardy, J., Wright, T., Frazer, C., Daniels, J., Kirkpatrick, J., Kirkpatrick, J., Wakamatsu, K., & Hill, J. T. (2019). SOX10 regulates multiple genes to direct eumelanin versus pheomelanin production in domestic rock pigeon. *Pigment Cell and Melanoma Research, 32*(5), 634–642. https://doi.org/10.1111/pcmr.12778

Dong, F., Li, S. H., Chiu, C. C., Dong, L., Yao, C. Te, & Yang, X. J. (2020). Strict allopatric speciation of sky island *Pyrrhula erythaca* species complex. *Molecular Phylogenetics and Evolution, 153*, 106941. https://doi.org/10.1016/j.ympev.2020.106941

Donoghue, M. J. (1985). A critique of the biological species concept and recommendations for a phylogenetic alternative. *Bryologist, 88*(3), 172–181.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dorsett, D., & Krantz, I. D. (2009). On the molecular etiology of Cornelia de Lange Syndrome. *Annals of the New York Academy of Sciences, 1151*(1), 22–37. https://doi.org/10.1111/j.1749-6632.2008.03450.x.On

Doxsey, S., McCollum, D., & Theurkauf, W. (2005). Centrosomes in cellular regulation. *Annual Review of Cell and Developmental Biology*, *21*, 411-434. https://doi.org/10.1146/annurev.cellbio.21.122303.120418

Doxsey, S., Zimmerman, W., & Mikule, K. (2005). Centrosome control of the cell cycle. *Trends in cell biology*, *15*(6), 303-311. https://doi.org/10.1016/j.tcb.2005.04.008

Drummond, A. J., & Bouckaert, R. R. (2015). *Bayesian evolutionary analysis with BEAST*. Cambridge University Press.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution, 29*(8), 1969-1973. https://doi.org/10.1093/molbev/mss075

Dubay, S. G., & Witt, C. C. (2014). Differential high-altitude adaptation and restricted gene flow across a mid-elevation hybrid zone in Andean tit-tyrant flycatchers. *Molecular Ecology, 23*(14), 3551–3565. https://doi.org/10.1111/mec.12836

Dubuc-Messier, G., Caro, S. P., Perrier, C., van Oers, K., Réale, D., & Charmantier, A. (2018). Gene flow does not prevent personality and morphological differentiation between two blue tit populations. *Journal of Evolutionary Biology, 31*(8), 1127-1137. https://doi.org/10.1111/jeb.13291

Ducrest, A. L., Neuenschwander, S., Schmid-Siegert, E., Pagni, M., Train, C., Dylus, D., … & Goudet, J. (2020). New genome assembly of the barn owl (*Tyto alba alba*). *Ecology and Evolution, 10*(5), 2284-2298. https://doi.org/10.1002/ece3.5991

Dutoit, L., Burri, R., Nater, A., Mugal, C. F., & Ellegren, H. (2017). Genomic distribution and estimation of nucleotide diversity in natural populations: perspectives from the collared flycatcher (*Ficedula albicollis*) genome. *Molecular Ecology Resources, 17*(4), 586–597. https://doi.org/10.1111/1755-0998.12602

Eberhard, J. R., & Bermingham, E. (2005). Phylogeny and comparative biogeography of Pionopsitta parrots and Pteroglossus toucans. *Molecular Phylogenetics and Evolution, 36*(2), 288-304. https://doi.org/10.1016/j.ympev.2005.01.022

Edelaar, P., & Bolnick, D. I. (2012). Non-random gene flow: An underappreciated force in evolution and ecology. *Trends in Ecology and Evolution, 27*(12), 659–665. https://doi.org/10.1016/j.tree.2012.07.009

Edelaar, P., Siepielski, A. M., & Clobert, J. (2008). Matching habitat choice causes directed gene flow: a neglected dimension in evolution and ecology. *Evolution, 62*(10), 2462–2472. https://doi.org/10.1111/j.1558-5646.2008.00459.x

Ekblom, R., Brechlin, B., Persson, J., Smeds, L., Johansson, M., Magnusson, J., … & Ellegren, H. (2018). Genome sequencing and conservation genomics in the Scandinavian wolverine population. *Conservation Biology, 32*(6), 1301–1312. https://doi.org/10.1111/cobi.13157

Eldredge, N., & Cracraft, J. (1980). *Phylogenetic patterns and the evolutionary process. Method and Theory in Comparative Biology*. New York Columbia Univ Press.

Ellegren, H. (2010). Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology and Evolution, 25*(5), 283–291. https://doi.org/10.1016/j.tree.2009.12.004

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution, 29*(1), 51-63. https://doi.org/10.1016/j.tree.2013.09.008

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature, 491*(7426), 756–760. https://doi.org/10.1038/nature11584

Elmer, K. R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A. F., Franchini, P., & Meyer, A. (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nature Communications, 5*(1), 1-8. https://doi.org/10.1038/ncomms6168

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE, 6*(5), 1–10. https://doi.org/10.1371/journal.pone.0019379

Emerson, B. C. (2002). Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. *Molecular Ecology, 11*(6), 951-966.

Emerson, B. C., Oromí, P., & Hewitt, G. M. (1999). MtDNA phylogeography and recent intra-island diversification among Canary Island Calathus beetles. *Molecular Phylogenetics and Evolution, 13*(1), 149-158. https://doi.org/10.1006/mpev.1999.0644

Endler, J. A. (1973). Gene flow and population differentiation: studies of clines suggest that differentiation along environmental gradients may be independent of gene flow. *Science, 179*(4070), 243–250.

Ericson, P. G. P., Qu, Y., Blom, M. P. K., Johansson, U. S., & Irestedt, M. (2017). A genomic perspective of the pink-headed duck *Rhodonessa caryophyllacea* suggests a long history of low effective population size. *Scientific Reports, 7*(1), 1–6. https://doi.org/10.1038/s41598-017-16975-1

Espındola-Hernandez, P., Mueller, J. C., Carrete, M., Boerno, S., & Kempenaers, B. (2020). Genomic evidence for sensorial adaptations to a nocturnal predatory lifestyle in owls. *Genome Biology and Evolution, 12*(10), 1895–1908. https://doi.org/10.1093/GBE/EVAA166

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*(3), 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution, 23*(7), 347–351. https://doi.org/10.1016/j.tree.2008.04.004

Fagny, M., & Austerlitz, F. (2021). Polygenic adaptation: integrating population genetics and gene regulatory networks. *Trends in Genetics, 37*(7), 631–638. https://doi.org/10.1016/j.tig.2021.03.005

Fang, Y. T., Yao, C. T., Hsu, Y. C., & Hung, C. M. (2022). Elevational plumage divergence in the Rufous-capped Babbler (*Cyanoderma ruficeps*) on a mountainous island. *Ibis, 164*(1), 151-167. https://doi.org/10.1111/ibi.13009

Faramarz, A., Balk, J. A., Oostra, A. B., Ghandour, C. A., Rooimans, M. A., Wolthuis, R. M. F., & de Lange, J. (2019). Non-redundant roles in sister chromatid cohesion of the DNA helicase DDX11 and the SMC3 acetyl transferases ESCO1/2. *BioRxiv*, 1–19. https://doi.org/10.1101/704635

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics, 28*(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Feder, J. L., & Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution, 64*(6), 1729–1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x

Felicísimo, Á. M., Muñoz, J., González-Solis, J. (2008). Ocean surface winds drive dynamics of transoceanic aerial movements. *PLoS One, 3*(8), e2928. https://doi.org/10.1371/journal.pone.0002928

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., …& Zhang, G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature, 587*(7833), 252–257. https://doi.org/10.1038/s41586-020-2873-9

Feng, W., Leach, S. M., Tipney, H., Phang, T., Geraci, M., Spritz, R. A., Hunter, L. E., & Williams, T. (2009). Spatial and temporal analysis of gene expression during growth and fusion of the mouse facial prominences. *PLoS ONE, 4*(12), e8066. https://doi.org/10.1371/journal.pone.0008066

Ferchaud, A. L., & Hansen, M. M. (2016). The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: Three-spine sticklebacks in divergent environments. *Molecular Ecology, 25*(1), 238–259. https://doi.org/10.1111/mec.13399

Fernández-Palacios, J. M. (2009). Laurisilvas macaronésicas (Laurus, Ocotea)(*). VV. AA., Bases ecológicas preliminares para la conservación de los tipos de hábitat de interés comunitario en Espana. Ministerio de Medio Ambiente, y Medio Rural y Marino.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics, 180*(2), 977-933. https://doi.org/10.1534/genetics.108.092221

Foote, A. D. (2018). Sympatric speciation in the genomic era. *Trends in Ecology & Evolution, 33*(2), 85–95. https://doi.org/10.1016/j.tree.2017.11.003

Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular ecology*, *25*(1), 104-120. https://doi.org/10.1111/mec.13476

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology, 27*(9), 2215–2233. https://doi.org/10.1111/mec.14584

Foster, J. B. (1964). Evolution of mammals on islands. *Nature, 202*(4929), 234-235.

Förstner, W., & Moonen, B. (2003). A Metric for Covariance Matrices. In: Grafarend, E.W., Krumm, F.W., Schwarze, V.S. (eds) Geodesy-The Challenge of the 3rd Millennium. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-05296-9_31

Frank, F. (1939). Die Färbung der Vogelfeder durch Pigment und Struktur. *Journal Für Ornithologie, 87*(3), 426–523. https://doi.org/10.1007/BF01916524

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: A review. *Genetics Research, 89*(5–6), 491–503. https://doi.org/10.1017/S0016672308009695

Frankham, R. (1997). Do island populations have less genetic variation than mainland populations? *Heredity, 78*(3), 311–327. https://doi.org/10.1038/hdy.1997.46

Frankham, R. (2005). Genetics and extinction. *Biological Conservation, 126*(2), 131–140. https://doi.org/10.1016/j.biocon.2005.05.002

Frankl-Vilches, C., Kuhl, H., Werber, M., Klages, S., Kerick, M., Bakker, A., ... & Gahr, M. (2015). Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. *Genome Biology, 16*(1), 1-25. https://doi.org/10.1186/s13059-014-0578-9.

Freed, L. A., Conant, S., & Fleischer, R. C. (1987). Evolutionary ecology and radiation of Hawaiian passerine birds. *Trends in Ecology & Evolution, 2*(7), 196-203. https://doi.org/10.1016/0169-5347(87)90020-6

Freeman, S., & Jackson, W. M. (1990). Univariate metrics are not adequate to measure avian body size. *The Auk, 107*(1), 69-74.

Friedman, N. R., & Remeš, V. (2017). Ecogeographical gradients in plumage coloration among Australasian songbird clades. *Global Ecology and Biogeography, 26*(3), 261–274. https://doi.org/10.1111/geb.12522

Friedrich, S. R., Lovell, P. V., Kaser, T. M., & Mello, C. V. (2019). Exploring the molecular basis of neuronal excitability in a vocal learner. *BMC Genomics, 20*(1), 1–26. https://doi.org/10.1186/s12864-019-5871-2

Friis, G., Aleixandre, P., Rodríguez-Estrella, R., Navarro-Sigüenza, A. G., & Milá, B. (2016). Rapid postglacial diversification and long-term stasis within the songbird genus *Junco*: phylogeographic and phylogenomic evidence.

Molecular Ecology, 25(24), 6175–6195. https://doi.org/10.1111/mec.13911

Friis, G., Fandos, G., Zellmer, A., McCormack, J., Faircloth, B. C., & Milá, B. (2018). Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. *Molecular Ecology, 27*(24), 5137–5153. https://doi.org/10.1111/mec.14946

Friis, G., & Milá, B. (2020). Change in sexual signaling traits outruns morphological divergence in a recent avian radiation across an ecological gradient. *Journal of Evolutionary Biology, 33*(9), 1276-1293. https://doi.org/10.1111/jeb.13671

Friis, G., Vizueta, J., Ketterson, E., & Milá, B. (2022). A high-quality genome assembly and annotation of the dark-eyed junco Junco hyemalis, a recently diversified songbird. G3: *Genes, Genomes, Genetics*, In press.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics, 147*(2), 915-925. https://doi.org/10.1093/genetics/147.2.915

Funk, V. A., & Wagner, W. L. (1995). *Biogeographic patterns in the Hawaiian Islands. Hawaiian biogeography: evolution on a hot spot archipelago* (ed. by W.L. Wagner and V.L. Funk), pp. 379– 419. Smithsonian Institution Press, Washington, DC.

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., … & Andelt, W. F. (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). *Molecular Ecology, 25*(10), 2176-2194. https://doi.org/10.1111/mec.13605

Gabrielli, M., Nabholz, B., Leroy, T., Milá, B., & Thébaud, C. (2020). Within-island diversification in a passerine bird. *Proceedings of the Royal Society B: Biological Sciences, 287*(1923), 20192999. https://doi.org/10.1098/rspb.2019.2999

Gamboa, M. P., Ghalambor, C. K., Scott Sillett, T., Morrison, S. A., & Chris Funk, W. (2022). Adaptive divergence in bill morphology and other thermoregulatory traits is facilitated by restricted gene flow in song sparrows on the

California Channel Islands. *Molecular Ecology, 31*(2), 603–619. https://doi.org/10.1111/mec.16253

Garant, D., Kruuk, L. E. B., Wilkin, T. A., McCleery, R. H., & Sheldon, B. C. (2005). Evolution driven by differential dispersal within a wild bird population. *Nature, 433*(7021), 60–65. https://doi.org/10.1038/nature03051

García-Verdugo, C., Caujapé-Castells, J., Illera, J.C., Mairal, M., Patiño, J., Reyes-Betancort, A., Scholz, S. (2019). Pleistocene extinctions as drivers of biogeographical patterns on the easternmost Canary Islands. *Journal of Biogeography, 46*(5), 845-859. https://doi.org/10.1111/jbi.13563

Gardner, P. P., Daub, J., Tate, J., Moore, B. L., Osuch, I. H., Griffiths-Jones, S., ... & Bateman, A. (2010). Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Research, 39*(suppl_1), D141-D145. https://doi.org/10.1093/nar/gkq1129.

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics, 201*(4), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources, 17*(1), 78–90. https://doi.org/10.1111/1755-0998.12634

Gazda, M. A. (2019). Genetic basis of simple and complex traits with relevance to avian evolution.

Geraldes, A., Farzaneh, N., Grassa, C. J., Mckown, A. D., Guy, R. D., Mansfield, S. D., … & Cronk, Q. C. B. (2014). Landscape genomics of populus trichocarpa: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution, 68*(11), 3260–3280. https://doi.org/10.1111/evo.12497

Gibbs, H. L., & Grant, P. R. (1987). Oscillating selection on Darwin's finches. *Nature, 327*(6122), 511–513. https://doi.org/10.1038/327511a0

Gill, F. B. (2014). Species taxonomy of birds: which null hypothesis? *The Auk: Ornithological Advances, 131*(4), 150-161. https://doi.org/10.1642/AUK-13-206.1

Gillespie, R. G., Baldwin, B. G., Waters, J. M., Fraser, C. I., Nikula, R., & Roderick, G. K. (2012). Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology and Evolution, 27*(1), 47–56. https://doi.org/10.1016/j.tree.2011.08.009

Gillespie, R. G., Bennett, G. M., De Meester, L., Feder, J. L., Fleischer, R. C., Harmon, L. J., ... & Wogan, G. O. (2020). Comparing adaptive radiations across space, time, and taxa. *Journal of Heredity, 111*(1), 1-20. https://doi.org/10.1093/jhered/esz064

Gillespie, R. G., Howarth, F. G., & Roderick, G. K. (2001). Adaptive radiation. *Encyclopedia of Biodiversity*, 1, 25–44.

Gittenberger, E. (1991). What about non-adaptive radiation? *Biological Journal of the Linnean Society, 43*(4), 263–272.

Gloger, C. W. L. (1833). *Das Abändern der Vögel durch Einflußdes Klima's: Nach zoologischen, zunächst von den europäischen Landvögeln entnommenen Beobachtungen dargestellt...* Schulz.

Glor, R. E., Gifford, M. E., Larson, A., Losos, J. B., Rodríguez Schettino, L., Chamizo Lara, A. R., & Jackman, T. R. (2004). Partial island submergence and speciation in an adaptive radiation: a multilocus analysis of the Cuban green anoles. *Proceedings of the Royal Society B: Biological Sciences, 271*(1554), 2257–2265. https://doi.org/10.1098/rspb.2004.2819

Glor, R. E., Kolbe, J. J., Powell, R., Larson, A., & Lossos, J. B. (2003). Phylogenetic analysis of ecological and morphological diversification in hispaniolan trunk-ground anoles (*Anolis cybotes* group). *Evolution, 57*(10), 2383–2397. https://doi.org/10.1111/j.0014-3820.2003.tb00250.x

Gomez, D., & Théry, M. (2004). Influence of ambient light on the evolution of colour signals: comparative analysis of a Neotropical rainforest bird community. *Ecology Letters, 7*(4), 279–284. https://doi.org/10.1111/j.1461-0248.2004.00584.x

Gompel, N., & Prud'homme, B. (2009). The causes of repeated genetic evolution. *Developmental Biology, 332*(1), 36–47. https://doi.org/10.1016/j.ydbio.2009.04.040

Gorukmez, O., Gorukmez, O., & Ekici, A. (2020). A novel nonsense FMN2 mutation in Nonsyndromic Autosomal Recessive Intellectual Disability Syndrome.

Fetal and Pediatric Pathology, 40(6), 702-706. https://doi.org/10.1080/15513815.2020.1737991

Gosselin, T. (2019). radiator: RADseq data exploration, manipulation and visualization using R. https://doi.org/10.5281/zenodo. 1475182.

Gould, S. J. (1990). *Wonderful life: the Burgess Shale and the nature of history*. WW Norton & Company.

Gould, S. J. (2002). *The structure of evolutionary theory*. Harvard University Press.

Grabherr, M. G., 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*(9), 1145– 1151. https://doi.org/10.1093/bioinformatics/btq102

Grant, P. R. (1976). Population variation on islands. *Proceedings of the International Ornithological Congress, 16*, 603–615.

Grant, P. R. (1979). Evolution of the chaffinch, *Fringilla coelebs*, on the Atlantic Islands. *Biological Journal of the Linnean Society, 11*(4), 301-332. https://doi.org/10.1111/j.1095-8312.1979.tb00042.x

Grant, P. R. (1980). Colonization of Atlantic islands by chaffinches (*Fringilla* sp.). *Bonner Zoologische Beiträge, 31*(3-4), 311–317.

Grant, P.R. (1986). *Ecology and evolution of Darwin's finches*. Princeton Univ. Press, Princeton, NJ.

Grant, P. R. (2001). Reconstructing the evolution of birds on islands: 100 years of research. *Oikos, 92*(3), 385–403. https://doi.org/10.1034/j.1600-0706.2001.920301.x

Grant, P. R, & Grant, B. R. (2002). Adaptive radiation of Darwin ' s finches. *American Scientist, 90*(2), 130–139. https://doi.org/10.1511/2002.2.130

Grant, P. R., Grant, B. R. (2008). How and Why Species Multiply: The Radiation of Darwin's Finches, Princeton University Press.

Grant, P. R., Grant, B. R., Markert, J. A., Keller, L. F., & Petren, K. (2004). Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution, 58*(7), 1588–1599. https://doi.org/10.1111/j.0014-3820.2004.tb01738.x

Greenberg, R., & Danner, R. M. (2012). The influence of the california marine layer on bill size in a generalist songbird. *Evolution, 66*(12), 3825–3835. https://doi.org/10.1111/j.1558-5646.2012.01726.x

Gregory, T. R. (2002). Animal genome size database. http://www. genomesize. com.

Gremme, G., Brendel, V., Sparks, M. E., & Kurtz, S. (2005). Engineering a software tool for gene structure prediction in higher organisms. *Information and Software Technology, 47*(15), 965-978. https://doi.org/10.1016/j.infsof.2005.09.005.

Gremme, G., Steinbiss, S., & Kurtz, S. (2013). GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 10*(3), 645-656. https://doi.org/10.1109/TCBB.2013.68.

Griffiths, R., Daan, S., & Dijkstra, C. (1996). Sex identification in birds using two CHD genes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 263*(1374), 1251-1256.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources, 18*(3), 691–699. https://doi.org/10.1111/1755-0998.12745

Guenther, C., Pantalena-Filho, L., & Kingsley, D. M. (2008). Shaping skeletal growth by modular regulatory elements in the Bmp5 gene. *PLoS Genetics, 4*(12), e1000308. https://doi.org/10.1371/journal.pgen.1000308

Gustafson, E. J., Miranda, B. R., De Bruijn, A. M. G., Sturtevant, B. R., & Kubiske, M. E. (2017). Do rising temperatures always increase forest productivity? Interacting effects of temperature, precipitation, cloudiness and soil texture on tree species growth and competition. *Environmental Modelling and Software, 97*, 171–183. https://doi.org/10.1016/j.envsoft.2017.08.001

Guy, L., Roat Kultima, J., & Andersson, S. G. (2010). genoPlotR: comparative gene and genome visualization in R. *Bioinformatics, 26*(18), 2334-2335. https://doi.org/10.1093/bioinformatics/btq413.

Hackett, S. J. (1996). Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). *Molecular Phylogenetics and Evolution, 5*(2), 368-382. https://doi.org/10.1006/mpev.1996.0032

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research, 27*(6), 1004-1015. https://doi.org/10.1101/gr.212522.116

Handbook of the Birds of the World and BirdLife International. (2019). Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 4. Available at: http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBW-BirdLife_Checklist_v4_Dec19.zip

Hanna, Z. R., Henderson, J. B., Wall, J. D., Emerling, C. A., Fuchs, J., Runckel, C., … & Dumbacher, J. P. (2017). Northern spotted owl (*Strix occidentalis caurina*) genome: Divergence with the barred owl (*Strix varia*) and characterization of light-associated genes. *Genome Biology and Evolution, 9*(10), 2522–2545. https://doi.org/10.1093/gbe/evx158

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., ... & Townshend, J. (2013). High-resolution global maps of 21st-century forest cover change. *Science, 342*(6160), 850-853. https://doi.org/10.1126/science.1244693

Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research, 16*(6), 730–737. https://doi.org/10.1101/gr.5045006

Harris, T. J. C., & Tepass, U. (2010). Adherens junctions: from molecules to morphogenesis. *Nature Reviews Molecular Cell Biology, 11*(7), 502–514. https://doi.org/10.1038/nrm2927

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 28*(1), 100-108. https://doi.org/10.2307/2346830

Harvey, P. H., & Pagel, M. D. (1991). *The Comparative method in evolutionary biology*. (Vol. 239). Oxford University Press.

Hedrick, P. W., Ginevan, M. E., & Ewing, E. P. (1976). Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics, 7*(1), 1–32. https://doi.org/10.1146/annurev.es.07.110176.000245

Hejase, H. A., Salman-Minkov, A., Campagna, L., Hubisz, M. J., Lovette, I. J., Gronau, I., & Siepel, A. (2020). Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. *Proceedings of the National Academy of Sciences of the United States of America, 117*(48), 30554–30565. https://doi.org/10.1073/pnas.2015987117

Henderson, E. C., & Brelsford, A. (2020). Genomic differentiation across the speciation continuum in three hummingbird species pairs. *BMC Evolutionary Biology, 20*(1), 1–11. https://doi.org/10.1186/s12862-020-01674-9

Hendry, A. P. (2004). Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evolutionary Ecology Research, 6*(8), 1219–1236.

Hendry, A. P. (2009). Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences, 66*(8), 1383–1398. https://doi.org/10.1139/F09-074

Hendry, A. P., Wenburg, J. K., Bentzen, P., Volk, E. C., & Quinn, T. P. (2000). Rapid evolution of reproductive isolation in the wild: evidence from introduced salmon. *Science, 290*(5491), 516–518. https://doi.org/10.1126/science.290.5491.516

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics, 169*(4), 2335–2352. https://doi.org/10.1534/genetics.104.036947

Hey, J. (2001). The mind of the species problem. *Trends in Ecology & Evolution, 16*(7), 326–328. https://doi.org/10.1016/0002-8703(78)90476-3

Hijmans, R. J. (2019). Introduction to the" geosphere" package (Version 1.5-10). Citeseer.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology, 25*(15), 1965–1978. https://doi.org/10.1002/joc.1276

Hill, G. E., Hill, G. E., McGraw, K. J., & Kevin, J. (Eds.). (2006). *Bird coloration, volume 1: mechanisms and measurements* (Vol. 1). Harvard University Press.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., ... & Emerson, J. J. (2005). Erratum: Sequence and comparative analysis of the chicken genome provide unique perspective on vertebrate evolution (*Nature* (2004) 432 (695-716)). *Nature, 433*(7027). https://doi.org/10.1038/nature0339409.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., … & Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist, 188*(4), 379–397. https://doi.org/10.1086/688018

Hobson, K. A. (1999). Stable-carbon and nitrogen isotope ratios of songbird feathers grown in two terrestrial biomes: implications for evaluating trophic relationships and breeding origins. *The Condor, 101*(4), 799–805. https://doi.org/10.2307/1370067

Hobson, K. A., Alisauskas, R. T., & Clark, R. G. (1993). Stable-nitrogen isotope enrichment in avian tissues due to fasting and nutritional stress: implications for isotopic analyses of diet. *The Condor, 95*(2), 388. https://doi.org/10.2307/1369361

Hobson, K. A., & Clark, R. G. (1992). Assessing avian diets using stable isotopes I: turnover of 13 C in tissues. *The Condor, 94*(1), 181–188. https://doi.org/10.2307/1368807

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science, 313*(5783), 101–104. . https://doi.org/10.1126/science.1126121

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics, 32*(5), 767-769. https://doi.org/10.1093/bioinformatics/btv661.

Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-Genome Annotation with BRAKER. In *Gene Prediction*, 65-95. Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9173-0_5.

Hoffmann, A. A., Sgrò, C. M., & Weeks, A. R. (2004). Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology and Evolution, 19*(9), 482–488. https://doi.org/10.1016/j.tree.2004.06.013

Hohenlohe, P. A., Bernatchez, L., Funk, W. C., & Catchen, J. M. (2017). Unbroken : RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources, 17*(3), 362-365. https://doi.org/10.1111/1755-0998.12669

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2020). Population genomics for wildlife conservation and management. *Molecular Ecology, 30*(1), 62-82. https://doi.org/10.1111/mec.15720

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70. http://www.jstor.org/stable/4615733.

Hooper, D. M., & Price, T.D. (2017). Chromosomal inversion differences correlate with range overlap in passerine birds. *Nature Ecology & Evolution, 1*(10), 1526-1534. https://doi.org/10.1038/s41559-017-0284-6.

Howell, T. & Cade, T. (1954). The birds of Guadalupe Island in 1953. *The Condor, 56*(5), 283–294.

Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology, 29*(14), 2535–2549. https://doi.org/10.1111/mec.15428

Humble, E., Dobrynin, P., Senn, H., Chuven, J., Scott, A. F., Mohr, D. W., … & Koepfli, K. P. (2020). Chromosomal-level genome assembly of the scimitar-horned oryx: Insights into diversity and demography of a species extinct in the wild. *Molecular Ecology Resources, 20*(6), 1668-1681. https://doi.org/10.1111/1755-0998.13181

Illera, J.C., Rando, J.C., Richardson, D.S., & Emerson, B.C. (2012). Age, origins and extinctions of the avifauna of Macaronesia: a synthesis of phylogenetic and fossil information. *Quaternary Science Reviews, 50*, 14-22. https://doi.org/10.1016/j.quascirev.2012.07.013

Illera, J. C., Rando, J. C., Rodriguez-Exposito, E., Hernández, M., Claramunt, S., & Martín, A. (2018). Acoustic, genetic, and morphological analyses of the

Canarian common chaffinch complex *Fringilla coelebs* ssp. reveals cryptic diversification. *Journal of Avian Biology, 49*(12). https://doi.org/10.1111/jav.01885

Illera, J. C., Spurgin, L. G., Rodriguez-Exposito, E., Nogales, M., & Rando, J. C. (2016). What are we learning about speciation and extinction from the Canary Islands? *Ardeola, 63*(1), 15-34. https://doi.org/10.13157/arla.63.1.2016.rp1

Irl, S. D. H., & Beierkuhnlein, C. (2011). Distribution of endemic plant species on an oceanic island - a geospatial analysis of La Palma (Canary Islands). *Procedia Environmental Sciences, 7*, 170–175. https://doi.org/10.1016/j.proenv.2011.07.030

Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., & Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology, 25*(18), 4488–4507. https://doi.org/10.1111/mec.13792

Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., … & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology, 27*(23), 4839-4855. https://doi.org/10.1111/mec.14858

Jackson, L., Kline, A. D., Barr, M. A., & Koch, S. De. (1993). de Lange syndrome: a clinical review of 310 individuals. *American Journal of Medical Genetics, 47*(7), 940–946. https://doi.org/10.1002/ajmg.1320470703

Jaiswal, S. K., Gupta, A., Shafer, A. B. A., Vishnu Prasoodanan, P. K., Vijay, N., & Sharma, V. K. (2021). Genomic insights into the molecular basis of sexual selection in birds. *Frontiers in Ecology and Evolution*, *9*, 538498. https://doi.org/10.3389/fevo.2021.538498

James, J. E., Lanfear, R., Eyre-Walker, A. (2016). Molecular evolutionary consequences of island colonization. *Genome Biology and Evolution, 8*(6), 1876-1888. https://doi.org/10.1093/gbe/evw120

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... & Avian Phylogenomics Consortium. (2015). Phylogenomic analyses data of the avian phylogenomics project. *GigaScience, 4*(1), s13742-014. https://doi.org/10.1186/s13742-014-0038-1.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... & Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science, 346*(6215), 1320-1331. https://doi.org/10.1126/science.1253451.

Jensen, J. D., Kim, Y., DuMont, V. B., Aquadro, C. F., & Bustamante, C. D. (2005). Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics, 170*(3), 1401–1410. https://doi.org/10.1534/genetics.104.038224

Jeon, D. J., Paik, S., Ji, S., & Yeo, J. S. (2021). Melanin-based structural coloration of birds and its biomimetic applications. *Applied Microscopy, 51*(1), 1-11. https://doi.org/10.1186/s42649-021-00063-w

Ji, W., Hou, li E., Yuan, X., Gu, T., Chen, Z. Y., Zhang, Y., … & Zhao, W. (2021). Identifying molecular pathways and candidate genes associated with knob traits by transcriptome analysis in the goose (*Anser cygnoides*). *Scientific Reports, 11*(1), 1–11. https://doi.org/10.1038/s41598-021-91269-1

Jolicoeur, P. (1963). 193. Note: the multivariate generalization of the allometry equation. *Biometrics*, *19*(3), 497-499. https://doi.org/10.2307/2527939

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... & Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55-61. https://doi.org/10.1038/nature10944

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics, 30*(9), 1236-1240. https://doi.org/10.1093/bioinformatics/btu031.

Juan, C., Emerson, B.vC., Oromí, P., & Hewitt, G. M. (2000). Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution, 15*(3), 104-109. https://doi.org/10.1016/S0169-5347(99)01776-0

Jühling, F., Mörl, M., Hartmann, R. K., Sprinzl, M., Stadler, P. F., & Pütz, J. (2009). tRNAdb 2009: compilation of tRNA sequences and tRNA genes. *Nucleic Acids Research, 37*(suppl_1), D159-D162. https://doi.org/10.1093/nar/gkn772.

Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., & Petrov, A. I. (2018). Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in Bioinformatics, 62*(1), e51. https://doi.org/10.1002/cpbi.51.

Kaplan, N. L., Hudson, R. R., & Langley, C. H. (1989). The" hitchhiking effect" revisited. *Genetics, 123*(4), 887-899. https://doi.org/10.1093/genetics/123.4.887

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., & Flouri, T. (2017). Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics, 33*(11), 1630-1638. https://doi.org/10.1093/bioinformatics/btx025

Kawakami, T., Mugal, C. F., 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*(16), 4158–4172. https://doi.org/10.1111/mec.14197

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters, 7*(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kearns, A. M., Campana, M. G., Slikas, B., Berry, L., Saitoh, T., Cibois, A., & Fleischer, R. C. (2022). Conservation genomics and systematics of a near-extinct island radiation. *Molecular Ecology*, *31*(7), 1995-2012. https://doi.org/10.1111/mec.16382

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., & Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics, 19*(1), 1-12. https://doi.org/10.1186/s12859-018-2203-5.

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research, 44*(9), e89-e89. https://doi.org/10.1093/nar/gkw092.

Kess, T., & Boulding, E. G. (2019). Genome-wide association analyses reveal polygenic genomic architecture underlying divergent shell morphology in Spanish *Littorina saxatilis* ecotypes. *Ecology and Evolution, 9*(17), 9427-9441. https://doi.org/10.1002/ece3.5378

Kirkpatrick, M., & Ravigné, V. (2002). Speciation by natural and sexual selection: models and experiments. *The American Naturalist, 159*(S3), S22-S35. https://doi.org/10.2307/3078919

Knief, U., Bossu, C. M., Saino, N., Hansson, B., Poelstra, J., Vijay, N., … & Wolf, J. B. W. (2019). Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nature Ecology and Evolution, 3*(4), 570–576. https://doi.org/10.1038/s41559-019-0847-9

Kolbe, J. J., Schoener, T. W., Spiller, D. a, & Losos, J. B. (2012). Founder effects persist despite adaptive differentiation: a field experiment with lizards. *Molecular Biology, 335*(6072), 1086-1089. https://doi.org/10.1126/science.1209566

Kondraskov, P., Schütz, N., Schüßler, C., de Sequeira, M. M., Guerra, A. S., Caujapé-Castells, J., ... & Thiv, M. (2015). Biogeography of mediterranean hotspot biodiversity: re-evaluating the'tertiary relict'hypothesis of macaronesian laurel forests. *PloS one*, *10*(7), e0132091. https://doi.org/10.1371/journal.pone.0132091

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources, 15*(5), 1179-1191. https://doi.org/10.1111/1755-0998.12387

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS : a review Self-fertilisation makes *Arabidopsis* particularly well suited to GWAS. *Plant Methods, 9*(1), 29. https://doi.org/10.1186/1746-4811-9-29

Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources, 21*(4), 1359–1368. https://doi.org/10.1111/1755-0998.13326

Kozak, K. M., McMillan, O., Joron, M., & Jiggins, C. D. (2018). Genome-wide admixture is common across the *Heliconius* radiation. *BioRxiv*, 414201. https://doi.org/10.1101/414201

Krueger, F. (2015). Trim Galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Bioinformatics.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research, 19*(9), 1639-1645. https://doi.org/10.1101/gr.092759.109.

Lachlan, R. F., Verzijden, M. N., Bernard, C. S., Jonker, P. P., Koese, B., Jaarsma, S., … & Ten Cate, C. (2013). The progressive loss of syntactical structure in bird song along an Island colonization chain. *Current Biology, 23*(19), 1896–1901. https://doi.org/10.1016/j.cub.2013.07.057

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research, 6*(1287), 1287. https://doi.org/10.12688/f1000research.12232.1.

Lambert, J. W., Reichard, M., & Pincheira-Donoso, D. (2019). Live fast, diversify non-adaptively: evolutionary diversification of exceptionally short-lived annual killifishes. *BMC Evolutionary Biology, 19*(1), 1–13. https://doi.org/10.1186/s12862-019-1344-0

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., … & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature, 518*(7539), 371–375. https://doi.org/10.1038/nature14181

Lamichhaney, S., Card, D. C., Grayson, P., Tonini, J. F. R., Bravo, G. A., Näpflin, K., … & Edwards, S. V. (2019). Integrating natural history collections and comparative genomics to study the genetic architecture of convergent evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 374*(1777), 20180248. https://doi.org/10.1098/rstb.2018.0248

Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M. S., Webster, M. T., ... & Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science, 352*(6284), 470-474. https://doi.org/10.1126/science.aad8786

Lande, R. (1982). Rapid origin of sexual isolation and character divergence in a cline. *Evolution*, *36*(2), 213–223. https://doi.org/10.2307/2408039

Lande, R., & Barrowclough, G. F. (1987). Effective population size, genetic variation, and their use in population management. In *Viable Populations for Conservation, 87*, 87-124. https://doi.org/10.1017/cbo9780511623400.007

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2016). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution, 34*(3), 772-773. https://doi.org/10.1093/molbev/msw260

Langin, K. M., Sillett, T. S., Funk, W. C., Morrison, S. A., Desrosiers, M. A., & Ghalambor, C. K. (2015). Islands within an island: repeated adaptive divergence in a single population. *Evolution, 69*(3), 653–665. https://doi.org/10.1111/evo.12610

Lapiedra, O., Sayol, F., Garcia-Porta, J., & Sol, D. (2021). Niche shifts after island colonization spurred adaptive diversification and speciation in a cosmopolitan bird clade. *Proceedings of the Royal Society B, 288*(1958), 20211022. https://doi.org/10.1098/rspb.2021.1022

Latch, E. K. (2020). Integrating genomics into conservation management. *Molecular Ecology Resources, 20*(6), 1455-1457. https://doi.org/10.1111/1755-0998.13188

Laurentino, T. G., Moser, D., Roesti, M., Ammann, M., Frey, A., Ronco, F., … & Berner, D. (2020). Genomic release-recapture experiment in the wild reveals within-generation polygenic selection in stickleback fish. *Nature Communications, 11*(1), 1–9. https://doi.org/10.1038/s41467-020-15657-3

Law, R., Dixon-Salazar, T., Jerber, J., Cai, N., Abbasi, A. A., Zaki, M. S., ... & Gleeson, J. G. (2014). Biallelic truncating mutations in FMN2, encoding the actin-regulatory protein Formin 2, cause nonsyndromic autosomal-recessive intellectual disability. *The American Journal of Human Genetics, 95*(6), 721-728. https://doi.org/10.1016/j.ajhg.2014.10.016

Lawson, L. P., & Petren ,K. (2017). The adaptive genomic landscape of beak morphology in Darwin's finches. *Molecular Ecology, 26*(19), 4978-4989. https://doi.org/10.1111/mec.14166.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genetics, 8*(1), e1002453. https://doi.org/10.1371/journal.pgen.1002453

Legendre, P., & Legendre, L. (1988). Numerical Ecology. *Developments in Environmental Modelling, 20*, 235-245. https://doi.org/10.1017/CBO9781107415324.004

Lehti, M. S., Henriksson, H., Rummukainen, P., Wang, F., Uusitalo-Kylmälä, L., Kiviranta, R., Heino, T. J., Kotaja, N., & Sironen, A. (2018). Cilia-related protein SPEF2 regulates osteoblast differentiation. *Scientific Reports, 8*(1), 1–11. https://doi.org/10.1038/s41598-018-19204-5

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution, 17*(4), 183–189. https://doi.org/10.1016/S0169-5347(02)02497-7

Lerner, H. R. L., Meyer, M., James, H. F., Hofreiter, M., & Fleischer, R. C. (2011). Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biology, 21*(21), 1838–1844. https://doi.org/10.1016/j.cub.2011.09.039

Leroy, T., Rousselle, M., Tilak, M. K., Caizergues, A. E., Scornavacca, C., Recuerda, M., … & Nabholz, B. (2021). Island songbirds as windows into evolution in small populations. *Current Biology, 31*(6), 1303-1310. https://doi.org/10.1016/j.cub.2020.12.040

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., … & Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences of the United States of America, 115*(17), 4325–4333. https://doi.org/10.1073/pnas.1720115115

Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics, 74*(1), 175–195. https://doi.org/10.1093/genetics/74.1.175

Li, A., Wang, J., Sun, K., Wang, S., Zhao, X., Wang, T., ... & Lu, Y. (2021). Two reference-quality sea snake genomes reveal their divergent evolution of adaptive traits and venom systems. *Molecular Biology and Evolution, 38*(11), 4867-4883.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature, 475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, H., & Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics, 211*(1), 289–304. https://doi.org/10.1534/genetics.118.301747

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, M., Zhu, F., & Hong, Y. (2013). Differential evolution of duplicated medakafish mitf genes. *International Journal of Biological Sciences, 9*(5), 496–508. https://doi.org/10.7150/ijbs.4668

Li, Y., Ren, Y., Zhang, D., Jiang, H., Wang, Z., Li, X., & Rao, D. (2019). Chromosome-level assembly of the mustache toad genome using third-generation DNA sequencing and Hi-C analysis. *GigaScience, 8*(9), giz114. https://doi.org/10.1093/gigascience/giz114

Lieberman-Aiden, E., Van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., ... & Dekker, J. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science, 326*(5950), 289-293. https://doi.org/10.1126/science.1181369.

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

Liu, W., Sun, X., Braut, A., Mishina, Y., Behringer, R. R., Mina, M., & Martin, J. F. (2005). Distinct functions for Bmp signaling in lip and palate fusion in mice. *Development, 132*(6), 1453–1461. https://doi.org/10.1242/dev.01676

Liu, X., Li, Y. I., & Pritchard, J. K. (2019). Trans effects on gene expression can drive omnigenic inheritance. *Cell, 177*(4), 1022-1034.e6. https://doi.org/10.1016/j.cell.2019.04.014

Lizarraga, G., Lichtler, A., Upholt, W. B., & Kosher, R. A. (2002). Studies on the role of Cux1 in regulation of the onset of joint formation in the developing limb. *Developmental Biology, 243*(1), 44–54. https://doi.org/10.1006/dbio.2001.0559

Lomolino, M. V. (2005). Body size evolution in insular vertebrates: generality of the island rule. *Journal of Biogeography, 32*(10), 1683–1699. https://doi.org/10.1111/j.1365-2699.2005.01314.x

Lomolino, M. V., Sax, D. F., Palombo, M. R., & van der Geer, A. A. (2012). Of mice and mammoths: Evaluations of causal explanations for body size evolution in insular mammals. *Journal of Biogeography, 39*(5), 842–854. https://doi.org/10.1111/j.1365-2699.2011.02656.x

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution, 65*(7), 1827–1840. https://doi.org/10.1111/j.1558-5646.2011.01289.x

Losos, J. B., Arnold, S. J., Bejerano, G., Brodie Iii, E. D., Hibbett, D., Hoekstra, H. E., … & Turner, T. L. (2013). Evolutionary Biology for the 21st Century. *PLoS Biology, 11*(1), e1001466. https://doi.org/10.1371/journal.pbio.1001466

Losos, J. B., Jackman, T. R., Larson, A., De Queiroz, K., & Rodríguez-Schettino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science, 279*(5359), 2115–2118. https://doi.org/10.1126/science.279.5359.2115

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature, 457*(7231), 830–836. https://doi.org/10.1038/nature07893

Losos, J. B, & Schluter, D. (2000). Analysis of an evolutionary species–area relationship. *Nature, 408*(6814, 847-850. https://doi.org/10.1038/35048558.

Louha, S., Ray, D. A., Winker, K., & Glenn, T. C. (2020). A high-quality genome assembly of the North American Song Sparrow, *Melospiza melodia*. *G3: Genes, Genomes, Genetics, 10*(4), 1159-1166.https://doi.org/10.1534/g3.119.400929.

Loureiro, L. O., Engstrom, M. D., & Lim, B. K. (2020). Single nucleotide polymorphisms (SNPs) provide unprecedented resolution of species boundaries, phylogenetic relationships, and genetic diversity in the mastiff bats (*Molossus*). *Molecular Phylogenetics and Evolution, 143*, 106690. https://doi.org/10.1016/j.ympev.2019.106690

Lovell, P. V., Clayton, D. F., Replogle, K. L., & Mello, C. V. (2008). Birdsong "transcriptomics": neurochemical specializations of the oscine song system. *PLoS ONE, 3*(10), e3440. https://doi.org/10.1371/journal.pone.0003440

Lowe, T. M., & Chan, P. P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research, 44*(W1), W54-W57. https://doi.org/10.1093/nar/gkw413.

Lundregan, S. L., Hagen, I. J., Gohli, J., Niskanen, A. K., Kemppainen, P., Ringsby, T. H., … & Jensen, H. (2018). Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Molecular Ecology, 27*(17), 3498–3514. https://doi.org/10.1111/mec.14811

Lynch, A., & Baker, A. J. (1991). Increased vocal discrimination by learning in sympatry in two species of chaffinches. *Behaviour,116*(1-2),109-126. https://www.jstor.org/stable/4534911

Lynch, A., & Baker, A. J. (1994). A population memetics approach to cultural evolution in chaffinch song: differentiation among populations. *Evolution, 48*(2), 351-359. https://doi.org/10.1111/j.1558-5646.1994.tb01316.x

MacArthur, R. H., & Wilson, E. O. (1967). The theory of island biogeography. Princeton: Princeton Univ. Press. 203 p.

Maddison, W., & Knowles, L. (2006). Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology, 55*(1), 21–30. https://doi.org/10.1080/10635150500354928

Magalhaes, I. S., Whiting, J. R., D'Agostino, D., Hohenlohe, P. A., Mahmud, M., Bell, M. A., … & MacColl, A. D. C. (2020). Intercontinental genomic parallelism in multiple three-spined stickleback adaptive radiations. *Nature Ecology & Evolution, 5*(2), 251-261.https://doi.org/10.1038/s41559-020-01341-8

Mahler, D. L., Ingram, T., Revell, L. J., & Losos, J. B. (2013). Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science, 341*(6143), 292-295. https://doi.org/10.1126/science.1232392

Maia, R., Gruson, H., Endler, J. A., & White, T. E. (2019). pavo 2: new tools for the spectral and spatial analysis of colour in R. *Methods in Ecology and Evolution*, *10*(7), 1097-1107. https://doi.org/10.1111/2041-210X.13174

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018a). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature Ecology and Evolution, 2*(12), 1940–1955. https://doi.org/10.1038/s41559-018-0717-x

Malinsky, M., Trucchi, E., Lawson, D.J., & Falush, D. (2018b). RADpainter and fineRADstructure: population inference from RADseq data. *Molecular Biology and Evolution, 35*(5), 1284-1290. https://doi.org/10.1093/molbev/msy023.

Mallarino, R., Grant, P. R., Grant, B. R., Herrel, A., Kuo, W. P., & Abzhanov, A. (2011). Two developmental modules establish 3D beak-shape variation in Darwin's finches. *Proceedings of the National Academy of Sciences of the United States of America, 108*(10), 4057–4062. https://doi.org/10.1073/pnas.1011480108

Mallet, J. (2008). Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1506), 2971–2986. https://doi.org/10.1098/rstb.2008.0081

Manceau, M., Domingues, V. S., Linnen, C. R., Rosenblum, E. B., & Hoekstra, H. E. (2010). Convergence in pigmentation at multiple levels: Mutations, genes and function. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1552), 2439–2450. https://doi.org/10.1098/rstb.2010.0104

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. *PLoS Computational Biology, 14*(1), e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

Marra, P. P., Hobson, K. A., & Holmes, R. T. (1998). Linking winter and summer events in a migratory bird by using stable- carbon isotopes. *Science, 282*(5395), 1884–1886. https://doi.org/10.1126/science.282.5395.1884

Marshall, H. D., & Baker, A. J. (1999). Colonization history of Atlantic island common chaffinches (Fringilla coelebs) revealed by mitochondrial DNA. *Molecular Phylogenetics and Evolution, 11*(2), 201-212. https://doi.org/10.1006/mpev.1998.0552

Martín, A., & Lorenzo, J. A. (2001). *Aves del archipiélago canario*. Francisco Lemus. Editor, La Laguna, Spain.

Martin, A., & Orgogozo, V. (2013). The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution, 67*(5), 1235–1250. https://doi.org/10.1111/evo.12081

Martín, J. L. (2009). Are the IUCN standard home-range thresholds for species a good indicator to prioritise conservation urgency in small islands? A case study in the Canary Islands (Spain). *Journal for Nature Conservation, 17*(2), 87–98. https://doi.org/10.1016/j.jnc.2008.10.001

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … & Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research, 23*(11), 1817–1828. https://doi.org/10.1101/gr.159426.113

Mathieson, I. (2021). The omnigenic model and polygenic prediction of complex traits. *American Journal of Human Genetics, 108*(9), 1558–1563. https://doi.org/10.1016/j.ajhg.2021.07.003

Matzke, N. J. (2013). Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography, 5*(4). https://doi.org/10.21425/F5FBG19694

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology, 63*(6), 951-970. https://doi.org/10.1093/sysbio/syu056

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

Mayr, E. (1963). *Animal species and evolution.* Harvard University Press. Cambridge, MA.

McCormack, J. E., & Smith, T. B. (2008). Niche expansion leads to small-scale adaptive divergence along an elevation gradient in a medium-sized passerine bird. *Proceedings of the Royal Society B: Biological Sciences, 275*(1647), 2155–2164. https://doi.org/10.1098/rspb.2008.0470

McCoy, M. W., Bolker, B. M., Osenberg, C. W., Miner, B. G., & Vonesh, J. R. (2006). Size correction: comparing morphological traits among populations and environments. *Oecologia*, *148*(4), 547-554. https://doi.org/10.1007/s00442-006-0403-6

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … & Depristo, M. A. (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research, 20*(9), 1297-1303. https://doi.org/10.1101/gr.107524.110.20

McLaughlin, J. F., Faircloth, B. C., Glenn, T. C., & Winker, K. (2020). Divergence, gene flow, and speciation in eight lineages of trans-Beringian birds. *Molecular Ecology, 29*(18), 3526-3542. https://doi.org/10.1111/mec.15574

McNaught, M. K., & Owens, I. P. F. (2002). Interspecific variation in plumage colour among birds: species recognition or light environment? *Journal of Evolutionary Biology, 15*(4), 505–514. https://doi.org/10.1046/j.1420-9101.2002.00431.x

McVean, G., & Auton, A. (2007). LDhat 2.1: a package for the population genetic analysis of recombination. Department of Statistics, Oxford, OX1 3TG, UK. www.stats.ox.ac.uk/~mcvean/LDhat.html

Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of parallel ecological speciation in Lake Victoria cichlids. *Molecular Biology and Evolution , 35*(6), 1489-1506. https://doi.org/10.1093/molbev/msy051

Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology and Evolution, 28*(11), 659-669. https://doi.org/10.1016/j.tree.2013.08.003.Population

Michener, R., & Lajtha, K. (Eds.). (2008). *Stable isotopes in ecology and environmental science*. John Wiley & Sons. https://doi.org/10.1002/9780470691854

Mikles, C. S., Aguillon, S. M., Chan, Y. L., Arcese, P., Benham, P. M., Lovette, I. J., & Walsh, J. (2020). Genomic differentiation and local adaptation on a microgeographic scale in a resident songbird. *Molecular Ecology, 29*(22), 4295-4307. https://doi.org/10.1111/mec.15647

Milá, B, Aleixandre, P., Alvarez-Nordström, S., McCormack, J., Ketterson, E., & Atwell, J. (2016). More than meets the eye: Lineage diversity and evolutionary history of Dark-eyed and Yellow-eyed juncos. *Snowbird: integrative biology and evolutionary diversity in the Junco*. University of Chicago Press, Chicago, Illinois, USA, 179, 198.

Milá, B., Wayne, R. K., Fitze, P., & Smith, T. B. (2009). Divergence with gene flow and fine-scale phylogeographical structure in the wedge-billed woodcreeper, *Glyphorynchus spirurus*, a neotropical rainforest bird. *Molecular Ecology, 18*(14), 2979–2995. https://doi.org/10.1111/j.1365-294X.2009.04251.x

Milá, B., Wayne, R. K., & Smith, T. B. (2008). Ecomorphology of migratory and sedentary populations of the Yellow-rumped warbler (*Dendroica coronata*). *The Condor, 110*(2), 335–344. https://doi.org/10.1525/cond.2008.8396

Miller, M., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. New Orleans, LA, USA: Gateway Computing Environments Workshop (GCE). http://dx.doi.org/10.1109/GCE.2010.5676129

Minvielle, F., Bed'hom, B., Coville, J. L., Ito, S., Inoue-Murayama, M., & Gourichon, D. (2010). The "silver" Japanese quail and the MITF gene: Causal mutation, associated traits and homology with the "blue" chicken plumage. *BMC Genetics, 11*(1), 1–7. https://doi.org/10.1186/1471-2156-11-15

Mishler, B. D., & Brandon, R. N. (1987). Individuality, pluralism, and the phylogenetic species concept. *Biology and Philosophy, 2*(4), 397–414. https://doi.org/10.1007/BF00127698

Moest, M., Van Belleghem, S. M., James, J. E., Salazar, C., Martin, S. H., Barker, S. L., … & Jiggins, C. D. (2020). Selective sweeps on novel and introgressed variation

shape mimicry loci in a butterfly adaptive radiation. *PLoS Biology, 18*(2), e3000597. https://doi.org/10.1371/journal.pbio.3000597

Møller, A. P. (2006). Sociality, age at first reproduction and senescence: comparative analyses of birds. *Journal of Evolutionary Biology, 19*(3), 682–689. https://doi.org/10.1111/j.1420-9101.2005.01065.x

Moloney, G. M., van Oeffelen, W. E. P. A., Ryan, F. J., van de Wouw, M., Cowan, C., Claesson, M. J., Schellekens, H., Dinan, T. G., & Cryan, J. F. (2019). Differential gene expression in the mesocorticolimbic system of innately high- and low-impulsive rats. *Behavioural Brain Research, 364*, 193–204. https://doi.org/10.1016/j.bbr.2019.01.022

Monroe, J., Srikant, T., Carbonell-Bejerano, P., Becker, C., Lensink, M., Exposito-Alonso, M., ... & Weigel, D. (2022). Mutation bias reflects natural selection in *Arabidopsis thaliana*. *Nature 602*, 101-105. https://doi.org/10.1038/s41586-021-04269-6

Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin, R. K. (2019). Genomic architecture of parallel ecological divergence: beyond a single environmental contrast. *Science Advances, 5*(12), eaav9963. https://doi.org/10.1126/sciadv.aav9963

Morinha, F., Bastos, R., Carvalho, D., Travassos, P., Santos, M., Blanco, G., ... & Cabral, J. A. (2017). A spatially-explicit dynamic modelling framework to assess habitat suitability for endangered species: The case of Red-billed Chough under land use change scenarios in Portugal. *Biological Conservation*, *210*, 96-106. https://doi.org/10.1016/j.biocon.2017.04.013

Morinha, F., Milá, B., Dávila, J. A., Fargallo, J. A., Potti, J., & Blanco, G. (2020). The ghost of connections past: A role for mainland vicariance in the isolation of an insular population of the red-billed chough (Aves: Corvidae). *Journal of Biogeography, 47*(12), 2567-2583. https://doi.org/10.1111/jbi.13977

Morris, S. C. (2010). Evolution: Like any other science it is predictable. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1537), 133–145. https://doi.org/10.1098/rstb.2009.0154

Murray, K. D., & Borevitz, J. O. (2018). Axe: rapid, competitive sequence read demultiplexing using a trie. *Bioinformatics, 34*(22), 3924–3925. https://doi.org/10.1093/bioinformatics/bty432

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1587), 409–421. https://doi.org/10.1098/rstb.2011.0249

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*(5), 1058–1072. https://doi.org/10.1111/mec.13540

Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., Ffrench-Constant, R. H., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences. 367*(1587), 343–353. https://doi.org/10.1098/rstb.2011.0198

Nathan, R. (2006). Long-distance dispersal of plants. *Science, 313*(5788), 786– 788. https://doi.org/10.1126/science.1124975

Nawrocki, E. P. (2014). *Annotating functional RNAs in genomes using Infernal. In RNA Sequence, Structure, and Function: Computational and Bioinformatic Methods* (pp. 163-197). Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-62703-709-9_9.

Near, T. J., MacGuigan, D. J., Parker, E., Struthers, C. D., Jones, C. D., & Dornburg, A. (2018). Phylogenetic analysis of Antarctic notothenioids illuminates the utility of RADseq for resolving Cenozoic adaptive radiations. *Molecular Phylogenetics and Evolution, 129*, 268-279. https://doi.org/10.1016/j.ympev.2018.09.001

Nelson, G., & Platnick, N. (1981). *Systematics and biogeography*. Harcourt, Brace and World.

Newman, D., & Pilson, D. (1997). Increased probability of extinction due to decreased genetic effective population size: experimental populations of *Clarkia pulchella*. *Evolution, 51*(2), 354–362. https://doi.org/10.1111/j.1558-5646.1997.tb02422.x

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics, 39*, 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research, 15*(11), 1566–1575. https://doi.org/10.1101/gr.4252305

Nieminen, M., Singer, M. C., Fortelius, W., Schöps, K., & Hanski, I. (2001). Experimental confirmation that inbreeding depression increases extinction risk in butterfly populations. *The American Naturalist, 157*(2), 237–244. https://doi.org/10.1086/318630

Nixon, K. C., & Wheeler, Q. D. (1990). An amplification of the phylogenetic species concept. *Cladistics, 6*(3), 211–223. https://doi.org/10.1111/j.1096-0031.1990.tb00541.x

Nomura, T., & Izawa, E. I. (2017). Avian brains: insights from development, behaviors and evolution. *Development Growth and Differentiation, 59*(4), 244–257. https://doi.org/10.1111/dgd.12362

Nosil, P. (2012). *Ecological speciation*. In Oxford University Press. https://doi.org/10.18960/seitai.66.3_561

Nosil, P., Egan, S. P., & Funk, D. J. (2008). Heterogeneous genomic differentiation between walking-stick ecotypes: "Isolation by adaptation" and multiple roles for divergent selection. *Evolution, 62*(2), 316–336. https://doi.org/10.1111/j.1558-5646.2007.00299.x

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1587), 332–342. https://doi.org/10.1098/rstb.2011.0263

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology, 18*(3), 375-402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology and Evolution, 24*(3), 145–156. https://doi.org/10.1016/j.tree.2008.10.011

Nosil, P., Vines, T. H., & Funk, D. J. (2005). Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, *59*(4), 705-719. https://doi.org/10.1111/j.0014-3820.2005.tb01747.x

Ödeen, A., & Björklund, M. (2003). Dynamics in the evolution of sexual traits: losses and gains, radiation and convergence in yellow wagtails (*Motacilla flava*). *Molecular Ecology, 12*(8), 2113-2130. https://doi.org/10.1046/j.1365-294X.2003.01883.x

Oikkonen, J., Onkamo, P., Järvelä, I., & Kanduri, C. (2016). Convergent evidence for the molecular basis of musical traits. *Scientific Reports, 6*(1), 1–10. https://doi.org/10.1038/srep39707

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., … & Wagner, H. (2019). Package "vegan": Community ecology package. http://vegan.r-forge.r-project.org/. https://doi.org/ISBN 0-387-95457-0

Okuyama, Y., Goto, N., Nagano, A. J., Yasugi, M., Kokubugata, G., Kudoh, H., … & Sugawara, T. (2020). Radiation history of Asian Asarum (sect. Heterotropa, Aristolochiaceae) resolved using a phylogenomic approach based on double-digested RAD-seq data. *Annals of Botany, 126*(2), 245–260. https://doi.org/10.1093/aob/mcaa072

Osogwa, N. C.,Mori, C, Sánchez-Valpuesta, M., Hayase, S. & Wada, K. (2018). Inter- and intra-species differences in muscarinic acetylcholine receptor expression in the neural pathways for vocal learning in songbirds. *The Journal of Comparative Neurology, 526*(17) 2856-2869. https://doi.org/10.1002/cne.24532.

Padial, J.M., Miralles, A., De la Riva, I., Vences, M. (2010). The integrative future of taxonomy. *Frontiers in Zoology, 7*(1), 1-14. https://doi.org/10.1186/1742-9994-7-16

Parchman, T. L., Buerkle, C. A., Soria-Carrasco, V., & Benkman, C. W. (2016). Genome divergence and diversification within a geographic mosaic of coevolution. *Molecular Ecology, 25*(22), 5705–5718. https://doi.org/10.1111/mec.13825

Patiño, J., Whittaker, R. J., Borges, P. A., Fernández-Palacios, J. M., Ah-Peng, C., Araújo, M. B., ... & Emerson, B. C. (2017). A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island

Biogeography. *Journal of Biogeography*, *44*(5), 963-983.https://doi.org/10.1111/jbi.12986

Pearman, W. S., Wells, S. J., Silander, O. K., Freed, N. E., & Dale, J. (2020). Concordant geographic and genetic structure revealed by genotyping-by-sequencing in a New Zealand marine isopod. *Ecology and Evolution, 10*(24), 13624–13639. https://doi.org/10.1002/ece3.6802

Pennings, P. S., & Hermisson, J. (2006). Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics, 2*(12), 1998–2012. https://doi.org/10.1371/journal.pgen.0020186

Peñalba, J. V., Deng, Y., Fang, Q., Joseph, L., Moritz, C., & Cockburn, A. (2020). Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (*Malurus cyaneus*). *Molecular Ecology Resources, 20*(2), 560-578. https://doi.org/10.1111/1755-0998.13124.

Peñaloza, C., Gutierrez, A. P., Eöry, L., Wang, S., Guo, X., Archibald, A. L., … & Houston, R. D. (2021). A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *GigaScience, 10*(3), giab020.

Perktaş, U., Peterson, A. T., & Dyer, D. (2017). Integrating morphology, phylogeography, and ecological niche modeling to explore population differentiation in North African Common Chaffinches. *Journal of Ornithology, 158*(1), 1-13. https://doi.org/10.1007/s10336-016-1361-3

Peterson, B. J., & Fry, B. (1987). Stable isotopes in ecosystem studies. *Annual Review of Ecology and Systematic, 18*(1), 293–320. https://doi.org/10.1146/annurev.es.18.110187.001453

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. *Physiology & Behavior, 176*(1), 100–106. https://doi.org/10.1038/ng.3464.Visualizing

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., … & Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science, 346*(6215), 1256846. https://doi.org/10.1126/science.1256846

Philipp, U., Lupp, B., Mömke, S., Stein, V., Tipold, A., Eule, J. C., … & Distl, O. (2011). A MITF mutation associated with a dominant white phenotype and bilateral

deafness in German Fleckvieh cattle. *PLoS ONE, 6*(12), 4–9. https://doi.org/10.1371/journal.pone.0028857

Poelstra, J. W., Richards, E. J., & Martin, C. H. (2018). Speciation in sympatry with ongoing secondary gene flow and a potential olfactory trigger in a radiation of Cameroon cichlids. *Molecular Ecology, 27*(21), 4270–4288. https://doi.org/10.1111/mec.14784

Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Müller, I., ... & Wolf, J. B. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science, 344*(6190), 1410-1414. https://doi.org/10.1126/science.1253226

Poelstra, J. W., Vijay, N., Hoeppner, M. P., & Wolf, J. B. W. (2015). Transcriptomics of colour patterning and coloration shifts in crows. *Molecular Ecology, 24*(18), 4617–4628. https://doi.org/10.1111/mec.13353

Porlier, M., Garant, D., Perret, P., & Charmantier, A. (2012). Habitat-linked population genetic differentiation in the blue tit *cyanistes caeruleus*. *Journal of Heredity, 103*(6), 781–791. https://doi.org/10.1093/jhered/ess064

Power, D. (1983). Variability in island populations of the house finch. *The Auk, 100*(1), 180–187. https://doi.org/10.1093/auk/100.1.180

Prasad, A., Lorenzen, E. D., & Westbury, M. V. (2021). Evaluating the role of reference-genome phylogenetic distance on evolutionary inference. *Molecular Ecology Resources, 22*(1), 45-55. https://doi.org/10.1111/1755-0998.13457

Prentice, M. B., Bowman, J., Khidas, K., Koen, E. L., Row, J. R., Murray, D. L., & Wilson, P. J. (2017). Selection and drift influence genetic differentiation of insular Canada lynx (*Lynx canadensis*) on Newfoundland and Cape Breton Island. *Ecology and Evolution, 7*(9), 3281–3294. https://doi.org/10.1002/ece3.2945

Price, T. (2008). *Speciation in birds*. Roberts and Company. greenwood Village, CO.

Price, T. D., Grant, P. R., Gibbs, H. L., & Boag, P. T. (1984). Recurrent patterns of natural selection in a population of Darwin's finches. *Nature, 309*(5971), 787–789. https://doi.org/10.1038/309787a0

Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation - Not by sweeps alone. *Nature Reviews Genetics, 11*(10), 665–667. https://doi.org/10.1038/nrg2880

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic sdaptation. *Current Biology, 23*(1), 1–7. https://doi.org/10.1016/j.cub.2009.11.055

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945-959. https://doi.org/10.1093/genetics/155.2.945

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics, 81*(3), 559–575. https://doi.org/10.1086/519795

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., ... & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Research, 26*(3), 342-350. https://doi.org/10.1101/gr.193474.115.

Qi, L. M., Mohr, M., & Wade, J. (2012). Enhanced expression of tubulin-specific chaperone protein a, mitochondrial ribosomal protein S27, and the DNA excision repair protein XPACCH in the song system of juvenile male zebra finches. *Developmental Neurobiology, 72*(2), 199–207. https://doi.org/10.1002/dneu.20956

Quattrini, A. M., Wu, T., Soong, K., Jeng, M. S., Benayahu, Y., & McFadden, C. S. (2019). A next generation approach to species delimitation reveals the role of hybridization in a cryptic species complex of corals. *BioRxiv*, 1–19. https://doi.org/10.1101/523936

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

R Core Team. (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Ralph, C. L. (1969). The control of color in birds. *American Zoologist, 9*(2), 521–530. https://doi.org/10.1093/icb/9.2.531

Rambaut, A. (2017). FigTree-version 1.4. 3, a graphical viewer of phylogenetic trees. Computer program distributed by the author, website: http://tree. bio. ed. ac. uk/software/figtree.

Rambaut, A., Suchard, M.A., Xie, D., & Drummond, A.J. (2014). Tracer v1. 6. Computer program and documentation distributed by the author, website http://beast.bio.ed.ac.uk/Tracer

Rancilhac, L., Goudarzi, F., Gehara, M., Hemami, M. R., Elmer, K. R., Vences, M., & Steinfarz, S. (2019). Phylogeny and species delimitation of near Eastern Neurergus newts (Salamandridae) based on genome-wide RADseq data analysis*. Molecular Phylogenetics and Evolution, 133*, 189-197. https://doi.org/10.1016/j.ympev.2019.01.003

Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., Iny Stein, T., Levitt, J., Gershoni, M., Morrey, C. P., Safran, M., & Lancet, D. (2017). MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Research, 45*(D1), D877-D887. https://doi.org/10.1093/nar/gkw1012

Ravi Kumar, D., Joel Devadasan, M., Surya, T., Vineeth, M. R., Choudhary, A., Sivalingam, J., … & Verma, A. (2020). Genomic diversity and selection sweeps identified in Indian swamp buffaloes reveals it's uniqueness with riverine buffaloes. *Genomics, 112*(3), 2385–2392. https://doi.org/10.1016/j.ygeno.2020.01.010

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., … & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology, 30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Recuerda, M., Illera, J. C., Blanco, G., Zardoya, R., & Milá, B. (2021a). Sequential colonization of oceanic archipelagos led to a species-level radiation in the common chaffinch complex (Aves : *Fringilla coelebs*). *Molecular*

Phylogenetics and Evolution, 164, 107291. https://doi.org/10.1016/j.ympev.2021.107291

Recuerda, M., Vizueta, J., Cuevas-Caballé, C., Blanco, G., Rozas, J., & Milá, B. (2021b). Chromosome-level genome assembly of the common chaffinch (Aves: *Fringilla coelebs*): a valuable resource for evolutionary biology. *Genome Biology and Evolution*, *13*(4), evab034. https://doi.org/10.1101/2020.11.30.404061

Ree, R. H., & Sanmartín, I. (2018). Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography, 45*, 741-749. https://doi.org/10.1111/jbi.13173

Reid, J. M., Bignal, E. M., Bignal, S., McCracken, D. I., & Monaghan, P. (2003). Age-specific reproductive performance in red-billed choughs *Pyrrhocorax pyrrhocorax*: Patterns and processes in a natural population. *Journal of Animal Ecology, 72*(5), 765–776. https://doi.org/10.1046/j.1365-2656.2003.00750.x

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology, 24*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Ren, S., Lyu, G., Irwin, D. M., Liu, X., Feng, C., Luo, R., Zhang, J., Sun, Y., Shang, S., Zhang, S., & Wang, Z. (2021). Pooled sequencing analysis of geese (*Anser cygnoides*) reveals genomic variations associated with feather color. *Frontiers in Genetics, 12*, 650013. https://doi.org/10.3389/fgene.2021.650013

Ribot, R. F. H., Berg, M. L., Schubert, E., Endler, J. A., & Bennett, A. T. D. (2019). Plumage coloration follows Gloger's rule in a ring species. *Journal of Biogeography, 46*(3), 584–596. https://doi.org/10.1111/jbi.13497

Rice, W. R., & Hostert, E. E. (1993). Laboratory experiments on speciation: what have we learned in 40 years? *Evolution, 47*(6), 1637–1653.

Richardson, J. L., Urban, M. C., Bolnick, D. I., & Skelly, D. K. (2014). Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology and Evolution, 29*(3), 165–176. https://doi.org/10.1016/j.tree.2014.01.002

Ricklefs, R. E., & Bermingham, E. (2007). The causes of evolutionary radiations in archipelagoes: passerine birds in the Lesser Antilles. *The American Naturalist, 169*(3), 285–297. https://doi.org/10.1086/510730

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., … & Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology and Evolution, 1*(4), 1–13. https://doi.org/10.1038/s41559-017-0082

Rising, J., & Somers, K. (1989). The measurement of overall body size in birds. *The Auk, 106*(4), 666–674. https://doi.org/10.1093/auk/106.4.666

Robert, A., Lengagne, T., Melo, M., Gomez, D., & Doutrelant, C. (2021). Evolution of vocal performance and song complexity in island birds. *Journal of Avian Biology, 2022*(1). https://doi.org/10.1111/jav.02726

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution, 66*(1), 1–17. https://doi.org/10.1111/j.1558-5646.2011.01486.x

Rodrigues, P., Lopes, R. J., Reis, S., Resendes, R., Ramos, J. A., & Tristão da Cunha, R. (2014). Genetic diversity and morphological variation of the common chaffinch *Fringilla coelebs* in the Azores. *Journal of Avian Biology, 45*(2), 167–178. https://doi.org/10.1111/j.1600-048X.2013.00229.x

Rosenblum, E. B., Parent, C. E., & Brandt, E. E. (2014). The molecular basis of phenotypic convergence. *Annual Review of Ecology, Evolution, and Systematics, 45*, 203–226. https://doi.org/10.1146/annurev-ecolsys-120213-091851

Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology, 14*(12). https://doi.org/10.1371/journal.pbio.2000234

Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology and Evolution, 24*(7), 394–399. https://doi.org/10.1016/j.tree.2009.02.007

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters, 8*(3), 336-352. https://doi.org/10.1111/j.1461-0248.2004.00715.x

Rundle, H. D., & Schluter, D. (2004). Natural selection and ecological speciation in sticklebacks. *Adaptive Speciation, 19*(3), 192-209. https://doi.org/10.1017/CBO9781139342179.011

Russello, M., Amato, G., DeSalle, R., & Knapp, M. (2020). Conservation genetics and genomics. *Genes, 11*(3), 11–12. https://doi.org/10.3390/genes11030318

Ryan, P. G., Bloomer, P., Moloney, C. L., Grant, T. J., & Delport, W. (2007). Ecological speciation in South Atlantic island finches. *Science, 315*(5817), 1420–1422. https://doi.org/10.1126/science.1138829

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., … & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature, 419*(6909), 832–837. https://doi.org/10.1038/nature01140

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., ... & Lander, E. S. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature, 449*(7164), 913-918. https://doi.org/10.1038/nature06250

Sackton, T. B., & Clark, N. (2019). Convergent evolution in the genomics era: new insights and directions. *Philosophical Transactions of the Royal Society B: Biological Sciences, 374*(1777), 24–27. https://doi.org/10.1098/rstb.2019.0102

Sackton, T. B., Grayson, P., Cloutier, A., Hu, Z., Liu, J. S., Wheeler, N. E., Gardner, P. P., Clarke, J. A., Baker, A. J., Clamp, M., & Edwards, S.V. (2019). Convergent regulatory evolution and loss of flight in paleognathous birds. *Science, 364*(6435), 74-78. https://doi.org/10.1126/science.aat7244

Sacoto, M. J. G., Tchasovnikarova, I. A., Torti, E., Forster, C., Andrew, E. H., Anselm, I., ... & Juusola, J. (2020). De novo variants in the ATPase module of MORC2 cause a neurodevelopmental disorder with growth retardation and variable craniofacial dysmorphism. *The American Journal of Human Genetics, 107*(2), 352-363. https://doi.org/10.1016/j.ajhg.2020.06.013

Salmón, P., Jacobs, A., Ahrén, D., Biard, C., Dingemanse, N. J., Dominoni, D. M., … & Isaksson, C. (2021). Continent-wide genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications, 12*(1), 1-14. https://doi.org/10.1038/s41467-021-23027-w

Salzburger, W. (2018). Understanding explosive diversification through cichlid fish genomics. *Nature Reviews Genetics, 19*(11), 705-717. https://doi.org/10.1038/s41576-018-0043-9

Salzburger, W., Ewing, G. B., & Von Haeseler, A. (2011). The performance of pylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology, 20*(9), 1952-1963. https://doi.org/10.1111/j.1365-294X.2011.05066.x

Samarasin-Dissanayake, P. (2010). Population differentiation, historical demography and evolutionary relationships among widespread common chaffinch populations (*Fringilla coelebs* ssp.) (Doctoral dissertation).

Samuk, K., Owens, G. L., Delmore, K. E., Miller, S. E., Rennison, D. J., & Schluter, D. (2017). Gene flow and selection interact to promote adaptive divergence in regions of low recombination. *Molecular Ecology, 26*(17), 4378–4390. https://doi.org/10.1111/mec.14226

Sangster, G. (2013). The application of species criteria in avian taxonomy and its implications for the debate over species concepts. *Biological Reviews, 89*(1), 199-214. https://doi.org/10.1111/brv.12051

Sangster, G., Rodríguez-Godoy, F., Roselaar, C. S., Robb, M. S., & Luksenburg, J. A. (2016). Integrative taxonomy reveals Europe's rarest songbird species, the Gran Canaria blue chaffinch *Fringilla polatzeki*. *Journal of Avian Biology, 47*(2), 159–166. https://doi.org/10.1111/jav.00825

Sanmartín, I., Van Der Mark, P., Ronquist, F. (2008). Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. *Journal of Biogeography, 35*(3), 428-449. https://doi.org/10.1111/j.1365-2699.2008.01885.x

Santure, A. W., & Garant, D. (2018). Wild GWAS—association mapping in natural populations. *Molecular Ecology Resources, 18*(4), 729–738. https://doi.org/10.1111/1755-0998.12901

Saranathan, V., & Finet, C. (2021). Cellular and developmental basis of avian structural coloration. *Current Opinion in Genetics and Development, 69*, 56–64. https://doi.org/10.1016/j.gde.2021.02.004

Sato, Y. U., Ogden, R., Kishida, T., Nakajima, N., Maeda, T., & Inoue-Murayama, M. (2020). Population history of the golden eagle inferred from whole-genome

sequencing of three of its subspecies. *Biological Journal of the Linnean Society, 130*(4), 826-838. https://doi.org/10.1093/biolinnean/blaa068/5865777

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics, 14*(11), 807-820. https://doi.org/10.1038/nrg3522

Sayol, F., Downing, P. A., Iwaniuk, A. N., Maspons, J., & Sol, D. (2018). Predictable evolution towards larger brains in birds colonizing oceanic islands. *Nature Communications, 9*(1), 1-7. https://doi.org/10.1038/s41467-018-05280-8

Schielzeth, H., & Hysby, A. (2014). Challenges and prospects in genome-wide QTL mapping of standing genetic variation in natural populations. *Annals of the New York Academy of Sciences, 1320*(1), 35-57. https://doi.org/10.1111/nyas.12397

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press. Oxford (UK).

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution, 16*(7), 372-380. https://doi.org/10.2307/3881825

Schluter, D., Clifford, E. A., Nemethy, M., & McKinnon, J. S. (2004). Parallel evolution and inheritance of quantitative traits. *The American Naturalist, 163*(6), 809–822. https://doi.org/10.1086/383621

Schmutz, S. M., Berryere, T. G., & Dreger, D. L. (2009). MITF and white spotting in dogs: A population study. *Journal of Heredity, 100*(S1), S66–S74. https://doi.org/10.1093/jhered/esp029

Schwarzer, J., Shabani, N., Esmaeili, H. R., Mwaiko, S., & Seehausen, O. (2017). Allopatric speciation in the desert: diversification of cichlids at their geographical and ecological range limit in Iran. *Hydrobiologia, 791*(1), 193–207. https://doi.org/10.1007/s10750-016-2976-3

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... & Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics, 15*(3), 176-192. https://doi.org/10.1038/nrg3644

Senar, J. C., Borras, A., Cabrera, J., Cabrera, T., & Björklund, M. (2006). Local differentiation in the presence of gene flow in the citril finch *Serinus*

citrinella. *Biology Letters, 2*(1), 85–87. https://doi.org/10.1098/rsbl.2005.0412

Senar, J. C., & Pascual, J. (1997). Keel and tarsus length may provide a good predictor of avian body size. *Ardea, 85*(2), 269–274.

Sendell-Price, A. T., Ruegg, K. C., Robertson, B. C., & Clegg, S. M. (2021). An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology, 30*(11), 2495–2510. https://doi.org/10.1111/mec.15898

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: assessing genome assembly and annotation completeness. *Gene Prediction* (pp. 227-245). Humana, New York, NY. https://doi.org/10.1007/978-1-4939-9173-0_14.

Sequeira, A. S., Lanteri, A. A., Albelo, L. R., Bhattacharya, S., & Sijapati, M. (2008). Colonization history, ecological shifts and diversification in the evolution of endemic Galápagos weevils. *Molecular Ecology, 17*(4), 1089-1107. https://doi.org/10.1111/j.1365-294X.2007.03642.x

Servedio, M. R. (2016). Geography, assortative mating, and the effects of sexual selection on speciation with gene flow. *Evolutionary Applications, 9*(1), 91–102. https://doi.org/10.1111/eva.12296

Shafer, A. B. A., & Wolf, J. B. W. (2013). Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters, 16*(7), 940–950. https://doi.org/10.1111/ele.12120

Shao, C., Li, C., Wang, N., Qin, Y., Xu, W., Liu, Q., ... & Chen, S. (2018). Chromosome-level genome assembly of the spotted sea bass, *Lateolabrax maculatus*. *GigaScience, 7*(11), giy114. https://doi.org/10.1093/gigascience/giy114

Shaw, K. L., & Mullen, S. P. (2014). Speciation continuum. *Journal of Heredity, 105*(S1), 741–742. https://doi.org/10.1093/jhered/esu060

Shi, Z., Zhang, Z., Schaffer, L., Huang, Z., Fu, L., Head, S., Gaasterland, T., Wang, X., & Li, X. (2021). Dynamic transcriptome landscape in the song nucleus HVC between juvenile and adult zebra finches. *Advanced Genetics, 2*(1), 1–13. https://doi.org/10.1002/ggn2.10035

Shirihai, H., & Svensson, L. (2018). *Handbook of Western Palearctic Birds, Volume 1: Passerines: Larks to Warblers*. Bloomsbury Publishing.

Shultz, A. J., & Burns, K. J. (2013). Plumage evolution in relation to light environment in a novel clade of Neotropical tanagers. *Molecular Phylogenetics and Evolution, 66*(1), 112–125. https://doi.org/10.1016/j.ympev.2012.09.011

Simpson, G. G. (1951). The species concept. *Evolution, 5*(4), 285–298.

Simpson, G. G. (1953). *The major features of evolution*. Columbia University Press.

Simpson, G. G. (1961). *Principles of animal taxonomy*. Columbia University Press.

Singhal, S., Leffler, E. M., Sannareddy, K., Turner, I., Venn, O., Hooper, D. M., ... & Przeworski, M. (2015). Stable recombination hotspots in birds. *Science, 350*(6263), 928-932. https://doi.org/10.1126/science.aad0843

Skinner, B. M., & Griffin, D. K. (2012). Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity, 108*(1), 37-41. https://doi.org/10.1038/hdy.2011.99.

Smeds, L., Qvarnström, A., & Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research, 26*(9), 1211–1218. https://doi.org/10.1101/gr.204669.116

Smit, A., & Hubley, R. (2019). RepeatModeler-1.0. 11. Institute for Systems Biology. http://www. repeatmasker. org.

Smit, A. F. A., Hubley, R., & Green, P. (2013). 2015. RepeatMasker Open-4.0, 1040.

Sorenson, M. D., Ast, J. C., Dimcheff, D. E., Yuri, T., & Mindell, D. P. (1999). Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. *Molecular Phylogenetics and Evolution, 12*(2), 105-114. https://doi.org/10.1006/mpev.1998.0602

Stace, C. A. (1991). *Plant taxonomy and biosystematics*. Cambridge University Press.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. https://doi.org/10.1093/bioinformatics/btu033

Stamatakis, A., Hoover, P., & Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology, 57*(5), 758-771. https://doi.org/10.1080/10635150802429642

Stamps, J. A., Krishnan, V. V., & Willits, N. H. (2009). How different types of natal experience affect habitat preference. *The American Naturalist, 174*(5), 623–630. https://doi.org/10.1086/644526

Stange, M., Sánchez-Villagra, M. R., Salzburger, W., & Matschiner, M. (2018). Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. *Systematic Biology, 67*(4), 681-699. https://doi.org/10.1093/sysbio/syy006

Stanke, M., Schöffmann, O., Morgenstern, B., & Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC bioinformatics, 7*(1), 1-11. https://doi.org/10.1186/1471-2105-7-62.

Stankowski, S., & Ravinet, M. (2021). Defining the speciation continuum. *Evolution, 75*(6), 1256–1273. https://doi.org/10.1111/evo.14215

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

Steiner, C. C., Römpler, H., Boettger, L. M., Schöneberg, T., & Hoekstra, H. E. (2009). The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Molecular Biology and Evolution, 26*(1), 35–45. https://doi.org/10.1093/molbev/msn218

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., ... & Lancet, D. (2016). The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics, 54*(1), 1-30. https://doi.org/10.1002/cpbi.5

Stensrud, E. (2013). Allopatric speciation and multi-trait variation in the common chaffinch (*Fringilla coelebs*) complex (Master's thesis).

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics, 14*(11), 751–764. https://doi.org/10.1038/nrg3483

Stern, D. L., & Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution, 62*(9), 2155–2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x

Stervander, M., Illera, J. C., Kvist, L., Barbosa, P., Keehnen, N. P., Pruisscher, P., Bensch, S., & Hansson, B. (2015). Disentangling the complex evolutionary history of the Western Palearctic blue tits (*Cyanistes* spp.)–phylogenomic analyses suggest radiation by multiple colonization events and subsequent isolation. *Molecular Ecology, 24*(10), 2477-2494. https://doi.org/10.1111/mec.13145

Stinchcombe, J. R., & Hoekstra, H. E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity, 100*(2), 158–170. https://doi.org/10.1038/sj.hdy.6800937

Stoddard, M. C., & Prum, R. O. (2008). Evolution of avian plumage color in a tetrahedral color space: a phylogenetic analysis of new world buntings. *The American Naturalist*, *171*(6), 755-776.

Storz, J. F. (2016). Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics, 17*(4), 239–250. https://doi.org/10.1038/nrg.2016.11

Suarez, N. M., Betancor, E., Klassert, T. E., Almeida, T., Hernandez, M., & Pestano, J. J. (2009). Phylogeography and genetic structure of the Canarian common chaffinch (*Fringilla coelebs*) inferred with mtDNA and microsatellite loci. *Molecular Phylogenetics and Evolution, 53*(2), 556–564. https://doi.org/10.1016/j.ympev.2009.07.018

Suarez, N. M., Betancor, E., & Pestano, J. J. (2009). Microsatellite loci isolation in the Canarian common chaffinch (*Fringilla coelebs*) and their utility in other Canarian finches. *Molecular Ecology Resources, 9*(4), 1162–1164. https://doi.org/10.1111/j.1755-0998.2009.02595.x

Sultana, H., Seo, D., Choi, N. R., Bhuiyan, M. S. A., Lee, S. H., Heo, K. N., & Lee, J. H. (2018). Identification of polymorphisms in MITF and DCT genes and their associations with plumage colors in Asian duck breeds. *Asian-Australasian Journal of Animal Sciences, 31*(2), 180–188. https://doi.org/10.5713/ajas.17.0298

Supple, M. A., Papa, R., Hines, H. M., Mcmillan, W. O., & Counterman, B. A. (2015). Divergence with gene flow across a speciation continuum of *Heliconius* butterflies. *BMC Evolutionary Biology, 15*(1), 1–12. https://doi.org/10.1186/s12862-015-0486-y

Suzuki, S., Marazita, M. L., Cooper, M. E., Miwa, N., Hing, A., Jugessur, A., … & Murray, J. C. (2009). Mutations in BMP4 Are Associated with subepithelial, microform, and overt cleft lip. *American Journal of Human Genetics, 84*(3), 406–411. https://doi.org/10.1016/j.ajhg.2009.02.002

Svensson, L. (2015). A new North African subspecies of Common Chaffinch *Fringilla coelebs*. *Bulletin of the British Ornithologists' Club, 135*(1), 69-76.

Szarmach, S. J., Brelsford, A., Witt, C. C., & Toews, D. P. L. (2021). Comparing divergence landscapes from reduced-representation and whole genome resequencing in the yellow-rumped warbler (*Setophaga coronata*) species complex. *Molecular Ecology, 30*(23), 5994–6005. https://doi.org/10.1111/mec.15940

Szukala, A., Lovegrove-Walsh, J., Luqman, H., Fior, S., Wolfe, T. M., Frajman, B., … & Paun, O. (2022). Polygenic routes lead to parallel altitudinal adaptation in *Heliosperma pusillum* (Caryophyllaceae). *Molecular Ecology, 0*, 1-16. https://doi.org/10.1111/mec.16393

Szulkin, M., Gagnaire, P. A., Bierne, N., & Charmantier, A. (2016). Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. *Molecular Ecology, 25*(2), 542–558. https://doi.org/10.1111/mec.13486

Tachibana, M., Takeda, K., Nobukuni, Y., Urabe, K., Long, J. E., Meyers, K. A., ... & Miki, T. (1996). Ectopic expression of MITF, a gene for Waardenburg syndrome type 2, converts fibroblasts to cells with melanocyte characteristics. *Nature Genetics, 14*(1), 50-54. https://doi.org/10.1038/ng0996-50

Taff, C. C., Campagna, L., & Vitousek, M. N. (2019). Genome-wide variation in DNA methylation is associated with stress resilience and plumage brightness in a wild bird. *Molecular Ecology, 28*(16), 3722–3737. https://doi.org/10.1111/mec.15186

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585–595. https://doi.org/PMC1203831

Tattersall, G. J., Arnaout, B., & Symonds, M. R. E. (2017). The evolution of the avian bill as a thermoregulatory organ. *Biological Reviews, 92*(3), 1630–1656. https://doi.org/10.1111/brv.12299

Tattersall, G. J., Chaves, J. A., & Danner, R. M. (2018). Thermoregulatory windows in Darwin's finches. *Functional Ecology, 32*(2), 358–368. https://doi.org/10.1111/1365-2435.12990

Terai, Y., Morikawa, N., & Okada, N. (2002). The evolution of the pro-domain of bone morphogenetic protein 4 (Bmp4) in an explosively speciated lineage of East African cichlid fishes. *Molecular Biology and Evolution, 19*(9), 1628–1632. https://doi.org/10.1093/oxfordjournals.molbev.a004225

Thompson, B. E., Freking, F., Pho, V., Schlinger, B. A., & Cherry, J. A. (2000). Cyclic AMP phosphodiesterases in the zebra finch: distribution, cloning and characterization of a PDE4B homolog. *Molecular Brain Research, 83*(1–2), 94–106. https://doi.org/10.1016/S0169-328X(00)00201-1

Thornton, I. (2007). *Island colonization: the origin and development of island communities*. Cambridge University Press.

Tietze, D. T. (2018). *Bird species: how they arise, modify and vanish* (p. 266). Springer Nature.

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology and Evolution, 29*(12), 673–680. https://doi.org/10.1016/j.tree.2014.10.004

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology, 25*(10), 2144-2164. https://doi.org/10.1111/mec.13606

Tobias, J. A., Donald, P. F., Martin, R. W., Butchart, S. H., & Collar, N. J. (2021). Performance of a points-based scoring system for assessing species limits in birds. *The Auk, 138*, ukab016. https://doi.org/10.1093/ornithology/ukab016

Tobias, J. A., Seddon, N., Spottiswoode, C. N., Pilgrim, J. D., Fishpool, L. D., & Collar, N. J. (2010). Quantitative criteria for species delimitation. *Ibis, 152*(4), 724-746. https://doi.org/10.1111/j.1474-919X.2010.01051.x

Toomey, M. B., Lopes, R. J., Araújo, P. M., Johnson, J. D., Gazda, M. A., Afonso, S., … & Carneiro, M. (2017). High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proceedings of the National Academy of*

Sciences of the United States of America, 114(20), 5219–5224. https://doi.org/10.1073/pnas.1700751114

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software, 3*(25), 731. https://doi.org/10.21105/joss.00731

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology, 3*(9), 1572–1578. https://doi.org/10.1371/journal.pbio.0030285

Tuttle, E. M., Bergland, A. O., Korody, M. L., Brewer, M. S., Newhouse, D. J., Minx, P., ... & Balakrishnan, C. N. (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology*, *26*(3), 344-350. https://doi.org/10.1016/j.cub.2015.11.069

UniProt Consortium. (2015). UniProt: a hub for protein information. *Nucleic Acids Research, 43*(D1), D204-D212. https://doi.org/10.1093/nar/gku989.

Valdar, W., Solberg, L. C., Gauguier, D., Cookson, W. O., Rawlins, J. N. P., Mott, R., & Flint, J. (2006). Genetic and environmental effects on complex traits in mice. *Genetics, 174*(2), 959–984. https://doi.org/10.1534/genetics.106.060004

Valente, L., Illera, J. C., Havenstein, K., Pallien, T., Etienne, R.S., & Tiedemann, R. (2017). Equilibrium bird species diversity in Atlantic islands. *Current Biology, 27*(11), 1660-1666. https://doi.org/10.1016/j.cub.2017.04.053

Valente, L., Phillimore, A. B., Melo, M., Warren, B. H., Clegg, S. M., Havenstein, K., ... & Etienne, R. S. (2020). A simple dynamic model explains the diversity of island birds worldwide. *Nature, 579*(7797), 92-96. https://doi.org/10.1038/s41586-020-2022-5

Valero, K. C. W., Pathak, R., Prajapati, I., Bankston, S., Thompson, A., Usher, J., & Isokpehi, R. D. (2014). A candidatemultimodal functional genetic network for thermal adaptation. *PeerJ 2*:e578. https://doi.org/10.7717/peerj.578

Van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology, 26*(15), 3982-3997. https://doi.org/10.1111/mec.14083

VanderWerf, E. A., Young, L. C., Yeung, N. W., & Carlon, D. B. (2010). Stepping stone speciation in Hawaii's flycatchers: Molecular divergence supports new

island endemics within the elepaio. *Conservation Genetics, 11*(4), 1283-1298. https://doi.org/10.1007/s10592-009-9958-1

Vernes, S. C., Spiteri, E., Nicod, J., Groszer, M., Taylor, J. M., Davies, K. E., … & Fisher, S. E. (2007). High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. *American Journal of Human Genetics, 81*(6), 1232–1250. https://doi.org/10.1086/522238

Verta, J. P., & Jacobs, A. (2021). The role of alternative splicing in adaptation and evolution. *Trends in Ecology and Evolution, 37*(4), 299-308. https://doi.org/10.1016/j.tree.2021.11.010

Vijay, N., Bossu, C. M., Poelstra, J. W., Weissensteiner, M. H., Suh, A., Kryukov, A. P., & Wolf, J. B. W. (2016). Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nature Communications, 7*(1), 1–10. https://doi.org/10.1038/ncomms13195

Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H., & Wolf, J. B. W. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular Ecology, 26*(16), 4284–4295. https://doi.org/10.1111/mec.14195

Vizueta, J., Macías-Hernández, N., Arnedo, M. A., Rozas, J., & Sánchez-Gracia, A. (2019). Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation. *Molecular Ecology, 28*(17), 4028–4045. https://doi.org/10.1111/mec.15199

Völker, M., Backström, N., Skinner, B. M., Langley, E. J., Bunzey, S. K., Ellegren, H., & Griffin, D. K. (2010). Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome research, 20*(4), 503-511. https://doi.org/10.1101/gr.103663.109.

Von Schirnding, Y., Van der Merwe, N. J., & Vogel, J. C. (1982). Influence of diet and age on carbon isotope ratios in ostrich eggshell. *Archaeometry, 24*(1), 3-20. https://doi.org/10.1111/j.1475-4754.1982.tb00643.x

vonHoldt, B. M., Kartzinel, R. Y., Huber, C. D., Le Underwood, V., Zhen, Y., Ruegg, K., … & Smith, T. B. (2018). Growth factor gene IGF1 is associated with bill size in the black-bellied seedcracker *Pyrenestes ostrinus*. *Nature Communications, 9*(1), 1-12. https://doi.org/10.1038/s41467-018-07374-9

Vujovic, D., Cornblath, D. R., & Scherer, S. S. (2021). A recurrent MORC2 mutation causes Charcot-Marie-Tooth disease type 2Z. *Journal of the Peripheral Nervous System, 26*(2), 184–186. https://doi.org/10.1111/jns.12443

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., … & Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology, 22*(3), 787–798. https://doi.org/10.1111/mec.12023

Wang, I. J. (2013). Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution, 67*(12), 3403–3411. https://doi.org/10.1111/evo.12134

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology, 23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

Wang, I. J., & Summers, K. (2010). Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology, 19*(3), 447–458. https://doi.org/10.1111/j.1365-294X.2009.04465.x

Wang, R., Chen, C. C., Hara, E., Rivas, M. V., Roulhac, P. L., Howard, J. T., … & Jarvis, E. D. (2015). Convergent differential regulation of SLIT-ROBO axon guidance genes in the brains of vocal learners. *Journal of Comparative Neurology, 523*(6), 892–906. https://doi.org/10.1002/cne.23719

Wang, W., Wang, F., Hao, R., Wang, A., Sharshov, K., Druzyaka, A., ... & Feng, S. (2020). First de novo whole genome sequencing and assembly of the bar-headed goose. *PeerJ, 8*, e8914. https://doi.org/10.7717/peerj.8914.

Wang, X., He, Z., Shi, S., & Wu, C. I. (2020). Genes and speciation: Is it time to abandon the biological species concept? *National Science Review, 7*(8), 1387–1397. https://doi.org/10.1093/nsr/nwz220

Wang, X., & Wang, L. (2016). GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in Plant Science, 7*, 1350. https://doi.org/10.3389/fpls.2016.01350.

Wang, Y., Li, S. M., Huang, J., Chen, S. Y., & Liu, Y. P. (2014). Mutations of TYR and MITF genes are associated with plumage colour phenotypes in geese. *Asian-*

Australasian Journal of Animal Sciences, 27(6), 778–783. https://doi.org/10.5713/ajas.2013.13350

Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., ... & Thébaud, C. (2015). Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters, 18*(2), 200-217. https://doi.org/10.1111/ele.12398

Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., ... & Heger, A. (2010). The genome of a songbird. *Nature, 464*(7289), 757. https://doi.org/10.1038/nature08819.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution, 38*(6), 1358–1370. https://doi.org/10.2307/2408641

Westfall, A. K., Telemeco, R. S., Grizante, M. B., Waits, D. S., Clark, A. D., Simpson, D. Y., ... & Schwartz, T. S. (2021). A chromosome-level genome assembly for the Eastern Fence Lizard (*Sceloporus undulatus*), a reptile model for physiological and evolutionary ecology. *GigaScience, 10*(10), giab066. https://doi.org/10.1093/gigascience/giab066

Wheeler, Q. D. (1999). Why the phylogenetic species concept? - Elementary. *Journal of Nematology, 31*(2), 134–141.

Whittaker, R. J., Araújo, M. B., Jepson, P., Ladle, R. J., Watson, J. E., & Willis, K. J. (2005). Conservation biogeography: assessment and prospect. *Diversity and Distributions, 11*(1), 3-23. https://doi.org/10.1111/j.1366-9516.2005.00143.x

Whittaker, R. J., & Fernández-Palacios, J. M. (2007). *Island biogeography: ecology, evolution, and conservation*. Oxford University Press.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of nature's laboratories. *Science, 357*(6354), eaam8326. https://doi.org/10.1126/science.aam8326

Wiley, E. O. (1978). The evolutionary species concept reconsidered. *Systematic Zoology, 27*(1), 17–26.

Williamson, S. H., Hubisz, M. J., Clark, A. G., Payseur, B. A., Bustamante, C. D., & Nielsen, R. (2007). Localizing recent adaptive evolution in the human

genome. *PLoS Genetics, 3*(6), 0901–0915. https://doi.org/10.1371/journal.pgen.0030090

Winker, K. (2010a). Is it a species?. *Ibis, 152*(4), 679-682.

Winker, K. (2010b). Chapter 1: Subspecies represent geographically partitioned variation, a gold mine of evolutionary biology, and a challenge for conservation. *Ornithological Monographs, 67*(1), 6-23. https://doi.org/10.1525/om.2010.67.1.6

Winker, K. (2021). An overview of speciation and species limits in birds. *The Auk, 138*(2), ukab006. https://doi.org/10.1093/ornithology/ukab006

Wirthlin, M., Lovell, P. V., Jarvis, E. D., & Mello, C. V. (2014). Comparative genomics reveals molecular features unique to the songbird lineage. *BMC Genomics, 15*(1), 1-20. https://doi.org/10.1186/1471-2164-15-1082.

Wittkopp, P. J., Williams, B. L., Selegue, J. E., & Carroll, S. B. (2003). *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. *Proceedings of the National Academy of Sciences of the United States of America, 100*(4), 1808–1813. https://doi.org/10.1073/pnas.0336368100

Wolf, J. B. W., & Ellegren, H. (2016). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics, 18*(2), 87-100. https://doi.org/10.1038/nrg.2016.133

Woolfit, M., & Bromham, L. (2005). Population size and molecular evolution on islands. *Proceedings of the Royal Society B: Biological Sciences, 272*(1578), 2277–2282. https://doi.org/10.1098/rspb.2005.3217

Wright, A. E., & Mank, J. E. (2013). The scope and strength of sex-specific selection in genome evolution. *Journal of Evolutionary Biology*, *26*(9), 1841-1853. https://doi.org/10.1111/jeb.12201

Wright, N. A., Steadman, D. W., & Witt, C. C. (2016). Predictable evolution toward flightlessness in volant island birds. *Proceedings of the National Academy of Sciences, 113*(17), 4765–4770. https://doi.org/10.1073/pnas.1522931113

Wright, S. (1931). Evolution in mendelian populations. *Bulletin of Mathematical Biology, 52*(1–2), 241–295. https://doi.org/10.1007/BF02459575

Wright, S. (1943). Isolation by distance. *Genetics, 28*(2), 114. https://doi.org/10.1109/IGARSS.2006.90

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology, 14*(6), 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x

Wu, P, Jiang, T. X., Suksaweang, S., Widelitz, R. B., & Chuong, C. M. (2004). Molecular Shaping of the Beak. *Science, 305*(5689), 1465–1466. https://doi.org/10.1126/science.1099191

Wu, P., Jiang, T. X., Shen, J. Y., Widelitz, R. B., & Chuong, C. M. (2006). Morphoregulation of avian beaks: comparative mapping of growth zone activities and morphological evolution. *Developmental Dynamics, 235*(5), 1400–1412. https://doi.org/10.1002/dvdy.20825

Xu, L., Auer, G., Peona, V., Suh, A., Deng, Y., Feng, S., ... & Irestedt, M. (2019). Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nature Ecology & Evolution, 3*(5), 834-844. https://doi.org/10.1038/s41559-019-0850-1.

Xu, S., Song, N., Zhao, L., Cai, S., Han, Z., & Gao, T. (2017). Genomic evidence for local adaptation in the ovoviviparous marine fish Sebastiscus marmoratus with a background of population homogeneity. *Scientific Reports, 7*(1), 1-12. https://doi.org/10.1038/s41598-017-01742-z

Yancey, C. (2015). Digital Commons @ Georgia Southern MyosinX is Required For Craniofacial Development in Danio Rerio.

Yang, J., Wan, W., Xie, M., Mao, J., Dong, Z., Lu, S., ... & Li, X. (2020). Chromosome-level reference genome assembly and gene editing of the dead-leaf butterfly *Kallima inachus*. *Molecular Ecology Resources, 20*(4), 1080-1092.

Yang, Y., Topol, L., Lee, H., & Wu, J. (2003). Wnt5a and Wnt5b exhibit distinct activities in coordinating chondrocyte proliferation and differentiation. *Development, 130*(5), 1003–1015. https://doi.org/10.1242/dev.00324

Yap, C. C., & Winckler, B. (2012). Harnessing the power of the endosome to regulate neural development. *Neuron, 74*(3), 440–451. https://doi.org/10.1016/j.neuron.2012.04.015

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., ... & Flicek, P. (2016). ensembl 2016. *Nucleic Acids Research, 44*(D1), D710–D716. https://doi.org/10.1093/nar/gkv1157

Ye, X., Yan, Z., Yang, Y., Xiao, S., Chen, L., Wang, J., ... & Ye, G. (2020). A chromosome-level genome assembly of the parasitoid wasp *Pteromalus puparum*. *Molecular Ecology Resources, 20*(5), 1384-1402.

Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America, 110*(19) , E1743-E1751. https://doi.org/10.1073/pnas.1219381110

Yeaman, S., & Otto, S. P. (2011). Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution, 65*(7), 2123–2129. https://doi.org/10.1111/j.1558-5646.2011.01277.x

Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution, 65*(7), 1897–1911. https://doi.org/10.1111/j.1558-5646.2011.01269.x

Yusnizar, Y., Wilbe, M., Herlino, A. O., Sumantri, C., Noor, R. R., Boediono, A., … & Andersson, G. (2015). Microphthalmia-associated transcription factor mutations are associated with white-spotted coat color in swamp buffalo. *Animal Genetics*, *46*(6), 676–682. https://doi.org/10.1111/age.12334

Zaharia, M., Bolosky, W. J., Curtis, K., Fox, A., Patterson, D., Shenker, S., ... & Sittler, T. (2011). Faster and more accurate sequence alignment with SNAP. *arXiv preprint*. https://doi.org/10.48550/arXiv.1111.5572.

Zan, Y., & Carlborg, Ö. (2019). A polygenic genetic architecture of flowering time in the worldwide *Arabidopsis thaliana* population. *Molecular Biology and Evolution, 36*(1), 141–154. https://doi.org/10.1093/molbev/msy203

Zhang, D. C., Guo, L., Guo, H. Y., Zhu, K. C., Li, S. Q., Zhang, Y., … & Li, J. T. (2019). Chromosome-level genome assembly of golden pompano (*Trachinotus ovatus*) in the family Carangidae. *Scientific Data, 6*(1), 1–11. https://doi.org/10.1038/s41597-019-0238-8

Zhang, G. (2015). Bird sequencing project takes off. *Nature, 522*(7554), 34-34. https://doi.org/10.1038/522034d

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., ... & Froman, D. P. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science, 346*(6215), 1311-1320. https://doi.org/10.1126/science.1251385.

Zhang, G., Parker, P., Li, B., Li, H., Wang, J., Parker, P., ... & Wang, J. (2012). The genome of Darwin's Finch (*Geospiza fortis*). *GigaScience, 1*, 13. https://doi.org/10.5524/100040.

Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M., & Andolfatto, P. (2012). Parallel molecular evolution in an herbivore community. *Science, 337*(6102), 1634–1637. https://doi.org/10.1126/science.1226630

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

Zhu, T., Qi, X., Chen, Y., Wang, L., Lv, X., Yang, W., … & Qu, L. (2021). Positive selection of skeleton-related genes during duck domestication revealed by whole genome sequencing. *BMC Ecology and Evolution, 21*(1), 1–8. https://doi.org/10.1186/s12862-021-01894-7

Zinzow-Kramer, W. M., Horton, B. M., Mckee, C. D., Michaud, J. M., Tharp, G. K., Thomas, J. W., Tuttle, E. M., Yi, S., & Maney, D. L. (2015). Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes, Brain and Behavior, 14*(8), 641–654. https://doi.org/10.1111/gbb.12252

Publications within the thesis dissertation

Recuerda, M., Illera, J. C., Blanco, G., Zardoya, R., & Milá, B. (2021). Sequential colonization of oceanic archipelagos led to a species-level radiation in the common chaffinch complex (Aves: Fringilla coelebs). *Molecular Phylogenetics and Evolution*, *164*, 107291.

Recuerda, M., Vizueta, J., Cuevas-Caballé, C., Blanco, G., Rozas, J., & Milá, B. (2021). Chromosome-level genome assembly of the common chaffinch (Aves: Fringilla coelebs): a valuable resource for evolutionary biology. *Genome Biology and Evolution*, *13*(4), evab034.

Molecular Phylogenetics and Evolution 164 (2021) 107291

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Sequential colonization of oceanic archipelagos led to a species-level radiation in the common chaffinch complex (Aves: *Fringilla coelebs*)

María Recuerda [a,*], Juan Carlos Illera [b], Guillermo Blanco [a], Rafael Zardoya [a], Borja Milá [a]

[a] National Museum of Natural Sciences, Spanish National Research Council (CSIC), Madrid 28006, Spain
[b] Biodiversity Research Unit (UO-CSIC-PA), Oviedo University, 33600 Mieres, Asturias, Spain

ARTICLE INFO

Keywords:
Islands
Phylogenomics
Phylogeography
Speciation
Species delimitation
Systematics

ABSTRACT

Oceanic archipelagos are excellent systems for studying speciation, yet inference of evolutionary process requires that the colonization history of island organisms be known with accuracy. Here, we used phylogenomics and patterns of genetic diversity to infer the sequence and timing of colonization of Macaronesia by mainland common chaffinches (*Fringilla coelebs*), and assessed whether colonization of the different archipelagos has resulted in a species-level radiation. To reconstruct the evolutionary history of the complex we generated a molecular phylogeny based on genome-wide SNP loci obtained from genotyping-by-sequencing, we ran ancestral range biogeographic analyses, and assessed fine-scale genetic structure between and within archipelagos using admixture analysis. To test for a species-level radiation, we applied a probabilistic tree-based species delimitation method (mPTP) and an integrative taxonomy approach including phenotypic differences. Results revealed a circuitous colonization pathway in Macaronesia, from the mainland to the Azores, followed by Madeira, and finally the Canary Islands. The Azores showed surprisingly high genetic diversity, similar to that found on the mainland, and the other archipelagos showed the expected sequential loss of genetic diversity. Species delimitation methods supported the existence of several species within the complex. We conclude that the common chaffinch underwent a rapid radiation across Macaronesia that was driven by the sequential colonization of the different archipelagos, resulting in phenotypically and genetically distinct, independent evolutionary lineages. We recommend a taxonomic revision of the complex that takes into account its genetic and phenotypic diversity.

1. Introduction

Oceanic archipelagos are excellent model systems to study evolution and have been crucial in advancing our understanding of species diversification and ecosystem assembly processes (Emerson, 2002; Losos and Ricklefs, 2009; Warren et al., 2015; Patiño et al., 2017; Whittaker et al., 2017)(Leroy et al., 2021). According to island biogeography theory, the number of species that can colonize and thrive on an oceanic island is a dynamic process primarily determined by the size of the island and its distance from the mainland (MacArthur and Wilson, 1967; Valente et al., 2017, 2020). Upon arrival, the original colonizers would start diverging from their mainland ancestors through neutral and/or selective processes (Warren et al., 2015). In many cases, the colonization of an archipelago is accompanied by an acceleration of net diversification rates (e.g. Delmore et al., 2020). This leads to species radiations in which phenotypic diversification could be driven either by adaptation to vacant ecological niches and available resources in the different islands (Schluter, 2000; Grant and Grant, 2008; Blanco et al., 2014), or by genetic drift and sexual selection in geographic isolation (Rundell and Price, 2009), although both types of processes can be at work within a single radiation (Gillespie et al., 2020).

Although evolutionary history is often simplified in oceanic archipelagos relative to continents, island colonization can be a complex process that can include multiple colonization and extinction events, back colonizations, as well as the maintenance of gene flow within and between archipelagos, and even with the continent (Illera et al., 2012; Morinha et al., 2020). When inferring the colonization history of oceanic archipelagos, it has been usually assumed that the original settlers originated from the closest mainland area (Grant, 1979; Thornton, 2007), subsequently following a chronological sequence of colonization consistent with a "stepping-stone model" (Funk and Wagner, 1995; Juan et al., 2000; Beheregaray et al., 2004; VanderWerf et al., 2010). However, this basic model is one of many possible ones (Sanmartín et al., 2008), and molecular phylogenetic analyses using exhaustive regional

* Corresponding author at: Museo Nacional de Ciencias Naturales, Calle José Gutiérrez Abascal 2, Madrid 28006, Spain.
E-mail address: mariarecuerdacarrasco@gmail.com (M. Recuerda).

https://doi.org/10.1016/j.ympev.2021.107291
Received 13 January 2021; Received in revised form 28 July 2021; Accepted 5 August 2021
Available online 9 August 2021

sampling are increasingly reporting counterintuitive colonization routes, suggesting that long distance migration events could be disrupted by a diverse range of factors (Emerson et al., 1999; Nathan, 2006; Felicísimo et al., 2008; Sequeira et al., 2008; Illera et al., 2012; Stervander et al., 2015; Morinha et al., 2020). Hence, in order to understand the evolutionary divergence of island biota, it is essential to set a robust phylogenetic framework to identify the closest living mainland relative, the phylogenetic relationships among insular species and populations, the timing and sequence of colonization (i. e., the order in which different islands were occupied), and the history of gene flow among insular populations within and between archipelagos (Whittaker and Fernández-Palacios, 2007; Losos and Ricklefs, 2009; Warren et al., 2015).

The common chaffinch (*Fringilla coelebs*) complex represents a sound system to study speciation processes on oceanic islands, as its broad geographic range includes Eurasia, Northern Africa, and the Atlantic Ocean archipelagos of Macaronesia, including Azores, Madeira and the Canary Islands, but not the Selvagens and Cabo Verde (Shirihai and Svensson, 2018). Common chaffinches on the mainland and the archipelagos differ genetically and in color pattern, morphology, and vocalizations (Grant, 1979; Lynch and Baker, 1994; Illera et al., 2018; Samarasin-dissanayake, 2010; Lachlan et al., 2013). Insular common chaffinches have characteristic dark blue-gray dorsal plumage, a larger body mass, shorter wings, as well as longer tarsi and bills compared to continental specimens (Grant, 1979). In addition, there are notable genetic and phenotypic differences among populations between and within the different archipelagos (see below). Although all common chaffinches are currently classified as a single species with several subspecific taxa, it has been suggested that mainland populations and the different archipelago radiations could be part of a multi-species complex (Illera et al., 2016).

Early proposals for the origin of Macaronesian chaffinches assumed the independent colonization of each archipelago from its nearest mainland, with phenotypic similarities among insular populations resulting from evolutionary convergence (Grant, 1979). In contrast, more recent studies based on mitochondrial DNA sequence data favored a single wave of colonization starting from Europe to Azores, Madeira, and finally the Canary Islands (Marshall and Baker, 1999), though limited genetic sampling and weak phylogenetic signal provided only tentative support for this hypothesis. Here, we tested these alternative hypotheses on the timing and colonization route of the common chaffinch radiation by building a robust phylogeny based on thousands of genome-wide loci. Genome-wide datasets based on SNP (single nucleotide polymorphism) loci have proven useful in resolving phylogenetic relationships at various evolutionary timescales, from deep nodes (Sackton et al., 2019) to very recent radiations (Stervander et al., 2015; Friis et al., 2016; Kozak et al., 2018; Meier et al., 2018). Based on a well-resolved phylogeny, we used biogeographical inference to estimate ancestral ranges using a dispersal-cladogenesis-extinction model that takes founder-event speciation into account and is thus particularly suited for oceanic island systems (de Queiroz, 2005; Gillespie et al., 2012; Matzke, 2013). Finally, in order to determine whether the colonization of oceanic archipelagos has resulted in a species-level radiation, we took an integrative taxonomy approach to determine the number of species in the complex according to different methods of species delimitation. This exercise has clear evolutionary and taxonomic implications, but also potentially major conservation impact for the taxa involved, most of which have restricted ranges and small population sizes (Whittaker et al., 2005).

2. Materials and methods

2.1. Study system and sample collection

The common chaffinch is currently considered to be a polytypic species composed of about 16 subspecies (Clement, 2020) which can be divided into three main geographic groups: a Eurasian group that includes the nominate form (*coelebs*) and related subspecies; a North African group that includes forms *africana*, *spodiogenys* and *harterti* (Svensson, 2015); and a Macaronesian group that includes *moreletti* from the Azores, *maderensis* from Madeira, and four subspecies on the Canary Islands, *canariensis* on Tenerife and La Gomera, *palmae* on La Palma, *ombriosa* on El Hierro, and the recently described *bakeri* on Gran Canaria (Martín and Lorenzo, 2001; Suárez et al., 2009; Illera et al., 2018).

For the present study we obtained blood samples from wild populations in Europe (Segovia, Spain), North West Africa (Ceuta, Spain), the Azores, Madeira and the Canary Islands, so that subspecies included were *coelebs*, *africana*, *moreletti*, *maderensis*, *canariensis*, *palmae*, *ombriosa* and *bakeri* (Fig. 1, Table S1). Birds were captured in the field using mist nets, and each individual was marked with a uniquely numbered Portuguese or Spanish aluminium band to avoid resampling. Birds were captured during the breeding season. Blood samples were obtained by venipuncture of the brachial vein and stored in absolute ethanol at −20 °C in the laboratory until DNA extraction.

2.2. SNP genotyping and analysis

High quality genomic DNA was extracted using a QIAGEN Blood and Tissue kit (Qiagen, Valencia, CA) following the manufacturer's protocol. SNP discovery was done using a genotyping-by-sequencing approach (Elshire et al., 2011) with restriction enzyme *PstI*, and sequencing was carried out on an Illumina Hiseq X Ten platform. Forward raw reads were trimmed to remove low quality ends using TrimGalore! V, 0.4.4 (http://www.bioinformatics. babraham.ac.uk/projects/trim_galore). We aligned the reads against the first version of the high-quality common chaffinch reference genome (GCA_015532645.1, Recuerda et al., 2021) using BWA 0.7.16 (Li and Durbin, 2009), using the "–mem" algorithm and default parameters. The reference genome was mapped against the zebra finch (*Taeniopygia guttata*) genome v87 available in Ensembl (Yates et al., 2016). We used the Chromosembler tool available in Satsuma (Grabherr et al., 2010) obtaining a final assembly 906.9 Mb in length and an N50 of 69.09 Mb. Variant calling was performed with GATK 3.6 HaplotypeCaller and GenotypeGVCFs tools (McKenna et al., 2010), calling all samples together with a minimum base and mapping quality score of 30. The variant dataset obtained was filtered using VCFtools version 0.1.15 (Danecek et al., 2011) keeping biallelic sites with a depth ranging between 4 and 60, a phred quality score over 30, and a minor allele frequency over 0.018. Indels were also removed along with sites with over 75% missing data and showing significant deviation from Hardy-Weinberg equilibrium (p-value $< 10^{-4}$). To recover the chromosomal coordinates of the scaffolds obtained with HiRiseTM, we mapped and oriented them against the zebra finch (*Taeniopygia guttata*) genome v87 available in ensembl (Yates et al., 2016). We used the Chromosembler tool available in Satsuma (Grabherr et al., 2010) resulting in a final genome assembly 955.9 Mb length and a N50 of 71.46 Mb.

In order to separate neutral loci from loci under divergent selection we used BayeScan v2.1 (Foll and Gaggiotti, 2008) to detect outlier loci in an F_{ST} distribution. We ran the program on a dataset of 159,534 loci with the default sample size of 5,000, a thinning interval of 200, a total of 20 pilot runs of 10,000 iterations each, and a burn-in of 100,000. We checked for convergence and set the false discovery rate (FDR) parameter at 0.1, obtaining 157,366 neutral SNPs and 2168 outliers. We filtered the neutral dataset for linkage disequilibrium (LD) using the snpgdsLDpruning function from the {SNPRELATE} package (Zheng et al., 2012) in R version 3.4.3 (R Core Team, 2017), resulting in a final dataset of 100,166 neutral SNPs.

2.3. Genetic diversity

Our final SNP dataset was composed of 81 individuals of the common chaffinch divided into two mainland and seven insular populations

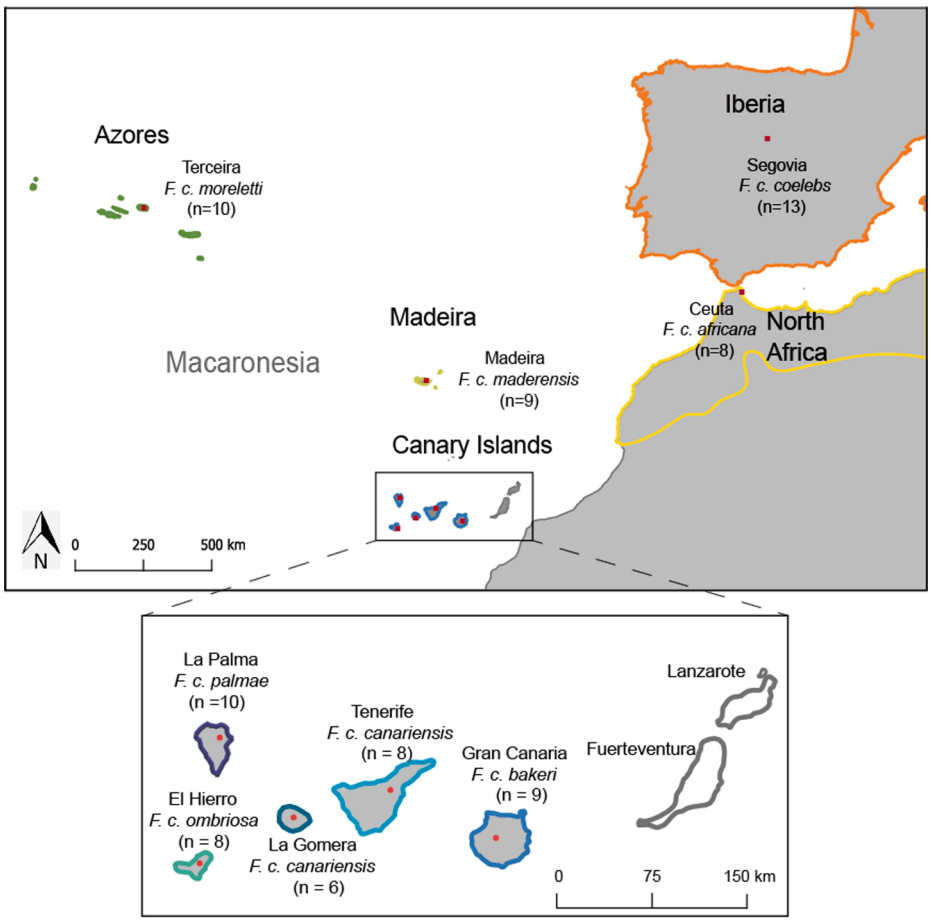


Fig. 1. Distribution map of the common chaffinch in the study area. Note the species is absent in the eastern Canary islands of Fuerteventura and Lanzarote. Red dots correspond to sampling sites and sample sizes are indicated in parentheses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Descriptive genetic statistics of the common chaffinch populations obtained with 159,534 SNPs: Locality, sample size (n), nucleotide diversity (π), observed heterozygosity (H_o) and expected heterozygosity (H_e). All F_{IS} values were significant (one sample *t*-test, $p < 0.0001$).

Region/Locality	n	π	H_o	H_e
Mainland	21	0.193	0.160	0.187
Africa (Ceuta)	8	0.177	0.160	0.165
Europe (Segovia)	13	0.188	0.159	0.177
Macaronesia	60	0.075	0.049	0.074
Azores (Terceira)	10	0.140	0.116	0.130
Madeira	9	0.051	0.047	0.048
Canary Islands	41	0.045	0.034	0.045
Gran Canaria	9	0.035	0.033	0.033
Tenerife	8	0.032	0.030	0.030
La Gomera	6	0.041	0.039	0.038
La Palma	10	0.031	0.031	0.029
El Hierro	8	0.042	0.039	0.039

(Table 1). Using the complete SNP dataset (159,534 loci), we calculated for each population: nucleotide diversity (π), the expected and observed heterozygosities (H_e and H_o) and pairwise F_{ST} among populations. All statistics were calculated using STACKS v 1.47 (Catchen et al., 2013). A one-sample *t*-test was used to determine whether the mean F_{IS} score in each population was statistically different from zero using R version 3.4.3 (R Core Team, 2017).

For comparative purposes, we also estimated genetic diversity and demographic parameters using coding regions from the mitochondrial genome (900 bp of the *atp8* and *atp6* genes, and 835 bp of the *nad2* gene), both individually and as a concatenated dataset (1,735 bp). The mitochondrial genes were amplified using primers L5215 (5′-TATCGGGCCCATACCCCGAAAAT-3′) (Hackett, 1996) and H6313 (5′-CTCTTATTTAAGGCTTTGAAGGC-3′) (Sorenson et al., 1999) for *nad2* and L8929 (5′-GGACAATGCTCAGAAATCTCGCGG-3′) (Eberhard and Bermingham, 2005) and H9855 (5′-ACGTAGGCTTGGATTATKGC-TACWGC-3′) (Sorenson et al., 1999) for *atp8* and *atp6*. PCR products were purified with an ethanol precipitation and sequenced by Sanger sequencing. Sequences were aligned using Sequencher 4.1.1 (Genecodes Inc., Ann Arbor, MI, USA) and the accuracy of variable sites was checked visually on the chromatograms. We calculated haplotype (*h*) and nucleotide (π) diversity indices per population, pairwise genetic distances and performed Fu's neutrality test (designed to detect changes in population growth; Fu, 1997) using Arlequin v. 3.5 (Excoffier and Lischer, 2010).

2.4. Phylogenetic analysis and estimation of divergence times.

To infer the evolutionary history of common chaffinches in the Macaronesian region we reconstructed a phylogenetic tree based on the neutral SNP dataset (100,166 loci), including a Tenerife blue chaffinch (*Fringilla teydea*) as outgroup. We built a maximum-likelihood (ML) tree using RAxML v8.1.16 (Stamatakis, 2014), using a GTR + GAMMA substitution model with the Lewis ascertainment bias correction as recommended. We implemented the rapid bootstrap algorithm (Stamatakis et al., 2008) and evaluated node support with 1000 replicates.

To estimate the timing of island colonization, we used three mitochondrial genes (*nad2*, *atp8* and *atp6*, Table S1) to reconstruct a

chronogram with Bayesian inference in BEAST v 1.8.4 (Drummond et al., 2012), using the CIPRES Science Gateway (Miller et al., 2010) and excluding the outgroup to avoid long-branch effects (Drummond and Bouckaert, 2015). We concatenated all genes (1,735 bp) and selected the best-fitting substitution model with Partitionfinder 2.1.1 (Lanfear et al., 2016), using the Akaike Information Criterion (AIC). The model selected for all markers and all codon positions was GTR + I. Based on results from preliminary runs, we implemented a strict molecular clock with a lognormal distribution of the mutation rate, setting mean values of 0.029 and 0.019 substitutions/site/My for *nad2* and *atp8&6* genes, respectively (Lerner et al., 2011). The haplotype networks for *nad2* and *atp8&6* genes were generated using Hapview (Salzburger et al., 2011) with maximum likelihood trees constructed using Geneious 10.2.2 (https://www.geneious.com) with default parameters.

We also estimated divergence times from a Bayesian phylogenetic tree using SNAPP, a template within BEAST version 2.5.1 (Bouckaert et al., 2018) using the CIPRES Science Gateway (Miller et al., 2010,). SNAPP infers the species tree from biallelic SNPs integrating over all possible gene trees by the implementation of the multispecies coalescent model. We used the neutral SNP dataset restricted to two individuals per population and allowing 5% of missing data, which resulted in 15,836 SNP loci. We used the script "snapp_prp.rb" (Stange et al., 2018) to generate the XML input file keeping the original settings, except that the MCMC chain was set to 2,000,000 generations. We used the RAxML tree as starting tree and set four constraints: (1) The monophyly of North Africa; (2) the monophyly of Europe; (3) the monophyly of the clade including all the insular populations; and (4) given the lack of common chaffinch fossil records in Macaronesia, we used a secondary calibration point based on our dating of the common chaffinch colonization of Macaronesia with mtDNA. We set a lognormal distribution for the divergence time of the insular clade with mean at 0.83 Ma (offset = 0, standard deviation = 0.1). A previous study based on a standard cyt-*b* calibration of 0.01 subs/site/lineage/ma obtained a similar date of 0.82 ma (Illera et al., 2018).

For both Bayesian analyses we checked for convergence using Tracer v 1.6 (Rambaut et al., 2014), ensuring that the estimated sample sizes (ESS) were over 200. Node ages and credible intervals (95% highest posterior density, HPD) were estimated, the best tree was generated using TREEANNOTATOR v1.8.4. (Drummond et al., 2012) and was displayed using FigTree v.1.4.3 (Rambaut, 2017).

2.5. Ancestral range estimation

Ancestral range estimation for the common chaffinch across Macaronesia was performed with the SNAPP phylogeny using the BioGeoBEARS package in R (Matzke, 2013). Among the dispersal-extinction-cladogenesis (DEC) models, we selected the DEC + J model. The "j" parameter allows for "founder-event speciation", which assumes that upon colonization of a remote locality, the founding population becomes instantly genetically distinct from the ancestral population (Matzke, 2014), a model that is appropriate for oceanic island systems, in which speciation takes place relatively quickly following colonization (De Queiroz, 2005; Cowie and Holland, 2006; Gillespie et al., 2012). Even though the DEC + J model may have a tendency to underestimate anagenetic events of dispersal and local extinction, which are probabilistic with respect to time, while inflating cladogenetic events of range expansion which are not time related (Ree and Sanmartín, 2018), we selected this model as the most biologically appropriate for our island scenario, where each taxon occupies a unique area that was likely sequentially colonized. We did not compare different models because according to Ree and Sanmartín (2018), their likelihoods are not statistically comparable, so that biological considerations are recommended for model selection instead. We set nine locations corresponding to the two continental areas (Europe and North Africa), which are also separated by sea, and the seven insular populations (Terceira in the Azores, Madeira, Gran Canaria, Tenerife, La Gomera, La Palma and El Hierro).

2.6. Genetic structure

To assess patterns of genetic structure and admixture between and within archipelagos, we used the program Structure (Pritchard et al., 2000) with the neutral SNP dataset, excluding the outgroup and filtered for missing data (5%), which resulted in a total of 16,416 loci. We used PGDSPIDER (Lischer and Excoffier, 2012) to convert the vcf file to the Structure format, ran preliminary analyses to infer the lambda value, and then ran analyses five times per K value, each one including 100,000 iterations and a burn-in of 50,000 iterations. The first analysis included individuals from all localities, with K values ranging from 2 to 9. To improve resolution in specific areas, we also ran separate region-specific analyses of the two mainland populations (K = 2–5), and the Canary Islands (K = 2–5). The structure plots were generated using CLUMPAK (Kopelman et al., 2015). The optimal K value was determined by the natural logarithm of the probability of the data [ln(Pr(X|K)] as described in the Structure manual. In order to check the robustness of results, we performed the same three analyses of population structure with Admixture v1.3.0 (Alexander et al., 2009) using the complete dataset of neutral SNPs (100,166 loci) with 200 bootstrap replicates.

To estimate fine-scale population structure and quantify the ancestry sources of each common chaffinch population, we used fineRADstructure (Malinsky et al., 2018), which uses information on haplotype linkage and common ancestry among individuals to produce a summary of nearest-neighbor haplotype relationships in the dataset in the form of a co-ancestry matrix. We converted the vcf file of neutral SNPs into fineRADstructure format using radiator (Gosselin, 2019) and we ran the pipeline using default parameters with 100,000 MCMC generations, sampling every 1,000 steps, and a burn-in of 100,000 steps. The tree was constructed with the fineSTRUCTURE algorithm (Lawson et al., 2012) with 10,000 iterations. The results obtained were plotted in R by adapting the scripts provided in http://cichlid.gurdon.cam.ac.uk/fineRADstructure.html.

2.7. Species delimitation

To estimate the number of species in the common chaffinch radiation, we applied the multi-rate Poisson Tree Processes (mPTP) method for species delimitation (Kapli et al., 2017). The mPTP method is based on a rooted phylogenetic tree obtained by probabilistic methods, and it attempts to differentiate speciation from coalescence processes, allowing different intraspecific coalescent rates and a constant speciation rate, assuming that branching events within species are more frequent than between species. For input, we used the RAxML tree based on neutral SNPs, and we ran 10 independent MCMC chains of 10^8 steps, logged every one million generations, with a burn-in of two million steps. We used the "-multi" option to allow variance in coalescent rates among species, and the minimum branch length used was 0.001831, as calculated with the tool "minbr_auto". Average node support values (AVS) were generated for each clade by the MCMC method, with values close to one indicating a robust ML delimitation. We set a conservative threshold for support values over 75 to consider clusters as different candidate species (Kapli et al., 2017). We ensured chain convergence using the Average Standard Deviation of Support Values (ASDDSV), which quantifies the similarity among independent MCMC runs.

In addition to the mPTP analysis we applied an integrative taxonomic approach to species delimitation (Padial et al., 2010). In addition to the genetic data, we took into account differences in plumage coloration (Fig. 7) as well as previously published morphological data (Grant, 1979) and bioacoustic data (Lachlan et al., 2013). Finally, we applied a scoring system for avian species delimitation proposed by Tobias et al. (2010) which is based on phenotypic and geographic data, and has been adopted by some major avian taxonomic systems (del Hoyo et al., 2020; Handbook of the Birds of the World and BirdLife International, 2019).

The method scores and combines the strongest differences in five types of avian traits: morphology, acoustics, plumage, ecology, behavior, and geographical relationships, and assigns species status if the total score reaches or exceeds an arbitrary threshold value (See supplementary Methods, File S1). This points-based scoring system has received some criticism due to the subjectivity involved in the scoring itself, and because the quantitative criteria are based on fairly arbitrary magnitudes of difference that are broadly applied across taxa (Winker, 2010b). However, the method has demonstrated to be useful when used for taxonomical purposes (Winker, 2021) and its performance has been found to be high when tested against recently accepted splits (Tobias et al., 2021).

3. Results

3.1. SNP genotyping

We obtained 27,052,300 reads from GBS, which resulted in 15,506,115 reads after trimming. The mapping using BWA resulted in 207,339,592 primary aligned reads mapped to the common chaffinch reference genome. The variant calling with GATK generated 1,988,317 variants and after filtering with VCFTOOLS we obtained 159,534 variants with an average depth per site of 16.5.

3.2. Genetic diversity and differentiation

Genetic diversity indices were lower on islands than on the mainland (Table 1). Nucleotide diversity and heterozygosity were highest in mainland populations, followed by Azores and Madeira, with the Canary Islands showing the lowest values (Table 1). Pairwise F_{ST} values among populations ranged from 0.07 to 0.16, with an average of 0.13 (Table 2). The lowest differentiation was found among the common chaffinches of Europe and North Africa, the latter being more differentiated from all insular populations than the former. The Azores population showed the lowest differentiation from mainland populations, and both Azores and Madeira showed similar values of differentiation with respect to the Canary Islands. Within the Canary Islands, F_{ST} values were generally consistent with geographic proximity among islands, with values ranging from 0.09 between Tenerife and La Gomera, and 0.14 between Gran Canaria and the other islands. Genetic distances calculated with the mtDNA dataset showed a similar pattern to that found for SNP markers (Tables 4, S3).

3.3. Phylogenetic analysis, colonization route and divergence times

The ML phylogenetic tree based on 100,166 neutral SNPs was highly resolved, with maximal node support for clades separating the different archipelagos and the different islands within the Canary archipelago (Fig. 2). The phylogenetic tree based on mitochondrial markers showed a similar topology to the genome-wide phylogeny, except for two relationships, which were not highly supported: (1) the two mainland populations (Europe and North Africa) formed a single clade (Fig. 3a); (2) the population of Gran Canaria was sister to a clade including two sister subclades: (a) the westernmost islands of La Palma and El Hierro, and (b) the geographically close islands of Tenerife and La Gomera. Individuals within the Tenerife-La Gomera clade showed an incomplete sorting of haplotypes despite a higher proportion of private haplotypes per island (Fig. 3a). Haplotype networks revealed that the *nad2* gene showed higher diversity than the *atp8&6* genes except in Madeira and La Palma, and showed better sorting of haplotypes between the two mainland populations and the Tenerife/La Gomera clade, yet neither marker showed complete lineage sorting relative to the genome-wide phylogeny (Fig. 3b,c, Tables 3 and S2), which provided higher phylogenetic resolution than the mtDNA data.

Dating estimates indicated that insular populations diverged from the mainland around 0.83 million years ago (HPD: 0.38–1.48 Ma), Madeira diverged from the Canary Islands about 0.70 Ma ago (HPD: 0.34–1.28), and the Canary Islands differentiated from each other within the last half million years (Fig. 3a).

The SNAPP phylogenetic tree recovered the same topology as the mitochondrial phylogenetic tree but separated the insular populations of Tenerife and La Gomera (Fig. 4). The ancestral range estimation confirmed that colonization of the Atlantic Islands started in Azores, then Madeira, and finally the Canary Islands (Fig. 4). However, the mainland starting point was not clear, with both Europe and North Africa showing similar probabilities. Within the Canary Islands, the analysis suggested that the first island to be colonized was Gran Canaria, but the ancestral range of the remaining islands was not resolved.

3.4. Genetic structure and admixture analysis

The Structure analysis based on the genome-wide SNP dataset revealed marked genetic structure across the region that was consistent with the ML phylogeny. The optimal number of genetic clusters was K = 6, with clusters corresponding to North Africa, Europe, Azores, Madeira, Gran Canaria and the remaining Canary Islands, respectively (Fig. 5a). An analysis restricted to the mainland individuals confirmed the separation of both populations as the best clustering (Fig. 5b, Fig. S2, Fig. S3), and a separate analysis of the Canarian archipelago yielded five clusters with high posterior probability of assignment of all individuals to each of the five islands at K = 5 (Fig. 5c). In the latter analysis, K = 2 separated Gran Canaria from the rest, K = 3 additionally separated the western islands (La Palma and El Hierro) and the central islands (Tenerife and La Gomera), and K = 4 and K = 5 separated these two pairs of islands from each other, although La Gomera showed a small proportion of admixture with Tenerife. The Admixture results were generally consistent with the Structure analysis, with the same optimal number of clusters but some differences in the sequence of population separation (Fig. S2, Table S4). In both analyses, the Azores shared some variance with the mainland at K = 2, and Gran Canaria shared some variance with Madeira, being the first island to separate from the rest within the Canary archipelago.

The FineRADstructure analysis showed consistent results with previous analyses and divided individuals into the same nine populations

Table 2
Fixation index (F_{ST}) values among populations of the common chaffinch obtained with 159,534 SNPs. EUR (Iberian Peninsula), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma).

	EUR	AZO	MAD	GC	TEN	GOM	PAL	HIE
AFR	0.069	0.127	0.152	0.148	0.150	0.133	0.152	0.143
EUR		0.094	0.113	0.109	0.111	0.097	0.113	0.105
AZO			0.155	0.157	0.159	0.143	0.161	0.151
MAD				0.159	0.163	0.147	0.158	0.150
GCA					0.135	0.127	0.140	0.136
TEN						0.089	0.134	0.126
GOM							0.118	0.113
PAL								0.096

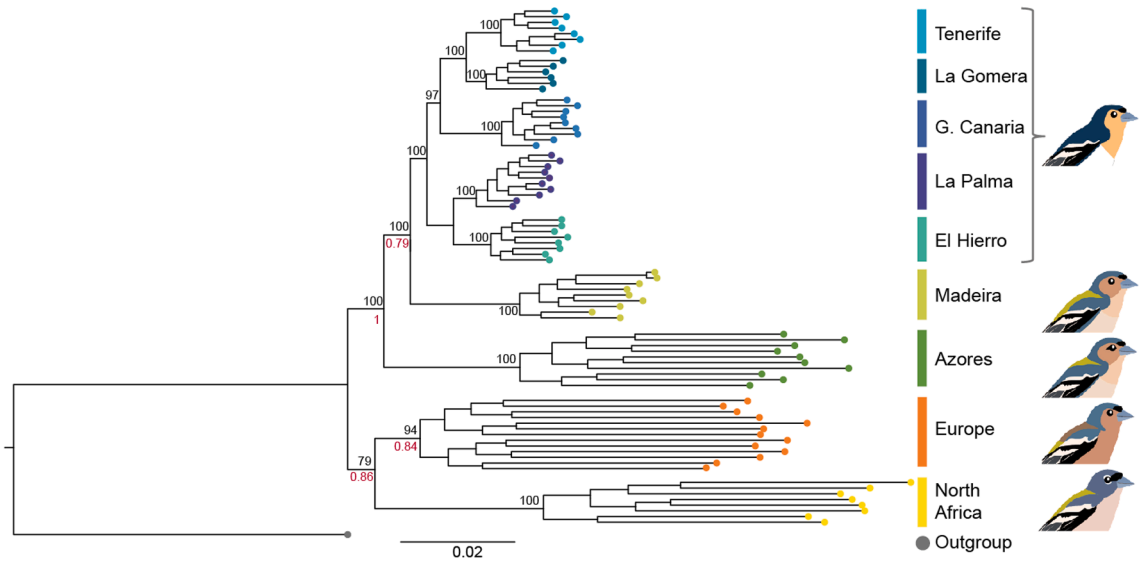


Fig. 2. Maximum likelihood phylogenetic tree based on 100,166 genome-wide neutral SNP loci performed using RAxML with 1000 rapid bootstraps and using the blue chaffinch (*Fringilla teydea)* as the outgroup. Figures in black are node support values. Figures in red correspond to Average support values (AVS) from the mPTP species delimitation method. Sketches on the right depict the main phenotypic differences between forms, with chaffinches from the Canary Islands represented by subspecies *palmae*. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

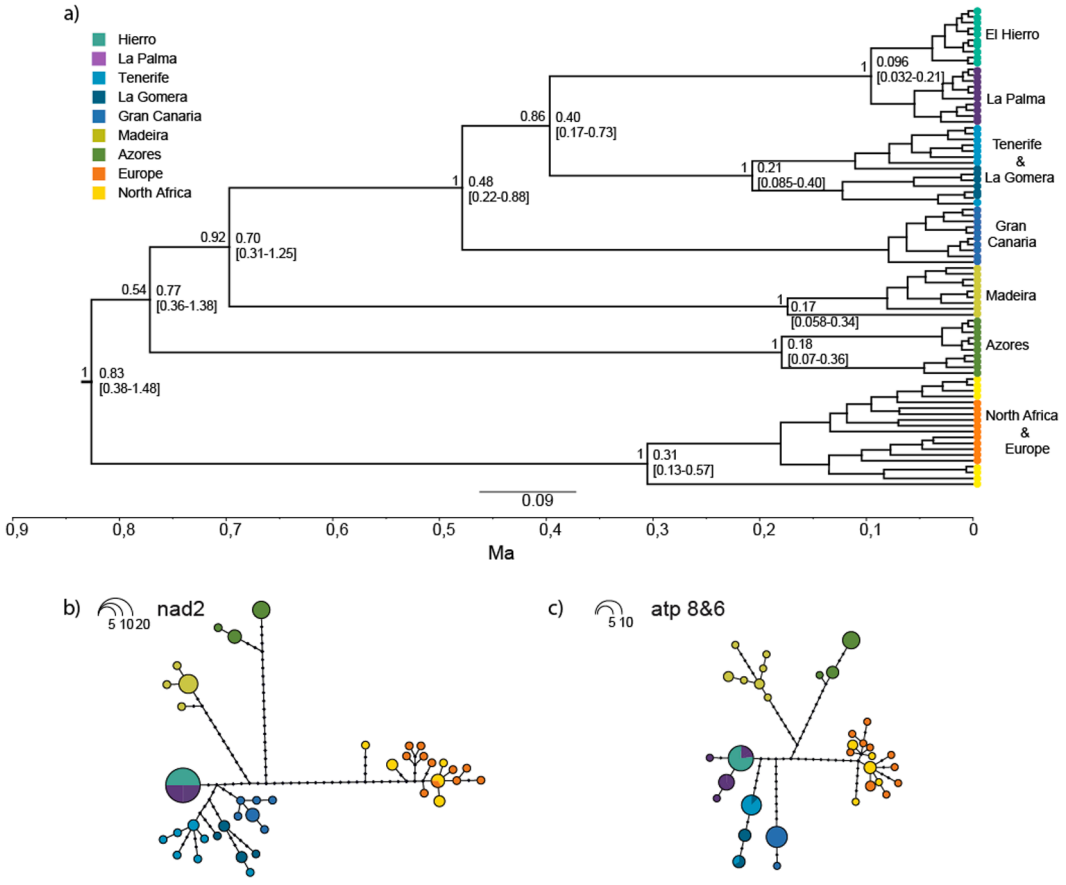


Fig. 3. (a) Ultrametric Bayesian tree based on three mitochondrial genes (*atp8*, *atp6* and *nad2*,) obtained with BEAST. Values on the left of each node represent posterior probability of node support. Values on the right of each node represent node age in million years, with confidence intervals (95% HPD) in brackets. (b) Haplotype networks based on *nad2* and (c) *atp8&6* genes. Circles correspond to haplotypes, and their size is proportional to the frequency of each haplotype in the population. Black dots along branches correspond to unsampled or extinct haplotypes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Genetic diversity and population expansion indices of common chaffinch populations. MtDNA genes used include *atp8* and *atp6* genes (900 bp), *nad2* (835 bp) concatenated (1,735 bp). Included are DNA marker, geographic region, sample size (n), number of haplotypes (No. haps), haplotype diversity (*h*), nucleotide diversity (π), Fu's neutrality test (F_S). Statistical significance of F_S values is indicated by asterisks (* $p = 0.05$, ** $p = 0.01$ and ***, $p = 0.001$). EUR (Iberian Peninsula), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma).

DNA marker	Region	n	No. haps.	$h \pm$ SD	$\pi \pm$ SD	F_S
atp8&6 + nad2						
	AFR	8	6	0.93 ± 0.084	0.0043 ± 0.0405	0.33
	EUR	11	11	1.00 ± 0.039	0.0045 ± 0.0394	−5.10 **
	AZO	9	4	0.69 ± 0.15	0.0036 ± 0.0430	3.27
	MAD	9	7	0.94 ± 0.07	0.0024 ± 0.0261	−1.67
	GCA	8	6	0.90 ± 0.11	0.0009 ± 0.1912	−3.44 ***
	TEN	8	7	0.96 ± 0.08	0.0024 ± 0.0293	−2.32
	GOM	6	4	0.87 ± 0.13	0.0039 ± 0.0408	1.78
	PAL	10	4	0.89 ± 0.08	0.0005 ± 0.0145	−1.02
	HIE	10	1	0.00 ± 0.00	0.0000 ± 0.0000	0.00

(Fig. 6). The plot also showed clear regional structure among populations with two main clusters, one formed by the continental individuals along with Azores, and the other including the remaining insular populations. Coancestry relationships among populations revealed that the Azores shares more ancestry with Europe than with North Africa. Within the insular cluster, two pairs within the Canary Islands show high coancestry (Tenerife and La Gomera, and La Palma and El Hierro, respectively).

3.5. Species delimitation

The 10 independent MCMC runs of mPTP suggested species-level designation for the five main clades in the ML phylogeny, corresponding to Europe, North Africa, Azores, Madeira and the Canary Islands, with support values ranging from 0.79 to 1 (Fig. 2, values in red). In addition, mPTP suggested one additional clade within Europe, with a support value of 0.84.

We integrated the molecular data from the mPTP analysis with phenotypic data and all five clades identified by mPTP showed congruent differentiation in phenotypic traits, mainly in terms of plumage color but also morphology and bioacoustics. When scoring differences in plumage coloration (Fig. 7) and morphology (Table S6) among pairs of subspecies using the five most prominent traits (Tobias et al., 2010), all comparisons reached the minimum threshold for species designation (Table S5).

4. Discussion

4.1. Colonization history in the common chaffinch radiation

Our results from molecular phylogenies, ancestral range estimation, and coancestry analyses, provide strong and consistent support for a colonization of Macaronesia by the common chaffinch that took place from the mainland, via the Azores and Madeira to the Canary Islands, and resulted in a rapid species-level radiation. This circuitous colonization route seems counterintuitive from a biogeographic perspective, given the large distance separating the Azores from the mainland (*ca.* 1300 km) compared to the other archipelagos, and suggests that factors other than mere geographic distance were at play in the common chaffinch radiation. Although the topologies of the phylogenetic trees do not allow determining whether the original colonizers of the Azores came from Europe or North Africa, the coancestry analysis with fineRADstructure, along with the genetic distances based on both datasets, suggests that a European origin is more likely. The estimation of the colonization time of these Atlantic islands by the common chaffinch obtained with BEAST coincides with previous estimates of about one million years before present (Illera et al., 2018), which is relatively recent compared to the age of most of the islands (Illera et al., 2012). The estimated colonization time falls within the last 3 million years, a period found to include most colonization events by Macaronesian bird taxa (Valente et al., 2017). This period coincides with the establishment of most Macaronesian laurel forests in the Plio-Pleistocene (2.6 Ma), and with the movement of the trade wind zone over the islands during the Pleistocene (2.6–0.01 Ma), which provided sufficient precipitation and moisture (Kondraskov et al., 2015). The phylogenomic tree obtained with ~ 100,000 neutral SNPs provided enough resolution to reveal independently evolving, monophyletic lineages of the common chaffinch on each archipelago. Results also suggest shared ancestry of all the Macaronesian islands, followed by divergence with restricted gene flow among islands. This single-wave colonization history is supported by shared phenotypic characters among insular populations. Macaronesian chaffinches show plumage patterns with blue-gray dorsal coloration and reduced green and red patches (Grant, 1980); longer tarsi and shorter wings than their mainland counterparts (Grant, 1979; Dennison and Baker, 1991), as documented for other passerines (Wright et al., 2016); and decreasing song complexity after each colonization event (Lynch and Baker, 1994; Lachlan et al., 2013). Overall, this pattern of shared traits among all insular populations is more consistent with common ancestry than convergence following independent colonizations from the nearest mainland (Marshall and Baker, 1999). Given the phylogenetic relationships among all insular populations, common ancestry is more parsimonious than the alternative hypothesis of repeated, independent evolution of these traits on each island under common selective pressures.

Table 4
Genetic distances between the different lineages of the common chaffinch using the *atp8* and *atp6* genes (900 bp) and *nad2* (835 bp) concatenated (1,735 bp). Above the diagonal: average number of pairwise differences between populations. Below the diagonal: corrected average pairwise differences. Along the diagonal (in italics): average number of pairwise differences within populations. EUR (Iberian Peninsula), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma).

	AFR	EUR	AZO	MAD	GCA	TEN	GOM	PAL	HIE
atp8&6 + nad2									
AFR	*7.46*	8.38	59.75	52.13	48.13	48.31	48.88	36.92	36.13
EUR	0.76	*7.76*	61.86	53.33	49.36	50.53	50.88	39.24	38.55
AZO	52.91	54.87	*6.22*	50.78	50.22	46.46	47.85	40.92	40.22
MAD	46.31	47.37	45.58	*4.17*	39.00	37.35	38.80	31.91	31.11
GCA	43.64	44.73	46.36	36.17	*1.50*	26.38	28.00	18.80	18.00
TEN	42.37	44.43	41.13	33.05	23.41	*4.43*	8.92	15.18	14.38
GOM	41.74	43.60	41.34	33.31	23.85	3.30	*6.80*	16.80	16.00
PAL	32.72	34.89	37.34	29.36	17.58	12.49	12.93	*0.93*	0.80
HIE	32.39	34.66	37.11	29.03	17.25	12.16	12.60	0.33	*0.00*

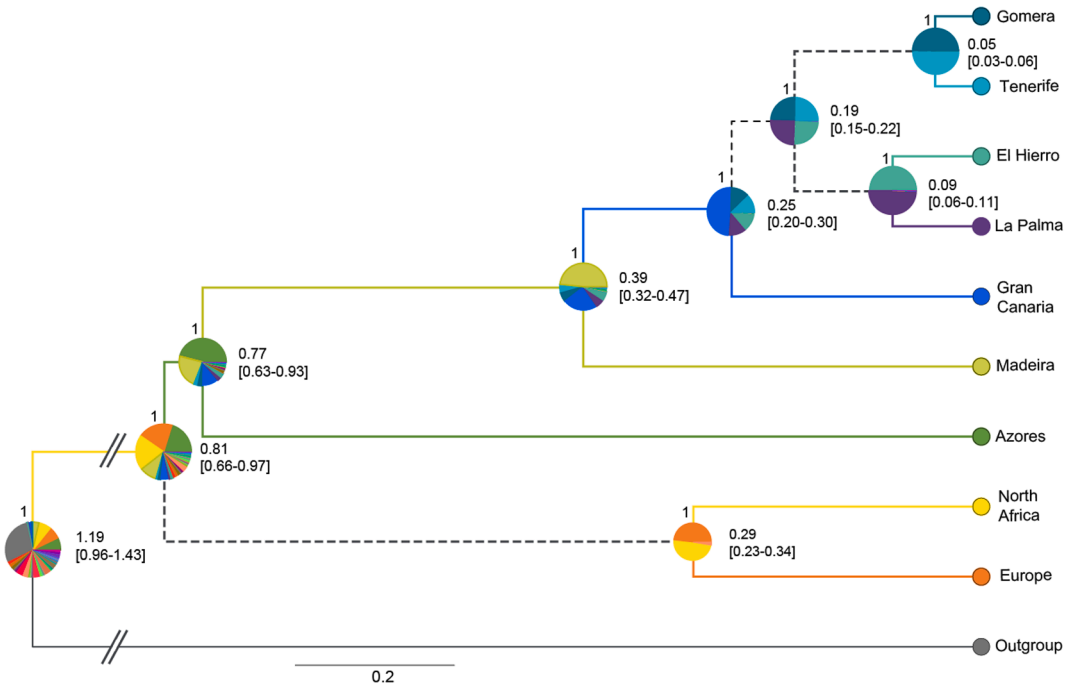


Fig. 4. Ancestral range estimation of common chaffinch populations. Inference based on a dispersal-extinction-cladogenesis model with founder event (DEC + J), with the Bayesian phylogeny based on 15,836 neutral SNPs. Pie diagrams at each node represent the inferred geographical ranges for each ancestral taxon, with the probability of each area indicated by its respective color. Branch color represents the most likely state for each branch. Dashed branches indicate that multiple states were tied. Figures above pies represent posterior probabilities of node support, and figures to the right of each node correspond to age in Ma, with confidence intervals (95% HPD) in brackets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

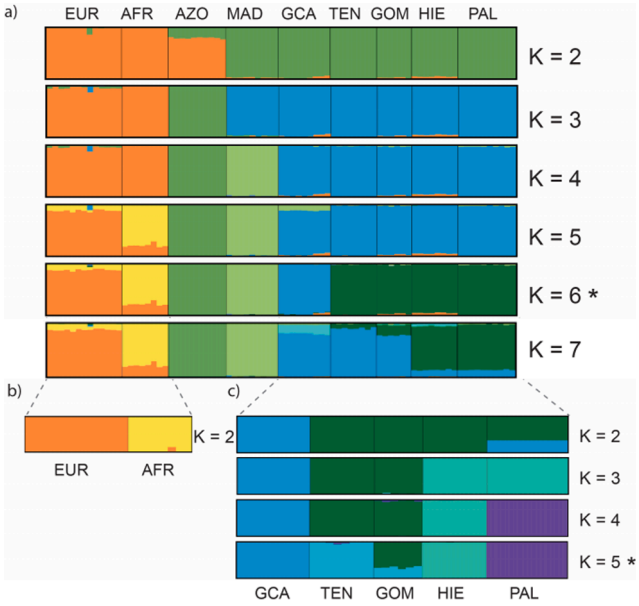


Fig. 5. STRUCTURE analysis plots for (a) all chaffinch populations with K ranging from 2 to 7 (plots for K = 8 and 9 are not shown as they do not differ from K = 7), (b) mainland populations only for K = 2, and (c) Canary Islands populations only with K ranging from 2 to 5. EUR (Iberia), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma). Asterisks (*) mark the optimal K value for each analysis.

Unlike the Azores, where gene flow appears to have prevented the differentiation of common chaffinch populations among islands (Baker et al., 1990; Rodrigues et al., 2014), those in the Canary Islands have diverged markedly from each other, giving rise to a range of phenotypes currently grouped into four different subspecific taxa (Illera et al., 2018). Partly because of this recent inter-island differentiation, inferring the specific order in which the Canary Islands were colonized is challenging (Marshall and Baker, 1999). The absence of the common chaffinch in the eastern-most islands of Lanzarote and Fuerteventura may be due to the current lack of suitable habitat, which is known to have varied widely over time due to the frequent extinction-recolonization events of their flora (García-Verdugo et al., 2019), but whether or not the common chaffinch was present there in the past cannot be determined from available data. For the islands where the common chaffinch is present, we obtained conflicting results and found evidence consistent with both an east-to-west and a west-to-east pattern of colonization. On one hand, our results support the eastward colonization because La Palma is closest to Madeira in the haplotype networks and shows lower genetic distance with Madeira than Gran Canaria. This route may have been favoured by the wind patterns that blow south-eastwards from the Azores in winter (Grant, 1980), as previously proposed (e.g., Grant, 1980; Marshall and Baker, 1999; Suárez et al., 2009; Lachlan et al., 2013). On the other hand, the mitochondrial DNA tree, the ancestral range estimation and the population structure analysis are more consistent with a westward colonization starting from Gran Canaria. More research will be needed to disentangle the specific common chaffinch colonization within the Canary Islands, an archipelago with a diverse range of avian colonization histories given its proximity to neighboring archipelagos and mainland (Illera et al., 2012; Morinha et al., 2020).

The progressive reduction of genetic diversity from the Azores to the Canary Islands is also consistent with the colonization route, and expected when islands are sequentially colonized from other islands by small groups of individuals from source populations of progressively smaller effective population size (Clegg et al., 2002). Genetic diversity in the Azores was similar to that found on mainland populations and an order of magnitude higher than that found on other archipelagos. This suggests that a relatively large group of original colonizers (or multiple colonization events in a short period of time), arrived to Azores, avoiding a major founder event (James et al., 2016), but also that

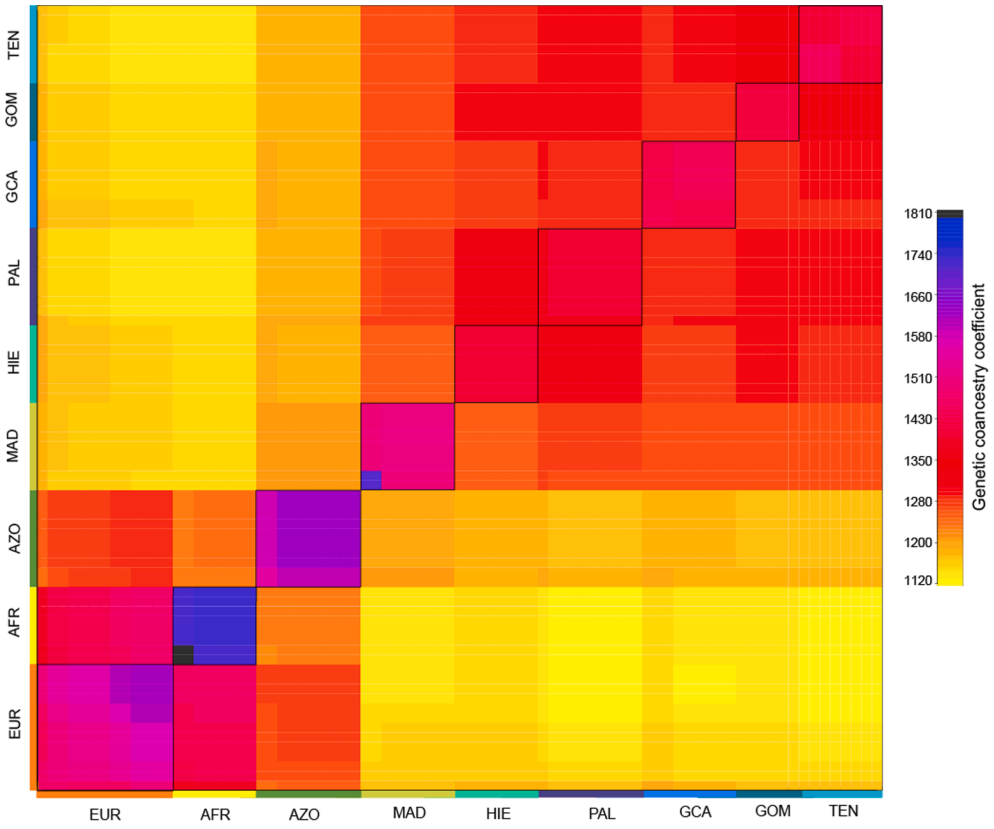


Fig. 6. Matrix of pairwise genetic co-ancestry values among chaffinch populations. Averaged co-ancestry coefficients per population are color-coded from low (yellow) to high (black). Individuals clustering into populations are shown along the diagonal (squares framed in black). EUR (Iberia), AFR (North Africa), AZO (Azores), MAD (Madeira), GCA (Gran Canaria), TEN (Tenerife); GOM (La Gomera), HIE (El Hierro) and PAL (La Palma). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

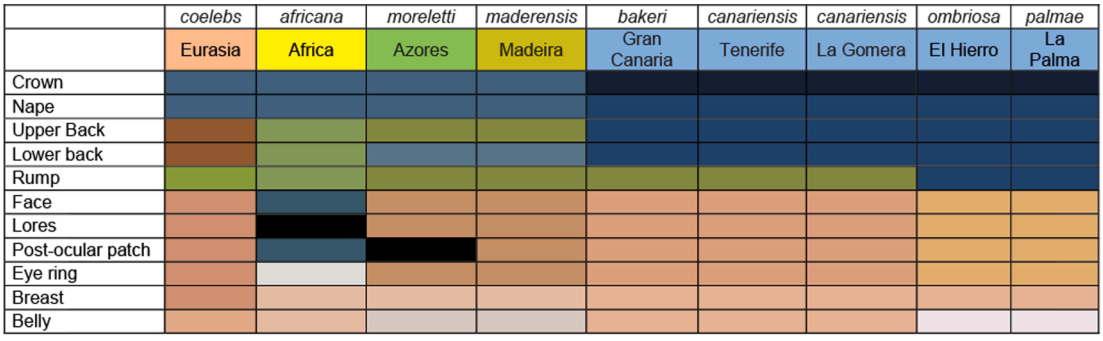


Fig. 7. Summary of the main phenotypic differences among males of the different chaffinch taxa. Colors depicted for the different body parts are approximate estimates of real colors obtained from photographs (see Methods). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

effective population size was maintained relatively large over time. Indeed, in addition to the magnitude of potential founder events, the surface area of suitable common chaffinch habitat in the different islands and the presence of gene flow among them are also likely to have influenced present levels of genetic diversity. Except for La Palma, where common chaffinches have stable breeding populations in dry pine forests, Macaronesian common chaffinches are largely restricted to *monteverde* humid habitats, from cloud forest to moist heaths, and the geographic area of these habitat types varies widely among islands (Martín and Lorenzo, 2001). While most of the Azores are humid enough to sustain common chaffinch populations, suitable habitat decreases markedly with latitude, becoming less abundant in Madeira, and restricted to small "islands within islands" in the Canaries, where humid habitats are more restricted than in the other archipelagos (Fernández-Palacios, 2009). In turn, gene flow among the Azores, which has prevented genetic differentiation among islands (Rodrigues et al., 2014), has favored the maintenance of high population sizes and genetic diversity, in contrast to the Canary Islands, where populations have become isolated from each other due to highly restricted gene flow.

The chaffinch taxa produced in the Macaronesian archipelagos differ from each other mostly in plumage coloration, and to a much lesser degree in morphological characters. This is similar to what has been observed in non-adaptive avian radiations, such as those in South American capuchino seedeaters (Campagna et al., 2012), North American juncos (Friis and Milá, 2020), or European wagtails (Ödeen and Björklund, 2003), where taxa differ in color traits with a simple genetic basis (Campagna et al., 2017; Abolins-Abols et al., 2018), yet are relatively uniform in morphology. This suggests that drift and sexual selection have been the main drivers of the phenotypic diversification, with morphological adaptation to local ecological conditions playing a relatively minor role (Rundell and Price, 2009), likely due to the ecological similarity between Macaronesia and its mainland. This is in contrast to well-studied adaptive radiations such as that of the Darwin's finches in the Galapagos Islands (Grant and Grant, 2008; Lamichhaney

et al., 2015), or the honeycreepers in Hawaii (Lerner et al., 2011). Within a similar time frame to that of the chaffinch diversification, these two radiations gave rise to markedly diverse beak morphologies as populations of the original colonizers adapted through strong directional selection to the food resources available in the different islands. In the case of the honeycreepers, which belong to the same family Fringillidae as chaffinches, morphological divergence was accompanied by a stunning diversification in color patterns and other ornamental traits (Freed et al., 1987), suggesting the combined action of natural and sexual selection (Gillespie et al., 2020). Even though the common chaffinches have not diversified bill morphology to that extent, natural selection has likely played a role in modifying their morphology, especially the size and shape of their beaks (Grant, 1979).

4.2. Systematics and taxonomy of the chaffinch radiation

Our species delimitation analyses suggest that the common chaffinch radiation has resulted in several species-level taxa. The genome-wide analysis of genetic variation revealed the existence of several distinct evolutionary lineages evolving independently from each other, and species delimitation analyses provided support for the existence of at least five different species within the complex. The mPTP method provided support for the five nodes corresponding to North Africa, Europe, Azores, Madeira and Canary Islands, respectively. The additional supported clade within Europe could be due to high genetic diversity of the European population, and does not seem to be associated with phenotypic differences or geographical limits. The Structure and Admixture analyses for the continental clades showed that for $K > 2$, some individuals of the Iberian population show some divergence, but do not correspond to the clades in the phylogenomic tree (Fig. S3). Marked phenotypic divergence among major lineages was confirmed by Tobias' et al. (2010) delimitation method, which was also consistent with the five-species hypothesis. Even though, plumage coloration and morphological differences among *F. c. moreletti* and *F. c. maderensis* were less prominent than between other members of the complex, they are known to differ in other characters relevant to reproductive isolation like territorial male song (Lachlan et al., 2013), that were not included in our analysis.

We concur with previous studies on this system (Marshall and Baker, 1999; Suárez et al., 2009; Rodrigues et al., 2014; Illera et al., 2016; Perktaş et al., 2017; Clement, 2018), on the need for a taxonomic revision of this group, and based on their and our results, we propose that the common chaffinch be divided into five different species, corresponding to Eurasia (*Fringilla coelebs*), North Africa (*Fringilla spodiogenys/africana*), Azores (*Fringilla moreletti*), Madeira (*Fringilla maderensis*) and the Canary Islands (*Fringilla canariensis*). *F. coelebs* would include all subspecies closely related and phenotypically similar to *F. c. coelebs* found across continental Eurasia. Although populations on the different Canary Islands are genetically distinct, their phenotypic differentiation is relatively minor, and we propose to maintain their current subspecific status within *F. canariensis*. Such a subspecific classification would be as follows: *F. canariensis canariensis* on Tenerife and La Gomera, *F. canariensis palmae* on La Palma, *F. canariensis ombriosa* on El Hierro, and *F. canariensis bakeri* on Gran Canaria.

North African subspecies *spodiogenys* and *harterti* were not included in this study, yet they are phenotypically similar to *africana* (Svensson, 2015; Perktaş et al., 2017). The early molecular study by Marshall and Baker (1999) reported *spodiogenys* as a divergent lineage that was basal to the *Fringilla coelebs* complex in a mtDNA phylogeny, yet more recent molecular analyses using nuclear DNA markers indicate that the two North African subspecies are indeed closely related sister taxa (Samarasin-Dissanayake, 2010). This result is consistent with both phenotype and geography, and suggests that mtDNA may not be suitable to recover the evolutionary history of these taxa. Based on this evidence, and since *spodiogenys* Bonaparte 1841 was described before *africana* Levaillant 1850 and *harterti* Svensson 2015, we recommend recognizing species *Fringilla spodiogenys* with three subspecies (*F. spodiogenys spodiogenys*, *F. spodiogenys africana*, and *F. spodiogenys harterti*).

Recognizing the new proposed species should be consistent with most species concepts that take into account evidence for independent evolving lineages and phenotypic differentiation (De Queiroz, 2007; Sangster, 2013; Gill, 2014). The taxonomic upgrade from subspecies to species is likely to have important conservation implications, as species tend to receive more conservation attention than subspecies (Winker, 2010a, 2010b; Sangster et al., 2016). Specifically, species status would guarantee that the conservation status of each chaffinch taxon is evaluated by the International Union for Conservation of Nature (IUCN), taking into account their distribution area and population size independently, making the difference especially for the more restricted insular populations (Martín, 2009). Hence, conservation biogeography (Whittaker et al., 2005), which includes the distribution of taxa in the conservation criteria by applying biogeographical analysis is important for the improvement of biodiversity conservation. This may in turn help preserve the genetic diversity of the species complex, which is crucial for the resilience to environmental change in the current scenario of climate change, especially given the reduced genetic variability found across the region.

4.3. Conclusions

The colonization of Macaronesia by the common chaffinch has resulted in an evolutionary radiation as populations differentiated phenotypically and genetically in the different archipelagos, and even between islands within the Canary archipelago. The molecular phylogeny was instrumental in revealing a circuitous colonization route from the mainland to the faraway Azores, and then south to Madeira and the Canary Islands. Relatively minor differences in morphology between insular and mainland chaffinches compared to differences in coloration, suggest that drift due to founder events, along with sexual selection acting on plumage coloration and song, are likely the major factors driving the common chaffinch radiation in Macaronesia. The sequential colonization of three Atlantic archipelagos and Northern Africa has led to the formation of at least four new species-level taxa in the genus *Fringilla*, and our results should help further our understanding of the evolutionary processes involved.

CRediT authorship contribution statement

María Recuerda: Conceptualization, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft. **Juan Carlos Illera:** Investigation, Resources, Writing – review & editing, Funding acquisition. **Guillermo Blanco:** Conceptualization, Investigation, Resources, Writing – review & editing, Funding acquisition. **Rafael Zardoya:** Conceptualization, Writing – review & editing. **Borja Milá:** Conceptualization, Investigation, Resources, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Mitochondrial markers sequences are deposited at Gen Bank, the accession numbers are (MW460715- MW460796) for *atp8&6* and (MW460797-MW460875) for *nad2* (see Table S1 for details). SNP raw data is deposited at NCBI under the SRA data project PRJNA692563 with accession numbers (SAMN17349018 - SAMN17349101), see Table S1 for details) and the vcf datasets are deposited in *Figshare* (https://doi.org/10.6084/m9.figshare.13562582). The *Fringilla*

coelebs reference genome is deposited at NCBI (Accession number: JADKPM000000000.1).

Acknowledgement

We are grateful to Félix Medina and the Cabildo of La Palma for logistical support during our work on La Palma. Garajonay National Park provided logistical support and permits for sampling within this protected area. The Cabildos of Tenerife and Gran Canaria provided the permits for working within the network of protected areas. Regional governments of the Canaries, Azores and Madeira issued the necessary research and sampling permits for collecting blood samples. Óscar Frías provided excellent help with fieldwork in Iberia, Canaries and Ceuta. Thanks to Pepe Navarrete for greatly facilitating field work and logistics while sampling in Ceuta. Andrea S. Messeguer provided useful advice on the biogeographical analysis. The research was funded by Spain's Ministry of Science grants CGL2015-66381P, PGC-2018-098897-B-I00, and PGC2018-097575-B-I00. MR was supported by a doctoral fellowship from the Spanish Ministry of Education, Culture, and Sport (FPU16/05724), and JCI was funded by a GRUPIN research grant from the Regional Government of Asturias (Ref.: IDI/2018/000151).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107291.

References

Abolins-Abols, M., Kornobis, E., Ribeca, P., Wakamatsu, K., Peterson, M.P., Ketterson, E., Milá, B., 2018. A role for differential gene regulation in the rapid diversification of melanic plumage coloration in the dark-eyed junco (*Junco hyemalis*). Mol. Ecol. 27, 4501–4515. https://doi.org/10.1111/mec.14878.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19 (9), 1655–1664. https://doi.org/10.1101/gr.094052.109.

Baker, A.J., Dennison, M.D., Lynch, A., Le Grand, G., 1990. Genetic divergence in peripherally isolated populations of chaffinches in the Atlantic islands. Evolution 44 (4), 981–999. https://doi.org/10.1111/evo.1990.44.issue-410.1111/j.1558-5646.1990.tb03819.x.

Beheregaray, L.B., Gibbs, J.P., Havill, N., Fritts, T.H., Powell, J.R., Caccone, A., 2004. Giant tortoises are not so slow: rapid diversification and biogeographic consensus in the Galápagos. Proc. Natl. Acad. Sci. 101, 6514–6519. https://doi.org/10.1073/pnas.0400393101.

Blanco, G., Laiolo, P., Fargallo, J.A., 2014. Linking environmental stress, feeding-shifts and the 'island syndrome': a nutritional challenge hypothesis. Popul. Ecol. 56 (1), 203–216. https://doi.org/10.1007/s10144-013-0404-3.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C.-H., Xie, D., Zhang, C., Stadler, T., Drummond, A.J., Pertea, M., 2018. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15 (4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650.

Campagna, L., Repenning, M., Silveira, L.F., Fontana, C.S., Tubaro, P.L., Lovette, I.J., 2017. Repeated divergent selection on pigmentation genes in a rapid finch radiation. Scientific Adv. 3 (5), e1602404. https://doi.org/10.1126/sciadv.1602404.

Campagna, L., Benites, P., Lougheed, S.C., Lijtmaer, D.A., Di Giacomo, A.S., Eaton, M.D., Tubaro, P.L., 2012. Rapid phenotypic evolution during incipient speciation in a continental avian radiation. Proc. R. Soc. B: Biol. Sci. 279 (1734), 1847–1856. https://doi.org/10.1098/rspb.2011.2170.

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

Clegg, S.M., Degnan, S.M., Kikkawa, J., Moritz, C., Estoup, A., Owens, I.P.F., 2002. Genetic consequences of sequential founder events by an island-colonizing bird. Proc. Natl. Acad. Sci. 99 (12), 8127–8132. https://doi.org/10.1073/pnas.102583399.

Clement, P., 2018. Common Chaffinch (Fringilla coelebs). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.), Handbook of the Birds of the World Alive. Lynx Edicions, Barcelona.

Clement, P., 2020. Common Chaffinch (Fringilla coelebs), version 1.0. In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana (Eds.), Birds of the World. Cornell Lab of Ornithology, Ithaca, NY, USA. DOI: 10.2173/bow.comcha.01.

Cowie, R.H., Holland, B.S., 2006. Dispersal is fundamental to biogeography and the evolution of biodiversity on oceanic islands. J. Biogeogr. 33 (2), 193–198. https://doi.org/10.1111/jbi.2006.33.issue-210.1111/j.1365-2699.2005.01383.x.

Delmore, K.E., Illera, J.C., Pérez-Tris, J., Segelbacher, G., Lugo Ramos, J.S., Durieux, G., Ishigohoka, J., Liedvogel, M., 2020. The evolutionary history and genomics of European blackcap migration. Elife 9, e54462. https://doi.org/10.7554/eLife.54462.

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

Queiroz, A., 2005. The resurrection of oceanic dispersal in historical biogeography. Trends Ecol. Evol. 20 (2), 68–73. https://doi.org/10.1016/j.tree.2004.11.006.

De Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. https://doi.org/10.1080/10635150701701083.

del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., Kirwan, G. (Eds.), 2020. Handbook of the Birds of the World Alive. Lynx Edicions, Barcelona (retrieved from http://www.hbw.com/ on 01 Abril 2020).

Dennison, Michael D., Baker, Allan J., 1991. Morphometric variability in continental and Atlantic island populations of chaffinches (Fringilla coelebs). Evolution 45 (1), 29–39. https://doi.org/10.1111/evo.1991.45.issue-110.1111/j.1558-5646.1991.tb05263.x.

Drummond, A.J., Bouckaert, R.R., 2015. Bayesian evolutionary analysis with BEAST. Cambridge University Press.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. https://doi.org/10.1093/molbev/mss075.

Eberhard, Jessica R., Bermingham, Eldredge, 2005. Phylogeny and comparative biogeography of *Pionopsitta* parrots and Pteroglossus toucans. Mol. Phylogenet. Evol. 36 (2), 288–304. https://doi.org/10.1016/j.ympev.2005.01.022.

Elshire, Robert J., Glaubitz, Jeffrey C., Sun, Qi, Poland, Jesse A., Kawamoto, Ken, Buckler, Edward S., Mitchell, Sharon E., Orban, Laszlo, 2011. A robust, simple genotyping-by-sequen- cing (GBS) approach for high diversity species. PLoS ONE 6 (5), e19379. https://doi.org/10.1371/journal.pone.0019379.

Emerson, Brent C., Oromí, Pedro, Hewitt, Godfrey M., 1999. MtDNA phylogeography and recent intra-island diversification among Canary Island Calathus beetles. Mol. Phylogenet. Evol. 13 (1), 149–158. https://doi.org/10.1006/mpev.1999.0644.

Emerson, B.C., 2002. Evolution on oceanic islands: molecular phylogenetic approaches to understanding pattern and process. Mol. Ecol. 11, 951–966.

Excoffier, Laurent, Lischer, Heidi E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10 (3), 564–567. https://doi.org/10.1111/men.2010.10.issue-310.1111/j.1755-0998.2010.02847.x.

Foll, M., Gaggiotti, O., 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180, 977–993. https://doi.org/10.1534/genetics.108.092221.

Freed, Leonard A., Conant, Sheila, Fleischer, Robert C., 1987. Evolutionary ecology and radiation of Hawaiian passerine birds. Trends Ecol. Evol. 2 (7), 196–203. https://doi.org/10.1016/0169-5347(87)90020-6.

Friis, G., Milá, B., 2020. Change in sexual signaling traits outruns morphological divergence in a recent avian radiation across an ecological gradient. J. Evol. Biol. 33, 1276–1293. https://doi.org/10.1111/jeb.13671.

Friis, Guillermo, Aleixandre, Pau, Rodríguez-Estrella, Ricardo, Navarro-Sigüenza, Adolfo G., Milá, Borja, 2016. Rapid postglacial diversification and long-term stasis within the songbird genus *Junco*: phylogeographic and phylogenomic evidence. Mol. Ecol. 25 (24), 6175–6195. https://doi.org/10.1111/mec.2016.25.issue-2410.1111/mec.13911.

Fu, Y.X., 1997. Statistical tests of neutrality of mutations.

Felicísimo, Ángel M., Muñoz, Jesús, González-Solis, Jacob, Figuerola, Jordi, 2008. Ocean surface winds drive dynamics of transoceanic aerial movements. PLoS ONE 3 (8), e2928. https://doi.org/10.1371/journal.pone.0002928.

Fernández-Palacios, J.M., 2009. 9360 Laurisilvas macaronésicas (Laurus, Ocotea). In: VVAA (Ed.), against population growth, hitchhiking and background selection. Genetics, 147, pp. 915–925.

Funk, V.A., Wagner, W.L., 1995. Biogeographic patterns in the Hawaiian Islands. In: Wagner, W.L., Funk, V.L. (Eds.), Hawaiian biogeography: evolution on a hot spot archipelago. Smithsonian Institution Press, Washington, DC, pp. 379–419.

Grabherr, M.G., 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 (9), 1145–1151. https://doi.org/10.1093/bioinformatics/btq102.

García-Verdugo, Carlos, Caujapé-Castells, Juli, Illera, Juan Carlos, Mairal, Mario, Patiño, Jairo, Reyes-Betancort, Alfredo, Scholz, Stephan, 2019. Pleistocene extinctions as drivers of biogeographical patterns on the easternmost Canary Islands. J. Biogeogr. 46 (5), 845–859. https://doi.org/10.1111/jbi.2019.46.issue-510.1111/jbi.13563.

Gill, Frank B., 2014. Species taxonomy of birds: which null hypothesis? The Auk: Ornithological Adv. 131 (2), 150–161. https://doi.org/10.1642/AUK-13-206.1.

Grant, P.R., 1979. Evolution of the chaffinch, *Fringilla coelebs*, on the Atlantic Islands. Biol. J. Linn. Soc. 11, 301–332. https://doi.org/10.1111/j.1095-8312.1979.tb00042.x.

Grant, P.R., 1980. Colonization of Atlantic islands by chaffinches (*Fringilla* spp.). Bonner Zoologische Beiträge 31, 311.

Grant, P.R., Grant, B. Rosemary, 2008. How and why species multiply: the radiation of Darwin's finches. Princeton University Press.

Gillespie, Rosemary G., Baldwin, Bruce G., Waters, Jonathan M., Fraser, Ceridwen I., Nikula, Raisa, Roderick, George K., 2012. Long-distance dispersal: a framework for hypothesis testing. Trends Ecol. Evol. 27 (1), 47–56. https://doi.org/10.1016/j.tree.2011.08.009.

Gillespie, R.G., Bennett, G.M., De Meester, L., Feder, J.L., Fleischer, R.C., Harmon, L.J., Hendry, a. P., Knope, M.L., Mallet, J., Martin, C., Parent, C.E., Patton, A.H., Pfennig, K.S., Rubinoff, D., Schluter, D., Seehausen, O., Shaw, K.L., Stacy, E., Stervander, M., Stroud, J.T., Wagner, C. & Wogan, G.O.U., 2020. Comparing adaptive radiations across space, time, and taxa. J. Heredity, 111, 1–20. https://doi.org/10.1093/jhered/esz064.

Gosselin, T., 2019. radiator: RADseq Data Exploration, Manipulation and Visualization using R. DOI: 10.5281/zenodo.1475182, https://thierrygosselin.github.io/radiator/.

Hackett, Shannon J., 1996. Molecular phylogenetics and biogeography of tanagers in the genusRamphocelus (Aves). Mol. Phylogenet. Evol. 5 (2), 368–382. https://doi.org/10.1006/mpev.1996.0032.

Handbook of the Birds of the World and BirdLife International, 2019. Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 4. Available at: http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBW-BirdLife_Checklist_v4_Dec19.zip.

Illera, J.C., Rando, J.C., Richardson, D.S., Emerson, B.C., 2012. Age, origins and extinctions of the avifauna of Macaronesia: a synthesis of phylogenetic and fossil information. Quat. Sci. Rev. 50, 14–22. https://doi.org/10.1016/j.quascirev.2012.07.013.

Illera, J.C., Spurgin, L.G., Rodriguez-Exposito, E., Nogales, M., Rando, J.C., 2016. What are we learning about speciation and extinction from the Canary Islands? Ardeola 63, 15–34. https://doi.org/10.13157/arla.63.1.2016.rp1.

Illera, Juan Carlos, Rando, Juan Carlos, Rodriguez-Exposito, Eduardo, Hernández, Mariano, Claramunt, Santiago, Martín, Aurelio, 2018. Acoustic, genetic, and morphological analyses of the Canarian common chaffinch complex *Fringilla coelebs* ssp. reveals cryptic diversification. J. Avian Biol. 49 (12) https://doi.org/10.1111/jav.2018.v49.i1210.1111/jav.01885.

James, J.E., Lanfear, R., Eyre-Walker, A., 2016. Molecular evolutionary consequences of island colonization. Genome Biology Evolution, 8, 1876–1888. DOI: 10.1093/gbe/evw120.

Juan, Carlos, Emerson, Brent C., Oromí, Pedro, Hewitt, Godfrey M., 2000. Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. Trends Ecol. Evol. 15 (3), 104–109. https://doi.org/10.1016/S0169-5347(99)01776-0.

Kapli, P., Lutteropp, S., Zhang, J., Kobert, K., Pavlidis, P., Stamatakis, A., Flouri, T., 2017. Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. Bioinformatics 33, 1630–1638. https://doi.org/10.1093/bioinformatics/btx025.

Kondraskov, Paulina, Schütz, Nicole, Schüßler, Christina, de Sequeira, Miguel Menezes, Guerra, Arnoldo Santos, Caujapé-Castells, Juli, Jaén-Molina, Ruth, Marrero-Rodríguez, Águedo, Koch, Marcus A., Linder, Peter, Kovar-Eder, Johanna, Thiv, Mike, Buerki, Sven, 2015. Biogeography of Mediterranean hotspot biodiversity: re-evaluating the'Tertiary relict'hypothesis of Macaronesian laurel forests. PLoS ONE 10 (7), e0132091. https://doi.org/10.1371/journal.pone.0132091.

Kopelman, N.M., Mayzel, J., Jakobsson, M., Rosenberg, N.A., Mayrose, I., 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15, 1179–1191. https://doi.org/10.1111/1755-0998.12387.

Kozak, K.M., McMillan, O., Joron, M., Jiggins, C.D., 2018. Genome-wide admixture is common across the *Heliconius* radiation. BioRxiv 414201. https://doi.org/10.1101/414201.

Lamichhaney, Sangeet, Berglund, Jonas, Almén, Markus Sällman, Maqbool, Khurram, Grabherr, Manfred, Martinez-Barrio, Alvaro, Promerová, Marta, Rubin, Carl-Johan, Wang, Chao, Zamani, Neda, Grant, B. Rosemary, Grant, Peter R., Webster, Matthew T., Andersson, Leif, 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518 (7539), 371–375. https://doi.org/10.1038/nature14181.

Lachlan, RobertF., Verzijden, MachteldN., Bernard, CarolineS., Jonker, PeterPaul, Koese, Bram, Jaarsma, Shirley, Spoor, Willemijn, Slater, PeterJ.B., ten Cate, Carel, 2013. The progressive loss of syntactical structure in bird song along an island colonization chain. Curr. Biol. 23 (19), 1896–1901. https://doi.org/10.1016/j.cub.2013.07.057.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. https://doi.org/10.1093/molbev/msw260.

Lawson, Daniel John, Hellenthal, Garrett, Myers, Simon, Falush, Daniel, Copenhaver, Gregory P., 2012. Inference of population structure using dense haplotype data. PLoS Genet. 8 (1), e1002453. https://doi.org/10.1371/journal.pgen.1002453.

Lerner, HeatherR.L., Meyer, Matthias, James, HelenF., Hofreiter, Michael, Fleischer, RobertC., 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Curr. Biol. 21 (21), 1838–1844. https://doi.org/10.1016/j.cub.2011.09.039.

Leroy, Thibault, Rousselle, Marjolaine, Tilak, Marie-Ka, Caizergues, Aude E., Scornavacca, Céline, Recuerda, María, Fuchs, Jérôme, Illera, Juan Carlos, De Swardt, Dawie H., Blanco, Guillermo, Thébaud, Christophe, Milá, Borja, Nabholz, Benoit, 2021. Island songbirds as windows into evolution in small populations. Curr. Biol. 31 (6), 1303–1310.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25 (14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

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

Losos, Jonathan B., Ricklefs, Robert E., 2009. Adaptation and diversification on islands. Nature 457 (7231), 830–836. https://doi.org/10.1038/nature07893.

Lynch, Alejandro, Baker, Allan J., 1994. A population memetics approach to cultural evolution in chaffinch song: differentiation among populations. Evolution 48 (2), 351–359. https://doi.org/10.1111/evo.1994.48.issue-210.1111/j.1558-5646.1994.tb01316.x.

MacArthur, R.H., Wilson, E.O., 1967. The theory of island biogeography. Princeton University Press, Princeton, N.J.

Malinsky, M., Trucchi, E., Lawson, D.J., Falush, D., 2018. RADpainter and fineRADstructure: population inference from RADseq data. Mol. Biology and Evolution, 35, 1284-1290. DOI: 10.1093/molbev/msy023.

Marshall, H.Dawn, Baker, Allan J., 1999. Colonization history of Atlantic island common chaffinches (*Fringilla coelebs*) revealed by mitochondrial DNA. Mol. Phylogenet. Evol. 11 (2), 201–212.

Martín, José L., 2009. Are the IUCN standard home-range thresholds for species a good indicator to prioritise conservation urgency in small islands? A case study in the Canary Islands (Spain). J. Nature Conserv. 17 (2), 87–98. https://doi.org/10.1016/j.jnc.2008.10.001.

Martín, A., Lorenzo, J.A., 2001. Aves del archipiélago canario. Francisco Lemus.Editor, La Laguna, Spain.

Matzke, N.J., 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Front. Biogeogr. 5 https://doi.org/10.21425/F5FBG19694.

Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63, 951–970. https://doi.org/10.1093/sysbio/syu056.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. http://www.genome.org/cgi/doi/10.1101/gr.107524.110.

Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L., Seehausen, O., 2018. Genomics of parallel ecological speciation in Lake Victoria cichlids. Mol. Biol. Evol. 35, 1489–1506. DOI: 10.1093/molbev/msy051.

Miller, M., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. New Orleans, LA, USA: Gateway Computing Environments Workshop (GCE). https://doi.org/10.1109/GCE.2010.5676129.

Morinha, Francisco, Milá, Borja, Dávila, José A., Fargallo, Juan A., Potti, Jaime, Blanco, Guillermo, 2020. The ghost of connections past: A role for mainland vicariance in the isolation of an insular population of the red-billed chough (Aves: Corvidae). J. Biogeogr. 47 (12), 2567–2583. https://doi.org/10.1111/jbi.v47.1210.1111/jbi.13977.

Nathan, R., 2006. Long-distance dispersal of plants. Science 313 (5788), 786–788. https://doi.org/10.1126/science:1124975.

Ödeen, A., Björklund, M., 2003. Dynamics in the evolution of sexual traits: losses and gains, radiation and convergence in yellow wagtails (*Motacilla flava*). Mol. Ecol. 12 (8), 2113–2130. https://doi.org/10.1046/j.1365-294X.2003.01883.x.

Padial, J.M., Miralles, A., De la Riva, I., Vences, M., 2010. The integrative future of taxonomy. Front. Zool. 7, 1–14. https://doi.org/10.1186/1742-9994-7-16.

Patiño, J., Whittaker, R.J., Borges, P.A., Fernández-Palacios, J.M., Ah-Peng, C., Araújo, M.B., Ávila, S.P., Cardoso, P., Cornuault, J., de Boer, E.J., de Nascimento, L., Gil, A., González-Castro, A., Gruner, D.S., Heleno, R., Hortal, J., Illera, J.C., Kaiser-Bunbury, C.N., Matthews, T.J., Papadopoulou, A., Pettorelli, N., Price, J.P., Santos, A.M.C., Steinbauer, M.J., Triantis, K.A., Valente, L., Vargas, P., Weigelt, P., Emerson, B.C., 2017. A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island Biogeography. J. Biogeogr. 44, 963–983. https://doi.org/10.1111/jbi.12986.

Perktaş, Utku, Peterson, A. Townsend, Dyer, Dale, 2017. Integrating morphology, phylogeography, and ecological niche modeling to explore population differentiation in North African Common Chaffinches. J. Ornithol. 158 (1), 1–13.

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

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1. 6. Computer program and documentation distributed by the author, website http://beast.bio.ed.ac.uk/Tracer.

Rambaut, A., 2017. FigTree-version 1.4. 3, a graphical viewer of phylogenetic trees. Computer program distributed by the author, website: http://tree.bio.ed.ac.uk/software/figtree.

Recuerda, M., Vizueta, J., Cuevas-Caballé, C., Blanco, G., Rozas, J., Milá, B., 2021. Chromosome-level genome assembly of the common chaffinch (Aves: Fringilla coelebs): a valuable resource for evolutionary biology. Genome Biol. Evol. 13 (4), evab034. https://doi.org/10.1093/gbe/evab034.

Ree, R.H., Sanmartín, I., 2018. Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. 45, 741–749. https://doi.org/10.1111/jbi.13173.

Rodrigues, Pedro, Lopes, Ricardo Jorge, Reis, Sandra, Resendes, Roberto, Ramos, Jaime Albino, Tristão da Cunha, Regina, 2014. Genetic diversity and morphological variation of the common chaffinch *Fringilla coelebs* in the Azores. J. Avian Biol. 45 (2), 167–178. https://doi.org/10.1111/jav.2014.45.issue-210.1111/j.1600-048X.2013.00229.x.

Rundell, Rebecca J., Price, Trevor D., 2009. Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. Trends Ecol. Evol. 24 (7), 394–399. https://doi.org/10.1016/j.tree.2009.02.007.

Sackton, Timothy B., Grayson, Phil, Cloutier, Alison, Hu, Zhirui, Liu, Jun S., Wheeler, Nicole E., Gardner, Paul P., Clarke, Julia A., Baker, Allan J., Clamp, Michele, Edwards, Scott V., 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. Science 364 (6435), 74–78. https://doi.org/10.1126/science:aat7244.

Salzburger, W., Ewing, G.B., Von Haeseler, A., 2011. The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. Mol. Ecol. 20, 1952–1963. https://doi.org/10.1111/j.1365-294X.2011.05066.x.

Samarasin-Dissanayake, P., 2010. Population differentiation, historical demography and evolutionary relationships among widespread common chaffinch populations (Fringilla coelebs ssp.). Master's Thesis. University of Toronto, Canada.

Sangster, George, 2013. The application of species criteria in avian taxonomy and its implications for the debate over species concepts. Biol. Rev. 89 (1), 199–214. https://doi.org/10.1111/brv.12051.

Sangster, George, Rodríguez-Godoy, Felipe, Roselaar, C.S., Robb, Magnus S., Luksenburg, Jolanda A., 2016. Integrative taxonomy reveals Europe's rarest songbird species, the Gran Canaria blue chaffinch *Fringilla polatzeki*. J. Avian Biol. 47 (2), 159–166. https://doi.org/10.1111/jav.00825.

Sanmartín, Isabel, van der Mark, Paul, Ronquist, Fredrik, 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. J. Biogeogr. 35 (3), 428–449. https://doi.org/10.1111/jbi.2008.35.issue-310.1111/j.1365-2699.2008.01885.x.

Schluter, D., 2000. The ecology of adaptive radiation. Oxford University Press, Oxford (UK).

Sequeira, A.S., Lanteri, A.A., Albelo, L.R., Bhattacharya, S., Sijapati, M., 2008. Colonization history, ecological shifts and diversification in the evolution of endemic Galápagos weevils. Mol. Ecol. 17, 1089–1107. https://doi.org/10.1111/j.1365-294X.2007.03642.x.

Shirihai, H., Svensson, L., 2018. Handbook of Western Palearctic Birds, Volume 1: Passerines: Larks to Warblers. Bloomsbury Publishing.

Sorenson, Michael D., Ast, Jennifer C., Dimcheff, Derek E., Yuri, Tamaki, Mindell, David P., 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol. Phylogenet. Evol. 12 (2), 105–114. https://doi.org/10.1006/mpev.1998.0602.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Systematic Biol. 57, 758–771. https://doi.org/10.1080/10635150802429642.

Stange, M., Sánchez-Villagra, M.R., Salzburger, W., Matschiner, M., 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. Systematic Biol. 67, 681–699. https://doi.org/10.1093/sysbio/syy006.

Stervander, Martin, Illera, Juan Carlos, Kvist, Laura, Barbosa, Pedro, Keehnen, Naomi P., Pruisscher, Peter, Bensch, Staffan, Hansson, Bengt, 2015. Disentangling the complex evolutionary history of the Western Palearctic blue tits (*Cyanistes spp.*)–phylogenomic analyses suggest radiation by multiple colonization events and subsequent isolation. Mol. Ecol. 24 (10), 2477–2494. https://doi.org/10.1111/mec.13145.

Suárez, Nicolás M., Betancor, Eva, Klassert, Tilman E., Almeida, Teresa, Hernández, Mariano, Pestano, José J., 2009. Phylogeography and genetic structure of the Canarian common chaffinch (*Fringilla coelebs*) inferred with mtDNA and microsatellite loci. Mol. Phylogenet. Evol. 53 (2), 556–564. https://doi.org/10.1016/j.ympev.2009.07.018.

Svensson, L., 2015. A new North African subspecies of Common Chaffinch *Fringilla coelebs*. Bulletin of the British Ornithologists' Club 135, 69–76.

Thornton, I., 2007. Island colonization: the origin and development of island communities. Cambridge University Press.

Tobias, J.A., Seddon, N., Spottiswoode, C.N., Pilgrim, J.D., Fishpool, L.D., Collar, N.J., 2010. Quantitative criteria for species delimitation. Ibis 152, 724–746. https://doi.org/10.1111/j.1474-919X.2010.01051.x.

Tobias, J.A., Donald, P.F., Martin, R.W., Butchart, S.H., Collar, N.J., 2021. Performance of a points-based scoring system for assessing species limits in birds. Auk 138, ukab016. https://doi.org/10.1093/ornithology/ukab016.

Valente, Luis, Illera, Juan Carlos, Havenstein, Katja, Pallien, Tamara, Etienne, Rampal S., Tiedemann, Ralph, 2017. Equilibrium bird species diversity in Atlantic islands. Curr. Biol. 27 (11), 1660–1666.e5. https://doi.org/10.1016/j.cub.2017.04.053.

Valente, Luis, Phillimore, Albert B., Melo, Martim, Warren, Ben H., Clegg, Sonya M., Havenstein, Katja, Tiedemann, Ralph, Illera, Juan Carlos, Thébaud, Christophe, Aschenbach, Tina, Etienne, Rampal S., 2020. A simple dynamic model explains the diversity of island birds worldwide. Nature 579 (7797), 92–96. https://doi.org/10.1038/s41586-020-2022-5.

VanderWerf, Eric A., Young, Lindsay C., Yeung, Norine W., Carlon, David B., 2010. Stepping stone speciation in Hawaii's flycatchers: Molecular divergence supports new island endemics within the elepaio. Conserv. Genet. 11 (4), 1283–1298.

Warren, Ben H., Simberloff, Daniel, Ricklefs, Robert E., Aguilée, Robin, Condamine, Fabien L., Gravel, Dominique, Morlon, Hélène, Mouquet, Nicolas, Rosindell, James, Casquet, Juliane, Conti, Elena, Cornuault, Josselin, Fernández-Palacios, José María, Hengl, Tomislav, Norder, Sietze J., Rijsdijk, Kenneth F., Sanmartín, Isabel, Strasberg, Dominique, Triantis, Kostas A., Valente, Luis M., Whittaker, Robert J., Gillespie, Rosemary G., Emerson, Brent C., Thébaud, Christophe, Courchamp, Franck, 2015. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecol. Lett. 18 (2), 200–217. https://doi.org/10.1111/ele.2015.18.issue-210.1111/ele.12398.

Whittaker, R.J., Araújo, M.B., Jepson, P., Ladle, R.J., Watson, J.E., Willis, K.J., 2005. Conservation biogeography: assessment and prospect. Divers. Distrib. 11, 3–23. https://doi.org/10.1111/j.1366-9516.2005.00143.x.

Whittaker, R.J., Fernández-Palacios, J.M., 2007. Island biogeography: ecology, evolution, and conservation. Oxford University Press.

Whittaker, Robert J., Fernández-Palacios, José María, Matthews, Thomas J., Borregaard, Michael K., Triantis, Kostas A., 2017. Island biogeography: taking the long view of nature's laboratories. Science 357 (6354), eaam8326. https://doi.org/10.1126/science:aam8326.

Winker, K., 2010a. Chapter 1: Subspecies represent geographically partitioned variation, a gold mine of evolutionary biology, and a challenge for conservation. Ornithol. Monogr., 67, 6–23. DOI: 10.1525/om.2010.67.1.6.

Winker, K., 2010b. Is it a species? Ibis 152, 679–682.

Winker, K., 2021. An overview of speciation and species limits in birds. Auk 138, ukab006. https://doi.org/10.1093/ornithology/ukab006.

Wright, Natalie A., Steadman, David W., Witt, Christopher C., 2016. Predictable evolution toward flightlessness in volant island birds. Proc. Natl. Acad. Sci. 113 (17), 4765–4770. https://doi.org/10.1073/pnas.1522931113.

Yates, Andrew, Akanni, Wasiu, Amode, M. Ridwan, Barrell, Daniel, Billis, Konstantinos, Carvalho-Silva, Denise, Cummins, Carla, Clapham, Peter, Fitzgerald, Stephen, Gil, Laurent, Girón, Carlos García, Gordon, Leo, Hourlier, Thibaut, Hunt, Sarah E., Janacek, Sophie H., Johnson, Nathan, Juettemann, Thomas, Keenan, Stephen, Lavidas, Ilias, Martin, Fergal J., Maurel, Thomas, McLaren, William, Murphy, Daniel N., Nag, Rishi, Nuhn, Michael, Parker, Anne, Patricio, Mateus, Pignatelli, Miguel, Rahtz, Matthew, Riat, Harpreet Singh, Sheppard, Daniel, Taylor, Kieron, Thormann, Anja, Vullo, Alessandro, Wilder, Steven P., Zadissa, Amonida, Birney, Ewan, Harrow, Jennifer, Muffato, Matthieu, Perry, Emily, Ruffier, Magali, Spudich, Giulietta, Trevanion, Stephen J., Cunningham, Fiona, Aken, Bronwen L., Zerbino, Daniel R., Flicek, Paul, 2016. ensembl 2016. Nucleic Acids Res. 44 (D1), D710–D716. https://doi.org/10.1093/nar/gkv1157.

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

Chromosome-Level Genome Assembly of the Common Chaffinch (Aves: *Fringilla coelebs*): A Valuable Resource for Evolutionary Biology

María Recuerda [1,*†], Joel Vizueta [1,2,†], Cristian Cuevas-Caballé[2], Guillermo Blanco [1], Julio Rozas [2], and Borja Milá [1]

[1]National Museum of Natural Sciences, Spanish National Research Council (CSIC), Madrid, Spain

[2]Departament de Genètica, Microbiologia i Estadística, Facultat de Biologia and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Barcelona, Spain

†These authors contributed equally to this work.

*Corresponding author: E-mail: mariarecuerdacarrasco@gmail.com.

Accepted: 16 February 2021

Abstract

The common chaffinch, *Fringilla coelebs*, is one of the most common, widespread, and well-studied passerines in Europe, with a broad distribution encompassing Western Europe and parts of Asia, North Africa, and the Macaronesian archipelagos. We present a high-quality genome assembly of the common chaffinch generated using Illumina shotgun sequencing in combination with Chicago and Hi-C libraries. The final genome is a 994.87-Mb chromosome-level assembly, with 98% of the sequence data located in chromosome scaffolds and a N50 statistic of 69.73 Mb. Our genome assembly shows high completeness, with a complete BUSCO score of 93.9% using the avian data set. Around 7.8% of the genome contains interspersed repetitive elements. The structural annotation yielded 17,703 genes, 86.5% of which have a functional annotation, including 7,827 complete universal single-copy orthologs out of 8,338 genes represented in the BUSCO avian data set. This new annotated genome assembly will be a valuable resource as a reference for comparative and population genomic analyses of passerine, avian, and vertebrate evolution.

Key words: common chaffinch, *Fringilla coelebs*, reference genome, whole genome assembly.

Significance

High-quality reference genomes of wild, nonmodel species are very useful tools to understand how organisms evolve. If genomes are annotated, so that the specific genes are identified, we can make progress toward associating specific physical or behavioral traits with the genes that code for them, and thus further understand the evolutionary process. Here, we provide a high-quality, annotated genome of the common chaffinch, a common and widespread Eurasian finch, that will be a useful resource in studies related to evolution, phylogenomics, biogeography, and adaptation genomics, among others.

Introduction

The decreasing costs of DNA sequencing, along with advances in computational genomics, are promoting a rapid increase in the availability of high-quality reference genomes of nonmodel species, which greatly improves our capacity to address a range of biological questions from a genomic perspective. Among them, the correct annotation of protein-coding genes in whole genomes allows to identify new genes involved in the process of evolutionary adaptation and provides a better understanding of the evolutionary mechanisms involved in the speciation process. Avian genomes are

particularly suited for studying the molecular basis of speciation as they have a relatively simple architecture and are among the smallest within amniotes, ranging from 0.91 to 1.3 Gb (Gregory 2002). In the last decade, the number of bird reference genomes has increased dramatically (e.g. Dalloul et al. 2010; Warren et al. 2010; Zhang et al. 2012; Jarvis et al. 2014; Poelstra et al. 2014; Frankl-Vilches et al. 2015; Friis et al. 2018; Louha et al. 2020; Peñalba et al. 2020; Ducrest et al. 2020, Wang et al. 2020), providing major scientific breakthroughs in phylogenetics (Alström et al. 2018; Braun et al. 2019; Jarvis et al. 2015), comparative genomics (Zhang et al. 2014, Feng et al. 2020), adaptation genomics (Wirthlin et al. 2014; Lawson and Petren 2017), and genomic architecture (Poelstra et al. 2014; Vijay et al. 2016), among others. Moreover, the Ten-Thousand Bird Genomes (B10K) consortium has generated and analyzed over 300 avian genomes from 92.4% of bird families, providing an unprecedent genomic resource for avian comparative studies (Zhang et al. 2015, Feng et al. 2020).

The common chaffinch (Aves, Passeriformes, Fringillidae, *Fringilla coelebs*) is a widely distributed species, ranging from across Eurasia to the north of Africa, and has colonized three Macaronesian archipelagos in the Atlantic Ocean (Azores, Madeira, and the Canary Islands) (Collar et al. 2020). With about 15 currently recognized subspecies, the common chaffinch is an ideal system for testing hypotheses on the evolutionary process given its distribution across the continent and the colonization of several oceanic islands, recognized as excellent natural laboratories for studying evolution (Brown et al. 2013). Island systems have inspired the development of biogeographical theories (MacArthur and Wilson 1967) and are of central importance for understanding the role of area and isolation in colonization, extinction, and speciation rates (Valente et al. 2020), which are processes influencing global patterns of species richness (Losos and Schluter 2000). Species that have colonized insular environments, like the common chaffinch, are also excellent systems for the study of demographic events, such as bottlenecks leading to small effective population size (N_e) (Leroy et al. 2021), or the roles of drift and selection in the divergence process (Barton 1996). The common chaffinch has been intensively studied using molecular tools, so that the availability of a reference genome represents a valuable resource to improve our understanding of avian evolution, biogeography, and demography (Illera et al. 2018).

Results and Discussion

Assembly and Quality Control

The total length obtained by the HiRise software for the common chaffinch assembly was 994.87 Mb. Nevertheless, the estimate from *k*-mer metrics is 1.2 Gb. The discrepancy between these estimates could be caused by the presence of repetitive elements given the assembly strategy used, which could have been improved including long-read sequencing technologies. This final assembly consists of 3,255 scaffolds, 3,239 over 1 kb, and an N50 of 69.73 Mb (see supplementary table S1, Supplementary Material online) with a sequence coverage of 249×. The use of Chicago and Hi-C libraries provided a clear improvement in quality by increasing 917 times the scaffold N50, reducing the number of scaffolds from 38,666 to 3,255 (Supplementary table S1, see supplementary methods for details, Supplementary Material online). In fact, 98% of the total genome sequence maps in the 30 described chromosomes.

The chaffinch genome showed high synteny with the zebra finch genome (fig. 1), evidencing the completeness of the assembly, with all micro-chromosomes and the Z chromosome present in the assembly. In addition, the alignment between these genomes suggests the presence of several inversions in chromosomes 1, 1 A, 2, 3, 5, 7, 8, and 9. Several studies have documented that inversions are very common in birds (Aslam et al. 2010; Völker et al. 2010; Skinner and Griffin 2012; Zhang et al. 2014). For instance, Hooper and Price (2017) identified 319 inversions on the 9 largest autosomes combined in 81 independent clades. No putative contaminations were detected and 89.6% of the reads were mapped in the genome assembly (see Supplementary fig. S1, Supplementary Material online). The mean GC content of the assembly was 41.86% (±11 SD). The common chaffinch genome assembly included 7,832 complete copies (93.9%) out of the 8,338 BUSCO data set from avian genomes, among which 7,816 were single-copy orthologs and 16 were duplicated. Only 1.8% of the gene models were fragmented, and 4.3% were missing in the genome. These few missing gene models could represent divergent or lost genes in our species, but also could be related with putative errors during the assembly process or missing data.

Repetitive Regions

Overall, 7.82% of the genome assembly are repeats (~78 Mb), of which 85.4% are transposable elements (TEs). The most abundant TEs are LINEs (53.5%) followed by LTR (29.4%), DNA elements (4.1%), and SINEs (1.4%), with the remaining 11.6% unclassified. The rest of repeats (14.6%) contained simple repeats (75.4%), low complexity repeats (18.5%), satellites (4.2%) and small RNA (1.9%) (see supplementary table S2, Supplementary Material online). The number of repetitive regions is within the expected range in birds, which is at 4–10% of the genome (Zhang et al. 2014).

A total of 111,076 microsatellites, with motif length ranging between 2 and 20 bp, were identified in the common chaffinch genome (see supplementary fig. S3, Supplementary Material online; their genomic locations are shown in supplementary file S1 in the Figshare repository). The most common *k*-mer sizes conforming the microsatellites were 2 (68.2%), 3 (15.9%) and 4 (8.2%) (see supplementary

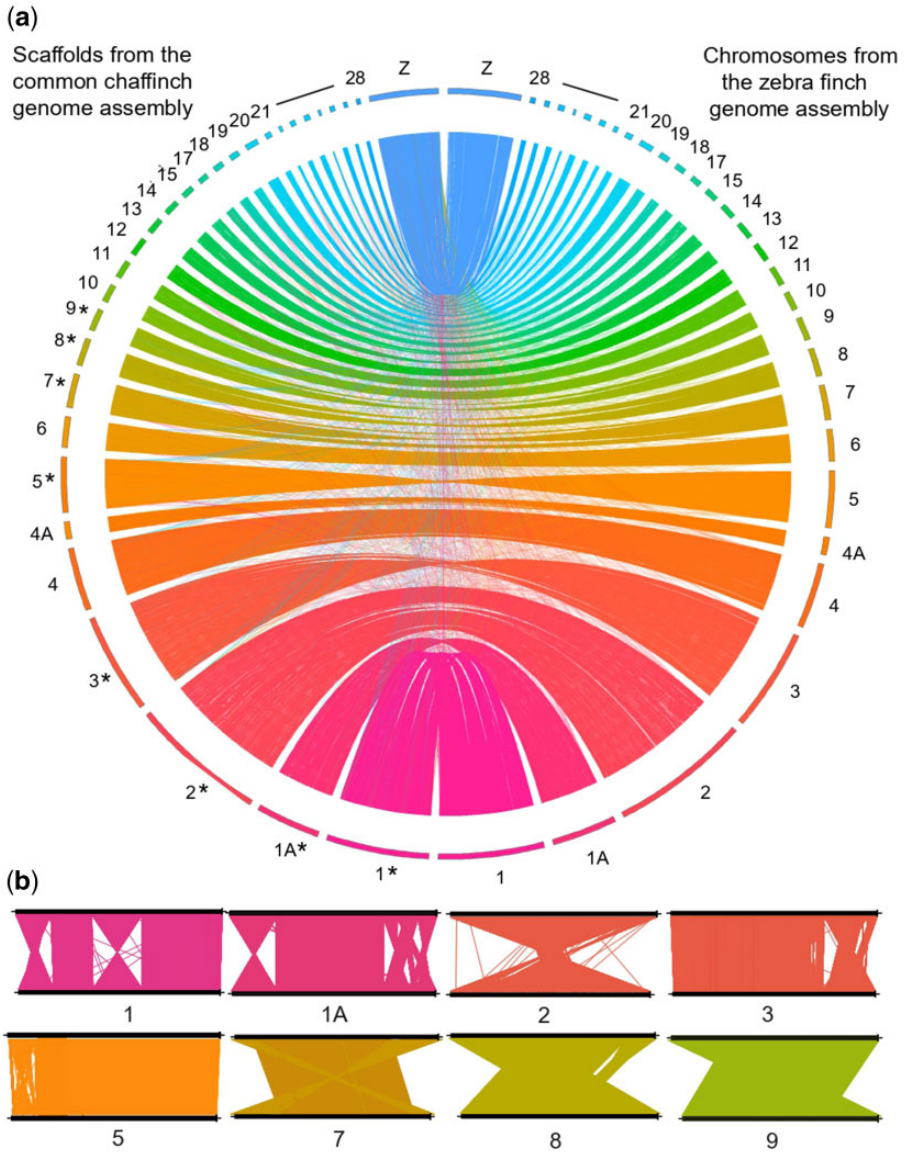


Fig. 1.—(*a*) Circos plot comparing the zebra finch (right hemisphere) and the common chaffinch (left hemisphere) genome assemblies. The common chaffinch chromosomes marked with an asterisk (*) show inversions with respect to the zebra finch assembly. (*b*) Linear synteny plots of the common chaffinch chromosomes showing inversions relative to the zebra finch generated with the R package genoPlotR (Guy et al. 2010). The zebra finch assembly (top) is compared with the common chaffinch assembly (bottom), and numbers designate specific chromosomes.

file S1). The most common length of the microsatellites was 10 bp (40.4%), followed by 12 bp (13%) and 15 bp (8.8%) (see supplementary file S1 for the length distribution of microsatellites). In addition, the number of microsatellites was positively correlated with the sequence length (Supplementary fig. S3, Supplementary Material online; see supplementary file S1 for the frequency of occurrence in every scaffold).

Gene Annotation and Function Prediction

Our annotation pipeline combining both *de novo* and homology-based predictions inferred 21,831 proteins encoded by 17,703 genes in the common chaffinch genome with a mean length of 15,818 bp (Table 1). The common chaffinch genome annotation (see supplementary file S2 in Figshare) included 7,850 complete copies (94.2%) out of the 8,338 of BUSCO avian data set used, retrieving all expected copies with a slight increase from that estimated in the unannotated genome (see above). Among the complete BUSCO genes, 7,827 were single-copy orthologs (99.7%) and 23 were duplicated (0.3%). Around 1.9% (162) of the gene models were fragmented and 3.9% showed no significant matches (326).

Over all predicted proteins, 19,458 (89.1%) provided positive BLASTP hits against the Uniprot SwissProt database, and

Table 1

Genome Statistics and Predicted ncRNAs of the *Fringilla coelebs* Genome Compared with Other Similarly Sized Avian Species (*Melospiza melodia, Taeniopygia guttata, Ficedula albicollis, Manacus vitellinus*, and *Geospiza fortis*), Modified from Louha et al. (2020).

	F. coelebs	*M. melodia*	*T. guttata*	*F. albicollis*	*M. vitellinus*	*G. fortis*
Number of genes	17,703	15,086	17,561	16,763	18,976	14,399
Mean gene length (bp)	15,818	14,457	26,458	31,394	27,847	30,164
Number of CDSs	17,703	15,086	17,561	16,763	18,976	14,399
Mean CDs length (bp)	1,679	1,325	1,677	1,942	1,929	1,766
Number of exons	221,872	131,940	171,767	189,043	190,390	164,721
Mean exon length (bp)	165	153	255	253	264	195
Mean number of exons/gene	10.16	8.67	10.25	12.22	11.51	11.41
Number of introns	200,041	116,724	153,909	171,236	171,089	149,563
Mean intron length (bp)	1,902	1,695	2,930	3,257	3,294	2,813
Total proteins	21,831					
ncRNA						
tRNA	325	267	184	179		
miRNA	140	166	302	510		
snRNA	18	16	44	32		
snoRNA	126	154	241	199		
rRNA	5	8	100	22		
lncRNA	17	20	908	1473		

19,617 (89.9%) against the annotated proteins from the zebra finch genome. In addition, InterproScan identified 18,551 (85%) specific protein-domain signatures in the predicted peptides. The combination of the annotation from these databases allowed assigning a functional annotation with GO terms to 19,425 proteins (89%) assigned to 15,309 genes (86.5%; supplementary file S3 in Figshare).

tRNAs and Other Noncoding RNA Prediction

The search by tRNAscan-SE (supplementary file S4 in Figshare) identified 325 tRNAs in the common chaffinch genome, of which 167 decode for the standard twenty amino acids. Among all the tRNAs detected, 131 presented low scores and therefore were categorized as pseudogenes (i.e. lacking tRNA-like secondary structures). There were no suppressor tRNAs, 1 had undetermined isotopes, 25 were chimeric and 15 included introns within their sequences. One of the tRNAs was predicted to code for selenocysteine (sequences and structures of the predicted tRNAS are available in File S5 in Figshare). In addition, the search against both tRNA databases (GtRNAdb and tRNAdb) yielded positive results in many other species, suggesting that tRNA prediction in our assembly was correct. Moreover, our searches using Infernal identified 354 ncRNAs, which were classified as follows: 39 CREs, 2 Ribozymes, 7 Gene, 140 miRNAs, 126 snoRNAs, 18 snRNAs, 5 rRNAs, and 17 lncRNAs (File S6 in Figshare). The number of tRNAs predicted in the common chaffinch genome is the highest when compared with other passerine species (i.e., *M. melodia*, *T. guttata*, and *F. albicollis*), but the other types of ncRNAs present similar values to the *M. melodia* genome and lower than the other two species (Table 1), probably because we applied a strict threshold to avoid an excess of false positives.

Conclusions

We provide here a high-quality assembly for the common chaffinch, a valuable resource as a reference genome to address a range of biological questions from a genomic perspective. Moreover, our annotation provides useful information to detect candidate genes involved in adaptation and divergence processes. The combination of the Chicago and shotgun sequencing with the HiRise assembly approach lead to a highly contiguous chromosome-level genome assembly. The genome assembly size was 994.87 Mb, with the 30 chromosomes accounting for 98% of it. Although the expected length of the genome was 1.2 Gb, closer to those obtained in other avian species by flow cytometry (Gregory 2002), the BUSCO analyses showed that both the assembly and structural annotation encode 93.9% and 94.2% complete copies out of the 8,338 orthologous conserved genes in avian species, respectively. This discrepancy of the genome size could be caused by the absence of large repetitive elements in the assembly. The structural annotation predicted 17,703 coding genes, with most of them (86.5%) assigned to functional annotation and GO terms.

Materials and Methods

Sample Collection and Genome Assembly

A blood sample was extracted from a common chaffinch female captured in Torreiglesias, Segovia, Spain, in 2017 and frozen immediately in liquid nitrogen. The sample was

handled by Dovetail Genomics for DNA extraction, sequencing and genome assembly using the HiRise pipeline (Putnam et al. 2016). The absence of a Z chromosome in our first assembly (GenBank assembly accession: GCA_015532645.1) led us to conduct a second assembly presented here, which includes sex-linked scaffolds used to reconstruct the chaffinch Z chromosome (see supplementary methods, Supplementary Material online, for details). Gene completeness in the chaffinch genome assembly (and in the annotated gene set) was assessed through BUSCO (Benchmarking Universal Single-Copy Orthologs) v4.0.5 (Seppey et al. 2019) by using the 8,338 single-copy orthologous genes in the Aves lineage group odb10, using chicken as the Augustus reference species.

Identification of Repetitive Regions and Gene Annotation

Repetitive regions were identified and masked prior to gene prediction. First, repeats were modelled ab initio using Repeat Modeler 1.0.11 (Smit and Hubley 2019) in scaffolds longer than 100 Kb with default options. The repeats obtained were merged with known bird repeat libraries from the RepBase database (RepBase-20181026) (Bao et al. 2015), Dfam_Consensus-20181026 and repeats from the zebra finch (obtained from B10K). The resulting repeat library was compared against the complete assembly with Repeat Masker 4.0.7 (Smit et al. 2015) and the identified regions were soft-masked. For the identification and description of microsatellites in the common chaffinch genome assembly we used GMATA v.2.01 (Wang and Wang 2016), with sequence motif length between 2 and 20 bp.

Gene prediction was conducted with BRAKER v2.1.5 (Hoff et al. 2016) and GeMoMa v1.7.1 (Keilwagen et al. 2016, 2018). First, the conserved orthologous genes from BUSCO Aves_odb10 were used as proteins from short evolutionary distance to train Augustus (Gremme et al. 2005; Stanke et al. 2006; see figure 3B in Hoff et al. 2019). The predicted proteins were combined with homology-based annotations using the zebra finch (GCF_008822105.2; Warren et al. 2010) and chicken (GCF_000002315.6; Hillier et al. 2014) annotated genes with GeMoMa pipeline, obtaining the final reported gene models. We applied a similarity-based search approach to assist the functional annotation of the chaffinch predicted proteins, using the UniProt SwissProt database, the annotated proteins from the zebra finch genome (Warren et al. 2010; UniProt Consortium 2014) and InterProScan v5.31 (Jones et al. 2014). The functional annotation, including Gene Ontology terms, was integrated from all searches providing a curated set of chaffinch coding genes (see supplementary methods, Supplementary Material online, for details).

Noncoding RNA Prediction and Identification

For the prediction and functional classification of Transfer RNAs (tRNAs) in the common chaffinch genome we used tRNAscan-SE v2.0 (Lowe and Chan 2016). The tRNA search across the genome and the identification of ncRNA (noncoding RNA) homologues was conducted using the software package Infernal v1.1.1 (Nawrocki 2014) (see supplementary methods, Supplementary Material online, for details). For comparative purposes, we added our results to those from Louha et al. (2020), which compared different genome assemblies of avian species.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This research was supported by the Spanish Ministry of Economy and Competitiveness (CGL2015-66381P to B.M. and G.B.) and the Spanish Ministry of Science and Innovation (PGC2018-098897-B-I00 from to B.M.). M.R. was supported by a doctoral fellowship from the Spanish Ministry of Education, Culture, and Sport (FPU16/05724).

Data Availability

The chaffinch genome assembly has been deposited at NCBI under BioProject PRJNA674347 with accession number JADKPM000000000, the raw data are available at SRA NCBI database with accession numbers SRR12998620–SRR12998622, and the annotation and all described data sets are publicly accessible in Figshare (https://doi.org/10.6084/m9.figshare.13296122.v3).

Literature Cited

Alström P, et al. 2018. Complete species-level phylogeny of the leaf warbler (Aves: Phylloscopidae) radiation. Mol Phylogenet Evol. 126:141–152.

Aslam ML, et al. 2010. A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes. BMC Genomics 11(1):647–611.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 6:11.

Barton NH. 1996. Natural selection and random genetic drift as causes of evolution on islands. Philos Trans R Soc Lond B Biol Sci. 351(1341):785–795.

Braun EL, Cracraft J, Houde P. 2019. Resolving the avian tree of life from top to bottom: the promise and potential boundaries of the phylogenomic era. In: Kraus R, editor. Avian genomics in ecology and evolution. Cham (Switzerland): Springer. p. 151–210.

Brown RM, et al. 2013. Evolutionary processes of diversification in a model island archipelago. Annu Rev Ecol Evol Syst. 44(1):411–435.

Collar N, Newton I, Bonan A. 2020. Finches (*Fringillidae*). In: del Hoyo J, Elliott A, Sargatal J, Christie DA, de Juana E, editors. Handbook of the birds of the world alive. Barcelona (Spain): Lynx Edicions. Available from: https://www.hbw.com/node/52376 [accessed 2020 Jan 15].

Dalloul RA, et al. 2010. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS Biol. 8(9):e1000475.
Ducrest AL, et al. 2020. New genome assembly of the barn owl (*Tyto alba alba*). Ecol Evol. 10(5):2284–2298.
Feng S, et al. 2020. Dense sampling of bird diversity increases power of comparative genomics. Nature 587(7833):252–257.
Frankl-Vilches C, et al. 2015. Using the canary genome to decipher the evolution of hormone-sensitive gene regulation in seasonal singing birds. Genome Biol. 16(1):19.
Friis G, et al. 2018. Genome-wide signals of drift and local adaptation during rapid lineage divergence in a songbird. Mol Ecol. 27(24):5137–5153.
Gregory TR. 2002. Animal genome size database. Available from: http://www. genomesize. com.
Gremme G, Brendel V, Sparks ME, Kurtz S. 2005. Engineering a software tool for gene structure prediction in higher organisms. Inf Softw Technol. 47(15):965–978.
Guy L, Roat Kultima J, Andersson SG. 2010. genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26(18):2334–2335.
Hillier LW, et al. 2014. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 423:695–777.
Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32(5):767–769.
Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. In: Kollmar M, editor. Gene prediction. New York: Humana. p. 65–95.
Hooper DM, Price TD. 2017. Chromosomal inversion differences correlate with range overlap in passerine birds. Nat Ecol Evol. 1(10):1526–1534.
Illera JC, et al. 2018. Acoustic, genetic, and morphological analyses of the Canarian common chaffinch complex *Fringilla coelebs* ssp. reveals cryptic diversification. J Avian Biol. 49(12):1–12.
Jarvis ED, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346(6215):1320–1331.
Jarvis ED, et al.; The Avian Phylogenomics Consortium. 2015. Phylogenomic analyses data of the avian phylogenomics project. GigaScience 4(1):s13742-014.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
Keilwagen J, et al. 2016. Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 44(9):e89.
Keilwagen J, Hartung F, Paulini M, Twardziok SO, Grau J. 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. BMC Bioinformatics 19(1):189.
Lawson LP, Petren K. 2017. The adaptive genomic landscape of beak morphology in Darwin's finches. Mol Ecol. 26(19):4978–4989.
Leroy T, et al. 2021. Forthcoming. Endemic island songbirds as windows into evolution in small effective population sizes. Curr Biol. Available from: 10.1016/j.cub.2020.12.040.
Losos JB, Schluter D. 2000. Analysis of an evolutionary species–area relationship. Nature 408(6814):847–850.
Louha S, Ray DA, Winker K, Glenn TC. 2020. A high-quality genome assembly of the North American Song Sparrow, *Melospiza melodia*. G3 (Bethesda) 10(4):1159–1166.
Lowe TM, Chan PP. 2016. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 44(W1):W54–W57.
MacArthur RH, Wilson EO. 1967. The theory of island biogeography. Princeton (NJ): Princeton University Press. 203 p.
Nawrocki EP. 2014. Annotating functional RNAs in genomes using Infernal. In: Gorodkin J, Ruzzo W, editors. RNA sequence, structure, and function: computational and bioinformatic methods. Totowa (NJ): Humana Press. p. 163–197.
Peñalba JV, et al. 2020. Genome of an iconic Australian bird: High-quality assembly and linkage map of the superb fairy-wren (Malurus cyaneus). Mol Ecol Resour. 20(2):560–578..
Poelstra JW, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. Science 344(6190):1410–1414.
Putnam NH, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26(3):342–350.
Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. Gene prediction. New York: Humana. p. 227–245.
Skinner BM, Griffin DK. 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. Heredity (Edinb). 108(1):37–41.
Smit A, Hubley R. 2019. RepeatModeler-1.0. 11. Institute for Systems Biology. Available from: http://www. repeatmasker. org/RepeatModeler/
Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013–2015. Available from: http://www.repeatmasker.org.
Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7:62.
UniProt Consortium. 2014. UniProt: a hub for protein information. Nucleic Acids Res. 43:D204–D212.
Valente L, et al. 2020. A simple dynamic model explains the diversity of island birds worldwide. Nature 579(7797):92–96.
Vijay N, et al. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat Commun. 7(1):13195.
Völker M, et al. 2010. Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. Genome Res. 20(4):503–511.
Wang X, Wang L. 2016. GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. Front Plant Sci. 7:1350.
Wang W, et al. 2020. First de novo whole genome sequencing and assembly of the bar-headed goose. PeerJ. 8:e8914.
Warren WC, et al. 2010. The genome of a songbird. Nature 464(7289):757–762.
Wirthlin M, Lovell PV, Jarvis ED, Mello CV. 2014. Comparative genomics reveals molecular features unique to the songbird lineage. BMC Genomics 15:1082.
Zhang G, et al. 2015. Genomics: bird sequencing project takes off. Nature 522(7554):34.
Zhang G, et al.; Avian Genome Consortium. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346(6215):1311–1320.
Zhang G, Parker P, Li B, Li H, Wang J. 2012. The genome of Darwin's finch (*Geospiza fortis*). GigaScience. *Database*. Available from: 10.5524/100040.

Associate editor: Bonnie Fraser

Publications outside the thesis

Leroy, T., Rousselle, M., Tilak, M. K., Caizergues, A. E., Scornavacca, C., Recuerda, M., Fuchs, J., Illera, J. C., Blanco, G., Thébaud, C., Milá, B., & Nabholz, B. (2021). Island songbirds as windows into evolution in small populations. *Current Biology*, *31*(6), 1303-1310.

Report

Current Biology

Island songbirds as windows into evolution in small populations

Highlights

- Island species have lower census and effective population sizes than continental ones
- Natural selection is less efficient in island species compared to continental ones
- Stronger differences in deleterious mutations in low recombination genomic regions
- Considerable empirical support for the nearly neutral theory of molecular evolution

Authors

Thibault Leroy, Marjolaine Rousselle, Marie-Ka Tilak, ..., Christophe Thébaud, Borja Milá, Benoit Nabholz

Correspondence

thibault.leroy@univie.ac.at (T.L.), benoit.nabholz@umontpellier.fr (B.N.)

In Brief

By studying the population genomic of 25 passerines, Leroy et al. find that island species have lower nucleotide diversity, higher ratios of non-synonymous to synonymous polymorphisms, and lower adaptive substitution rates than continental species. Their results are consistent with the nearly neutral theory of molecular evolution.

Leroy et al., 2021, Current Biology *31*, 1303–1310
March 22, 2021
https://doi.org/10.1016/j.cub.2020.12.040

CellPress

Current Biology

CellPress

Report

Island songbirds as windows into evolution in small populations

Thibault Leroy,[1,2,*] Marjolaine Rousselle,[1,3] Marie-Ka Tilak,[1] Aude E. Caizergues,[1,4] Céline Scornavacca,[1] María Recuerda,[5] Jérôme Fuchs,[6] Juan Carlos Illera,[7] Dawie H. De Swardt,[8] Guillermo Blanco,[9] Christophe Thébaud,[10] Borja Milá,[5] and Benoit Nabholz[1,11,12,*]

[1]ISEM, Université de Montpellier, CNRS, IRD, EPHE, Montpellier, France
[2]Department of Botany & Biodiversity Research, University of Vienna, Vienna, Austria
[3]Bioinformatics Research Centre, Aarhus University, C.F. Møllers Alle 8, 8000 Aarhus, Denmark
[4]CEFE, Université de Montpellier, CNRS, University Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France
[5]National Museum of Natural Sciences (MNCN), Spanish National Research Council (CSIC), Madrid, Spain
[6]Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, CP51, 57 rue Cuvier, 75005 Paris, France
[7]Biodiversity Research Unit (UO-CSIC-PA), Oviedo University, 33600 Mieres, Asturias, Spain
[8]Department of Ornithology, National Museum, Bloemfontein, South Africa
[9]Department of Evolutionary Ecology, National Museum of Natural Sciences (MNCN), Spanish National Research Council (CSIC), Madrid, Spain
[10]Laboratoire Evolution et Diversité Biologique (EDB), UMR 5174 CNRS - Université Paul Sabatier - IRD, Toulouse, France
[11]Institut Universitaire de France (IUF), Paris, France
[12]Lead contact
*Correspondence: thibault.leroy@univie.ac.at (T.L.), benoit.nabholz@umontpellier.fr (B.N.)
https://doi.org/10.1016/j.cub.2020.12.040

SUMMARY

Due to their limited ranges and inherent isolation, island species have long been recognized as crucial systems for tackling a range of evolutionary questions, including in the early study of speciation.[1,2] Such species have been less studied in the understanding of the evolutionary forces driving DNA sequence evolution. Island species usually have lower census population sizes (*N*) than continental species and, supposedly, lower effective population sizes (*Ne*). Given that both the rates of change caused by genetic drift and by selection are dependent upon *Ne*, island species are theoretically expected to exhibit (1) lower genetic diversity, (2) less effective natural selection against slightly deleterious mutations,[3,4] and (3) a lower rate of adaptive evolution.[5–8] Here, we have used a large set of newly sequenced and published whole-genome sequences of Passerida species (14 insular and 11 continental) to test these predictions. We confirm that island species exhibit lower census size and *Ne*, supporting the hypothesis that the smaller area available on islands constrains the upper bound of *Ne*. In the insular species, we find lower nucleotide diversity in coding regions, higher ratios of non-synonymous to synonymous polymorphisms, and lower adaptive substitution rates. Our results provide robust evidence that the lower *Ne* experienced by island species has affected both the ability of natural selection to efficiently remove weakly deleterious mutations and also the adaptive potential of island species, therefore providing considerable empirical support for the nearly neutral theory. We discuss the implications for both evolutionary and conservation biology.

RESULTS

To assemble our dataset, we used population-level sequencing data (Table 1) from 25 passerine bird species or subspecies, consisting of 14 insular and 11 continental, with a total of 295 individual whole-genome sequences (89 newly sequenced). All species belong to the Passerida lineage, a species-rich clade of songbirds with fairly similar life-history traits. Our dataset includes at least 4 independent continental-island transitions that occurred across the songbird phylogeny (Figure S1) enabling us to efficiently account for phylogenetic structure in all statistical tests reported below (phylogenetic generalized least square [PGLS]; see also Table S1 for additional tests).

Do island species exhibit genomic signatures consistent with low *Ne*?

Past effective population sizes were inferred using the pairwise sequentially Markovian coalescent (PSMC) approach for one randomly selected individual from each species (Figure S2) and were then averaged over the last one million years. The analyses confirmed that island species exhibit a significantly lower mean *Ne* than continental species over the last one million years (mean *Ne* = 362,456 and 94,944 for continental and island species, respectively, Figure 1A; log-transformed *Ne*, PGLS $p = 1.0 \times 10^{-4}$). Specifically, inferred mean *Ne* values over the last million years range from 6.1×10^4 for the Tenerife blue chaffinch (*Fringilla teydea*) to 1.2×10^6 for a continental population of

Table 1. Sequencing data used in this study

	Species	Clade	Range	Individuals	Data
1	*Certhidea olivacea (S)*	Darwin's finches	island	5	[9]
2	*Certhidea fusca (E)*	Darwin's finches	island	10	[9]
3	*Certhidea fusca (L)*	Darwin's finches	island	10	[9]
4	*Platyspiza crassirostris (Z)*	Darwin's finches	island	5	[9]
5	*Camarhynchus pallidus (Z)*	Darwin's finches	island	5	[9]
6	*Pinaroloxias inornata (C)*	Darwin's finches	island	8	[9]
7	*Geospiza difficilis (P)*	Darwin's finches	island	10	[9]
8	*Geospiza septentrionalis (W)*	Darwin's finches	island	8	[9]
9	*Geospiza conirostris (E)*	Darwin's finches	island	10	[9]
10	*Ficedula albicollis*	Ficedula flycatchers	continental	20	[10]
11	*Ficedula hypoleuca*	Ficedula flycatchers	continental	20	[10]
12	*Ficedula speculigera*	Ficedula flycatchers	continental	20	[10]
13	*Ficedula semitorquata*	Ficedula flycatchers	continental	20	[10]
14	*Zosterops borbonicus*	white-eyes	island	6	[11]
				1	[12]
				18	this study
15	*Zosterops olivaceus*	white-eyes	island	15	this study
16	*Zosterops mauritianus*	white-eyes	island	9	this study
17	*Zosterops pallidus*	white-eyes	continental	2	this study
18	*Zosterops virens*	white-eyes	continental	11	this study
19	*Fringilla coelebs*	chaffinches	continental	9	this study
20	*Fringilla coelebs palmae*	chaffinches	island	15	this study
21	*Fringilla teydea*	chaffinches	island	10	this study
22	*Taeniopygia guttata castanotis*	Estrildidae	continental	19	[13]
23	*Poephila acuticauda acuticauda*	Estrildidae	continental	10	[13]
24	*Parus major*	Paridae	continental	10	[14]
25	*Phylloscopus trochilus*	Phylloscopidae	continental	9	[15]

The abbreviation in parenthesis following Darwin's finches names indicates the island of origin (C, Coco; E, Española; L, San Cristobal; P, Pinta; S, Santiago; W, Wolf; Z, Santa Cruz). See also Methods S1.

the common chaffinch (*F. coelebs*), representing an ~20 fold difference (Figure 1A; Table S2).

Such long-term differences in *Ne* between insular and continental species are expected to generate differences in nucleotide diversity levels, because genetic variation is determined by both mutation rate and effective population size. By estimating nucleotide diversity at synonymous (π_S) and at non-synonymous sites (π_N), we find marked differences between island and continental species. Using 6,499 orthologous genes on average (range: 5,018–7,514, among 8,253 orthogroups[16]), we find that π_S varies from 0.07% in the Tenerife blue chaffinch to 1.25% in the willow warbler (*Phylloscopus trochilus*), representing a 17-fold difference between these island and continental species (Table S2; Figure S3). Island species exhibit significantly lower mean π_S than continental species (mean π_S = 0.59% and 0.18% for continental and island species, respectively, Figure 1B; PGLS, $p = 5.29 \times 10^{-3}$).

In addition to strong evidence for lower *Ne* in island species, we also find lower census population sizes in the island species (island: 7 species, median: 1.1×10^4 [range: $6.2 \times 10^2 - 3.0 \times 10^5$]; continental: 6 species, median: 2.5×10^8 [range: $2.0 \times 10^5 - 5.7 \times 10^8$]; log-transformed census sizes, PGLS, $p = 8.29 \times 10^{-5}$). Furthermore, both $\log_{10}$-transformed current census population sizes and geographical range in square kilometers are positively correlated with π_S (Figures 2A and 2C; PGLS, $p < 0.01$; Table S1). Taken all together, these results provide strong support for the view that long-term restrictions on census population sizes due to the limited surface area available to island species constrains the upper bound of effective population size.

Are deleterious mutations segregating more in island species?

Based on the nearly neutral theory of molecular evolution, the higher level of genetic drift associated with lower *Ne* is expected to contribute to an accumulation of slightly deleterious mutations in island species relative to their continental counterparts. Using the ratio of non-synonymous to synonymous mutations (π_N/π_S) as a proxy for the proportion of these slightly deleterious mutations, we recover, on average, a 40% higher π_N/π_S in island

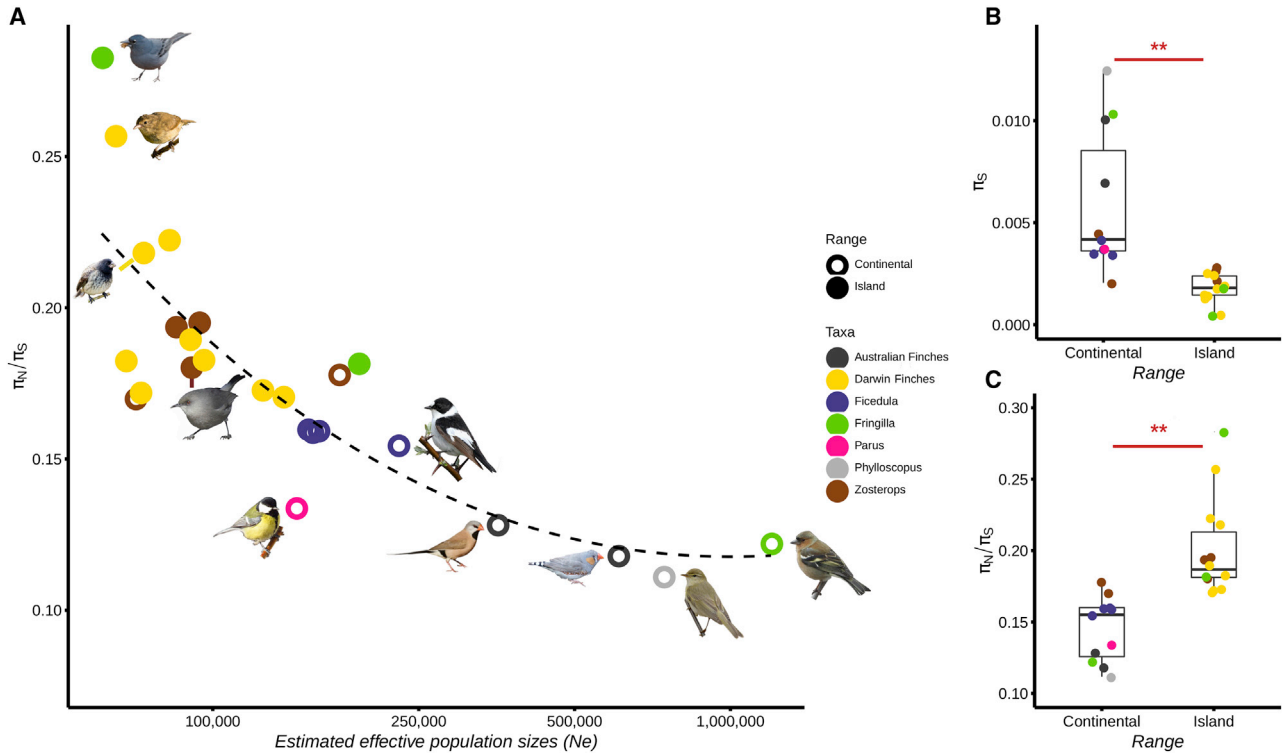


Figure 1. Island species as models for evolution in small effective population sizes

(A) Local polynomial regression (LOESS with span = 1.25) between the ratio of nonsynonymous to synonymous nucleotide diversity (π_N/π_S) and the mean effective population sizes over the last million years (Ne), as inferred using PSMC (see Figure S3 for a log-log regression between π_N/π_S and π_S estimates, respectively).

(B and C) Variation in nucleotide diversity (π_S, B) and π_N/π_S between island endemic and continental species (C). Red asterisks indicate significance of the PGLS test (**p < 0.01).

Photo credits: A. Chudý, F. Desmoulins, E. Giacone, G. Lasley, Lianaj, Y. Lyubchenko, B. Nabholz, J.D. Reynolds, K. Samodurov, A. Sarkisyan (iNaturalist.org); M. Gabrielli (personal communication). See also Figures S1–S3 and Tables S1 and S2.

species than in continental species (Figure 1C; mean π_N/π_S = 0.145 and 0.201 for continental and island species, respectively, PGLS p = 4.57 × 10^{-3}).

In addition, we find substantial within-genome variation in the accumulation of slightly deleterious mutations, as well as in the levels of nucleotide diversity, in such a way that π_S and π_N/π_S are, respectively, positively and negatively correlated to the GC content at the third codon position (GC3). GC3 provides a robust proxy of recombination rate in birds[17,18] (Methods S1). By comparing sets of genes exhibiting the lowest and highest GC3, we found a more marked π_N/π_S differences in genes exhibiting low GC3 ($\Delta mean_{continental\ versus\ Island}$ = 0.107, 95% CI: 0.077–0.194) than in those exhibiting high GC3 (Δmean = 0.029, 95% CI: 0.011–0.048). The stronger effect of recombination for island species is captured by the significant interaction between GC3 and insularity in the linear model: π_N/π_S ~GC3+insularity+GC3:insularity (R^2 = 0.72, p value model <2.2 × 10^{-16}, including p < 2.2 × 10^{-16}, 2.8 × 10^{-10}, and 1.33 × 10^{-05} for GC3, insularity, and the interaction, respectively). These correlations are found to be stronger in island species relative to their continental counterparts, with a particularly pronounced difference in π_N/π_S in genes exhibiting a low GC3. These results are robust to a control for GC-biased gene conversion (Methods S1). Recombination limits genetic interactions between selected mutations and can therefore improve the efficiency of selection.[19,20] These results suggest that the intensity of the differences between island and continental species in the effectiveness of purifying selection relies heavily on the local genomic context.

We found strong negative correlations between (1) π_N/π_S and the $\log_{10}$-transformed *Ne* averaged over the last one million years (Figure 1A; PGLS, p = 1.0 × 10^{-4}) and (2) the non-transformed π_N/π_S and π_S values (PGLS, p = 4.85 × 10^{-5}). $\log_{10}$-transformed current census population sizes, as well as geographical range sizes, significantly correlate with π_N/π_S (Figure 2). In contrast, the IUCN red list assessments have no effect on π_N/π_S or π_S (Table S1; PGLS p >> 0.05). Taken together, our results provide strong empirical evidence that differences in census population sizes between island and continental species translate into differences in *Ne,* and that these differences have a marked influence on genetic diversity and the efficiency of natural selection. These findings fit remarkably well with the expectation from the nearly neutral theory.

Do insular species show lower adaptive potential?

Theory predicts that lower *Ne* in island species should lead to a lower rate of adaptive substitutions than in continental species, if

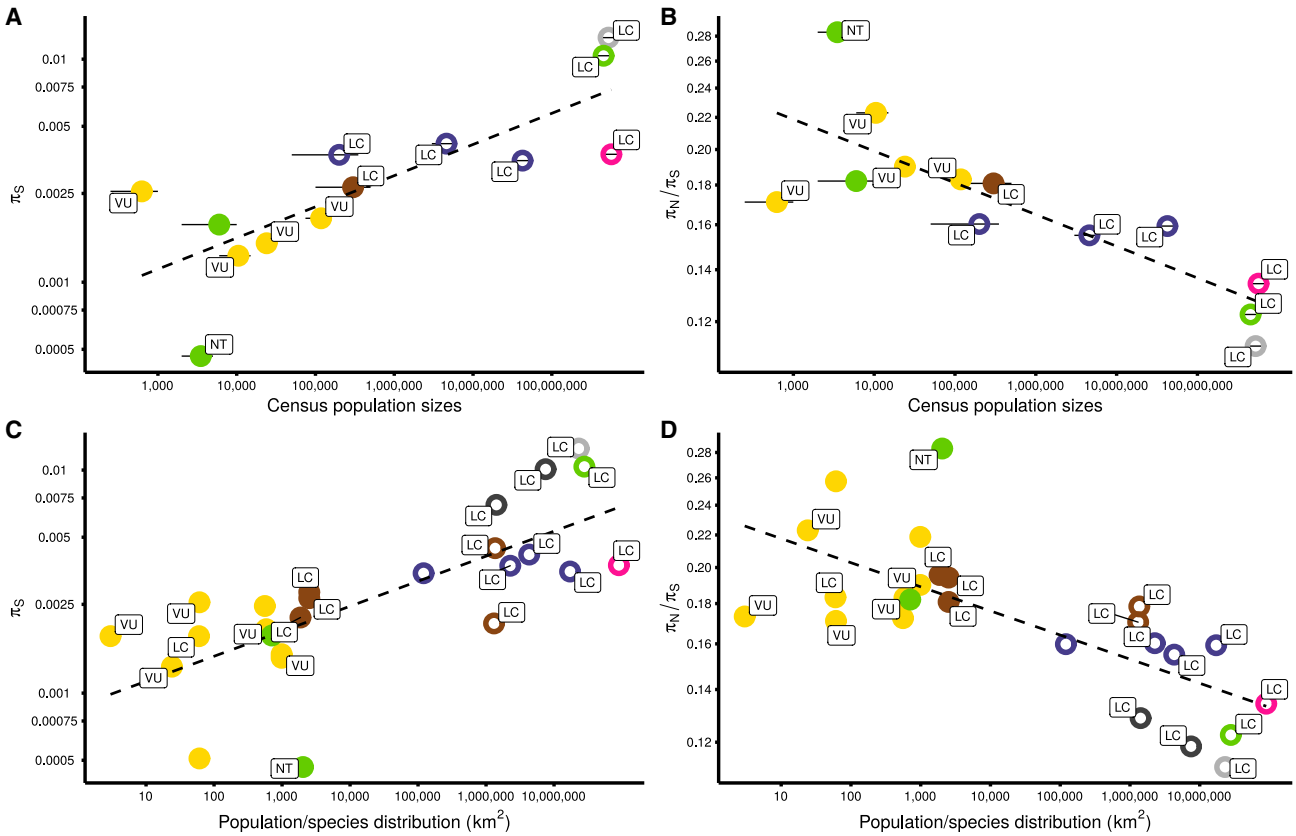


Figure 2. Ecological-evolutionary correlations based on the variables investigated in this study

π_S and π_N/π_S are used as proxies of *Ne* and the efficiency of natural selection to remove deleterious variants and are correlated with both the median estimates of the current census population sizes (A and B) and the geographical range sizes (C and D). Both ecological and evolutionary parameters are log-transformed. Filled and open dots represent the island and the continental species, respectively (Figure 1 for details). Only the 13 species with estimates of the current census population sizes are included for the (A) and (B) (with ranges shown with a thin black line). Where known, the IUCN conservation status of the investigated species is indicated (LC, least concerned; NT, near threatened; VU, vulnerable). See also Figure S3 and Tables S1 and S2.

adaptation is limited by the supply of new mutations[8] and/or if slightly advantageous mutations become effectively neutral in low *Ne* species.[6] For taxa with at least two species (i.e., all except *Parus* and *Phylloscopus*), we used the maximum likelihood method implemented in Grapes[21] to estimate non-adaptive rate of substitution (ω_{NA}) and adaptive rate of substitution (ω_A) with ω (i.e., d_N/d_S) being the sum of $\omega_{NA} + \omega_A$. No significant difference in ω was observed between island and continental species (ω island = 0.194 and ω continental = 0.187). By contrast, island species showed a higher ω_{NA} (Δmean $_{\text{continental } versus \text{ Island}}$ = 0.063) and a lower ω_A (Δmean $_{\text{continental } versus \text{ Island}}$ = 0.056) (Figure 3; see Figure S4 for α estimates) than continental counterparts. However, these differences are only significant for tests that did not explicitly take phylogenetic structure into account (PGLS, p = 0.257 and p = 0.237; non-PGLS, p = 0.014 and p = 0.002 for ω_A and ω_{NA}, respectively; Table S1), and therefore they should be interpreted with caution.

We found that ω_A was positively correlated with $\log_{10}$-transformed π_S (PGLS, p = 0.029; Figure 3A) and negatively correlated with the $\log_{10}$-transformed π_N/π_S (PGLS, p = 0.034; Table S1). Reciprocally, ω_{NA} is significantly negatively correlated with log10-transformed π_S (PGLS, p = 0.020; Figure 3B) and positively with $\log_{10}$-transformed π_N/π_S (PGLS, p = 0.025; Figure S4).

Overall, our analysis suggests that a lower *Ne* doubly affects island species relative to continental species, because (1) relatively fewer adaptive mutations can reach fixation, and (2) the lower efficiency of natural selection allows a greater proportion of weakly deleterious variants to reach fixation in insular species.

DISCUSSION

Our analysis of whole-genome resequencing data has allowed us to find lower nucleotide diversity, a higher frequency of slightly deleterious mutations, and lower adaptive substitution rates in the island species than in the continental ones. These results provide important insights for evolutionary biology, and they also have major implications for the conservation of species with small populations.

Island species as models for studying the evolutionary consequences of small *Ne*

The smaller land area available on oceanic islands should constrain the upper bound of both census and effective

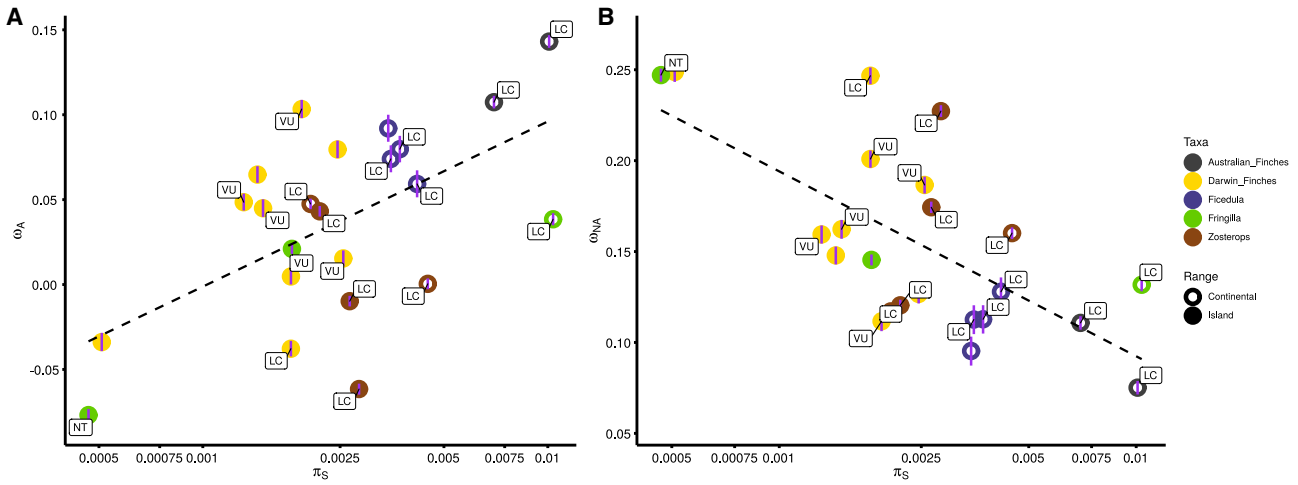


Figure 3. Relationships between adaptive and non-adaptive susbtitution rates and synonymous genetic diversity

Proportion of adaptive (A) and non-adaptive (B) substitutions along the neutral genetic diversity gradient (π_S) as estimated by comparing the observed and the expected d_N/d_S under near neutrality assuming the polymorphism data using the DFE-α method (α shown Figure S4; with $\omega_A = \alpha(d_N/d_S)$ & $\omega_{NA} = (1-\alpha)(d_N/d_S)$). Estimates were performed using all sites and the GammaExpo model. Error bars (purple line) represent the 95% confidence intervals of each estimate under this model. Where known, the IUCN conservation status of the investigated species is indicated (LC, least concerned; NT, near threatened; VU, vulnerable). See also Table S1.

population sizes of insular species, to such an extent that demography affects the ability of purifying selection to remove weakly deleterious mutations. Our results are largely consistent with this general hypothesis and suggest that contemporary census sizes provide information on long-term *Ne* (but see also Díez-Del-Molino et al.[22] and Peart et al.[23]). For most population genomic estimates we investigated, including π_S, π_N/π_S and PSMC-inferred *Ne*, we observed significant differences between continental and island species that are consistent with theoretical expectations.

Previous taxon-specific studies have reported low *Ne* in a diverse range of island organisms (e.g., giant Galápagos tortoises,[24] woolly mammoths,[25] island foxes,[26,27] and *Corvus*[28]). Therefore, it is very likely that island species predominantly exhibit lower *Ne* than their more abundant, broadly distributed, mainland relatives, and this pattern may not be restricted to some specific animal clades such as birds or mammals but may also be true for a large range of taxa (e.g., Hamabata et al.[29] for plants). More broadly, this result opens up new opportunities for using island species as models to understand the impact of *Ne* on genome evolution in natural populations, including genome size, or of natural selection on non-coding genomic regions.

Broad support for the nearly neutral theory of molecular evolution

Fifty years after the introduction of the neutral theory of molecular evolution by Kimura[30] and King and Jukes,[31] and, after being extended into the nearly neutral theory,[4] the neutralist-selectionist controversy remains one of the sharpest and most polarized debates in biology. Based on our large genome-scale empirical data, our results match theoretical expectations of the nearly neutral theory remarkably well. This is consistent with the strength of this theory in explaining patterns of DNA sequence evolution, allowing us to affirm that the nearly neutral theory is overwhelmingly supported by our dataset. Slightly deleterious mutations are frequent and become effectively neutral when the effect of genetic drift increases, as is typically observed in insular species.

Selective processes, including positive selection on beneficial alleles and background selection, play an important role in the sequence evolution of the investigated species but cannot be used to reject the theory as a whole. Empirical investigations found that the proportion of adaptive substitutions does not overall scale with *Ne* when distant taxa are considered all together (e.g., Galtier[21]), but taxa-specific investigations were able to find such a relationship, with a lower proportion of adaptive substitutions in species with a lower *Ne*, as recently reported for several groups of animals.[8] First, our analyses provide additional evidence for such a relationship in passerine birds. Second, we indeed observe that local recombination rates influence both local levels of nucleotide diversity and the number of deleterious mutations, which is consistent with heterogeneous landscapes of *Ne* throughout genomes.[18] However, significant differences between island and continental species were similarly recovered in both lowly and highly recombining regions of the genome, supporting the claim that background selection does not fundamentally change the predictions that can be drawn from the theory.

Ecological-evolutionary ties and perspectives

At the macroevolutionary scale, strong correlations between life-history traits and both levels of polymorphism and ratios of non-synonymous to synonymous mutations have been reported in the literature for both animals and plants,[32–34] suggesting that determinants of genetic diversity are mostly ecologically driven. We found that nucleotide diversity scales positively with species range, which therefore suggests a gradual transition between species restricted to small islands and species widely distributed over continents. Recently, Peart et al.[23] proposed that conservation priorities should be defined based on the ratio of census size

to *Ne*. However, whether population genomic estimates of *Ne* are informative enough to assess conservation status is questionable. A general outcome is that animal species classified as threatened generally exhibit lower genetic diversity than those classified as non-threatened, including birds (at least at microsatellite loci[35,36]). Based on our whole-genome analyses, we can report no obvious contrast between the four island species classified as threatened (vulnerable status) and the species classified as non-threatened, neither for the levels of nucleotide diversity nor for their efficiency of natural selection (but see Brüniche-Olsen et al.[37]). Díez-del-Molino et al.[22] were also unable to recover a significant effect of the IUCN assessment on the levels of nucleotide diversity in birds and mammals. Using 78 mammal species, Brüniche-Olsen et al.[38] only recovered this pattern when the animals' diets were explicitly taken into account. Consequently, it seems that we still have a long way to go toward precisely describing whether these genomic features are completely independent or are correlated to some extent with the current conservation status.

Another open question is whether population genomics can provide information so that short-term IUCN objectives can be extended over a longer time frame? Even if some island species accumulate slightly deleterious mutations,[39] supposedly leading to increased maladaptation, we can question whether this burden of slightly deleterious mutations can lead to species extinction. This hypothesis holds true only if these deleterious mutations are neither purged nor opposed by compensatory or beneficial mutations.[40] Remarkably, the four species classified as threatened are not those exhibiting the lowest proportion of adaptive substitutions (mean ω_A = 0.053 compared to 0.036 for the 15 species with a least concern status). Recent macroevolutionary investigations, however, provide support for this increased risk of (1) being endangered depending on the time since the species colonized the island[41] or (2) becoming extinct depending on the island size.[42] Age-dependent processes such as ecological specialization were proposed, but the accumulation of deleterious mutations might explain this phenomenon as well. Rogers and Slatkin[25] propose that, after a tipping point, this mutational meltdown might contribute to the ultimate steps in the road to extinction. Endemic island species therefore represent taxa of high interest in the evaluation of the long-term consequences of evolution under low effective population sizes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Species included in the study
  - Species distribution and IUCN red list status
- METHOD DETAILS
  - DNA extraction and sequencing (*Zosterops* and *Fringilla* species)
  - Publicly available sequencing data
  - Variant identification
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Gene models & orthology prediction
  - Effective population size estimates
  - Summary statistics of the polymorphic data
  - Summary statistics of the divergence data
  - Statistical analyses

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.12.040.

ACKNOWLEDGMENTS

This research was funded by the French ANR (BirdIslandGenomic project, ANR-14-CE02-0002). The analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform services, the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul), and the Biogenouest BiRD core facility (Université de Nantes). We are grateful to Quentin Rougemont for providing feedback on a previous version of the manuscript, colleagues from the phylogeny and molecular evolution team at ISEM Montpellier for fruitful discussions throughout this project, and Neil McNair (University of Vienna) for proofreading the manuscript, as well as three anonymous reviewers for their constructive comments, which helped us to substantially improve the manuscript. A previous manuscript version also benefited from detailed discussions on Twitter; we would thus like to thank Graham Coop, Jenny James, Nicolas Rode, among others, for their helpful suggestions. The authors wish to thank all the farmers and private and national nature reserves where fieldwork was conducted. For South African samples, the handling and sampling protocols were approved by the Comité Cuvier (68-055 to J.F.). We are grateful to the provincial authorities in the Eastern Cape and Free State provinces of South Africa, and Eastern Cape Parks (Alan Southwood, Cathy Dreyer, Gavin Shaw, Sizwe Mkhulise) for granting permission to collect samples (permit numbers RA-190, CRO144/14CR, and 01-24158). We would also like to acknowledge the Percy FitzPatrick Institute (University of Cape Town), the National Museum Bloemfontein (Free State), and R.C.K. Bowie, P.-H. Fabre, E. Kolarova, and G. Oatley for help during field work and logistical support. For Reunion samples, we thank the Reunion National Park for granting us permission to conduct fieldwork in Pas de Bellecombe, Reunion, France and the field station of Marelongue, funded by the P.O.E., Reunion National Park and OSU Reunion, for logistical support. For Spanish samples, the Canary government gave permission to perform the sampling work to J.C.I. (permit 01-24158). J.C.I. was funded by the Spanish Ministry of Science, Innovation and Universities (Ref.: PGC2018-097575-B-I00) and by a GRUPIN research grant from the Regional Government of Asturias (Ref.: IDI/2018/000151). B.M. and M.R. were partly funded by the Spanish Ministry of Science and Innovation (grant PGC2018-098897-B-I00), and M.R. was supported by a doctoral fellowship from the Spanish Ministry of Education, Culture, and Sport (FPU16/05724). This is publication ISEM 2020-321.

AUTHOR CONTRIBUTIONS

T.L. designed the research, conducted the analyses, and wrote the paper. B.N. designed the research, contributed to the analyses, supervised the work, and wrote the paper. M.R. contributed to the analysis of the divergence data. M.-K.T. performed the wet lab work. A.E.C. performed preliminary investigations using the publicly available data and performed the analysis to identify the Darwin's finch populations and species to consider for the paper. C.S. conducted the analyses of orthologous assignation. M.R., G.B., and B.M. provided the sequence data of *Fringilla coelebs* species and the reference *Fringilla* genome. J.C.I. provided the *Fringilla* samples used in this study. D.H.D.S. and J.F. helped in field work to collect mainland and island *Zosterops* samples. C.T. and B.M. helped during field work in Reunion and revised the manuscript. All the authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 14, 2020
Revised: October 12, 2020
Accepted: December 23, 2020
Published: January 20, 2021

REFERENCES

1. Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (John Murray).
2. Mayr, E. (1999). Systematics and the Origin of Species, from the Viewpoint of a Zoologist (Harvard University Press).
3. Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. Nature *246*, 96–98.
4. Ohta, T. (1992). The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. *23*, 263–286.
5. Gossmann, T.I., Keightley, P.D., and Eyre-Walker, A. (2012). The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. Genome Biol. Evol. *4*, 658–667.
6. Lanfear, R., Kokko, H., and Eyre-Walker, A. (2014). Population size and the rate of evolution. Trends Ecol. Evol. *29*, 33–41.
7. Nam, K., Munch, K., Mailund, T., Nater, A., Greminger, M.P., Krützen, M., Marquès-Bonet, T., and Schierup, M.H. (2017). Evidence that the rate of strong selective sweeps increases with population size in the great apes. Proc. Natl. Acad. Sci. USA *114*, 1613–1618.
8. Rousselle, M., Simion, P., Tilak, M.-K., Figuet, E., Nabholz, B., and Galtier, N. (2020). Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. PLoS Genet. *16*, e1008668.
9. Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature *518*, 371–375.
10. Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L.Z., 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.
11. Bourgeois, Y.X.C., Delahaie, B., Gautier, M., Lhuillier, E., Malé, P.G., Bertrand, J.A.M., Cornuault, J., Wakamatsu, K., Bouchez, O., Mould, C., et al. (2017). A novel locus on chromosome 1 underlies the evolution of a melanic plumage polymorphism in a wild songbird. R. Soc. Open Sci. *4*, 160805.
12. Leroy, T., Anselmetti, Y., Tilak, M.-K., Bérard, S., Csukonyi, L., Gabrielli, M., Scornavacca, C., Milá, B., Thébaud, C., and Nabholz, B. (2019). A bird's white-eye view on neosex chromosome evolution. bioRxiv. https://doi.org/10.1101/505610.
13. Singhal, S., Leffler, E.M., Sannareddy, K., Turner, I., Venn, O., Hooper, D.M., Strand, A.I., Li, Q., Raney, B., Balakrishnan, C.N., et al. (2015). Stable recombination hotspots in birds. Science *350*, 928–932.
14. Corcoran, P., Gossmann, T.I., Barton, H.J., Slate, J., and Zeng, K.; Great Tit HapMap Consortium (2017). Determinants of the efficacy of natural selection on coding and noncoding variability in two passerine species. Genome Biol. Evol. *9*, 2987–3007.
15. Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S., and Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. Evol. Lett. *1*, 155–168.
16. Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., et al. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. Science *346*, 1320–1331.
17. Backström, N., Forstmeier, W., Schielzeth, H., Mellenius, H., Nam, K., Bolund, E., Webster, M.T., Ost, T., Schneider, M., Kempenaers, B., and Ellegren, H. (2010). The recombination landscape of the zebra finch Taeniopygia guttata genome. Genome Res. *20*, 485–495.
18. Kawakami, T., Smeds, L., Backström, N., Husby, A., Qvarnström, A., Mugal, C.F., Olason, P., and Ellegren, H. (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.
19. Hill, W.G., and Robertson, A. (1966). The effect of linkage on limits to artificial selection. Genet. Res. *8*, 269–294.
20. Rousselle, M., Laverré, A., Figuet, E., Nabholz, B., and Galtier, N. (2019). Influence of recombination and GC-biased gene conversion on the adaptive and nonadaptive substitution rate in mammals versus birds. Mol. Biol. Evol. *36*, 458–471.
21. Galtier, N. (2016). Adaptive protein evolution in animals and the effective population size hypothesis. PLoS Genet. *12*, e1005774.
22. Díez-Del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M.T.P., and Dalén, L. (2018). Quantifying temporal genomic erosion in endangered species. Trends Ecol. Evol. *33*, 176–185.
23. Peart, C.R., Tusso, S., Pophaly, S.D., Botero-Castro, F., Wu, C.-C., Aurioles-Gamboa, D., Baird, A.B., Bickham, J.W., Forcada, J., Galimberti, F., et al. (2020). Determinants of genetic variation across eco-evolutionary scales in pinnipeds. Nat. Ecol. Evol. *4*, 1095–1104.
24. Loire, E., Chiari, Y., Bernard, A., Cahais, V., Romiguier, J., Nabholz, B., Lourenço, J.M., and Galtier, N. (2013). Population genomics of the endangered giant Galápagos tortoise. Genome Biol. *14*, R136.
25. Rogers, R.L., and Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. PLoS Genet. *13*, e1006601.
26. Robinson, J.A., Brown, C., Kim, B.Y., Lohmueller, K.E., and Wayne, R.K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Curr. Biol. *28*, 3487–3494.e4.
27. Robinson, J.A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B.Y., vonHoldt, B.M., Marsden, C.D., Lohmueller, K.E., and Wayne, R.K. (2016). Genomic flatlining in the endangered island fox. Curr. Biol. *26*, 1183–1189.
28. Kutschera, V.E., Poelstra, J.W., Botero-Castro, F., Dussex, N., Gemmell, N.J., Hunt, G.R., Ritchie, M.G., Rutz, C., Wiberg, R.A.W., and Wolf, J.B.W. (2020). Purifying selection in corvids is less efficient on islands. Mol. Biol. Evol. *37*, 469–474.
29. Hamabata, T., Kinoshita, G., Kurita, K., Cao, P.-L., Ito, M., Murata, J., Komaki, Y., Isagi, Y., and Makino, T. (2019). Endangered island endemic plants have vulnerable genomes. Commun. Biol. *2*, 244.
30. Kimura, M. (1968). Evolutionary rate at the molecular level. Nature *217*, 624–626.
31. King, J.L., and Jukes, T.H. (1969). Non-Darwinian evolution. Science *164*, 788–798.
32. Chen, J., Glémin, S., and Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. Mol. Biol. Evol. *34*, 1417–1428.
33. Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K., Le Provost, G., et al. (2018). Oak genome reveals facets of long lifespan. Nat. Plants *4*, 440–452.
34. Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., et al. (2014). Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature *515*, 261–263.
35. Doyle, J.M., Hacking, C.C., Willoughby, J.R., Sundaram, M., and DeWoody, J.A. (2015). Mammalian genetic diversity as a function of habitat, body size, trophic class, and conservation status. J. Mammal. *96*, 564–572.
36. Willoughby, J.R., Sundaram, M., Wijayawardena, B.K., Kimble, S.J.A., Ji, Y., Fernandez, N.B., Antonides, J.D., Lamb, M.C., Marra, N.J., and DeWoody, J.A. (2015). The reduction of genetic diversity in threatened

vertebrates and new recommendations regarding IUCN conservation rankings. Biol. Conserv. *191*, 495–503.

37. Brüniche-Olsen, A., Kellner, K.F., and DeWoody, J.A. (2019). Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. Mol. Ecol. *28*, 4914–4925.
38. Brüniche-Olsen, A., Kellner, K.F., Anderson, C.J., and DeWoody, J.A. (2018). Runs of homozygosity have utility in mammalian conservation and evolutionary studies. Conserv. Genet. *19*, 1295–1307.
39. Kondrashov, A.S. (1988). Deleterious mutations and the evolution of sexual reproduction. Nature *336*, 435–440.
40. Lynch, M., and Gabriel, W. (1990). Mutation load and the survival of small populations. Evolution *44*, 1725–1737.
41. Warren, B.H., Hagen, O., Gerber, F., Thébaud, C., Paradis, E., and Conti, E. (2018). Evaluating alternative explanations for an association of extinction risk and evolutionary uniqueness in multiple insular lineages. Evolution *72*, 2005–2024.
42. Valente, L., Phillimore, A.B., Melo, M., Warren, B.H., Clegg, S.M., Havenstein, K., Tiedemann, R., Illera, J.C., Thébaud, C., Aschenbach, T., and Etienne, R.S. (2020). A simple dynamic model explains the diversity of island birds worldwide. Nature *579*, 92–96.
43. Chapman, J.A., Ho, I., Sunkara, S., Luo, S., Schroth, G.P., and Rokhsar, D.S. (2011). Meraculous: de novo genome assembly with short paired-end reads. PLoS ONE *6*, e23501.
44. Putnam, N.H., O'Connell, B.L., Stites, J.C., Rice, B.J., Blanchette, M., Calef, R., Troll, C.J., Fields, A., Hartley, P.D., Sugnet, C.W., et al. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. *26*, 342–350.
45. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
46. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv, 1303.3997. https://arxiv.org/abs/1303.3997.
47. Broad Institute (2019). Picard Toolkit. http://broadinstitute.github.io/picard/.
48. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
49. Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., and Delsuc, F. (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol. Ecol. Resour. *20*, 892–905.
50. Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., and Delsuc, F. (2018). MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol. Biol. Evol. *35*, 2582–2584.
51. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
52. She, R., Chu, J.S.-C., Uyar, B., Wang, J., Wang, K., and Chen, N. (2011). genBlastG: using BLAST searches to build homologous gene models. Bioinformatics *27*, 2141–2143.
53. Eddy, S.R. (2011). Accelerated profile HMM searches. PLoS Comput. Biol. *7*, e1002195, e1002195.
54. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496.
55. R Core Team (2018). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
56. Lamichhaney, S., Han, F., Berglund, J., Wang, C., Almén, M.S., Webster, M.T., Grant, B.R., Grant, P.R., and Andersson, L. (2016). A beak size locus in Darwin's finches facilitated character displacement during a drought. Science *352*, 470–474.
57. Zink, R.M., and Vázquez-Miranda, H. (2019). Species limits and phylogenomic relationships of Darwin's finches remain unresolved: potential consequences of a volatile ecological setting. Syst. Biol. *68*, 347–357.
58. Parker, P., Li, B., Li, H., and Wang, J. (2012). The genome of Darwin's finch (Geospiza fortis). Gigascience.
59. Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., et al. (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature *491*, 756–760.
60. Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., et al. (2010). The genome of a songbird. Nature *464*, 757–762.
61. Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., Garroway, C.J., Madsen, O., Verhoeven, K.J.F., de Jager, V., Megens, H.-J., Warren, W.C., Minx, P., et al.; Great Tit HapMap Consortium (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. Nat. Commun. *7*, 10474.
62. Recuerda, M., Vizueta, J., Cuevas-Caballé, C., Blanco, G., Rozas, J., and Milá, B. (2020). Chromosome-level genome assembly of the common chaffinch (Aves: Fringilla coelebs): a valuable resource for evolutionary biology. bioRxiv. https://doi.org/10.1101/2020.11.30.404061.
63. Scornavacca, C., Belkhir, K., Lopez, J., Dernat, R., Delsuc, F., Douzery, E.J.P., and Ranwez, V. (2019). OrthoMaM v10: scaling-up orthologous coding sequence and exon alignments with more than one hundred mammalian genomes. Mol. Biol. Evol. *36*, 861–862.
64. Guéguen, L., Gaillard, S., Boussau, B., Gouy, M., Groussin, M., Rochette, N.C., Bigot, T., Fournier, D., Pouyet, F., Cahais, V., et al. (2013). Bio++: efficient extensible libraries and tools for computational molecular evolution. Mol. Biol. Evol. *30*, 1745–1750.
65. Nadachowska-Brzyska, K., Burri, R., Olason, P.I., Kawakami, T., Smeds, L., and Ellegren, H. (2013). Demographic divergence history of pied flycatcher and collared flycatcher inferred from whole-genome re-sequencing data. PLoS Genet. *9*, e1003942.
66. Kim, S., Cho, Y.S., Kim, H.-M., Chung, O., Kim, H., Jho, S., Seomun, H., Kim, J., Bang, W.Y., Kim, C., et al. (2016). Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. Genome Biol. *17*, 211.
67. Smeds, L., Qvarnström, A., and Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. Genome Res. *26*, 1211–1218.
68. Brommer, J.E., Gustafsson, L., Pietiäinen, H., and Merilä, J. (2004). Single-generation estimates of individual fitness as proxies for long-term genetic contribution. Am. Nat. *163*, 505–517.
69. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).
70. Wilke, C. (2016). cowplot: Streamlined Plot Theme and Plot Annotations for "ggplot2".
71. Bolívar, P., Mugal, C.F., Nater, A., and Ellegren, H. (2016). Recombination rate variation modulates gene sequence evolution mainly via GC-biased gene conversion, not Hill-Robertson interference, in an avian system. Mol. Biol. Evol. *33*, 216–227.
72. Eyre-Walker, A., and Keightley, P.D. (2009). Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol. Biol. Evol. *26*, 2097–2108.
73. Eyre-Walker, A., Woolfit, M., and Phelps, T. (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics *173*, 891–900.
74. Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D.; R Core Team (2020). nlme: linear and nonlinear mixed effects models. https://CRAN.R-project.org/package=nlme.
75. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.
76. Kassambara, A. (2018). ggplot2 (Based Publication Ready Plots).
77. Slowikowski, K. (2019). ggrepel: automatically position non-overlapping text labels with "ggplot2".
78. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.-Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
18 *Zosterops borbonicus* samples	Field sampling in Reunion , France	See Table S3
15 *Zosterops olivaceus* samples	Field sampling in Reunion , France	See Table S3
9 *Zosterops mauritinaus* samples	Field sampling in Mauritius	See Table S3
1 *Zosterops pallidus* sample	Field sampling in South Africa	See Table S3
11 *Zosterops virens*samples	Field sampling in South Africa	See Table S3
9 continental *Fringilla coelebs* samples	Field sampling in continental Spain	See Table S3
15 *Fringilla coelebs palmae* samples	Field sampling in La Palma, Spain	See Table S3
10 *Fringilla teydae* samples	Field sampling in Tenerife, Spain	See Table S3
Deposited data		
Raw reads *Taeniopygia* & *Poephila*	[13]	BioProject PRJEB10586
Raw reads *Parus*	[14]	BioProject PRJNA381923
Raw reads *Phylloscopus*	[15]	BioProject PRJNA319295
Raw reads Darwin's finches	[9]	BioProject PRJNA263122
Raw reads *Ficedula*	[10]	BioProject PRJEB7359
Raw reads *Zosterops*	[11,12] and this study	BioProjects PRJEB18566, PRJNA530916, PRJNA661201
Raw reads *Fringilla*	This study	BioProject PRJNA661201
Software and algorithms		
Meraculous v. 2.2.2.5	[43]	https://jgi.doe.gov/data-and-tools/meraculous/
HiRise	[44]	https://github.com/DovetailGenomics/HiRise_July2015_GR
Trimmomatic	[45]	http://www.usadellab.org/cms/?page=trimmomatic
BWA mem	[46]	http://bio-bwa.sourceforge.net/
Picard v. 1.140	[47]	http://broadinstitute.github.io/picard/
GATK v. 3.7	[48]	https://gatk.broadinstitute.org/hc/en-us
MitoFinder v.1.1	[49]	https://github.com/RemiAllio/MitoFinder
Macse v.2	[50]	https://bioweb.supagro.inra.fr/macse/index.php?menu=releases
IQTREE	[51]	http://www.iqtree.org/
genBlastG	[52]	http://genome.sfu.ca/genblast/download.html
HMMER toolkit	[53]	http://hmmer.org/
PSMC	[54]	https://github.com/lh3/psmc
Grapes. v1.0	[21]	https://github.com/BioPP/grapes
R v3.6.3	[55]	https://cran.r-project.org/
Scripts used	This study	https://osf.io/uw6mb/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Benoit Nabholz (benoit.nabholz@umontpellier.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

All raw sequencing data have been deposited in the Sequence Read Archive (SRA) under BioProject:PRJNA661201. All scripts and programs used are available at the following Open Science Framework repository: https://osf.io/uw6mb/

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Species included in the study

In this study, we both reanalyzed publicly available data and generated our own sequencing data from 25 passerine species (Table S3). By generating new sequencing data, our objective was to target taxa containing both island and continental relatives (chaffinches and white-eyes) in order to increase our statistical power. More broadly, our comparison is only based on species with relatively similar body-mass, longevity and clutch-size. This control was introduced to reduce the risk of some confounding factors that could correlate with *Ne*[34] in order to be able to truly assess the effect of insularity.

Species distribution and IUCN red list status

Species range sizes were obtained from BirdLife (http://datazone.birdlife.org/home) or estimated based on the information shown on the IUCN-red list webpage using CalcMaps (https://www.calcmaps.com/map-area/). For endemic island species, we considered the total island area as a maximum bound for the population range. The IUCN red list conservation status is given at the species level and not at below-species level. As a consequence, we either considered this information to be missing for both populations (e.g., *Certhidea fusca* E and *C. fusca* L) or we only used the status for the most widely distributed species (e.g., the Least Concern (LC) status for the population with a large continental distribution rather than for the island one as in *F. coelebs palmae*). For *Ficedula speculigera*, a species with a $D_A > 0.002$ (Methods S1) and recognized as a distinct species from *F. hypoleuca*, no information is yet available in the IUCN red list database.

METHOD DETAILS

DNA extraction and sequencing (*Zosterops* and *Fringilla* species)

All *Zosterops* and *Fringilla* individuals were captured using mist nets. With the exception of African *Zosterops* species (*Z. pallidus* and *Z. virens*, see below), we collected blood samples for each bird by venipuncture of the brachial vein and stored blood in absolute ethanol at −20°C until DNA extraction. For African species, *Z. pallidus* and *Z. virens* individuals, DNA was extracted from liver, muscle or blood. For these samples, voucher specimens are stored at the Museum National d'Histoire Naturelle (MNHN), Paris, France and a tissue duplicate is deposited in the National Museum Bloemfontein (South Africa). For all *Zosterops* and *Fringilla* samples, total genomic DNA was extracted using the DNeasy Blood and Tissue kit (QIAGEN, Valencia, CA) following the manufacturer's instructions. Library preparation (1.0 μg DNA used per sample) and Illumina high-throughput sequencing using a paired-end 150 bp (PE150) strategy were performed at Novogene (Cambridge, UK) to a minimum sequencing yield of 18 Gb per sample (i.e., ~15X coverage). Details on samples are available in Table S3. For these species, we used exactly the same approach as for the publicly available data for variant identification and sequence reconstruction strategies, as described in Methods S1. All newly sequenced raw reads are available under the SRA BioProject : PRJNA661201.

Publicly available sequencing data

We collected publicly available raw sequencing data on SRA from a large range of studies (Table S3). The phylogenetic relationships among Darwin's finches are not fully resolved,[9,56,57] so we first evaluate the net divergence between all pairs of species to delimit 9 groups of species with a net divergence ($D_A > 0.1\%$; Methods S1). Within each group, we selected a single population based on the number of sequenced individuals that were publicly available.[9] Variant identification and sequence reconstruction steps are described in Methods S1.

Variant identification

We used Trimmomatic (v.0.33[45]) to remove adapters, stringently trim and filter reads using the following set of parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50. All trimmed reads were then mapped against the reference genome for each clade (see above) with BWA mem (v. 0.7.12[46]) using default settings. Unmapped reads and mapped reads with a quality (MQ) below 20 were then discarded. Potential PCR duplicates were then flagged using MarkDuplicates v. 1.140 (Picard tools[47]). Variant calling was then performed using GATK (v. 3.7[48]). First, we used HaplotypeCaller on single samples (gVCF) to call SNPs using default parameters. For each species, we then performed a joint genotyping ("GenotypeGVCFs"). To ensure high quality in our dataset, we filtered out low-quality SNPs using several settings: a quality by depth (QD) < 2.0, a Fisher Strand (FS) bias > 60, a mapping quality (MQ) < 40, a MQranksum < −2 or a ReadPosRankSum < −2 or a Raw Mapping Quality (Raw_MQ) < 45,000. SNPs satisfying one or more of these conditions were discarded. For every group of species, we performed principal component analyses (PCA) based on a random sampling of SNPs over the genome (50-200k) to capture additional levels of population structure or an unfortunate misnaming of an individual that could have occurred at some point between the bird sampling campaign and the analysis of the raw sequencing data.

QUANTIFICATION AND STATISTICAL ANALYSIS

Gene models & orthology prediction

We used one reference genome for all species belonging to the same clade (Table S3). We used the genome and gene models of a medium ground-finch individual (*Geospiza fortis*; assembly GeoFor_1.0; GCF_000277835[58]) for all Darwin's finches, a collared

flycatcher (*Ficedula albicollis*, GCF_000247815; assembly FicAlb_1.4[59]) for all *Ficedula*, a zebra finch (*Taeniopygia guttata*; GCF_000151805; assembly taeGut3.2.4[60]) for the Estrildidae, a Reunion grey white-eye (*Zosterops borbonicus*; GCA_007252995; assembly ZoBo_15179_v2.0[12]) for all *Zosterops*. We also used the genome of the willow warbler (*Phylloscopus trochilus*; GCA_002305835; assembly ASM230583v1[15]) and the great tit (*Parus major*; GCF_001522545.2; assembly Parus_major1.1[61]). For all the investigated chaffinches, we used a newly generated assembly of *Fringilla coelebs* (Methods S1, version "HiRise" of Recuerda et al. [62]). For this latter species, as well as for the willow warbler (*P. trochilus*), no gene models were available and we therefore first performed a protein homology detection and intron resolution using genBlastG[52] (http://genome.sfu.ca/genblast/download.html) with the following options "-p genblastg -c 0.8 -r 3.0 -gff -e 1e-10."

To analyze the same orthologous sequences in all species, we used the set of 8253 orthologs identified by Jarvis et al.[16] (http://gigadb.org/dataset/101041). Then, we added the sequence of our species to this set of orthogroups using the method described in Scornavacca et al.[63] Briefly, each orthogroup was used to build an HMM profile using the HMMER toolkit.[53] Then, for each new sequence, hmmscan was used on the HMM database to get the best hits among the orthogroups. For each orthogroup, the most similar sequences for each species were then detected via hmmsearch. Outputs from hmmsearch and hmmscan were considered to be accurate if the first hit score was substantially better than the second best one (in order to limit the risk of paralogy), following a best-reciprocal-hit approach when the results of both programs were compared.[63]

Effective population size estimates

Historical demographic variations in *Ne* were estimated using the Pairwise Sequentially Markovian Coalescent (PSMC) model implemented in the software PSMC.[48] Fasta sequences were converted to the PSMC fasta format using a C++ program (Fasta2PSMCFasta: https://osf.io/uw6mb/) written using BIO++ library.[64] Only scaffolds longer than 500Kb were considered. We used block length of 100bp, with no more than 20% of missing data per block, as implemented in "fq2psmcfa" (https://github.com/lh3/psmc).

For each species, PSMC analyses were run using two randomly selected individuals. To identify suitable parameters, several -t and -p parameters were tested including -p "4+30*2+4+6+10" (as in Nadachowska-Brzyska et al.[65]) and -p "4+25*2+4+6" (as in Kim et al.[66]) but also -p "4+10*3+4" and -p "5*1+25*2+6." The best combination (t15 -r4 -p "5*1+25*2+6") was manually chosen after excluding other parameter values leading to large differences between the two individuals from the same species. Then, we randomly selected one individual and excluded the first four atomic time intervals to exclude the noisy estimates generally generated by PSMC for very recent times and therefore strengthen the reliability of the average estimates of *Ne* over the last million years.

Time was scaled assuming a mutation rate of 4.6×10^{-9} mutation/site/generation as estimated[67] and a generation time of 2 years.[65,68] Results were plotted in R (v3.6.3[55]) using the function "psmc.results" [69; http://datadryad.org/resource/doi:10.5061/dryad.0618v/4] and with the R packages ggplot2[69] and cowplot.[70]

Summary statistics of the polymorphic data

π_S and π_N/π_S ratios were computed using seq_stat_coding from reconstructed fasta sequences (Methods S1) using a publicly available Bio++ script and a procedure previously described[42] (https://osf.io/uw6mb/). We empirically validated that our π_S and π_N/π_S estimates were not impacted by the variable number of samples per species. In addition, we used the π_N/π_S estimates based on the site frequency spectra at both non-synonymous and synonymous sites as described in Rousselle et al.[18] to check the accuracy of these estimates (see also Table S2). Guanine-Cytosine (GC) content at third-codon positions of protein-coding genes (hereafter GC3), an excellent proxy of the local recombination rate in birds[71] was also computed under seq_stat_coding. To estimate the within-genome variation in the efficacy of selection, we estimated π_N/π_S on sets of genes representing a total concatenated coding alignment of 2 Mb, after sorting genes by ascending values of GC3. The last window corresponding to genes exhibiting the highest GC3 values was only considered if this window contained at least 1 Mb of coding sequence.

Summary statistics of the divergence data

We used the method implemented by Galtier[21] (Grapes. v1.0) to estimate α, ω_A and ω_{NA} using the approach introduced by Eyre-Walker & Keightley.[72] Briefly, we fitted both a negative Gamma distribution and an exponential distribution to the synonymous and non-synonymous Site Frequency Spectrum (SFS) (the so-called GammaExpo model[21]) to model the distribution of fitness effect (DFE). Fitted parameters of the DFE were then used to compute the expected d_N/d_S under near neutrality (i.e., without adaptive substitutions but including weakly deleterious substitutions), which was compared to the observed d_N/d_S to estimate the adaptive substitution rate (ω_A) and the proportion of adaptive substitutions (α) [with $\omega_A = \alpha(d_N/d_S)$ & $\omega_{NA} = (1-\alpha)(d_N/d_S)$]. Potential recent changes in population size that affect the SFS were taken into account via the use of nuisance parameters capturing distortions of the SFS optimized alongside the DFE parameters.[73]

Statistical analyses

All statistical analyses were performed using R.[55] We only considered models with a similar number of observations and compared these models based on the Akaike information criterion with a correction for small sample sizes (AICc). To test for the influence of the explanatory variables on π_S and π_N/π_S, we used Phylogenetic Generalized Least Square (PGLS) models. Explanatory variables were always $\log_{10}$-transformed as this violated less frequently the assumption of normality, heteroscedasticity and independence of the residuals using a simple linear model. For PGLS, we used the model implemented in the "nlme" package.[74] The mitochondrial phylogeny was considered as the species tree (Figure S1, newick file is available at https://osf.io/uw6mb/) taken into account assuming a

Brownian correlation structure (using ''corBrownian'' from the ''ape'' package[75]). P value and AICc were computed using the anova.gls function. The rationale of the phylogenetic control is to account for the shared polymorphisms (part of species similarity that is explained by the inheritance from a common ancestor). The level of polymorphism of a given species is dynamically controlled by drift, mutation rate and natural selection. As soon as two species do not share a significant fraction of their polymorphism, there is no need to account for their phylogenetic proximity because their polymorphisms evolved independently. Therefore, the results of all the tests including with and without phylogenetic controls, transformed and untransformed π_S and π_N/π_S are presented in Table S1.

R plots were generated using a series of R packages: cowplot,[70] ggplot2,[69] ggpubr,[76] ggrepel[77] and ggtree.[78]


UAM
Universidad Autónoma
de Madrid


mncn


CSIC
Consejo Superior de Investigaciones Científicas