POPULATION GENOMICS OF THE GALAPAGOS ISLAND MOCKINGBIRDS AND IMPLICATION FOR CONSERVATION

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

in Partial Fulfilment of the Requirements

for a Doctor of Philosophy Degree

in the Department of Biology

University of Saskatchewan

Saskatoon, Saskatchewan

By

SEBASTIAN A. ESPINOZA-ULLOA

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Biology,
University of Saskatchewan,
Collaborative Science Research Building, 112 Science Place
Saskatoon, Saskatchewan, S7N 5E2, Canada

OR

Dean of the College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan, S7N 5C9, Canada

ABSTRACT

Islands are considered as natural laboratories for the understanding of the evolutionary process of speciation. The very first muses of Darwin's insights into evolution by natural selection were the Galapagos mockingbirds (*Mimus* spp.), a monophyletic group of four endemic species. Three species are restricted to a single island each whereas the fourth species occurs on (almost) all the other islands of the archipelago. These birds, known for their limited long-distance flying capabilities, are considered terrestrial species and serve as a clear example of allopatric evolution occurring on islands. The aim of my PhD research has been to unveil the evolutionary history of the Galapagos mockingbird species and its conservation implications using a whole-genome approach. Therefore, my research focused on generating a *de novo* reference genome within this monophyletic group in order to establish an adequate framework for subsequent genome-wide analyses (Chapter 2), and with it unveil the natural history of contrasting Galapagos mockingbird populations along the archipelago (Chapter 3). My findings have revealed that after the common ancestor of these species diverged, there was a systematic and directional spread of these species to the islands, which is directly related to the age of the islands. The geological history of the islands and anthropogenic factors have had different impacts on the demography and genetic variability of these species. Typically, smaller populations are more inbred and have higher rates of non-synonymous mutations becoming fixed. However, despite their extremely small sizes, the populations on Darwin, Wolf, and Floreana islands have maintained stable population sizes over many generations, indicating that the accumulation of these mutations has not had any impact on the mean fitness of these populations.

ACKNOWLEDGEMENTS

Firstly, I want to thank the University of Saskatchewan, which gave me the opportunity to be part of this prestigious institution and has fostered the development of my research for all these years. I would like to acknowledge that the University of Saskatchewan is located on Treaty 6 Territory and the Homeland of the Métis. I pay my respect to the First Nations and Métis ancestors of this place and reaffirm our relationship with one another. At the University of Saskatchewan, I particularly want to thank the Department of Biology and to its directors (Dr. Ken Wilson in the past and Dr. Chris Todd currently) that granted me access to the department scholarships, which let me to sustain myself as an international student and continue with my studies. I also want to thank the Pontificia Universidad Católica del Ecuador for its outstanding financial and motivational support to continue my doctoral studies abroad. I want to thank the institutions that made it possible for my research to materialize directly or indirectly, such as the Galapagos National Park, the San Francisco de Quito University (USFQ), the Charles Darwin Foundation and Cornell University.

I want to thank my supervisors Dr. Jose Andres and Dr. Neil Chilton, without a doubt, without their guidance and support, I would not have been able to get this far. Specifically, I want to thank Jose for his academic teachings, his scientific support, and his affection. In the same way, I want to thank Neil for his academic and moral support, for his care, motivation and attention, and for his great friendship. I am very grateful to all the members of my committee, Dr. Christy Morrisey, Dr. Luana Maroja, Dr. Andres Posso-Terranova, and Dr. Catherine Soos, whose advice and support have greatly inspired, enhanced, and refined my ideas and knowledge. I would like to thank my collaborators who worked in parallel in the development of my research: Dr. Jaime

Chaves (USFQ), Dr. Jan Stêfka and Dr. Jakub Vlček (University of South Bohemia), M.Sc. Cathy Coutu (Agriculture and Agri-Food Canada) and Dr. Luis Ortiz-Catedral (Massey University). Without them this research would not have been possible. Thanks to my friend Mauricio Rivadeneira-Corral for his outstanding help digitalizing different illustrations presented in this thesis.

Finally, I want to thank my family and friends who have been my strength to keep going and ease my burdens. I am deeply grateful to my dear wife Evelyn, and my beloved children Theo and Alessia, for being the driving force behind my determination to persevere, being my greatest source of love and motivation. Infinite thanks to my mother Patricia Ulloa-Almeida, who has consistently inspired me to aim higher and who I have been painfully losing since I left my home country to pursue this doctorate. I want to dedicate this doctorate to her, to her memory and her infinite love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IBT – Island biogeography theory

LD – Linkage disequilibrium

LDD – Linkage disequilibrium decay

WGS – Whole-genome sequencing

MSMC - Multiple sequentially markovian coalescent

ROHs – Runs of homozygosity

G-PhoCS – Generalized phylogenetic coalescent sampler

TWISST – Topology weightying by iterative sampling of sub-trees

CHAPTER I: GENERAL INTRODUCTION

1.1. ISLANDS AND EVOLUTION

Islands have been described as magnificent natural laboratories to study the evolutionary process of speciation. Island organisms have had an extraordinary relevance for the understanding of evolutionary mechanisms. This particular appreciation was born with the observations of Darwin (1859) and Wallace (1902). Years later, with Mendelian genetics and the novel perspectives of Fisher, Haldane, and Wright (1910-1930), a heritability-based evolutionary theory was formally conceived. Then, together with the molecular biology advances from the second half of the 20th century, island species (and models) gave rise to much deeper inferences of the evolutionary mechanisms involved in the divergence and genesis (or formation) of new species, which has been molding the theory of modern evolution (Goodnight *et al.*, 2004; Reynolds, 2011; Warren *et al.*, 2015).

1.1.1. *Island biogeography meets genetics*

According to the island biogeography theory (IBT) proposed by MacArthur & Wilson (1967), colonization and extinction rates are determined by both the island size and its isolation (distance from the source population of colonizers) (Figure 1.1). Although the IBT was originally proposed to be a proxy prediction of the species richness (number of species) on islands, the assumptions of the model also can be extrapolated to the number of individuals of the same species (population size) (Ronce & Olivieri, 2004; Warren *et al.*, 2015). Thus, more distant islands will have a smaller inflow of migrants of the same species, while islands closer to the source will have

a greater inflow. Furthermore, smaller islands would have less capacity to retain individuals of the same species, while larger islands would increase this capacity. IBT model assumptions and predictions remarkably fit in a source-island scenario; however, in a complex island system scenario (e.g., archipelago), colonization/extinction rates may be affected by additional interactive and stochastic events (Reynolds, 2011; Ronce & Olivieri, 2004). Therefore, the life history of an archipelago species can be understood from a metapopulation model (islands as demes). This model would offer a more complex multi-dependent scenario that would adjust rates of (re)colonization and extinction based not only on the island size and its distance, but also on the physical position it occupies in the landscape (position with respect to the other islands) (Reynolds, 2011; Warren *et al.*, 2015).

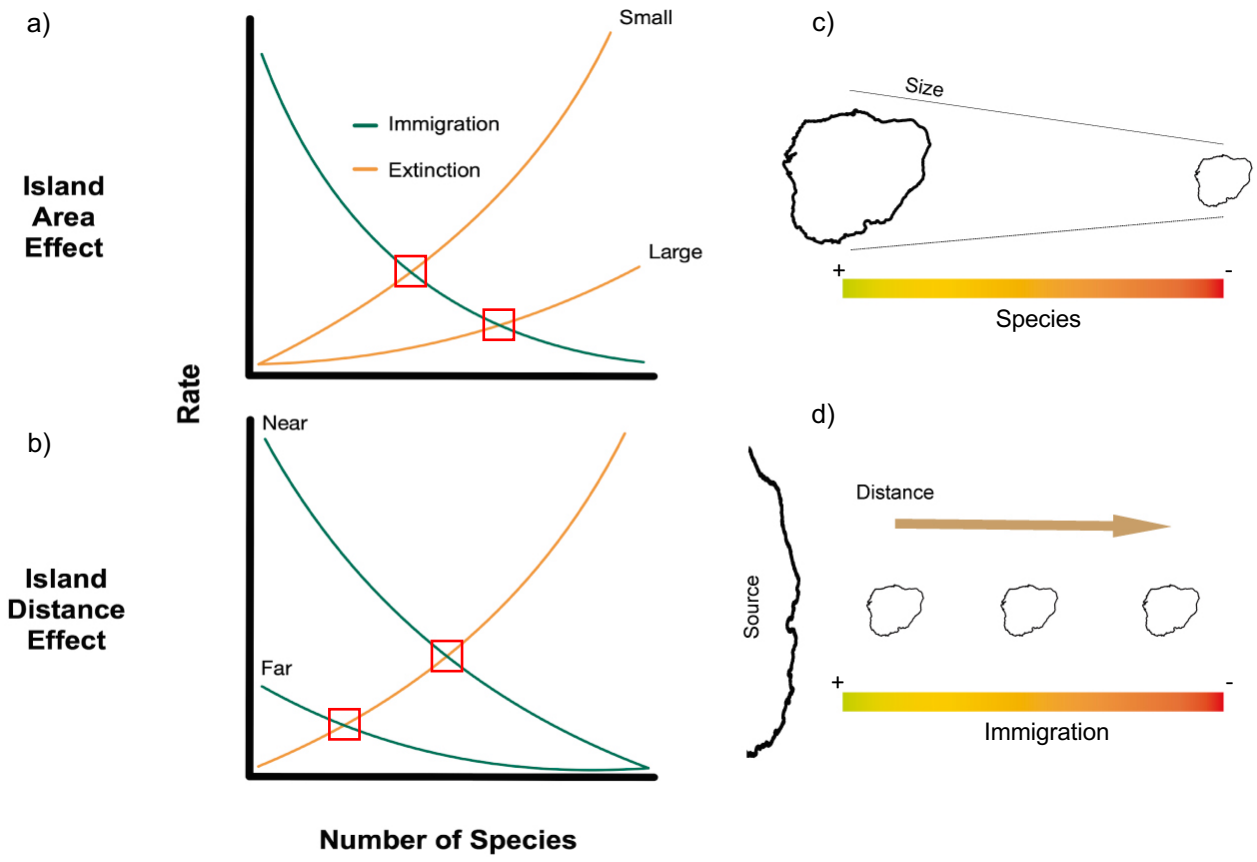


FIGURE 1.1. Island biogeography theory (IBT) model according to MacArthur & Wilson (1963, 1967). The figure shows: a) how island area affects extinction rate, and b) how distance from the mainland affects immigration rate. The predicted number of species is represented on the x-axis, and the equilibria between immigration and extinction rates are highlighted with a red square. The IBT model explains that the island area effect is characterized by a proportional relationship with the number of species present on the island. Thus, c) as the area of the island decreases, the number of species also decreases. On the other hand, the model also predicts an inverse proportional relationship between migration and the distance of the island from the source of migrants. Specifically, d) the probability of migration decreases as the distance between the island and the source increases.

Islands are clearly delimited and isolated areas, and according to their characteristics, determine the size of the populations throughout the time, and allow different migration rates according to their degree of isolation; they can be easily coupled to population genetics models to explain the evolutionary events of the species/populations that occur on them (Ronce & Olivieri, 2004; Warren *et al.*, 2015). Thus, because the size of a population is a proxy of the genetic variability contained within it, the repercussions of the IBT model on genetic variability would be easily inferred from the island size and its distance from the source (Santiago & Caballero, 2016; Watterson, 1975; Wright, 1931). Regarding the size of the island, the number of individuals of a given species would be determined by the size of its ecological niche, which in turn would be determined by the carrying capacity of the island (Reynolds, 2011; Wolf & Ellegren, 2017). Therefore, the size of the population will have a close relationship with the size of the island. In a genetic context, population size is directly related to the number and distribution of alleles in the population (Goodnight, 2004; Santiago & Caballero, 2016; Woolfit & Bromham, 2005). Large populations would have a greater number of alleles, which would decrease as the population becomes smaller. Likewise, large populations will have an allelic distribution (of neutral alleles) close to the equilibrium (i.e., same proportion) between homozygotes and heterozygotes, while in smaller populations consanguinity would cause an imbalance towards more homozygotes (Goodnight, 2004; James *et al.*, 2016; Santiago & Caballero, 2016; Welles & Dlugosch, 2018). Regarding the distance of an island from its source, this plays a role in determining the likelihood of reaching the island, as the probability decreases gradually as the distance increases. This probability can be extrapolated to the number of individuals and, consequently, to the number of alleles. Therefore, the number of alleles reaching the island is inversely proportional to the distance between the island and the source. In this way, more distant islands will suffer more extreme

bottlenecks due to the founder effect and lack of subsequent migrants, than the closest islands (Goodnight, 2004; James *et al.*, 2016, Welles & Dlugosch, 2018; Woolfit & Bromham, 2005). Roughly summarizing, it is expected that a reduction in island size leads to a decrease in genetic diversity and an increase in inbreeding due to a smaller population size. In addition, it is also expected that as the island becomes more distant, the probability of reaching it decreases, exacerbating the effects of reduced genetic diversity and increased inbreeding (more pronounced bottlenecks). Accordingly, in a metapopulation model context, in addition to the island size and distance, the island relative position within the archipelago would also be playing an important role that shapes the population genetic pool (Reynolds, 2011; Woolfit & Bromham, 2005). Consequently, the influence and interaction of these factors would adjust the genetic variability of each population.

1.1.2. *Island settlement evolutionary consequences*

The evolutionary consequences in island populations would be influenced by the characteristics of the island. Island size, isolation (distance) and relative position (in the archipelago) are factors that define the variations and interactions of population genetic evolutionary forces. In this way, each of the spatially structured populations (archipelago metapopulation) will have specific degrees of drift, selection and migration that will shape the population gene pool (James *et al.*, 2016; Whitlock, 2004; Woolfit & Bromham, 2005).

Genetic drift refers to the population probability that a variable position will randomly experience the fixation of a single allele (total loss of allelic diversity) from one generation to another. Therefore, the drift can be estimated based on the allelic frequency of alleles contained in the population gene pool (Figure 1.2). Since population size has a positive relationship with the

number of alleles, then an inverse relationship with allele loss (decrease in diversity) would also be expected; consequently, the drift would be greater towards small islands and would decrease towards larger islands (Goodnight, 2004; Santiago & Caballero, 2016; Whitlock, 2004; Wright, 1931).

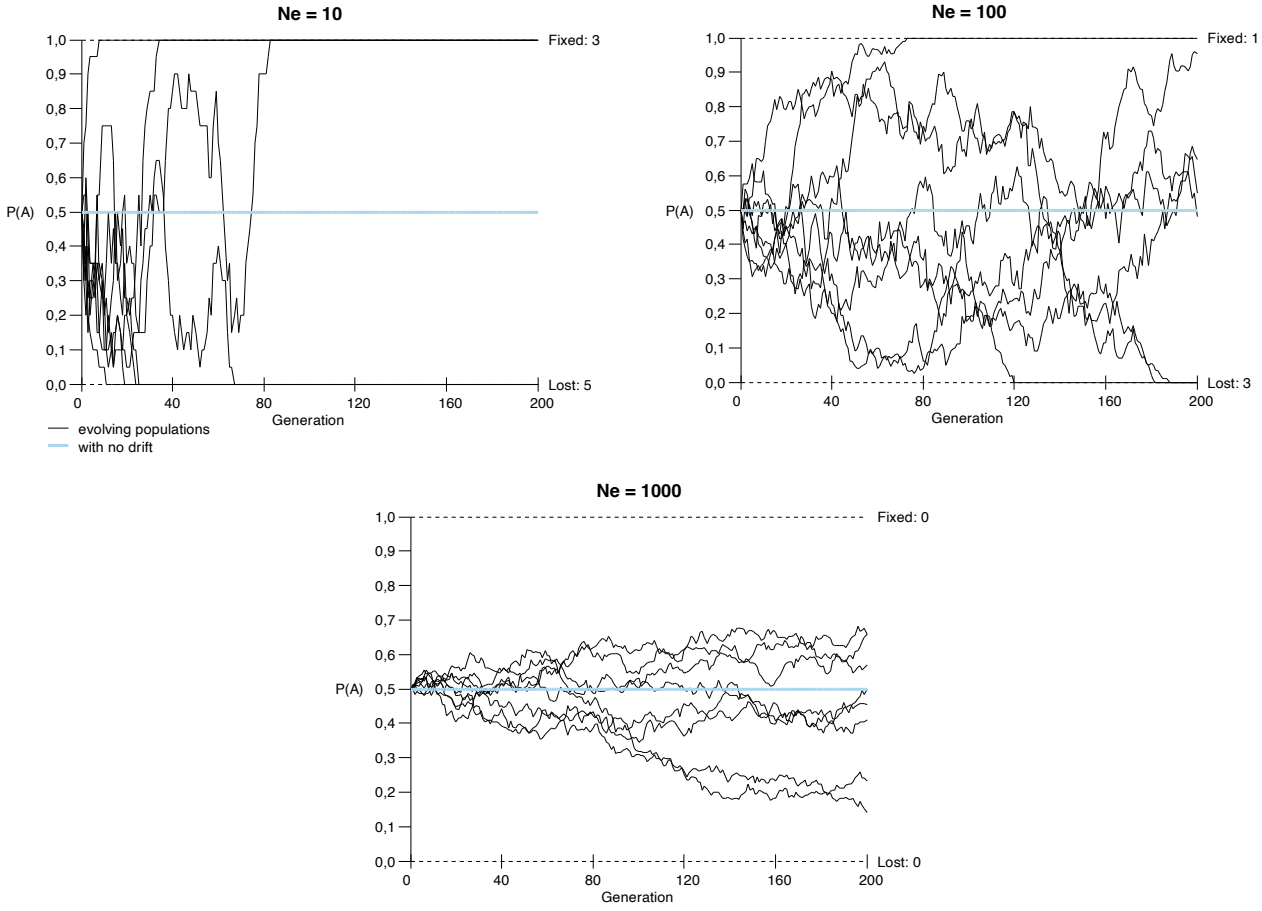


FIGURE 1.2. Genetic drift effect versus population size (*Ne*) based on PopG (ver. 4, Felsenstein, 1993-2016). Simulations shows the frequency variations of a single neutral locus-allele in eight independent populations. Each simulation was generated by me (the author) and is based on three distinct population sizes.: a) *Ne* = 10, b) *Ne* = 100, and c) *Ne* = 1,000. The smaller the population,

the more evident the drift effect, which results in an accelerated fixation or loss of the assessed allele.

Selection is a phenomenon that promotes the accumulation (positive selection) or discarding (negative selection) of mutations (alleles) that have effects on the mean population fitness. Therefore, selection would be estimated as the generational variation of the frequency of an allele that is increased (or decreased) by its effect on the fitness of the individuals in the population (Figure 1.3). Larger populations are expected to be more efficient in undergoing selection due to a greater number of alleles and higher levels of random mating, which results in reduced allele fixation by drift. In contrast, smaller populations tend to have higher levels of inbreeding, which reduces the effects of random mating, and consequently enhances the effect of genetic drift. As a result, the rate of random allele fixation in small populations can be high enough to surpass the effect of selection, leading to a reduced influence of selection on the population. In this way, it can be said that drift and selection behave as inverse related forces as function of the island size (population size). Thus, small islands are more likely to experience high levels of genetic drift and exhibit an apparent absence of selection, whereas larger islands are characterized by reduced genetic drift and a higher likelihood of selection takes place on the population (Goodnight, 2004; Reynolds, 2011; Ronce & Olivieri, 2004; Santiago & Caballero, 2016; Whitlock, 2004).

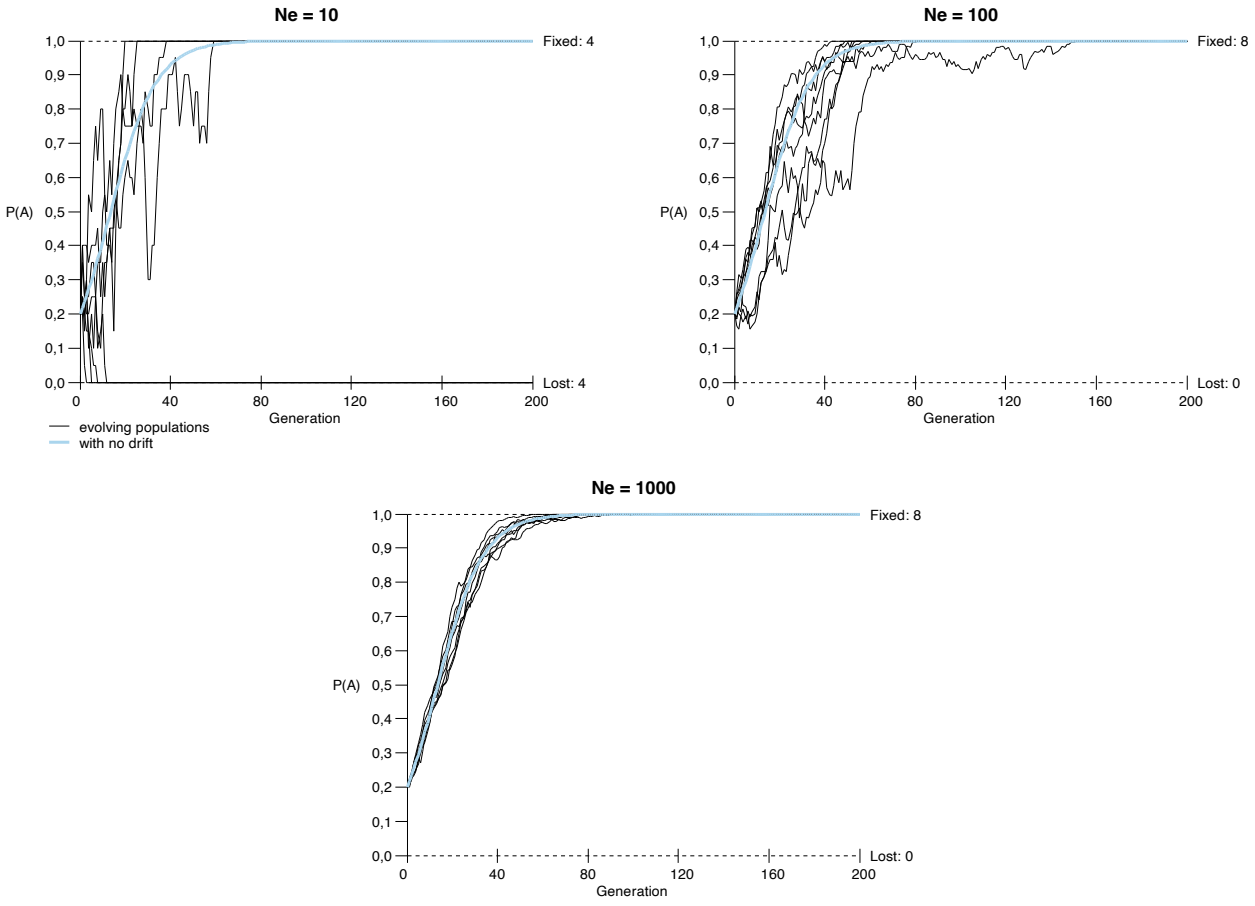


FIGURE 1.3. Selection effect versus population size based on PopG (ver. 4, Felsenstein, 1993-2016). Simulations show the frequency variations of a single locus-allele under positive selection in eight independent populations. Each simulation was generated by me (the author) and is based on three different population sizes: a) *Ne* = 10, b) *Ne* = 100, and c) *Ne* = 1,000. As population size decreases, selection becomes less effective, leading to an increased influence of genetic drift. In some cases, this can result in the loss of alleles that would otherwise be positively selected for.

In an evolutionary context, migration refers to the movement of alleles from one population to another (i.e., gene flow). Migration would be reflected as the covariance of the allelic frequencies between two populations, which can be understood as a measure of the relationship

(or structuring) between populations (Figure 1.4). The rate (and magnitude) of this phenomenon would have modifying repercussions on the individual population theoretical drift/selection. The migration rate will depend on the distance between populations defined by the dispersal capacity of individuals. Thus, nearby populations will have a higher rate of migration than distant populations (Reynolds, 2011; Whitlock, 2004). Likewise, higher migration rates mean higher relatedness (lower divergence rate) between populations, while lower migration rates mean less relatedness (higher divergence rate) (Nielsen & Wakeley, 2001; Ronce & Olivieri, 2004; Wolf & Ellegren, 2017). In a metapopulation model, the gene pool of strictly isolated populations will be shaped by their unique conditions of drift and selection leading to imminent differentiation within a few generations from other populations. While the gene pool of a population influenced by migration will have a certain shared identity with the immigrants' source population, reducing the rate of divergence between these populations. Therefore, the degree of shared identity should be given by the migration rate, which ultimately would be directly influencing the rate of divergence of the populations (Nielsen & Wakeley, 2001; Reynolds, 2011; Ronce & Olivieri, 2004; Wolf & Ellegren, 2017). In this way, isolated islands will have a low migration rate and their divergence rate will be higher, while towards closer islands the migration will be higher, and their divergence will be lower.

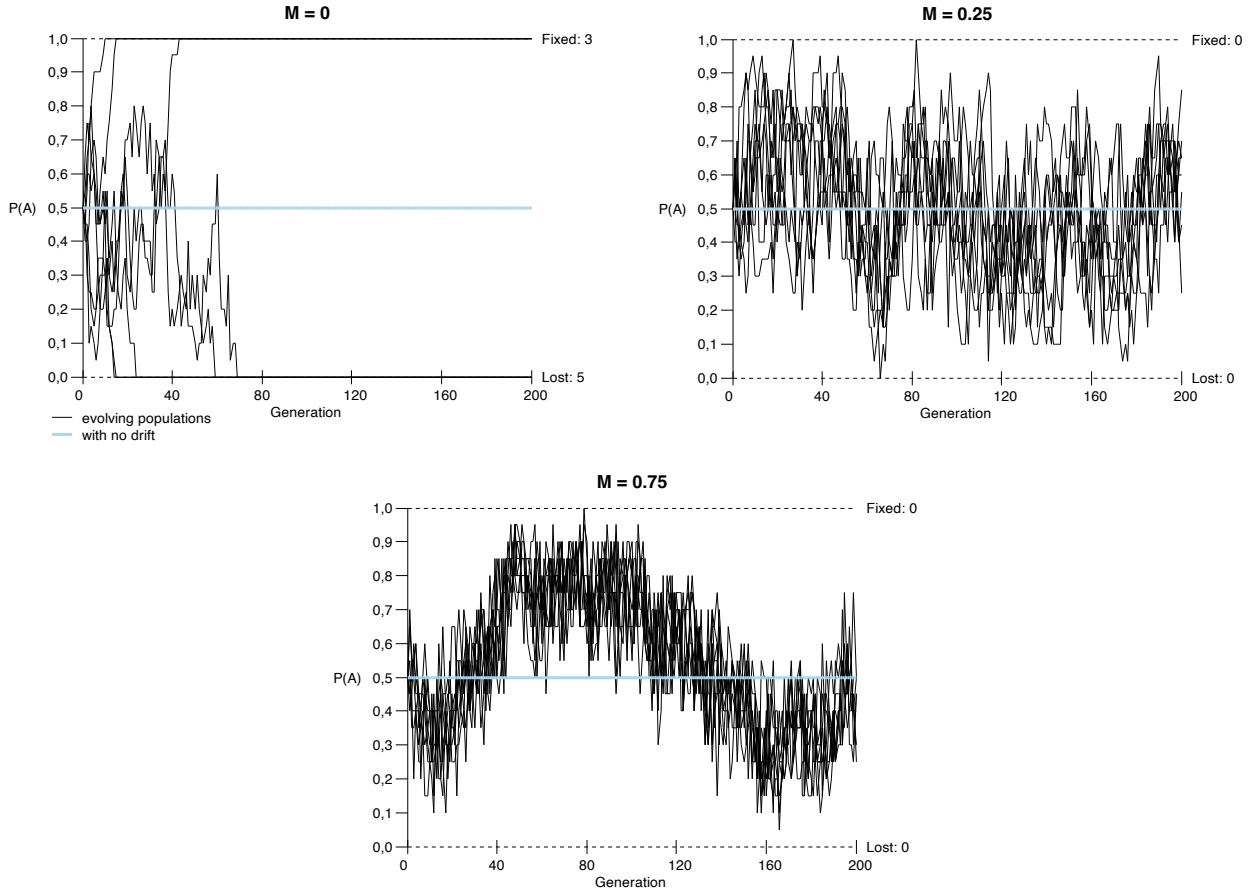


FIGURE 1.4. Migration effect as gene pool homogenizing force based on PopG (ver. 4, Felsenstein, 1993-2016) simulations showing the frequency variations of a single neutral locus-allele in eight populations in three different migration scenarios. Each simulation was generated by me (the author). All simulations were conducted using a constant population size (*Ne* = 10) and were run under three different migration scenarios: a) No migration (M = 0), b) moderate migration rate (M = 0.25), and c) high migration rate (M = 0.75). With no migration, the simulated allele frequency varies independently (full drift) towards to the rapid allele loss or fixation. Under moderate migration, the allele fluctuation is maintained along all generations despite some populations have lost or fixed the allele (in which the variation is recovered in next generations). In high migration scenario, the allele fluctuation is maintained along all generations (far from allele

loss or fixation) and its frequency strongly tends to vary similarly in populations, giving rise to a single general pattern of variation as if it were a single palmitic population.

Thus, the genetic variability of each population has been shaped by the influence and interaction of these evolutionary forces throughout the population generations. Consequently, in an archipelago (metapopulation) context, the genetic variability analysis would allow us to identify and quantify the magnitude of the structuring and divergence between populations (islands), and each population as a unit would reveal the general evolutionary pattern of the species in the archipelago.

1.2. THE GALAPAGOS ARCHIPELAGO

1.2.1. *Location and origin*

The Galapagos archipelago, also known as the archipelago of Colon (Columbus), lies on the eastern equatorial Pacific ocean, about 1000 kilometers west off the coast of South America (Ecuador). It comprises 19 islands and ~107 islets (Figure 1.5). This archipelago is the result of a volcanic hotspot (mantle plume) in the confluence site of three tectonic plates: the Pacific, Nazca and Cocos. The Galapagos hotspot volcanic activity began ~56 million years ago. Over time, continuing eruptions caused underwater mountains to build up. The mountain bases slowly merged to form the Galápagos platform, a basaltic submarine plateau located between 200 to 500 meters below sea level. Over time, the platform grew in height until the highest points eventually emerged above the sea surface, giving rise to the first islands of the Galapagos archipelago. In this way, the archipelago arose ~8 million years ago, and from then its configuration has been changing constantly (ceaseless replacement of islands) due to the permanent movement of the plates and the high volcanic activity. Since it is known that the Galapagos platform rests on the Nazca plate and

the movement of the plates over the hotspot is determined by the relative motion between the Pacific plate and the Cocos and Nazca plates, this interaction results in a net east-southeast crustal movement (~6-7 cm/year) for the Galapagos platform (Figure 1.5; Anderson, 1975; Morgan, 1972; Wilson, 1963). This southeasterly movement of the Galapagos platform has given to the islands a similar directionality from their genesis to their disappearance, in this sense, the newest islands are to the northwest (formed ~0.5 MYA) while the oldest to the southeast (formed ~2.5 MYA) (Figure 1.5; Geist *et al.*, 2014; Percy *et al.*, 2016). Therefore, islands show an age gradient similar to that of other hotspots associated oceanic archipelagos (e.g., Hawaii). Thus, the isolation and the high dynamics of the archipelago's configuration over time have been critical factors that have had direct effects on biodiversity, and the abundance and evolution of organisms.

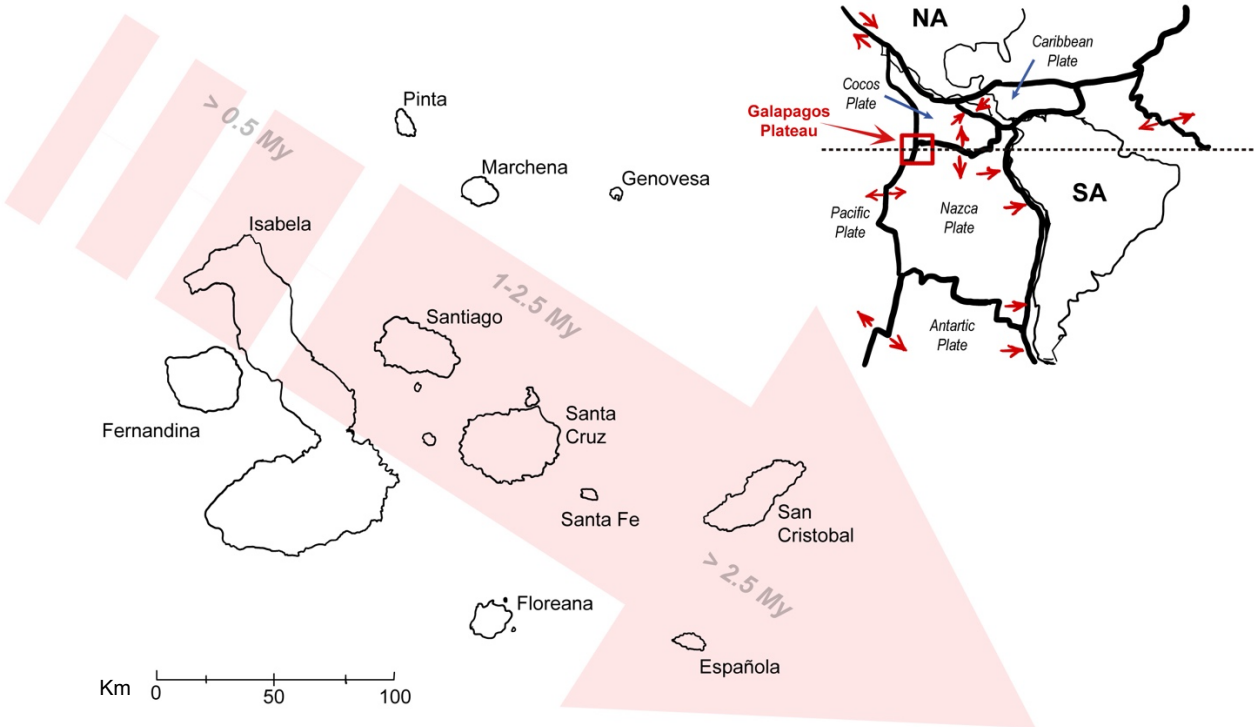


FIGURE 1.5. Current configuration of the Galapagos archipelago with approximate geological ages in million years scale (My). The faint red arrow illustrates the direction in which the

archipelago is moving due to the interaction of tectonic plates. This movement causes the formation of new islands towards the northwest while the existing islands become progressively older towards the southeast. Upper-right panel, the location of the Galapagos platform (hot-spot) is indicated by a red square, while the relative movement of the tectonic plates is shown by the red arrows.

1.2.2. *Colonization and biodiversity*

The natural arrival of terrestrial organisms to remote oceanic islands represents a great challenge with an exceptionally low probability of success. It is assumed that the colonizing individuals should have overcome more than 1,000 kilometers of open ocean to reach the Galapagos archipelago. Under this scenario, the net probability for a potential colonizing organism to reach the archipelago is low due to random or stochastic factors (e.g., storms, or changes in the winds or in ocean currents) that could occur during the transit. However, organism vagility (dispersal ability) can be a key factor to considerably increase the probability of reaching an island. Vagility depends on the intrinsic ability of dispersal (e.g., organism features that allow it to be transported by the wind or by other organism, or that allow it to withstand environmental conditions outside its comfort range) and the ability to survive for extended periods of time (in this case, the time needed to survive the long trip to the islands).

The Galapagos archipelago is located at the confluence of three East to West coastal-influenced currents (the warm South Equatorial and Panamic currents from the North, and the cold Humboldt current from the South) with the Eastward-flowing Cromwell-counter–current (Figure 1.6; Grehan, 2001; Kislik *et al.*, 2017). The interactions of these currents, and their associated trade winds, have had a profound effect on the species that exist on different islands of the Galapagos

archipelago. Because the trade winds and sea currents have been the conditional factors that allowed the colonization of the archipelago, it has been proposed that terrestrial organisms reached the Galapagos Islands from the Americas by 1) rafting (i.e., floating adrift), 2) carried by the trade winds from the southeast to the northwest, or 3) carried by another animal (Valle, 2013). Porter (1976) inferred that ~60% of the ancestral taxa of the islands could have been contributed by birds as vectors, ~32% could have been contributed by the trade winds and only ~9% could have arrived by the ocean.

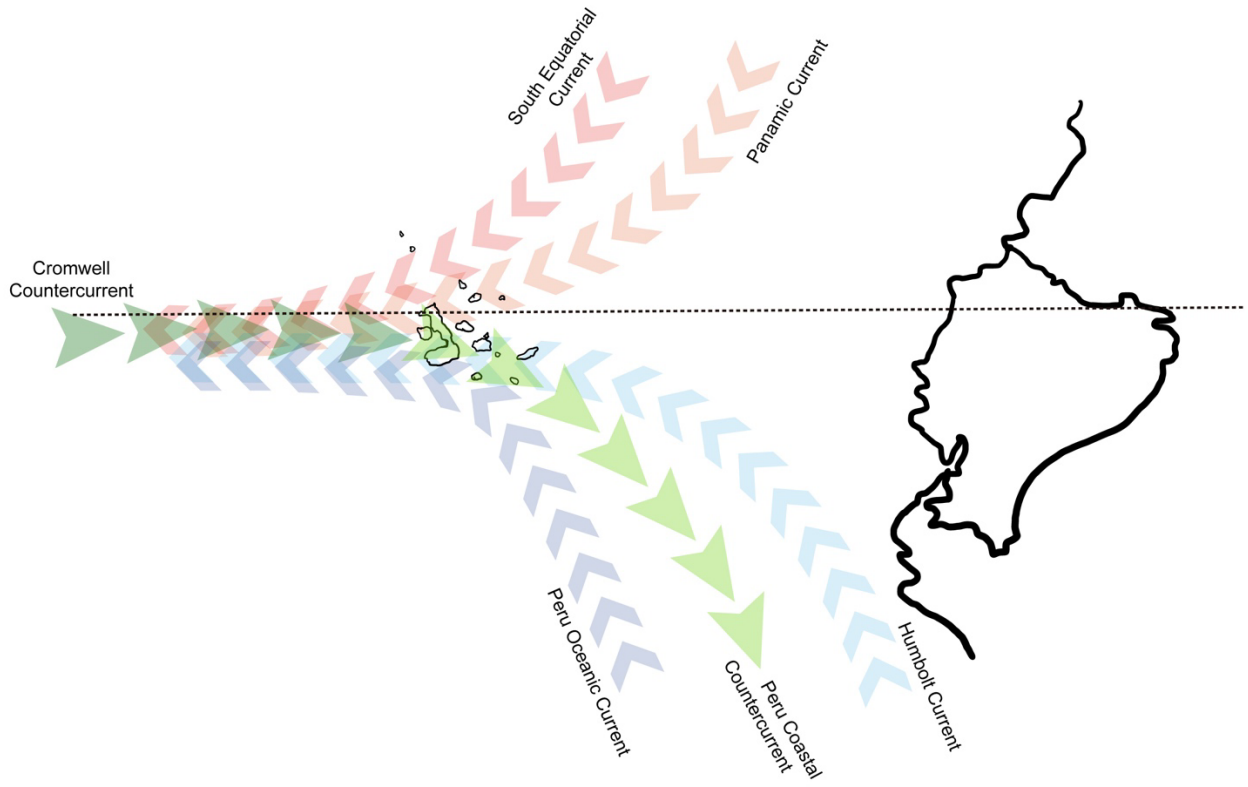


FIGURE 1.6. The directions of marine currents that interact with the Galapagos archipelago. The current Panamic and South Equatorial are sources of warm water, whereas the Humboldt Current, Peru Ocean Current and the Cromwell Countercurrent are sources of cold water.

The first organisms that colonized the Galapagos archipelago reached a territory astoundingly diverse in microhabitat which was the booster for species evolution. The main factors that have given rise to the impressive number of microhabitats have been the geographical position and the associated environmental seasonal events, and the island altitude gradient. The archipelago is situated at the confluence of three major ocean currents, whose interactions determine the seasonal weather patterns in the area. As a result, there are significant fluctuations in humidity levels and rainfall frequencies across the islands, and also have varying impacts on the availability of marine resources in the region. These variations can significantly influence the survival of the terrestrial species. Two seasons are recognized in the Galapagos archipelago. There is a hot-rainy season from January to June. This occurs when currents from the north have more influence on the archipelago, bringing warm waters and tropical winds loaded with moisture. The dry-cold season (July to December) occurs when the Humboldt current has a greater influence on the archipelago, bringing substantial amounts of cold water and dry, cold winds (Kislik *et al.*, 2017; Percy *et al.*, 2016). These two seasons are under influence of the El Niño Southern Oscillation (ENSO), which is a cyclic (2-7 years) pacific-wide tropical phenomenon that comprises two main phases, warm and cold. During the warm phase (El Niño), the trade winds weaken and the warm surface water reaches the coastal region. In these years, the Galapagos terrestrial life flourishes from the heavy rainfall in what is otherwise an arid region. Conversely, in the cold phase (La Niña) trade winds become particularly strong, bringing more cold-water currents reach the coastal region of South America. As a result, during these cycles, marine life flourishes, whereas terrestrial organisms must cope with drought and scarcity (Kislik *et al.*, 2017). Regarding the island altitude gradient, the prominent volcanic activity has given rise to islands with great elevations, volcano craters represent the highest elevations in the islands in some cases reaching over 1,700 meters

above the sea being another source for life diversification (Geist *et al.*, 2014). Thus, altitudinal gradients can be translated in terms of temperature (~5 x 10-3 °C/m) and moisture differences, giving rise to a vast diversity of specific microenvironments at different heights (Geist *et al.*, 2014; Grehan, 2001; Percy *et al.*, 2016). Consequently, the microhabitat megadiversity (and their associated niches) given by the highly dynamic and heterogeneous environment allowed many colonizing organism populations to initiate a single or multiple independent evolutionary processes, which resulted in an astonishing species diversity and extreme endemism. In this way, some ancestral species on the Galapagos islands gave rise to a single endemic species, as is the case for the Galapagos penguin (*Spheniscus mendiculus*) and the flightless cormorant (*Phalacrocorax harrisi*), other ancestral species underwent adaptive radiation resulting in a complex of endemic species. Two examples of this adaptive radiation are the Galapagos islands finches, commonly known as Darwin's finches (15 species), and the Galapagos giant tortoises (~ 15 subspecies). Today, more than 2000 species of terrestrial invertebrate, 530 species of fish and 119 species of vertebrate (i.e., birds, mammals, and reptiles) are recognized as non-migrant native species of the Galapagos Islands (Bungartz *et al.* 2009). There are ~1,861 endemic species of eukaryotes in the archipelago that include ~48 (29 terrestrial) species of birds, ~38 (37 terrestrial) species of reptiles, and 10 (8 terrestrial) species of mammals (Acharya, 2000; IUCN, 2022; FCD, 2022).

1.2.3. *Galapagos islands human history and conservation challenges*

Over five centuries of human history in the Galapagos archipelago has left a tremendous footprint on the ecology and biodiversity of the islands. The first documented visit to the Galapagos was by the Bishop of Panama, Fray Tomas de Berlanga, in 1535, being the official discovery of

the archipelago. Shortly after its discovery, the Galápagos archipelago was used by English pirates as a hiding place from the Spanish fleet. Through this time, as occasional visitors to the islands, the pirates began to hunt giant tortoises as a food source impacting their natural populations. In addition, they were also the first to introduce goats to the islands. Thus, the goats were the first invasive species that was introduced on purpose to the islands, as opposed to the simultaneous introduction of the black rat that was a stowaway in the ships (Latorre, 1999; Gonzales *et al.*, 2008). In this way, both the goats and the rats initiated an imbalance in the island ecosystems since then. Although during this time the human impact was still very restricted and specific, this would increase greatly some decades later. In the 17th century, James Colnett, a British officer, described the archipelago as a group of islands with plenty of flora and fauna. His flamboyant description reached the high English spheres and attracted the attention of the first whalers and merchants interested in whales, sperm whales, sea lions and in the new-known Galapagos (the giant tortoises). Quickly, the whalers' activities began to diminish the giant tortoise populations in some islands, in addition to promoting the extractivism of other native species and reinforcing the introduction of foreign species (Gonzales *et al.*, 2008).

By the 19th century, the first permanent crops were established on the island of Floreana. Later, these crops were dispersed to San Cristobal and Santa Cruz together with new human settlements. The first settlers cultivated fruits and vegetables, raised cattle, pigs, and goats, and did a brisker business trading with whalers, in this way a lot of foreigner species were introduced to these islands (Latorre, 1999; Taylor *et al.*, 2006). By the time of Darwin's visit in 1835, tortoises were already disappearing from Floreana (being the most affected island by human activities at that time). Years later (1869), the naturalist Berthold Seeman reported that giant tortoises were absent on Floreana, and instead the tortoise habitats were occupied by ~2000 cattle. In addition,

wild dogs roamed freely around the island and were responsible of aggressive attacks to visitors (Latorre, 1999; Gonzales *et al.*, 2008). By the end of 19th century, two other giant tortoise populations also become extinct (Santa Fe and Rabida tortoises). Potentially many other species (plants, reptiles, birds, and insects) that were not formally described had disappeared during this same time frame.

At the beginning of the 20th century, in 1924, William Beebe published the book Galapagos: World's End, and with it the archipelago was revealed to the world. This book impelled the curiosity and interest of North America and Europe people about these islands seeding the beginnings of ecotourism that today is the predominant economic activity in Galapagos (Epler, 2007; Latorre, 1999; Taylor *et al.*, 2006). Later, in 1935, the Ecuadorian government recognized the biological importance of the islands and decreed them as wildlife preserves (Epler, 2007; Latorre, 1999). By this time, the adverse effects of four centuries of human presence were beginning to be recognized (Grenier, 2000; Gonzales *et al.*, 2008). From 15 subspecies of giant tortoises, three subspecies were extinct, and populations of some other species were vastly reduced. One native mammal, rice rat (*Aegialomys galapagoensis*), had become extinct on many islands (Gonzales *et al.*, 2008). On the inhabited islands, introduced plants began to replace native species. Feral goats, pigs, donkeys, and cattle were defoliation agents on some islands. Introduced rats and feral cats, dogs, and pigs ate the eggs or young of the native birds and reptiles (Gonzales *et al.*, 2008; Snell *et al.*, 2002). Finally, environmental problems of the islands were being recognized and the concern for their conservation began. To date, the management policies of the archipelago have tried to control and mitigate the human impact, however, the growth of human settlements in the islands, the demand of visitors and the growth of economic activities, exert significant pressures on the unique species of the islands.

1.3. GALAPAGOS MOCKINGBIRDS: A MODEL SYSTEM FOR EVOLUTIONARY AND CONSERVATION GENOMIC STUDIES

The Galapagos mockingbirds (*Mimus* spp.) represent an endemic complex of four species with phenotypic variation across islands. During the description of species collected after the passage of the Beagle through the Galapagos islands, Charles Darwin noted the diversity of mockingbirds. He described three (uncertainly four) different morphotypes of Galapagos mockingbirds. While he recognized that they were all similar to those he had seen in South America he also noticed slight, consistent, variation among islands with the mockingbirds of Floreana being particularly distinct:

"These birds are closely allied in appearance to the Thenca of Chile (2169) or Callandra of la Plata (1216). In their habits I cannot point out a single difference; — They are lively inquisitive, active run fast, frequent houses to pick the meat of the Tortoise, which is hung up, — sing tolerably well; are said to build a simple open nest. — are very tame, a character in common with the other birds: I imagined however its note or cry was rather different from the Thenca of Chile? — Are very abundant, over the whole Island; are chiefly tempted up into the high & damp parts, by the houses & cleared ground. I have specimens from four of the larger Islands; the two above enumerated, and (3349: female. Albermarle Isd.) & (3350: male: James Isd). — The specimens from Chatham & Albermarle Isd appear to be the same; but the other two are different. In each Isld. each kind is exclusively found: habits of all are indistinguishable.

When I recollect, the fact that the form of the body, shape of scales & general size, the Spaniards can at once pronounce, from which Island any Tortoise may have been brought. When I see these Islands in sight of each other, & [but del.] possessed of but a scanty stock of animals,

tenanted by these birds, but slightly differing in structure & filling the same place in Nature, I must suspect they are only varieties.

The only fact of a similar kind of which I am aware, is the constant asserted difference between the wolf-like Fox of East & West Falkland Islds. ***If there is the slightest foundation for these remarks the zoology of Archipelagoes will be well worth examining; for such facts [would inserted] undermine the stability of Species****".*

Source: Darwin's Ornithological Notes – Galapagos (1835)

After his first observations, Darwin assumed that these mockingbird morphotypes were singular varieties of the same species. However, after Darwin arrival to London, the naturalist and ornithologist John Gould (1837) studied in detail such mockingbird morphotypes. Gould revealed that they were four distinct species and grouped them into a single genus that he called *Orpheus*. Afterwards, Gould and Darwin realized that these species (or at least three of them) were restricted to only one island each. At the first instance, this fact allowed Darwin to deduce that a species can be replaced geographically by a similar one. Subsequently, this also triggered the insight in Darwin's mind that different varieties (morphotypes) in isolation could in somehow evolve into different species. Thus, these thoughts would have been shaping the fundamentals in Darwin's theory of evolution.

Mockingbirds have flexible behaviours, and these are linked to the habitat conditions and their ecological interactions. They usually are found living in small groups (not necessarily restricted to relatives) and have cooperative breeding. Mockingbirds do not show an evident sexual dimorphism. They are normally distributed in the lowlands close to the coast, in arid landscapes

of small deciduous shrubs and arborescent cacti. Mockingbirds are omnivores and their diet is mainly based on small invertebrates (mostly insects), however they can also feed on fruits, nectar, pollen, seeds and even small vertebrates (e.g., lizards or fish). Additionally, in particular cases there are records of blood consumption from other living animals (*M. macdonaldi*, Curry & Anderson, 1987). Because of living in coastal areas, the effects of seasonality become quite evident, which also makes the availability of resources change, therefore the diet of the mockingbirds also varies seasonally. Their nests are simple open cups built on the emergent vegetation of the habitat and nesting seasons generally occur with the entry of the wet season (December - January). Finally, there are no specific predators for mockingbirds, however, the nests are vulnerable to predation by both short-eared owls and yellow-crowned night-herons, and there have also been occasional reports of predation on adult birds by the Galapagos hawk (Curry & Grant, 1990; Curry & Grant, 1991).

Each population is restricted to the conditions and size of the island on which it is located. Being landbirds with limited flying ability, no migration events between islands has been reported, therefore each mockingbird population is assumed to be in isolation (Grant *et al.*, 2000). One of the most important peculiarities of the archipelago is that each island has unique conditions depending on size, geological age, altitude, and location. These conditions will shape the availability of resources and the effects of seasonality (Larrea & Di Carlo, 2011). Regarding mockingbirds, we can highlight some adaptations linked to these variations, such as: 1) Española is the most eroded island of the archipelago and, therefore, the driest of all depending on the season, consequently *M. macdonaldi* modified its behavior to obtain water taking advantage of the blood of other species (Curry & Anderson, 1987); or 2) it has been observed that on islands where the vegetation canopy is taller and more exuberant, such as Santa Cruz, San Cristobal, and Isabela,

flock sizes are smaller but their territories are larger. In contrast, on islands with scattered vegetation and open spaces like Española, Genovesa, Champion, and Gardner-by-Floreana, birds form larger flocks but occupy smaller territories. This difference in behavior may be related to resource availability and competition (Curry, 1989; Curry & Grant, 1990). In conclusion, although in general the behaviour of all Galapagos' mockingbirds is similar, each population has developed modifications and unique characteristics to face the conditions of each island.

1.3.1. *Diversity and evolution*

Galapagos mockingbirds are a group of four species, all endemic to the archipelago. *Mimus melanotis* (San Cristobal), *M. trifasciatus* (Floreana) and *M. macdonaldi* (Española) are restricted to a single island each. The fourth species, *M. parvulus*, is widely distributed in the rest the archipelago. Their evolutionary process could be described as an allopatric model (Figure 1.7; Arbogast *et al.*, 2006; Curry, 1989). All four species share relatively conserved morphological characters, where phenotypic variations lie essentially in slight differences in coloration patterns and size (Table 1.1; Abbott & Abbott, 1978). Phylogenetically, molecular analyses revealed that the four species of Galapagos mockingbirds are comprising a well-defined single monophyletic clade. However, *M. parvulus* encompass several subspecies: *M. p. parvulus*, found in Santa Cruz, North Seymour, Isabela and Fernandina; *M. p. personatus*, found on Pinta; *M. p. barringtoni*, found on Santa Fe; *M. p. bauri,* found on Genovesa; *M. p. bindloei*, found on Marchena, Santiago and Rabida; *M. p. wenmani*, found on Wolf; and *M. p. hulli,* found on Darwin Island (Figure 1.8, Arbogast *et al.*, 2006). This whole monophyletic group is nested within the traditional genus *Mimus*, suggesting a single colonization of the archipelago followed by diversification. Mitochondrial analyses suggested that the colonization event happened about 3.5 MYA (Hoeck *et*

al., 2010; Lovette *et al*., 2012). Two independent phylogenetic reconstructions of the genus *Mimus* support a sister relationship between the Galapagos mockingbirds and the Bahamas mockingbirds (*M. gundlachi*) of the West Indies, rather than the long-tailed mockingbird (*M. longicaudatus*) or any of the other species presently found on the South American mainland. Within the Galapagos archipelago, four distinct mitochondrial DNA clades were identified. However, these clades differ from the above nominal species in two respects: firstly, the birds in the eastern islands (Española, San Cristobal, and Genovesa) have very similar mitochondrial DNA sequences, despite belonging to three different taxa. Secondly, the mitochondrial haplotypes of Isabela are clearly divergent from all other *M. parvulus* haplotypes (Figure 1.8). Genealogical analyses of mitochondrial genes suggest that diversification of mockingbirds within Galapagos proceeded primarily along two independent tracks, both generally south or southeast to north or northwest following the postulated directionality of prevailing winds (Arbogast *et al.*, 2006). These phylogenetic relationships suggest a colonization history similar to that proposed for Darwin's finches (Burns *et al.*, 2002), in which dispersal of mockingbird ancestors located in Central America and the Caribbean resulted in the colonization of Galapagos and a continental expansion in the Americas (Arbogast *et al.*, 2006).

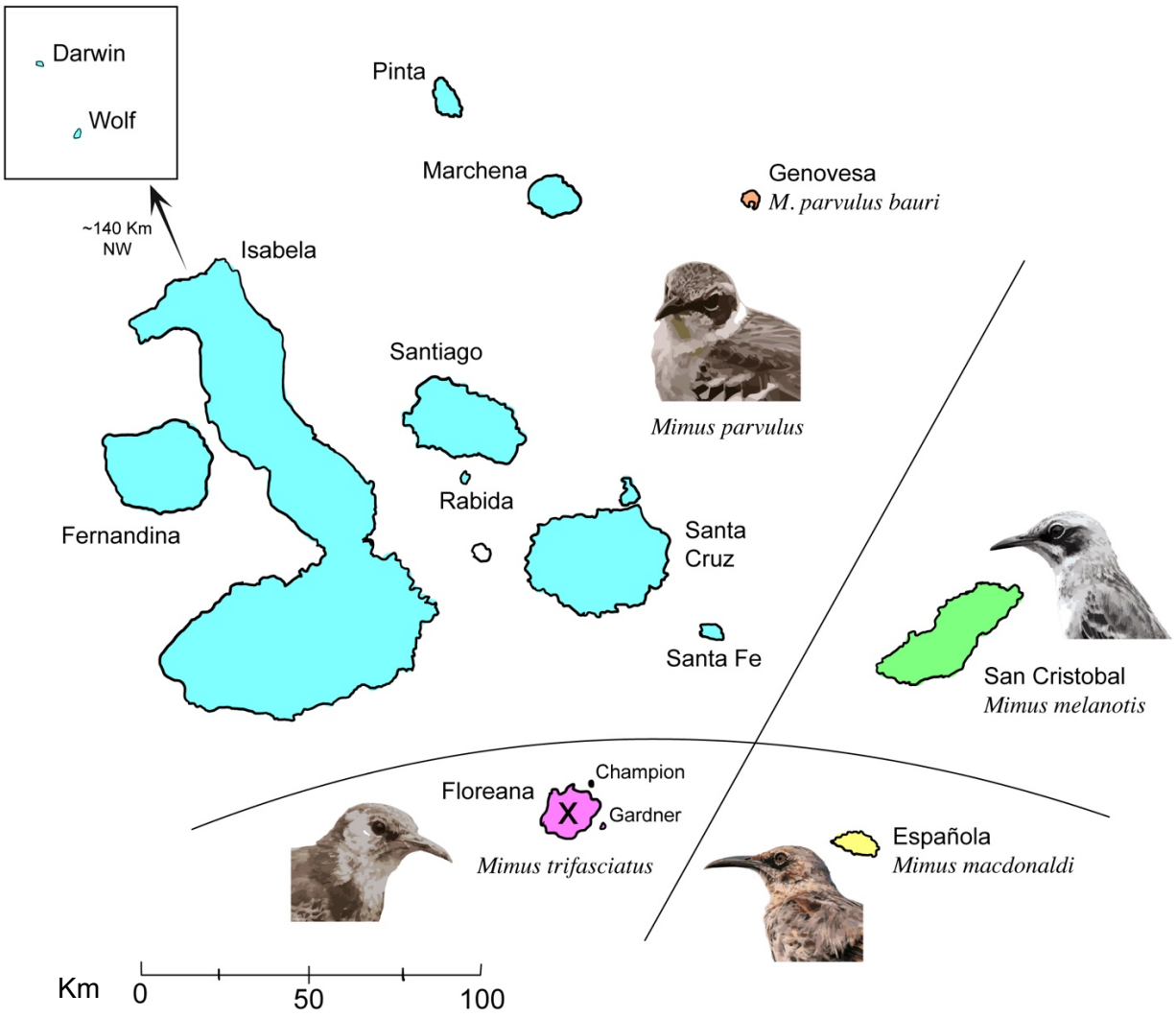


FIGURE 1.7. Distribution of mockingbird species within the Galapagos archipelago. Species are color-coded for improved reference. In the mid-1800s, *Mimus trifasciatus* went extinct on Floreana, but the species managed to survive on two small islets (Champion and Gardner) that are located near the main island.

TABLE 1.1. Morphology and diet comparison of the four Galapagos mockingbird species (Abbott & Abbott, 1978; Kleindorfer *et al.*, 2018). Species illustrations creation by myself from my own material and copyright free images.

Mimus parvulus

M. trifasciatus

M. melanotis

M. macdonaldi

Distribution	Whole archipelago except in Pinzon and where the other mockingbird species occur.	Formerly the Floreana mockingbird but it was extinct by the end of XIX century and the species survive in two Floreana satellite islets, Champion and Gardner.	San Cristobal	Española
Bill	M: 26.68 F: 25.43	33.40 31.26	28.83 27.20	37.83 35.82
Tarsus	M: 36.17 F: 34.32	40.60 38.33	37.52 35.89	38.54 36.52
Wing	M: 111.9 F: 103.9	124.7 117.6	113.5 105.4	124.7 115.4
Body Mass	~50-60g	~55-70g	~45-55g	~60-80g
Diet	Generalist, this broadly includes invertebrates, smaller vertebrates, carrion, fruits, pollen, nectar, eggs.			
Diet Peculiarity	On Santa Fe, the diet may include blood from iguanas.			Diet includes blood from seabirds, iguanas, and sea lions.

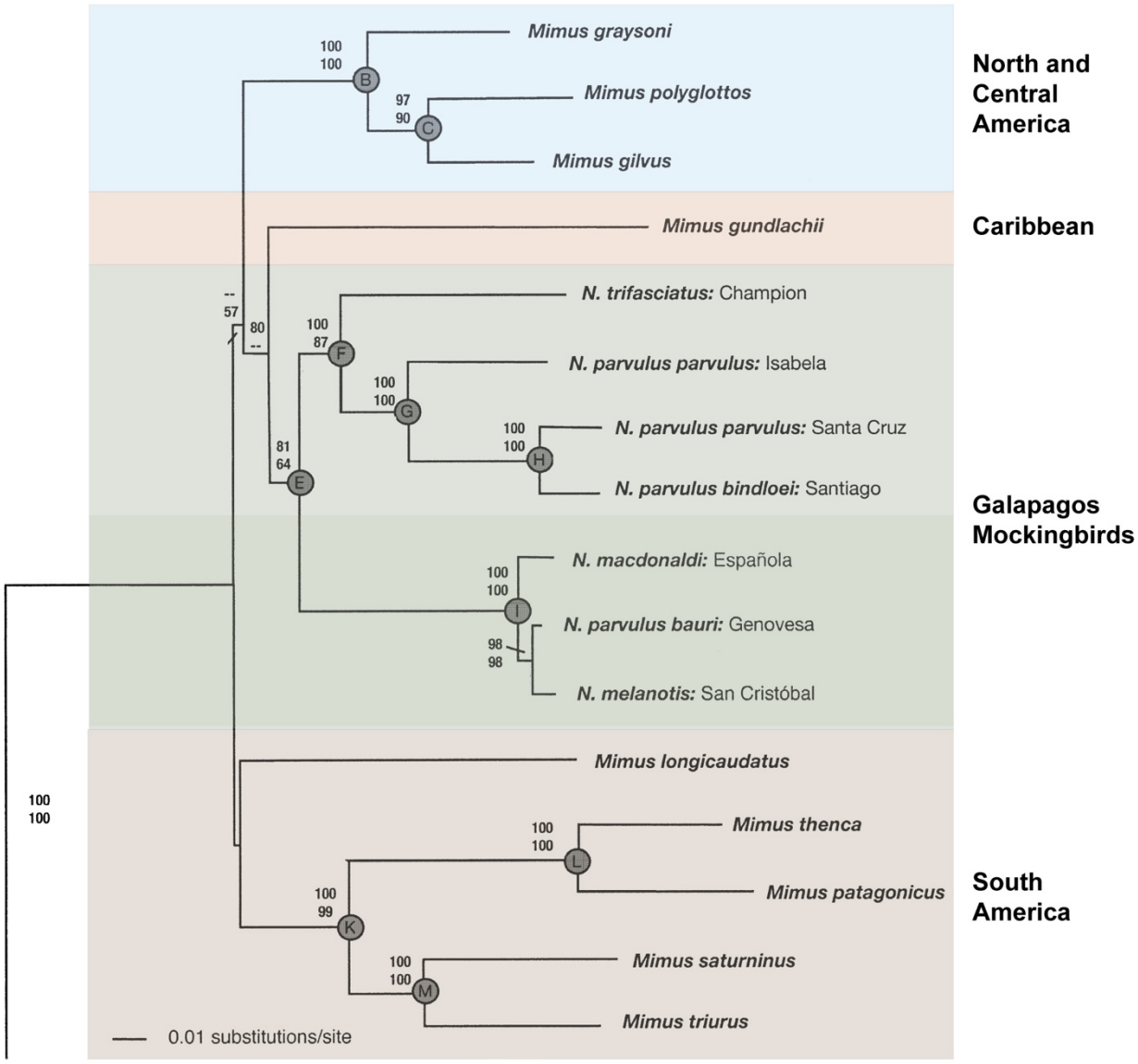


FIGURE 1.8. Rooted maximum-likelihood (ML) phylogeny of mockingbirds based on the combined ND2, COI, COII, tRNA-Lys, ATPase 6, and ATPase 8 mitochondrial genes. *Toxostoma curvirostre* and *Toxostoma rufum* were the taxa used as outgroup. This figure was modified from Arbogast *et al.* (2006).

1.3.2. *Conservation and significance*

Globally, one in eight bird species is threatened, and more specifically one-quarter of terrestrial bird species are threatened. These proportions represent an even more dramatic scenario in restricted and isolated habitats (BirdLife International, 2017). Thus, research, management, and conservation of the unique bird species in the Galapagos Islands have become increasingly critical to ensure their survival and population viability in face of growing human impact on the archipelago.

The terrestrial birds of the Galapagos Islands are mainly passerines and many of these species are threatened. The main threats to these bird species have been caused by introduced predators and diseases, which together with the loss of natural habitats and incidental accidents typical of human activities, have led to a decrease in the natural population sizes and the extinction of several species (Jiménez-Uzcátegui *et al.*, 2019). According to the IUCN (2017), of the 28 terrestrial bird species, 14 are classified as threatened. The condition of threatened species has become even more critical due to isolation and population size. The isolation generates a barrier for the replacement of individuals and long-term population maintenance. On the other hand, population size will depend on the island area and resources availability that the species needs, and populations restricted to small islands will be more sensitive to environmental and demographic changes. The most threatened Galapagos passerine species are mainly finches and mockingbirds. Of the 17 species of finch that exist in Galápagos, eight are threatened, and of the four currently recognized species of mockingbirds, three are threatened (Jiménez-Uzcátegui *et al.*, 2019; Wiedenfeld & Jiménez-Uzcátegui, 2008).

The four species of the Galapagos mockingbirds have contrasting population sizes. The most widely distributed species, *M. parvulus*, has the largest effective population size (*Ne*), and for this reason, this species is classified as Least Concern by the IUCN (International Union for Conservation of Nature). Out of the remaining three species, *M. macdonaldi* has the biggest population (IUCN: Vulnerable status), followed by *M. melanotis* (IUCN: endangered) and *M. trifasciatus* (IUCN: Critical Endangered). The population history of this latter species, the Floreana mockingbird, is particularly concerning. The last specimen was collected in 1852, and it is assumed that the last individuals disappeared around 1880, when it was thought to have become extinct. However, a population of *M. trifasciatus* was rediscovered on Gardner-by-Floreana in 1897, and later, on Champion. Both are two small satellite islets surrounding Floreana. Even though these two populations are surviving remnants of the Floreana mockingbird species, these are still on the verge of extinction. For example, Champion has the most reduced population with an annual mean of approximately only 25 breeding pairs (Dvorak *et al.*, 2017; Jiménez-Uzcátegui *et al.*, 2010).

Population genetic theories suggest that the fixation of slightly deleterious mutations is determined by the product of effective population size (*Ne*) and selection coefficient (*s*) (Santiago & Caballero, 2016). However, to date, little attention has been paid to decipher the genomic landscape of deleterious mutations, an insidious consequence of population fragmentation and the preponderance of genetic drift in small populations. Within populations, the genetic diversity is strongly correlated with island size, suggesting that Galapagos mockingbirds experience very little gene flow among islands, and therefore, are an ideal system to study the conservation consequences of extremely small population sizes (Hoeck *et al.*, 2010).

Although hybrids have traditionally been undervalued in conservation and are often viewed as a threat to pure species, the conservation value of hybrids and hybridization is gaining recognition (Jackiw *et al.*, 2015). Nietlisbach *et al*. (2013) proposed that the Genovesa population of *M. parvulus* may have undergone introgressive hybridization with other mockingbird species in the Galapagos. Despite this potential hybridization, the Genovesa population still exhibits the same physical characteristics as other populations of *M. parvulus*. This discovery implies that the population on Genovesa may represent a hybrid lineage that is currently undergoing a distinct process of speciation. Without a genome wide study of introgression patterns, it is not possible to assess to what extent the introgression has contributed to the morphological distinctiveness of Genovesa mockingbirds. If morphological distinctiveness and reproductive isolation of Genovesa mockingbirds from the parental species was increased by hybridization, this could potentially be a case of incipient homoploid hybrid speciation, the process in which hybridization results in a stable, fertile, and reproductively isolated hybrid lineage where there is no change in ploidy. The extent to which this mechanism contributes to the origin of new bird species remains to be determined. So far, only a recent lineage of Darwin's finches on the island of Daphne Major seems to fit the definition of a hybrid species where reproductive isolation is directly due to hybridization (Ottensburgh, 2018). In this case, hybridization has resulted in a transgressive bill morphology that separates this hybrid species ecologically and reproductively (through differences in song) from its parental species (Lamichhaney *et al*. 2018).

Galapagos mockingbirds are an ideal group for the understanding of evolutionary processes in islands. The differentiation between populations (and, later, between species) is based on reproductive isolation. This isolation can occur under different biological processes (e.g., disruptive selection) or physical conditions (e.g., geographical barrier) (Grehan, 2001; Hermansen

et al., 2011; Wolf & Ellegren, 2017; Woolfit & Bromham, 2005). Galapagos mockingbirds are generalist organisms with versatile behavioral plasticity. Therefore, it is assumed that selection pressures are not the primary driving force behind evolutionary differentiation processes. Instead, the fundamental evolutionary explanation for the population differentiation and speciation of these organisms on the islands rests on geographic isolation and restricted migration (Curry & Grant, 1990; Hoeck *et al.*, 2010; Kleindorfer *et al.*, 2019; Štefka *et al.*, 2011). Therefore, the evolution of mockingbird populations would be linked to the geographic characteristics of each island (e.g., size and isolation) and its historical geo-morphological variations (e.g., land growth-contraction and land connectivity). Consequently, current mockingbird populations show contrasting sizes which are the result of their past demographic history, which have been shaped by temporal geographic variations that could alter the population size and/or the gene flow (i.e., migration) probabilities between populations. In this way, Galapagos mockingbirds are not only an excellent system to understand the effect of drift on long-term survival of small populations but also to study the role of hybridization in adaptive introgression and speciation.

In conclusion, Galapagos mockingbirds provide a valuable system for investigating the impact of drift and inbreeding across varying population sizes, as well as for elucidating the patterns of colonization and the role of geographic dynamics in the evolutionary history of not only mockingbirds, but also other species on the archipelago. This thesis utilized a novel whole-genome approach, never before applied to this species system, to comprehensively address various questions related to the demographic and evolutionary factors, and provide unprecedented insight into the natural evolutionary history of these species. The application of genomics in studying the evolution of these species represents a groundbreaking advance, providing an unprecedented level of detail and reliability in our understanding of the factors driving their evolution. This novel

approach has allowed for a more thorough exploration of the complex interplay between demographic and evolutionary factors in shaping the natural history of these species. In these terms, my specific objectives were:

- To generate the first reference genome for the Galapagos mockingbird species (and for all genus *Mimus*), which will serve as a tool for further genomic analyses within this taxonomic group. By doing so, a robust framework can be established to reconstruct demographic parameters and test evolutionary hypotheses (**Chapter 2**). A reference genome will provide the possibility to test hypotheses relating to the evolutionary history and demographic parameters of mockingbird populations (e.g., examine the patterns of colonization and the relative effects of drift and inbreeding in different population sizes). Specifically, it is predicted that the reference genome would provide an opportunity to reveal previously unknown relationships between mockingbird populations and provide a more complete understanding of the genomic and evolutionary factors that have shaped this unique avian radiation. The availability of a reference genome will also enable future studies on the functional genomics of mockingbirds, including the identification of genes involved in adaptation, speciation, and other biological processes. Ultimately, the generation of a mockingbird reference genome will provide a valuable resource for the scientific community, and will help to advance our understanding of the natural evolutionary history of this iconic group of birds.

- To investigate the relationship between demographic history and effective population size (*Ne*) of mockingbird species and populations in the Galapagos archipelago (**Chapter 3**). Based on the hypothesis that the contrasting effective population sizes of mockingbirds are

the result of unique past demographic histories and evolutionary adaptations to current island constraints. It is predicted that demographic history analyses of eight mockingbird populations with differing sizes will reveal differences that correlate with spatio-temporal factors such as island size and isolation, island age, past connectivity between islands, and human impact. It is expected that some mockingbird species and populations may have experienced population size fluctuations or founder events, leading to reduced effective population sizes. Conversely, other species and populations may have larger effective population sizes due to a more stable demographic history. By identifying and analyzing these differences, this study aims to provide valuable insights into the evolutionary history of mockingbirds in the Galapagos archipelago.

PREFACE TO CHAPTER 2: TRANSITION STATEMENT

A reference genome of a Galapagos mockingbird is needed to accurately reconstruct the evolutionary history and demographic parameters of mockingbird populations in the Galapagos archipelago, including patterns of colonization and the effect of drift and inbreeding in different population sizes. However, there is no suitable reference genome available for this purpose. Therefore, in the next chapter, a *de novo* reference genome will be developed. This will be the first reference genome for the mockingbird species in the genus *Mimus* and will serve as a tool for further genomic analyses within this taxonomic group and establish a robust framework to reconstruct demographic parameters and test evolutionary hypotheses. The generation of this mockingbird reference genome will represent a significant contribution to the scientific community by advancing our understanding of the natural evolutionary history of this iconic group of birds.

CHAPTER 2: *DE NOVO* ASSEMBLY AND ANNOTATION OF THE GALAPAGOS MOCKINGBIRD GENOME

2.1. INTRODUCTION

The instruction manual of every living organism, as well as its history and even its most hidden secrets are written in its genome. From this conception, the need to reveal and understand the information written nucleotide by nucleotide in the different genomes has become one of the greatest ambitions of science. To this end, the development of tools that would allow the determination of the nucleotide sequence in a given fragment began. This is how the Sanger or Maxam – Gilbert sequencing, or what is now known as first-generation sequencing, was established in 1975 (Wink, 2019). In first-generation sequencing, the determination of the sequence was based on the interrupted replication of a DNA fragment for each nucleotide (sequence-by-synthesis; SBS), that then had to be encoded by an analysis of fragment sizes in a gel of high resolution of polyacrylamide, which was the biggest limitation when working with fragments of more than 100 bp (Hohenlohe *et al.*, 2018; Weissensteiner *et al.*, 2017). About 10 years later, another SBS method was developed that was known as pyrosequencing. This second-generation sequencing method is based on the light reaction produced by each pyro-nucleotide when reacting in replication on a luciferase substrate. The light reactions could be recorded and ordered in real time to obtain the sequence of the fragment, which later led to the creation of the first automated sequencing equipment (Wink, 2019). By the early 2000s, a new approach to sequencing based for the first time on single molecule sequencing (SMS) emerged, leading to third-generation sequencing. This sequencing method is characterized by the real-time reading of each nucleotide without the need for a DNA replication reaction (DNA amplification). This

technology is based on the automated reading of specific fluorescent markers for each nucleotide throughout DNA template. It has been extensively refined and optimized during the last decade allowing to generate reads of hundreds of base pairs in less time and at a very low cost. The development of this optimization is led to what is called next-generation DNA sequencing (NGS) (Weissensteiner *et al.*, 2017; Wink, 2019). Today, NGS technology has led to unprecedented advancements that were unimaginable just 10 years ago. In this way, during the last decade the largest number of completely sequenced genomes has been produced, which has ushered in the Whole-Genome Sequencing (WGS) era and the development of a new generation of statistical and bioinformatics tools. Specifically, population genomics based on WGS is allowing to obtain much more detailed and precise estimates and models in different fields of study such as ecology, evolution, conservation biology, among others (Dodgson, 2003; Hohenlohe *et al.*, 2018; Weissensteiner *et al.*, 2017). The ability to simultaneously analyze a large number of loci throughout the entire genome in several individuals, makes it possible to clarify evolutionary processes such as mutations, gene drift, gene flow or selection patterns with more precision (Black *et al.*, 2001; Dodgson, 2003; Luikart *et al.*, 2003).

Along these huge technological advances in whole-genome sequencing, birds were one of the focal taxonomic groups in the implementation of these technologies. Birds are the most diverse group of organisms among the vertebrates with more than 10,000 species (Wink, 2019). However, the era of genomics for this group began with a single species, the chicken (*Gallus gallus*). The chicken is a world-wide commercial species and a model organism in several research fields, and an ideal candidate for genome sequencing (Wallis *et al.*, 2004; Vignal & Eory, 2019). The knowledge achieved in its genetics due to basic, breeding and production research, established the foundations for the understanding of the particularities and organization of the bird genome. This

began in the early 1990s with the long sequence construction (contigs of ~650 Kb) of the chicken genome using BAC library mapping. Later, in 2004, with combined efforts under the ENCODE project, the whole-genome sequence of *G. gallus* was released (Dodgson, 2003; Wallis *et al.*, 2004; Vignal & Eory, 2019). Consequently, the chicken genome became the model genome for birds, and an indispensable tool for the assembly of new avian genomes, and to examine questions relating to the evolution, population genetics and conservation of different bird species. However, after this remarkable achievement and its wide use, during the last decade the concept of model genome started to be abandoned little by little, since the novel NGS technology led to produce longer and more reliable sequences that can be assembled without using a reference genome (Wink, 2019; Vignal & Eory, 2019). The first bird genomes belonged to different taxonomic groups focused mostly on domestic or species of commercial interest. However, initiatives like the B10K Genomes project and other individual efforts have increasingly boosted the availability of new avian genomes (including wild species) (Zang, 2015). Currently, 835 genomes representing 557 bird species have been published, which would represent about 5% of the total diversity of birds. Of these published genomes, 35 (4%) have been published at the contig assembly level, 665 (80%) at the scaffold level, and only 135 (16%) have been published at the chromosome level (NCBI Genomes database: Sayers *et al.*, 2022). Accordingly, even though the complete sequences of several genomes have now been published, the vast majority (~84%) are still halfway to being considered fully assembled genomes.

Bird genomes share unique features in that they have been conserved since their early evolutionary origin. It is estimated that the origin of modern birds was 150 million years ago, during the Jurassic period, but their diversification began during the Cretaceous period (Damas *et al.*, 2019). During their diversification period, after the great extinction of the dinosaurs (~66

MYA), modern birds spread rapidly throughout the world. Their accelerated expansion gave rise to high rates of diversification and with it to extreme morphological changes. Despite their evolutionary history, high diversification and morphological differentiation, modern birds are the group with the smallest genomes among vertebrates, and their genomes are highly conserved (high synteny). The mean size of bird genomes is on average 1.35 Gbp, ranging in size from 0.9 Gbp (black-chinned hummingbird, *Archilochus alexandri*) to 2.1 Gbp (ostrich, *Struthio camelus*). It has been proposed that the reduction in size of bird genomes is related to the high metabolic requirements that are necessary for flight, which is supported by the fact that flightless birds have the largest avian genomes (Gregory, 2005; Hughes & Friedman, 2008). Furthermore, the same pattern can also be evidenced in bats that have substantially reduced the size of their genomes compared to other mammals (Gregory, 2002; Gregory, 2005; Hughes & Friedman, 2008). However, this hypothesis remains under active controversy since apparently the reduction of their genomes occurred prior to their ability to fly (Tiersch & Wachtel, 1991; Organ *et al.*, 2007). Nevertheless, the reduced size of the bird genomes compared to other amniotes (such as mammals or reptiles), is due to 1) they have shorter repetitive elements and, in less quantity, 2) their introns are shorter (consequently, their genes are also shorter), and also 3) they have fewer transposable elements (TE) (Hughes & Hughes, 1995; Hughes & Piontkivska, 2005; Organ *et al.*, 2007). Furthermore, it was proposed that birds had fewer genes than mammals or reptiles, since it was reported that approximately 274 gene blocks identified by synteny in tortoises, lizards and humans were not present in birds (Lovell *et al.*, 2014). However, a later study reported that these “missing genes” were present, first discovered by their expression in RNAseq data and later locating them "hidden" in GC-rich regions (which are difficult to access regions and unorthodox gene location), which explains why they were not found before using gene seek traditional parameters (Botero-

Castro *et al.*, 2017). Another quasi-unique feature of an avian genome is its karyotype, both due to its high number of chromosomes, as well as its chromosomal arrangement. Bird karyotype shows a remarkable consistency among all species, which apparently has not varied much from the same common ancestor of all birds. In the same way, recombination sites are highly conserved and the degree of synteny is high among all avian genomes (Burt, 2002; Ellegren, 2010). The number of chromosomes in birds is generally 2n = 80, and these can be divided into two groups, macrochromosomes (> 20Mbp) and microchromosomes. Macrochromosomes generally represent 1/4 of the total chromosomes, while microchromosomes are the remaining 3/4. Apart from size, another thing that differentiates macro- and microchromosomes is that microchromosomes have a higher density of genes and a higher content of GC, in addition to having much higher recombination rates (Damas *et al.*, 2018; Burt, 2002; Smith *et al.*, 2000).

Bird genomes are known for their high synteny, which leads to limited variability among taxa. The coding regions in these genomes are typically well-conserved and arranged in a consistent structure, while the number of neutral (non-coding) regions is relatively low. However, relying on only a few genomic regions or markers can introduce significant biases or generate artifacts that depend on the type of data collected (known as "data-type effects", Hohenlohe *et al.*, 2018). This issue is particularly relevant in organisms or populations with low genetic variability, where the use of a small number of genetic markers can lead to misleading results (Braun *et al.*, 2019; Nadeau & Kawakami, 2018). The Galapagos mockingbirds are a good example of this problem, where systematic genetic studies show various incongruities in the phylogenetic relationships based on specific mitochondrial and nuclear markers (Arbogast *et al.*, 2006; Hoeck *et al.*, 2010; Nietlisbach *et al.*, 2013). Furthermore, the genetic analyses carried out on this species system have not achieved an optimal resolution to discriminate the genetic profiles of each

population (island) or species. Therefore, the inclusion of tools such as population genomics would make it possible to cover all the variants present in the genome, and thus obtain a clearer picture of the relationships between species, populations and even between individuals. The evolutionary history of this system is still little known, as well as its demography and its conservation status, therefore, the use of genomics becomes essential to elucidate the many unknowns about this group. However, by not having a close species with a completely assembled and annotated genome, it is also essential to generate a reference genome to obtain more precise variants between species within the same clade. This reference genome starts from a WGS of one individual, where it is necessary to have high quality short-reads, and a high sequencing coverage is prioritized (several repetitions of short-reads per genome region). For the assembly of this genome there are two methodologies, 1) reference-based genome assembly and 2) *de novo* genome assembly. For reference-based assembly, as its name indicates, it requires a genome already described to build the new genome based on that template. On the other hand, the *de novo* assembly refers to the assembly itself without any reference in between, where the short-reads will be paired and aligned to build large scaffolds and thus generate the new genome. The advantage of generating a *de novo* genome, compared to reference-based assembly, is that the introduction of artifacts or bias regarding the creation or elimination of genome regions that could be influenced by reference is avoided. Thereby, with access to the latest advances in genomics, I set out to generate the first-ever reference *de novo* genome for the *Mimus* genus. I propose that this could be achieved due to the high synteny of bird genomes, which would facilitate the validation of the assembly and annotations. By taking advantage of the high synteny between bird genomes, it would be ensured a comprehensive validation of the genome. Therefore, the objective of this chapter was to create a fully assembled and annotated *Mimus* genome and validate it using comparative genomics

analyses with other bird species genomes. This validated reference *de novo* genome can serve as a reliable baseline for future genomic analyses to uncover the evolutionary history and ecology of the Galapagos mockingbirds.

2.2. METHODS

2.2.1. *Sample collection*

The sampling for this study was carried out under permission MAE-DBN-CM-2016-0041 from the Ministry of Environment of Ecuador between 2017 and 2018. To ensure representative sampling, a target sample size of 30 individuals per population (island) was set, however, the actual sample size was adjusted based on factors such as population size and accessibility. The individuals were captured using tomahawk traps, which proved to be more effective and precise compared to traditional techniques like mist nets. Galapagos mockingbirds are naturally inquisitive and forage on the ground, making them susceptible to simple baiting techniques, such as using shiny objects, bread or containers with water, which were used to attract the birds without causing any harm. Each captured individual was measured, banded, and had a blood sample taken for subsequent analysis. Blood samples for genomic DNA and total RNA were collected *in situ* (and *in vivo*) by a venipuncture of trapped birds on several expeditions to Galapagos Islands. Detail information about the capture and handling of specimens, including sample collection, can be found in Animal Use Protocol #20180088. All animal procedures were conducted in accordance with the NIH Guide for the Principles of Animal Care and were approved by the Animal Ethics Committee of the University of Saskatchewan. A total of 139 individual Galapagos mockingbirds were sampled from six different islands, with 13 individuals from San Cristobal (*Mimus melanotis*), 31 from Santa Cruz (*M. parvulus*), 28 from Genovesa (*M. parvulus*), 30 from Española

(*M. macdonaldi*), 14 from Champion (*M. trifasciatus*), and 23 from Gardner-by-Floreana (*M. trifasciatus*). The sample used for genome assembly and annotation belongs to a single San Cristobal Mockingbird (*M. melanotis*), from the first set of samples collected in 2017 (sample code SK5, Supplementary Table 2.1). The sample was stored in two different tubes, one with ethanol 70% and other with RNA-later, prior to nucleic acid extraction procedures.

2.2.2. *Sample processing and genomic and transcriptomic libraries sequencing*

The high molecular weight DNA extraction was performed using the MagAttract HMW DNA Kit (Qiagen), whereas the Tri-Reagent (Molecular Research Center, Cincinnati, OH) protocol was used for total RNA extraction. For the DNA libraries, I used Chromium technology by 10x Genomics. Chromium is an emulsion-tagging library creation platform uses >1,000,000 unique barcodes for highly parallel sample partitioning and molecular barcoding to generate Illumina-ready libraries. DNA Chromium creates "linked reads" (barcoded sets of reads) from single, ultra-long molecules; the barcodes are used to link sequencing reads to the originating molecule (Weisenfeld *et al.*, 2017). These linked reads facilitate whole-genome phasing, structural variant detection, and *de novo* genome assembly. The sample was distributed in four lanes for barcoding, and the libraries were constructed using a nominal coverage of 56X. The obtained libraries were sequenced at the Cornell Sequencing center (BRC) on a NextSeq500 creating short-reads of 150 bp in both senses (forward and reverse). For the RNA libraries, after checking the quality of the RNA extraction (Agilent Bioanalyzer Agilent Technologies, Wilmington, DE), the RNA sample was used to generate an RNAseq library using the Illumina NEBNext Ultra II Directional RNA kit, with polyA+ selection (Illumina, San Diego, CA). Libraries were cleaned using AMPure XP and sequenced on a single Illumina NextSeq500 (TruSeq barcode 33). The

Chromium library construction process is thoroughly described and accompanied by visual aids in Weisenfeld *et al.* (2017) publication. To understand the downstream process after obtaining the raw-reads sequenced, please refer to Figure 2.1, which outlines each step of the process.

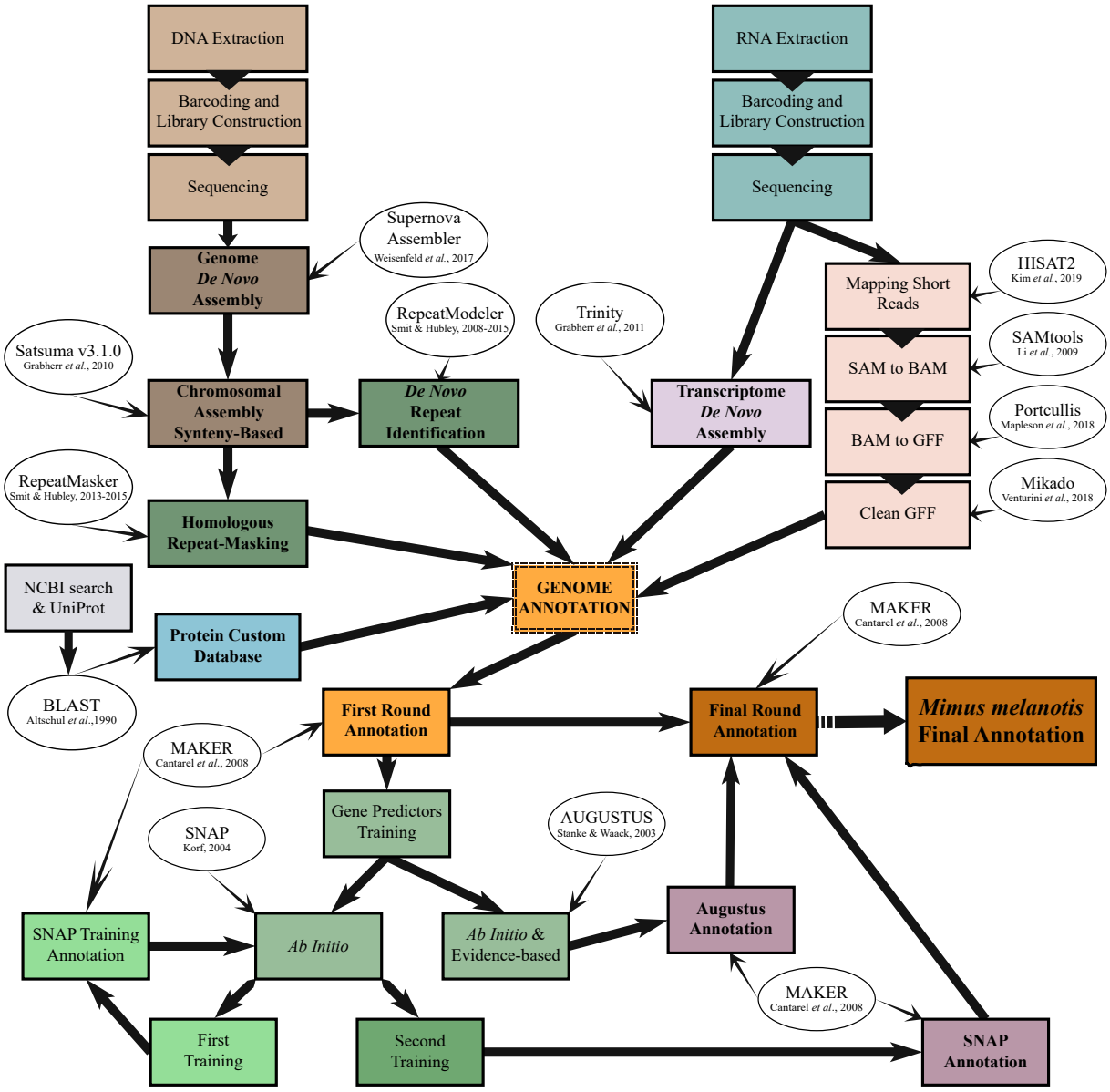


FIGURE 2.1. Methodological process map developed from obtaining the high molecular weight DNA sample to the final version of the assembled and annotated *Mimus melanotis* genome. Each

box describes the process carried out and in the attached circles are the software used for each process.

2.2.3. *Genome assembly*

Sequences obtained in four Chromium libraries were assembled using the Supernova assembler ver. 1.2.2 and the default parameters (Weisenfeld *et al.*, 2017). For the final genome sequence, I used the supernova argument --style=pseudohap to obtain a single consensus sequence for each pair-ended scaffold. Once the final scaffolds were obtained, I performed the chromosomal assembly based on synteny using Satsuma (ver 3.1.0; Grabherr *et al.*, 2010). Specifically, I used the Chromosembler module, which maps the scaffolds based on synteny, and the collared flycatcher's (*Ficedula albicollis*) genome (FicAlb1.5) as reference. The additional parameter used in Chromosembler was -s to run the full analysis of synteny coordinates between both species. Then, the genome completeness of the resulting chromosomal assembly was assessed with the BUSCO tool (ver. 5.1.0, Simão *et al*., 2015) using the *Aves Odb9* lineage dataset from Metazoa database (OrthoDB v10). Finally, for comparative purposes of the final chromosomal assembly, I carried out further simplified synteny analyses with Satsuma (SatsumaSynteny module) using three other bird species with fully assembled genomes: zebra finch (*Taeniopygia guttata*), hooded crow (*Corvus cornix*) and chicken (*Gallus gallus*). In addition, I ran a pairwise whole-genome comparison using Chromeister (Pérez-Wohlfeil, 2019), where I also included one reptilian (*Chrysemys picta*) and two mammalian (*Mus musculus* and *Homo sapiens*) genomes in the analyses.

2.2.4. *Genome annotation*

I used the MAKER genome annotation pipeline (ver. 3 Cantarel *et al.,* 2008) to carry out the whole genome annotation. For the first run of the analyses, I included several databases as reference sources to annotate all probable elements in the genome. The included databases were: 1) custom reference protein database, 2) repetitive elements database, and 3) own genomic expression databases (RNA-seq and transcriptome) and "hidden" genes transcripts (Botero *et al.*, 2017). The custom reference protein database was created using the Swiss-Prot database (598,558 proteins including isoforms; The UniProt Consortium, 2018) and a custom protein database of closely-related species (*Ficedula albicollis*: 27,008 proteins; *Taeniopygia guttata*: 23,445 proteins; *Corvus* spp.: 75,840 proteins). All proteins were merged into a single FASTA file to use it as one of the input files in MAKER pipeline.

For analyses conducted using the repetitive elements database, I performed a first annotation of repetitive regions in the genome following both, *ab initio* and reference-based methodologies. For the *ab initio* predictions, using RepeatModeler (ver. 2.0.1, Smit *et al.*, 2008-2015), I created a custom database looking for all possible repetitive motifs in the genome using the BuildDatabase module with NCBI search engine. Then, using RepeatModeler and the custom database, I obtained a final library with all repetitive elements found in the genome. For the evidence-based predictions, using RepeatMasker (ver. 4.1.0, Smit *et al.*, 2013-2015), I carried out three independent runs to mask repetitive elements to *M. melanotis* genome by homology, one of each using the following libraries: a) amniote (Repbase library), b) chicken (Repbase library), and c) the previous custom library generated with RepeatModeler for *M. melanotis*. Then, output files of the three independent runs were merged and were processed to create a single annotation file

using the module ProcessRepeats from RepeatMasker and tetrapoda library from Repbase as the complete containing library. The resulting annotation (GFF) file with all repetitive regions was another input file for MAKER pipeline.

Finally, for the genomic expression database, the RNA-seq short reads obtained from the blood sample were processed using two different approaches, I) mapping the RNA short reads to the genome and II) assembling a transcriptome. Mapping was performed using HISAT2 (Kim *et al.*, 2019), and the resulting SAM file was transformed to a BAM file using SAMtools (Li *et al.*, 2009). Then, using the Portcullis python script (Mapleson *et al.*, 2018), the BAM file was converted to a GFF file. Lastly, the GFF file was filtered using Mikado (Venturini *et al.*, 2018) to eliminate false junctions and precisely identify splicing sites and coding regions. The clean GFF was used as another input file for the MAKER pipeline. For the transcriptome assembly, I ran a *de novo* transcriptome assembly analysis using TRINITY v.2.8.4 (Haas *et al.*, 2013) with default settings. Then, the assembly control quality was carried out using Bowtie2 (ver. 2.3.0; Langmead *et al.*, 2009) under default parameters to map the RNA-reads to the assembled transcriptome to test their belonging. I assessed transcript completeness using the BUSCO tool (ver. 5.1.0, Simão *et al*., 2015) with the same genome arguments. The *de novo* transcriptome FASTA file was used as another input file for the MAKER pipeline. Finally, I included other transcripts from the database corresponding to the "hidden" bird genes reported by Botero *et al.* (2017), in the input files for the MAKER pipeline analyses to obtain the first-round annotation file.

The first MAKER GFF output file (first-round annotation file) formed the basis for gene predictor training. Gene prediction was carried out using *ab initio* and reference-based approaches to adjust the homology-based gene models from the first-round and to predict unidentified/missing

genes. For the *ab initio* gene prediction, using SNAP (Korf, 2004), I adjusted all the annotation discrimination patterns from the first *M. melanotis* annotation to generate a SNAP parameter (HMM) file. Using the HMM file as a sole reference input, I ran the data using the MAKER pipeline to obtain a new GFF file (SNAP training annotation - first round). Then, using the SNAP first-round annotation, I repeated the previous step to get a new HMM file to run MAKER once again and improve the gene predictions (SNAP training annotation - second round), obtaining the SNAP final annotation. Additionally, for the *ab initio* and reference-based gene predictions, I used the program AUGUSTUS (Stanke & Waack, 2003), which was trained using the first *M. melanotis* annotation to create an *ab initio* prediction library (*M. melanotis* library). Then, using *M. melanotis* (*ab initio*) and chicken (reference) as input references in the AUGUSTUS library, I ran MAKER to get the annotation file with AUGUSTUS gene predictions (AUGUSTUS annotation). Finally, using first-round annotation, SNAP final annotation and AUGUSTUS annotation files as input, I ran MAKER once again to generate the final annotation for *M. melanotis*.

2.2.5. *Annotation post-processing*

For the quality control of the *M. melanotis* final annotation, I assessed the Annotation Edit Distance (AED) to know how the level of support for the annotations. It is generally accepted that AED values > 0.5 indicate that an annotation is highly supported, with a high confidence in the annotation location. I also tested the completeness of the resulting transcripts and proteins using BUSCO. Ultimately, I carried out the MAKER post-processing pipeline to format all annotation names under the prefix Nmel and get the putative functions of annotated genes based on homology. For putative functions, I included in the pipeline analyses, a protein BLAST search using the following parameters: -evalue 1e-6 -max_hsps 1 -max_target_seqs 1 -outfmt 6, to get the putative

functions of the genes. I also used InterProScan under the following parameters: -appl pfam -dp -f TSV -goterms -iprlookup -pa -t p, to get the putative functional domains of each protein/gene.

2.3. RESULTS AND DISCUSSION

2.3.1. *Genomic DNA sequencing and genome assembly*

After construction of the 10X Genomics library and conducting the Illumina sequencing, I obtained 830.42 million reads with a mean length of 139.50 bp, resulting in an effective coverage of 42.43x (Table 2.1). For a 1 Gb size bird genome, the results showed an optimal coverage in terms of quantity, quality, and integrity (mean length of reads) of whole genome according to 10x Genomics standards (Weisenfeld *et al.*, 2017)

TABLE 2.1. Illumina sequencing statistics summary for *Mimus melanotis* (SK5) genome.

Parameter	**Estimation**	**10x Genomics Reference**
Number of Reads	830.42 M	Ideal 800M-1200M for human
Mean Read Length (after triming)	139.50 bp	Ideal 140
Effective Coverage	42.43x	Ideal ~42 for nominal 56x coverage
Phred Quality Score Q30	74.03%	Ideal 75-85
Proper Pairs	87.86%	Ideal >= 75
N50 reads per barcode	964	NA
Duplicated reads	35.09%	NA

Using the generated reads, Supernova assembled 830 long scaffolds (≥10 kb), and the N50 for scaffold size was 7.55 Mb (Table 2.2). Using only long scaffolds, the total assembly size was 979.35 Mb, which is within the range of known bird genome sizes (Vignal & Eory, 2019). For comparative quality proposes, the Contig N50 is a metric used to evaluate the quality of genome

assembly, with a higher N50 indicating a more contiguous assembly. B10K genomes is a project aimed at sequencing and assembling the genomes of all known bird species. Their analysis shows that the Contig N50 mean from all bird genomes assembled at scaffold level is 67.49 Kb. In contrast, the Contig N50 of the *M. melanotis* genome is 74.47 Kb (Figure 2.2), indicating a more contiguous assembly than the average bird genome. This difference may suggest that *M. melanotis* that the assembly process was more successful for this particular genome compared to the mean other B10K bird genomes. Additionally, the comparison with full-assembled model genomes, shows that the parameters obtained for the *M. melanotis* assembly were not far apart from the parameters reported for these fully develop genomes (Table 2.3). Furthermore, the completeness assessment using BUSCO showed that the genome completeness score was ~90%, where the remaining percentage were ~4% of fragmented elements and ~6% of missing elements. For eukaryotic genomes, a good quality assembly should have a completeness score of least 85-90% (Simão *et al.*, 2015; Veeckman *et al.*, 2016). Therefore, my results showed that the assembled *M. melanotis* genome is reliable with good readability, low fragmentation, and contains few ambiguities or missing elements (Waterhouse *et al.*, 2018).

TABLE 2.2. Supernova assembler stats summary from *Mimus melanotis de novo* genome assembly.

INPUT
- 830.42 M = READS   = number of reads; ideal 800M-1200M for human
- 139.50 b = MEAN READ LEN = mean read length after trimming; ideal 140
- 42.43 x = EFFECTIVE COV = effective read coverage; ideal ~42 for nominal 56x cov
- 74.03 % = READ TWO Q30 = fraction of Q30 bases in read 2; ideal 75-85
- 394.00 b = MEDIAN INSERT = median insert size; ideal 0.35-0.40
- 87.86 % = PROPER PAIRS = fraction of proper read pairs; ideal >= 75
- 55.83 Kb = MOLECULE LEN = weighted mean molecule size; ideal 50-100
-   1.83 Kb = HETDIST   = mean distance between heterozygous SNPs
-   6.55 % = UNBAR   = fraction of reads that are not barcoded
- 964.00   = BARCODE N50   = N50 reads per barcode
- 35.09 % = DUPS   = fraction of reads that are duplicates
- 29.47 % = PHASED   = nonduplicate and phased reads; ideal 45-50
OUTPUT
- 830.00   = LONG SCAFFOLDS = number of scaffolds >= 10 kb
- 16.56 Kb = EDGE N50   = N50 edge size
- 74.47 Kb = CONTIG N50   = N50 contig size
-   2.76 Mb = PHASEBLOCK N50 = N50 phase block size
-   7.55 Mb = SCAFFOLD N50   = N50 scaffold size
- 979.35 Mb = ASSEMBLY SIZE = assembly size (only scaffolds >= 10 kb)

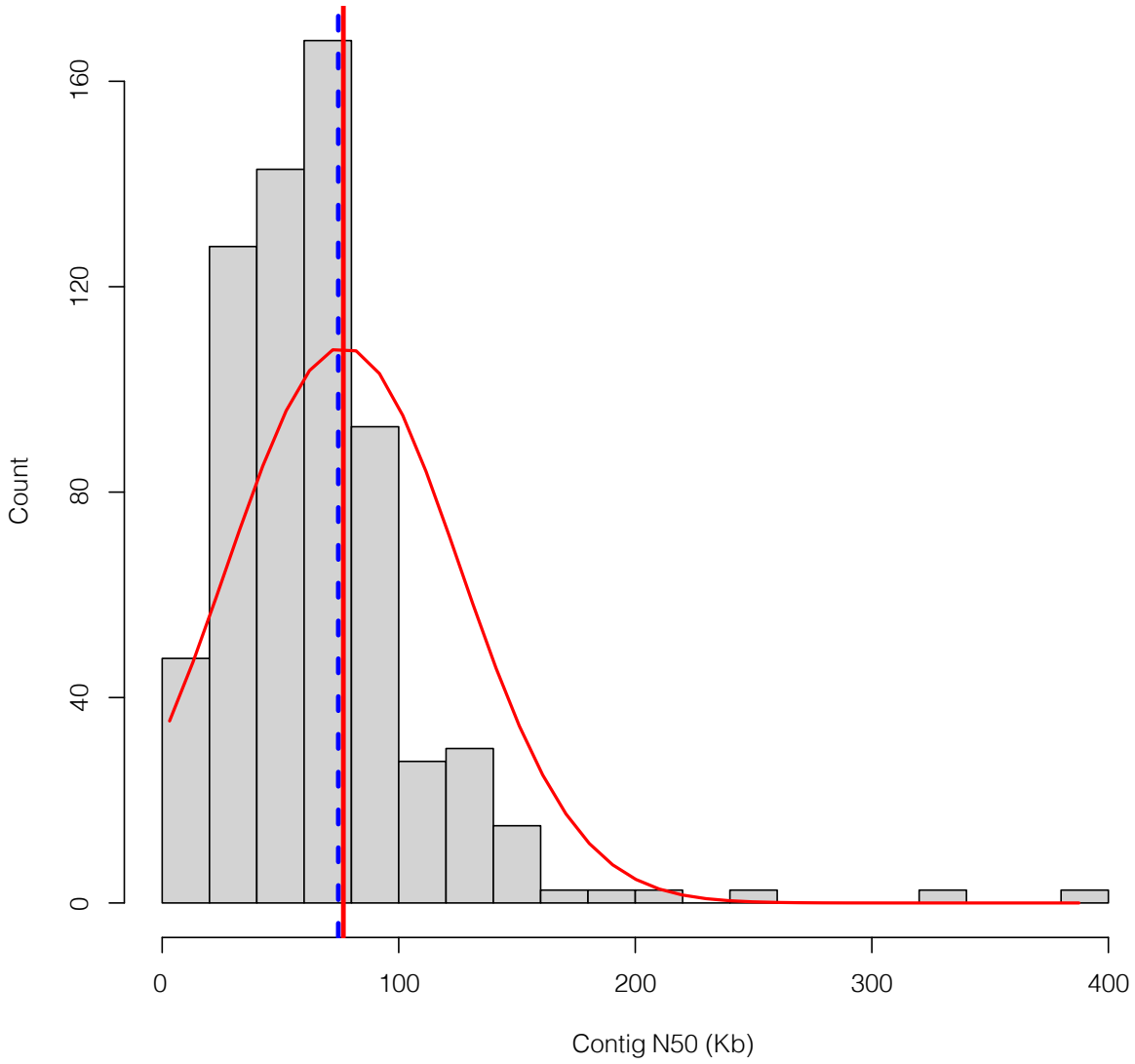


FIGURE 2.2. Contig N50 histogram with data recovered from B10K genomes database. The data considers only genomes assembled at scaffold level. Contig N50 mean (67.49 Kb) is represented with the blue dashed line, while the red line is the Contig N50 value (74.47 Kb) obtained from the *Mimus melanotis de novo* assembly.

TABLE 2.3. *Mimus melanotis* genome assembly statistics compared with other full-assembled bird genomes submitted at the NCBI.

Parameter	*Mimus melanotis*	*Ficedula albicollis* (GCA_000247815.2)[1]	*Corvus cornix* (GCA_002023255.2)[2]	*Taeniopygia guttata* (GCA_003957525.2)[3]	*Gallus gallus* (GCA_016699485.1)[4]
Coverage	42.43x	60x	52x	88.2x	102.01x
Assembly size (Mb)	1,013.37	1,118.34	1,050.11	887.57	1,053.33
Scaffold N50 (Mb)	7.55	6.54	18.37	3.56	90.86
Scaffolds (≥10 Kb)	830	1,168	145	199	214

[1]Ellegren *et al.*, 2012; [2]Weissensteiner *et al.*, 2017; [3]Warren *et al.*, 2010; [4]Vertebrate Genomes Project, 2021

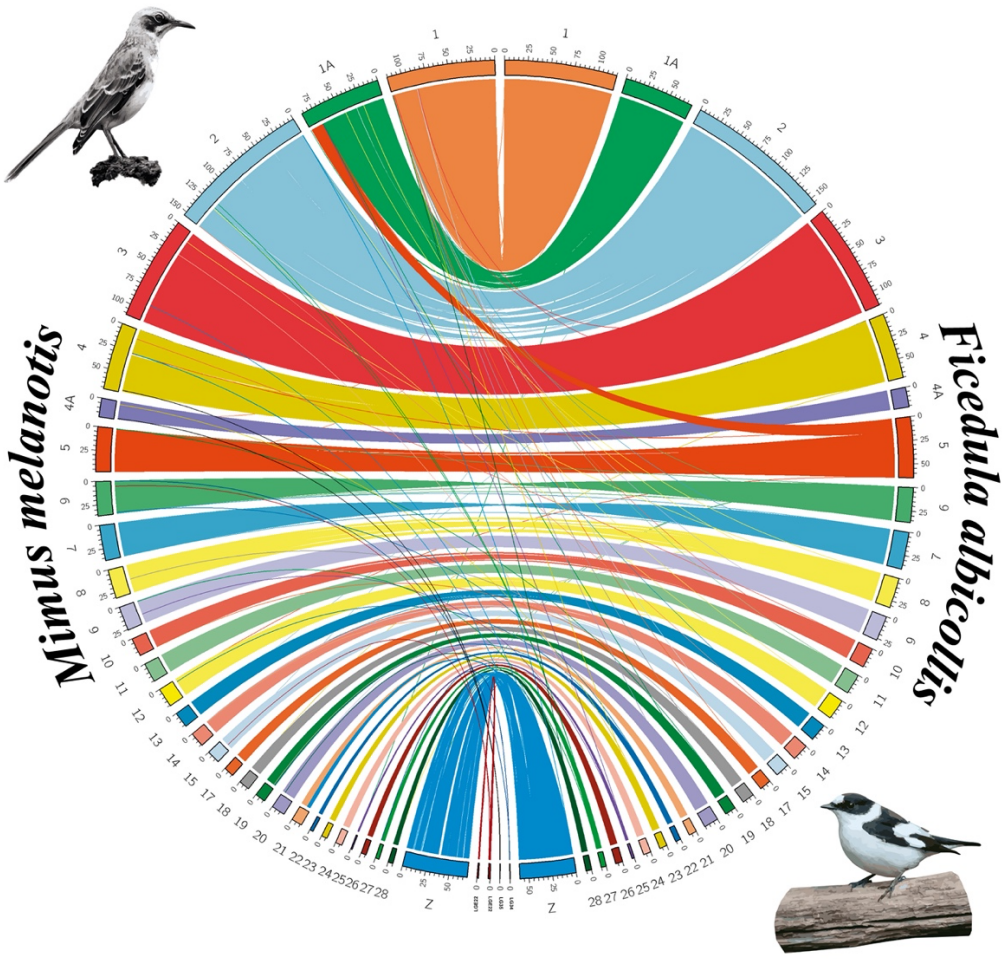
The Satsuma pseudochromosomal assembly built 30 pseudochromosomes, four superscaffolds and left five apart unplaced scaffolds based on synteny with *F. albicollis* (Table 2.4). Synteny analysis between *F. albicollis* and *M. melanotis* showed few rearrangements, with few chromosomal translocations and some inversions (Figure 2.3). As expected, given the assumption that birds have highly conserved genomes, the results of the additional synteny analyses performed with the other bird species revealed a high degree of synteny between these genomes. Specifically, the most chromosomes in the other bird species showed the same sequence arrangement and structure as the corresponding chromosomes in the *M. melanotis* genome (Figure 2.3-5). However, despite the high synteny between genomes, the results also showed that there were a few different chromosomal rearrangements for each species regarding the total length of translocated loci (Figure 2.4 and 2.5). Apparently, the number of chromosomal rearrangements increased when comparing more evolutionarily distant species from *M. melanotis*. In this way, the total length of translocated chromosomal regions in whole genome decreases from the most distant to the closest related species in relation to *M. melanotis* (chicken [2.6%] > zebra finch [2.1%] >

hooded crow [1.7%] > collared flycatcher [0.5%]) (Figure 2.4). In the same way, the Chromeister genome dissimilarity matrix analysis shows an evident correlation between the phylogenetic relatedness and the synteny between genomes. Accordingly, the dissimilarity score is clearly low between bird species (<0.3), where the same order of relationship between taxa is evident as it was seen for the total length of translocated loci (chicken [0.280] > zebra finch [0.257] > hooded crow [0.188] > collared flycatcher [0.129]). While the dissimilarity scores with respect to organisms of other orders (Reptilia and Mammalia) exceed 0.9, which proves that at this level the genomes are highly different (Figure 2.6). These results support the assumption that the degree of synteny is correlated with the degree of phylogenetic relatedness between species, since the degree of synteny between two genomes decreases as the evolutionary distance between the two species increases. This is because as species diverge from each other over time, their genomes accumulate differences, including inversions, deletions, duplications, and rearrangements, which can alter the order and orientation of genes and other genomic features. (Ellegren, 2010).

TABLE 2.4. Chromosome length comparison between the actual zebra finch (*Ficedula albicollis*) chromosomes and San Cristobal mockingbird (*Mimus melanotis*) pseudochromosomes generated by synteny.

Ficedula albicollis			***Mimus melanotis***	
Chromosome	**Code**	**Length (bp)**	**Pseudochromosome**	**Length (bp)**
Chr1	NC_021671	120,002,344	Mmel_Chr1	115,798,444
Chr1A	NC_021672	74,947,036	Mmel_Chr1A	87,609,823
Chr2	NC_021673	157,563,209	Mmel_Chr2	153,799,878
Chr3	NC_021674	115,844,353	Mmel_Chr3	111,983,055
Chr4	NC_021675	70,439,523	Mmel_Chr4	71,597,153
Chr4A	NC_021676	21,182,716	Mmel_Chr4A	19,968,083
Chr5	NC_021677	64,724,594	Mmel_Chr5	47,752,317
Chr6	NC_021678	37,227,452	Mmel_Chr6	36,241,581
Chr7	NC_021679	39,412,007	Mmel_Chr7	37,569,041
Chr8	NC_021680	32,100,816	Mmel_Chr8	30,711,958
Chr9	NC_021681	26,793,321	Mmel_Chr9	25,617,771
Chr10	NC_021682	21,346,708	Mmel_Chr10	19,561,496
Chr11	NC_021683	21,727,166	Mmel_Chr11	21,132,581
Chr12	NC_021684	21,938,106	Mmel_Chr12	20,951,658
Chr13	NC_021685	18,641,552	Mmel_Chr13	17,215,816
Chr14	NC_021686	17,374,186	Mmel_Chr14	15,961,263
Chr15	NC_021687	14,943,019	Mmel_Chr15	13,772,992
Chr17	NC_021688	12,378,331	Mmel_Chr17	11,163,097
Chr18	NC_021689	13,163,162	Mmel_Chr18	11,527,533
Chr19	NC_021690	11,933,672	Mmel_Chr19	11,259,476
Chr20	NC_021691	15,675,940	Mmel_Chr20	15,132,754
Chr21	NC_021692	8,073,070	Mmel_Chr21	11,735,686
Chr22	NC_021693	5,733,621	Mmel_Chr22	4,007,987
Chr23	NC_021694	7,944,683	Mmel_Chr23	6,045,787
Chr24	NC_021695	8,009,359	Mmel_Chr24	7,125,346
Chr25	NC_021696	2,802,420	Mmel_Chr25	1,594,678
Chr26	NC_021697	7,653,694	Mmel_Chr26	5,706,761
Chr27	NC_021698	5,572,044	Mmel_Chr27	4,287,364
Chr28	NC_021699	6,182,350	Mmel_Chr28	5,193,359
ChrZ	NC_021700	59,856,998	Mmel_ChrZ	69,700,058
LG34	NC_021701	476,164	NA	--
LG35	NC_021702	248,039	NA	--
LGE22	NC_021703	2,153,636	Mmel_LGE22	1,351,029
NA	--	--	Mmel_UnplacedScaffold1	152,949
NA	--	--	Mmel_UnplacedScaffold2	10,220
NA	--	--	Mmel_UnplacedScaffold3	13,907
NA	--	--	Mmel_UnplacedScaffold4	10,735
NA	--	--	Mmel_UnplacedScaffold5	25,144
TOTAL		**1,044,065,291**		**1,013,368,168**

A)



B)

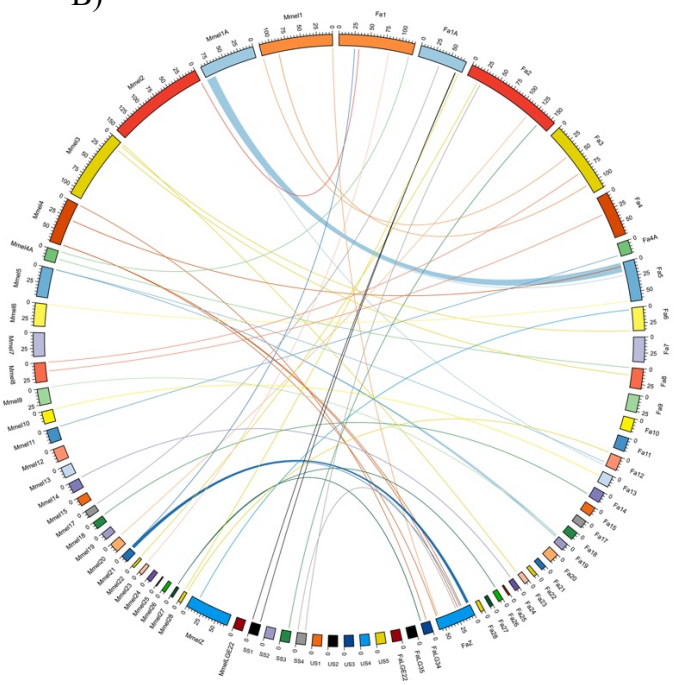


FIGURE 2.3. A) Whole chromosomal arrangement correspondence based on synteny between *Ficedula albicollis* and *Mimus melanotis*. B) Only syntenic translocated loci between *F. albicollis* (Fa) and *M. melanotis* (Mmel) genomes, the total length of these loci comprises ~0.5% of *M. melanotis* genome. Species illustrations creation by myself from my own material and copyright free images.

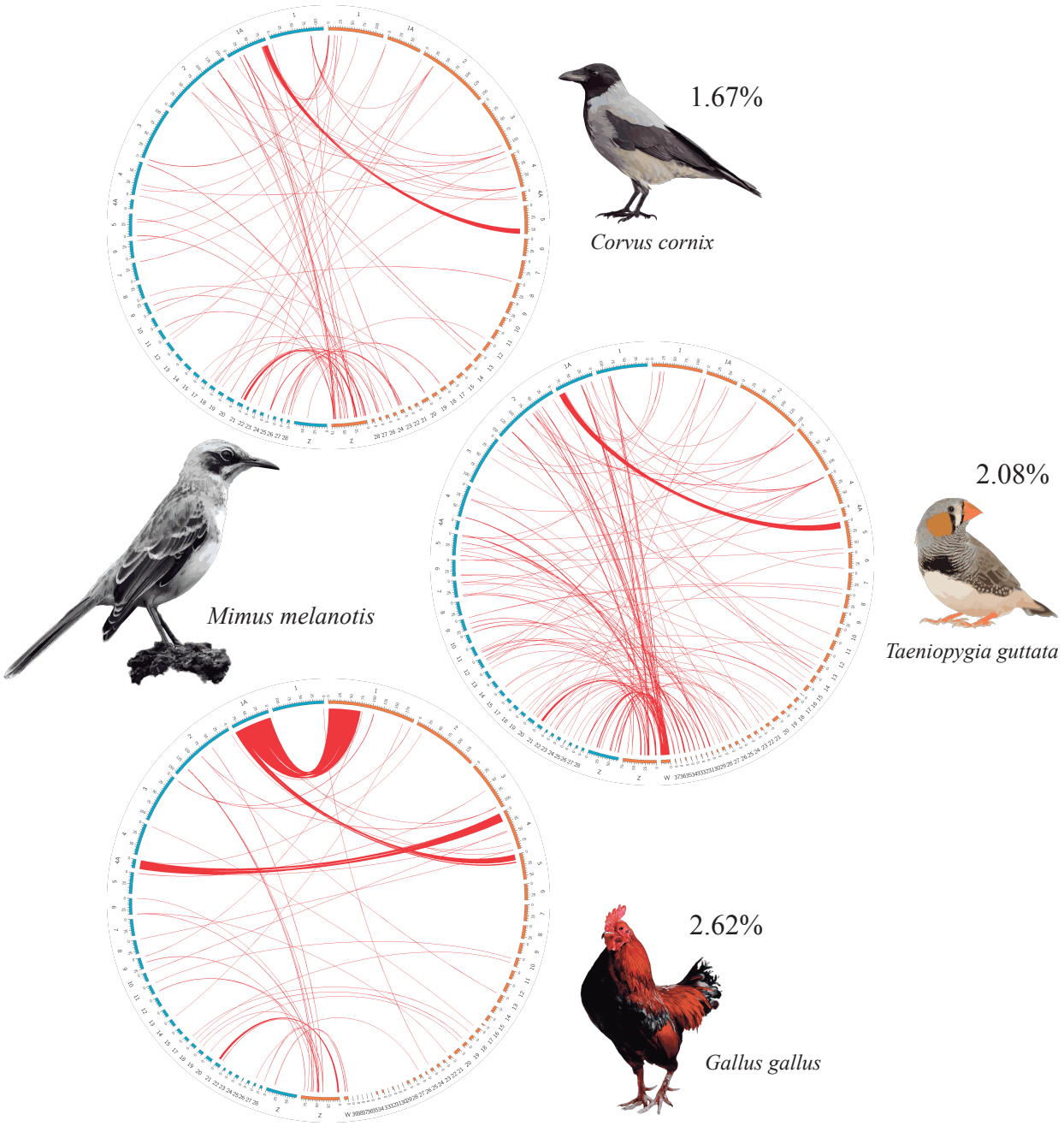


FIGURE 2.4. Synteny-based translocation correspondence between genomes of *Mimus melanotis* (San Cristobal mockingbird, blue karyotype) and: a) *Corvus cornix* (hooded crow, orange karyotype), b) *Taeniopygia guttata* (zebra finch, orange karyotype), and c) *Gallus gallus* (chicken, orange karyotype). Red links show the homologous loci translocated between genomes. For each

comparison, the percentage values refer to the total length of translocated sites in *M. melanotis* whole-genome length. For this graphic representation, only the defined chromosomes were considered, leaving aside Linkage Groups (LGs) and unplaced scaffolds, however, for the total estimation of translocated sites (percentage), the whole-genome was included. Species illustrations creation by myself from my own material and copyright free images.

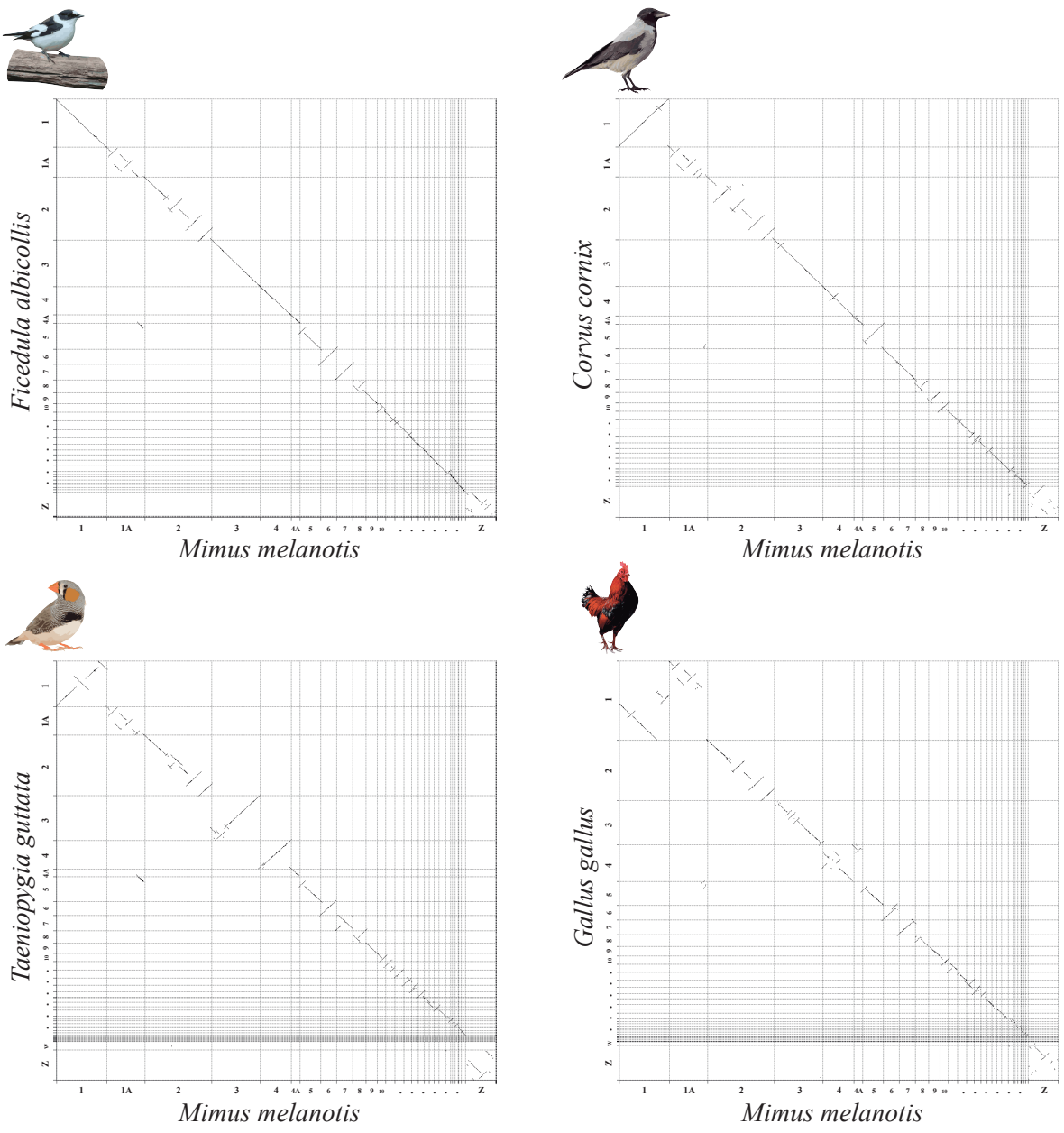


FIGURE 2.5. Pairwise genome-wide comparison of homologous sites between *Mimus melanotis* and: a) *Ficedula albicollis* (collared flycatcher), b) *Corvus cornix* (hooded crow), c) *Taeniopygia guttata* (zebra finch), and d) *Gallus gallus* (chicken). On the X-axis, the chromosomes of *M. melanotis* are arranged following the order of the autosomal chromosomes and ends with the sex chromosomes, while on the Y-axis, the chromosomes of each compared species are arranged in

the same way. The grid shows the limits of each chromosome, and the points show the homologous sites between both genomes. Descent patterns to the right show homologous regions with conserved order, ascent patterns to the left show inversions, and broken patterns with dots distributed in different chromosomes show translocations. Species illustrations creation by myself from my own material and copyright free images.

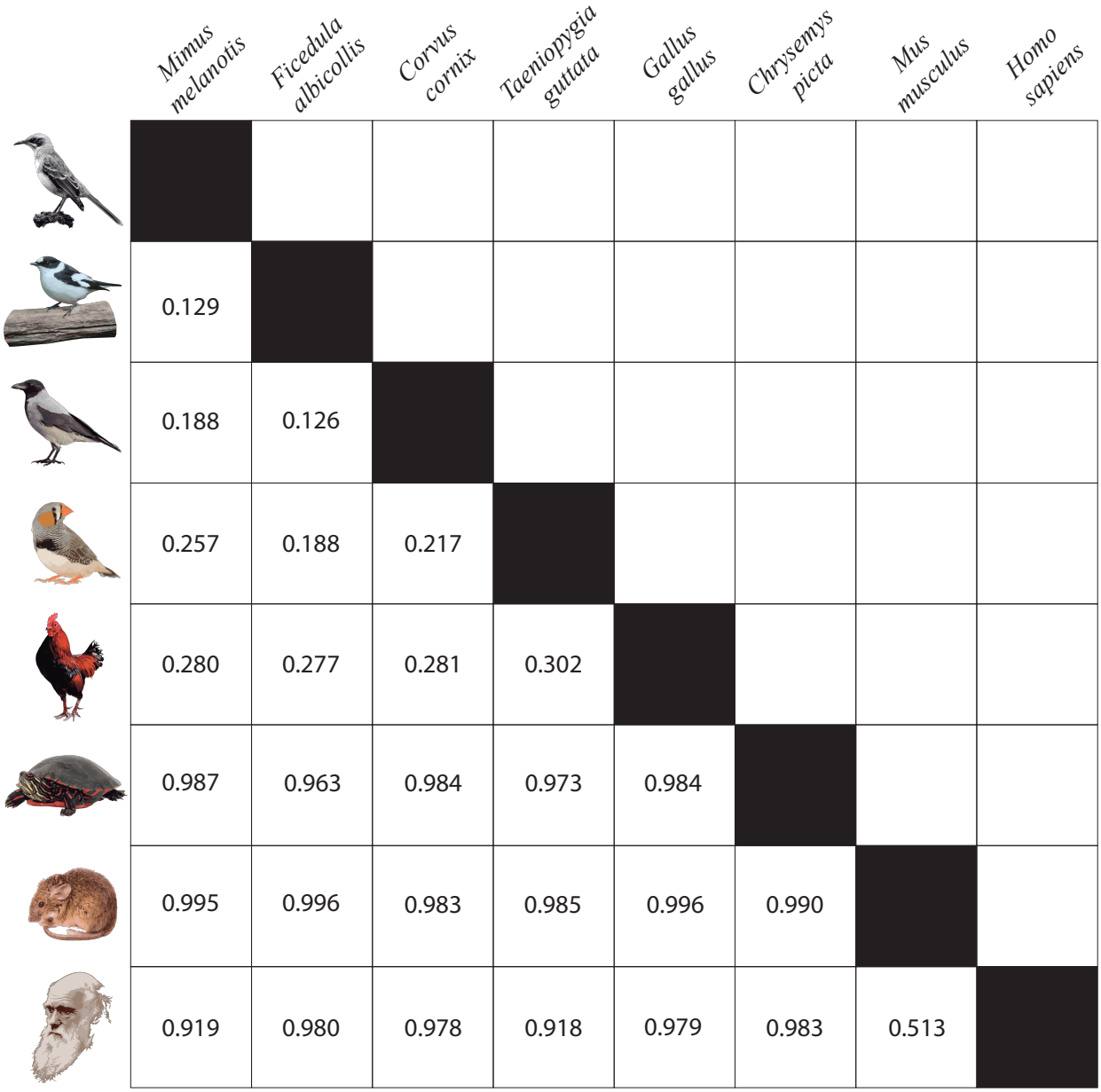


FIGURE 2.6. Whole-genome pairwise dissimilarity score between different organisms based on Chromeister (Pérez-Wohlfeil *et al.*, 2019) matrix estimation. Where a value of 0 corresponds to a comparison with exactly the same genome and a value of 1 corresponds to two completely different genomes. The compared species correspond to organisms with fully assembled genomes at the chromosomal level, which include five species of birds (including *Mimus melanotis*), one species

of reptile, and two species of mammals. Species illustrations creation by myself from my own material and copyright free images.

2.3.2. *Repetitive elements, RNA short-reads mapping and transcriptome assembly*

The final RepeatMasker run revealed that all repetitive elements that were found in the *M. melanotis* genome just represent over 6% of the genome (Table 2.5). Using data for four additional bird species genomes (*Ficedula albicollis*, *Taeniopygia guttata*, *Corvus cornix* and *Gallus gallus*), one reptile genome (*Chrysemys picta*) and two mammals genomes (*Mus musculus* and *Homo sapiens*), I ran a subsequent comparative analyses with RepeatMasker with Amniota Repbase (as a basal reference base for all these taxa). The results of these analyses consistently showed that bird genomes contain reduced repetitive regions covering less than 13% of whole genome. On the other hand, for the genomes of the reptile and two mammal species, repetitive regions exceed 40% of the whole genome (Table 2.6). Commonly, bird genomes have a low quantity of repetitive elements compared to other vertebrates. While the repetitive regions in birds represent, on average, less than 20%, those in mammals exceed 45% (Weissensteiner & Suh, 2019).

TABLE 2.5. Details of repetitive elements found in *Mimus melanotis* genome.

Repetitive Element	*Number of Elements*	*Length Occupied (bp)*	*Percentage (%)*
Retroelements	***153,825***	***52,652,847***	***5.20***
SINEs	6,827	765,802	0.08
Penelope	159	30,529	0.00
LINEs	118,543	36,491,270	3.60
L2/CR1/Rex	116,996	36,285,143	3.58
R1/LOA/Jockey	1	51	0.00
R2/R4/NeSL	26	10,320	0.00
RTE/Bov-B	309	91,713	0.01
L1/CIN4	1,022	69,355	0.01
LTR elements	28,455	15,395,775	1.52
BEL/Pao	11	746	0.00
Ty1/Copia	9	582	0.00
Gypsy/DIRS1	373	53,824	0.01
Retroviral	27,886	15,313,905	1.51
DNA transposons	***14,946***	***2,271,699***	***0.22***
hobo-Activator	2881	439,963	0.04
Tc1-IS630-Pogo	498	82,640	0.01
PiggyBac	3	163	0.00
Tourist/Harbinger	3658	343,628	0.03
Small RNA	***2,407***	***253,624***	***0.03***
Satellites	***750***	***62,451***	***0.01***
Simple repeats	***202,444***	***8,797,542***	***0.87***
Low complexity	***44,380***	***2,246,109***	***0.22***
Unclassified	***2,433***	***426,844***	***0.04***
TOTAL		***66,552,543***	***6.57***

TABLE 2.6. Repetitive regions masked with RepeatMasker (Smit *et al.*, 2013-2015) in seven genomes of different species (four birds, one reptile and two mammals) and in *Mimus melanotis* genome using the Amniota Repbase for comparative purposes.

Species	Common name	Assembly	Total Length	GC%	Total of repetitive regions masked (%)	Retroelements (%)	DNA Transposons (%)
Mimus melanotis	San Cristobal mockingbird	PRJNA655505	1,013,368,168	41.06	6.14	4.9	0.11
Corvus cornix	Hooded crow	ASM73873v5	1,028,128,602	42.06	8	6.5	0.4
Gallus gallus	Chicken	bGalGal1.mat.broiler.GRCg7b	1,041,122,857	42.09	12.68	9.27	1.15
Ficedula albicollis	Collared flycatcher	FicAlb1.5	1,044,065,291	43.61	11.39	5.83	0.34
Taeniopygia guttata	Zebra finch	bTaeGut1.4.pri	1,049,956,260	41.81	10.41	8.44	0.32
Chrysemys picta	Painted turtle	Chrysemys_picta_BioNano-3.0.4	2,481,351,664	43.11	40.41%	22.4	16.35
Mus musculus	Mouse	GRCm39	2,723,414,844	41.67	42.53	36.85	0.89
Homo sapiens	Human	GRCh38.p13	3,088,269,832	40.87	48.56	40.04	3.36

Out of the total number of RNA short-reads, 86% (n=179,693,291) were aligned successfully to *M. melanotis* genome. After the downstream process, the final clean GFF file created kept 42,503 superloci (set of short-reads in a locus) annotations. On the other hand, Trinity assembled a transcriptome consisted of 92,651 contigs ("genes") with an average N50 length of 1,831 bp. Mapping the RNA short-reads against the created transcriptome using Bowtie2, showed that there was 97% overall alignment rate. Instead, mapping the transcriptome against the *M. melanotis* genome revealed an overall alignment rate of 89%. Finally, after transcripts filtering by size removing all transcripts smaller than 1000 bp, the results of the BUSCO analyses gave a transcriptome completeness score of ~88%, with ~4% indicating fragmented elements and ~8%

suggesting missing elements. These results suggest that there is a high reliability in the RNA information obtained, being one of the keystones for genome annotation.

2.3.3. *Genome annotation*

The first-round annotation (based in all evidence and reference databases) established the homologous location of 15,750 genes (mean length ~11.5 Kb). Almost 99% of the annotations were within 0.5 AED score, revealing a high accuracy in the prediction of gene models based on homology. SNAP, after the two consecutives training runs, annotated 18,948 genes (mean length ~11.2 Kb). Over 94% of the SNAP-based gene models had AED scores <0.5. AUGUSTUS, after its training and independent MAKER run, annotated 14,752 genes (mean length ~14.1 Kb). Aproximately 96% of the AUGUSTUS-based gene models showed AED score <0.5. The final annotation was performed combining all previous MAKER annotation outputs to take into account all homology-, SNAP-, and AUGUSTUS-based gene models. The final annotation run located 18,594 genes with a mean length of ~10.7 Kb, and more than 95% of the genes were under 0.5 for AED score (Table 2.7, Figure 2.8). The annotated genes *M. melanotis* genome match the expected amount, length, and distribution of genes found in other bird genomes. Normally, the number of genes reported for bird genomes are within 15,000 to 30,000, and gene mean length is about 10 to 12 Kb (Vignal & Eory, 2019). Additionally, I ran a BUSCO analysis to assess the completeness of the final transcripts and proteins from the last annotation, which resulted in a completeness score of 94% (2.5% fragmented and 3.5% missing) for transcripts and 91% (4.5% fragmented and 5% missing) for proteins. The completeness results showed that most of the annotated genes are well supported by orthologous references, which means these genes have all (or most) of the expected functional elements (Waterhouse *et al.*, 2018). Finally, using the post-processing MAKER scripts,

I updated the final annotation with the putative functions of 14,432 genes and I also included 25,815 putative protein functional domains distributed across the different gene locations.

TABLE 2.7. Number of model genes annotated and their mean length after each MAKER annotation round.

	First-round	*SNAP*	*AUGUSTUS*	*Final*
Model Genes	15,750	18,948	14,752	18,594
Mean Gene Length (bp)	11,541.8	11,193.7	14,100.4	10,711.5

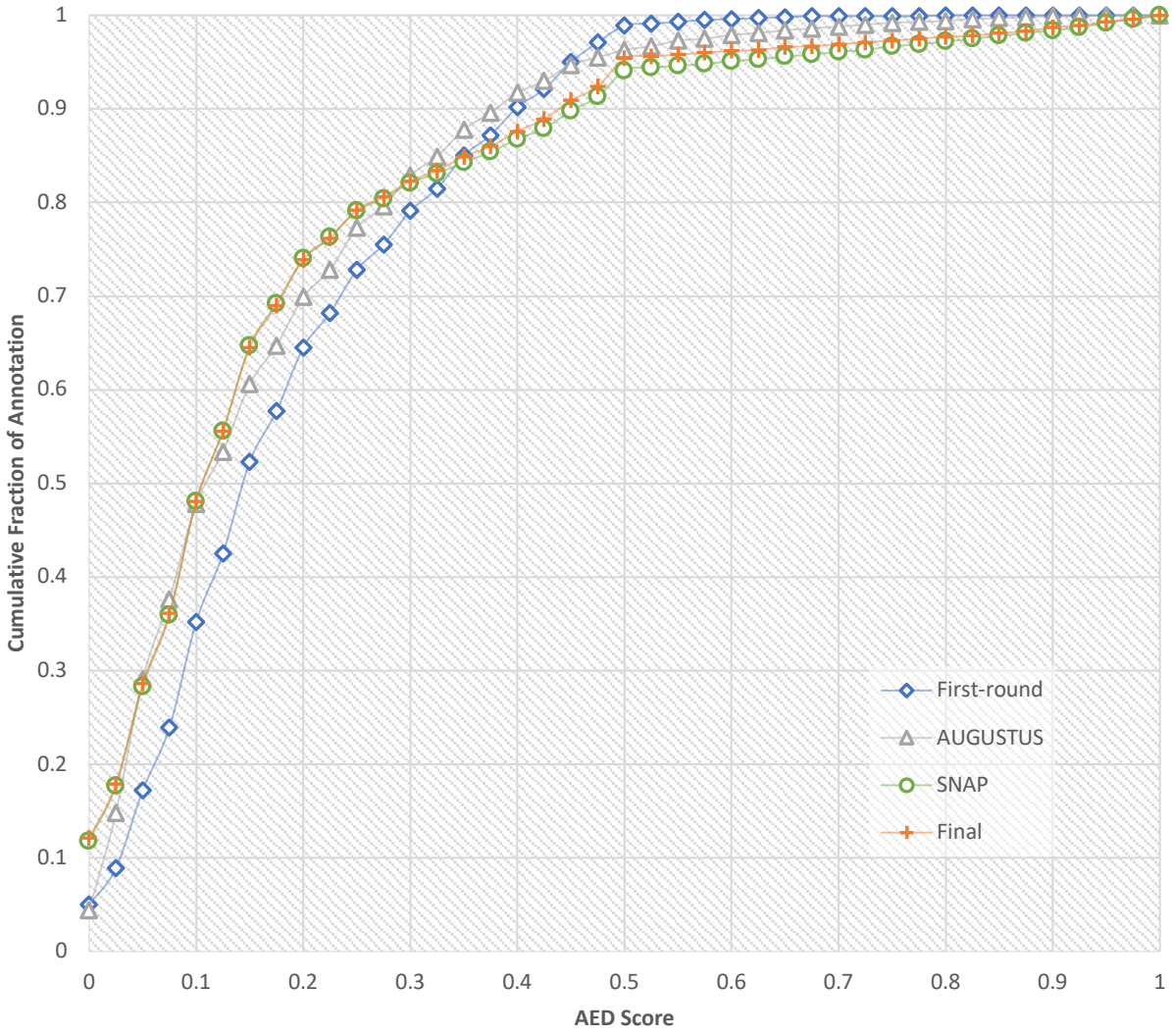


FIGURE 2.7. Cumulative fraction of annotations by Annotation Edit Distance (AED) scores for each annotation round.

2.4. CONCLUSIONS

The construction of the *M. melanotis* genome is a significant milestone for avian genomics, providing a valuable resource for studying Galapagos mockingbirds and their relatives. The initial assembly size was consistent with other bird genomes, and subsequent pseudochromosomal

assembly revealed a high degree of synteny with other bird species, enabling the reconstruction of nearly all chromosomes for *M. melanotis*. Achieving a successful assembly at the (pseudo)chromosomal level is a notable accomplishment, particularly given that only 4% of passerine species have sequenced genomes, and of these, only 30 have assembled genomes at this level. The Mimidae family, which includes the Galapagos mockingbirds and 33 other species, has only two sequenced genomes, both of which are assembled at the scaffold level. In context, from the Passeriformes, which exceed half of the total bird diversity (>5,000 species), only 234 species (~4% of passerines) have sequenced genomes. However, from these 234 species, only 30 have their genome assembled at the (pseudo)chromosomal level. As for the Mimidae family (consisting of 10 genera and ~34 species), the narrowest taxonomic group with sequenced genomes that contains the Galapagos mockingbirds, has only two species (*Toxostoma redivivum* and *Donacobius atricapilla*) with sequenced genomes, however, both are assembled at scaffold level.

The sequencing of the *M. melanotis* genome using 10X Genomics technology represents a significant advancement in our understanding of the evolution, ecology, and conservation of Galapagos mockingbirds. The use of 10x Genomics technology facilitated efficient library construction with high coverage, leading to a genome assembly with long, well-supported (high-covered) scaffolds and low fragmentation. The quality of the *de novo* assembly was exceptional, producing remarkably long genomic regions with minimal ambiguities. This allowed for a highly precise annotation of most genome elements, including coding and control regions, as well as repetitive elements.

The final version of the *M. melanotis* genome was validated through a rigorous process of synteny and comparative genomics analyses, which demonstrated exceptional levels of reliability

and support. The extensive validation process, which involved comparing the *M. melanotis* genome with other bird genomes, ensured that the final version of the genome was accurately annotated and contained minimal errors. The number of genome elements identified in the *M. melanotis* genome is similar to that found in other chromosome-level bird genomes, with approximately 20,000 genes and more than 15% repetitive elements. The lengths of these elements are also comparable to those reported for other birds. The high quality of the genome assembly and annotation ensures that all annotated elements are well-supported. As a result of this validation, I can confidently conclude that the genome is a highly dependable resource for downstream genomic analyses. Thus, the reliability and accuracy of this genome make it a valuable tool for future research on Galapagos mockingbirds, and will contribute to our understanding of their evolution, ecology, and conservation.

The generated reference genome of *M. melanotis* is of particular significance in Galapagos mockingbird research, as it provides a comprehensive view of this species' genetic makeup. With this genome, previously unknown relationships between mockingbird populations can be uncovered, offering a more complete understanding of the genomic and evolutionary consequences given by demographic, historical and environmental events. Moreover, the reference genome will allow for future studies on the functional genomics of mockingbirds, including the identification of genes involved in adaptation, speciation, and other biological processes. This resource will be invaluable to the scientific community, helping to advance our understanding of the natural evolutionary history of this iconic group of birds and to ease conservation efforts.

PREFACE TO CHAPTER 3: TRANSITION STATEMENT

Establishing the mockingbird reference genome (Chapter 2) now provides the opportunity to examine the previously unknown relationships among mockingbird populations, and to gain a more complete understanding of the genomic and evolutionary factors that have shaped this distinctive group of birds. In Chapter 3, building on the *de novo* genome generated in Chapter 2, I undertake a comprehensive population genomics analysis to investigate the population genetic dynamics and relationships of mockingbirds across the Galapagos archipelago. Through the identification and analysis of differences in effective population sizes, this study provides valuable insights into the natural evolutionary history of Galapagos mockingbirds.

CHAPTER 3: GENOMIC FOOTPRINTS OF ISLAND COLONIZATION AND ANTHROPOGENIC EFFECTS IN GALAPAGOS MOCKINGBIRDS

3.1. INTRODUCTION

Species inhabiting islands are expected to have small effective population sizes (N_e) because of the founder events associated with the initial colonization and by the relatively small area of islands imposing long-term restrictions on census population sizes (James *et al.* 2016). In such cases, demographic stochasticity causes variations in vital rates (i.e., changes the population birth and death rates) which can lead to an increased risk of extinction (Hastings & Harrison, 1994; Woolfit & Bromham, 2005). Beyond its pure demographic effects, a small population size, may also cause island endemic species to be genetically vulnerable because of their reduced adaptive potential to environmental change. Standing genetic variation can buffer extinction risks (*e.g.*, Bitter *et al.* 2019; Gignoux-Wolfsohn *et al*. 2021; Forester *et al.*, 2022), and island endemics (typically characterized by reduced genetic diversity; Frankham, 1997) tend to have lower fitness than counterparts found in mainland populations (Furlan *et al.* 2012).

Genetic factors cannot be discounted as a cause of higher extinction rates on islands, even in the absence of future stochastic harmful changes. In large populations, and/or when selection intensity is very strong, natural selection is an effective determinant of allelic fate. However, in small populations and/or when selection is weak (e.g., small-effect alleles), genetic drift is more pronounced, and allelic fate is more stochastic (Santiago & Caballero, 2016; Whitlock, 2004; Wright, 1931). Thus, the reduction in fitness due to segregating deleterious alleles (*i.e*., the genetic expressed load), is expected to be higher in small, isolated, island populations. At the same time, small populations are at a risk of high rates of inbreeding, which can result in a further reduction

of fitness because recessive deleterious (*i.e.*, the inbreeding load) will be expressed in homozygotes. For example, the ancient population of woolly mammoths inhabiting Wrangel island ~4,300 years ago showed an excess of detrimental mutations, consistent with genomic meltdown just prior to their local extinction. Compared to mainland populations from the time that mammoths were plentiful (~45,000 Ya), the genomes of individuals of the Wrangel island population had a greater number of deletions affecting gene sequences and an increased number of premature stop codons (Rogers and Slatkin 2017). As illustrated by this example, the continuous fixation of deleterious mutations in islands can lead to an extinction vortex. However, this is not necessarily the case because the genetic load is a dynamic property influenced by a number of past and present factors.

In stable populations, the inbreeding load is expected to be larger for populations with larger effective size N (García-Dorado 2007), the increase being much more dramatic for more recessive deleterious alleles (García-Dorado 2003; Hedrick and García-Dorado 2016). Thus, "healthy" continental population still disperse individuals that are heterozygous for many rare (partially) recessive deleterious alleles. Upon arrival to a newly colonized island, inbreeding is expected to increase the expression of recessive deleterious effects in homozygosis, which not only produces inbreeding depression but also triggers an increase of selection against the deleterious alleles (genetic purging). It is this balance between inbreeding depression and purging that might, ultimately, determine the fate of island populations. Theoretically, this scenario can be approximated by the Inbreeding-Purging model of García-Dorado (2012). According to this model, after a reduction of N, the fitness expected by generation t (W_t) is predicted by:

$$W_t = W_0 \exp(-B g_t) \qquad \text{Eq. 1}$$

Where W_0 and B are the expected fitness and the inbreeding load (respectively) in the continental (*i.e.,* initial) non-inbred population, and where g_t can be predicted as a function of the population size N and the purging coefficient (d), which for a given homozygous deleterious effect s and dominance coefficient h, amounts:

$$d = s(1 - 2\,h) \qquad \textit{Eq. 2}$$

where s is the deleterious homozygous effect and h is the dominance coefficient (h=0 represents recessive expression, h=1 represents full dominance, and h=0.5 implies additive expression). The corresponding inbreeding load at generation t (B_t) can be predicted as:

$$Bt = Bgt(1 - Ft)/Ft \qquad \textit{Eq. 3}$$

an expression that accounts for the joint reduction of the inbreeding load ascribed to drift (F) and purging (g), that is faster than under drift alone.

The model shows that when d approaches 0, the role of purging preventing deleterious fixation becomes negligible compared to that of drift. In contrast, when Nd increases, drift becomes irrelevant and the deleterious alleles responsible for B in the original non-inbred population are expected to be virtually removed by purging (Figure 3.1).

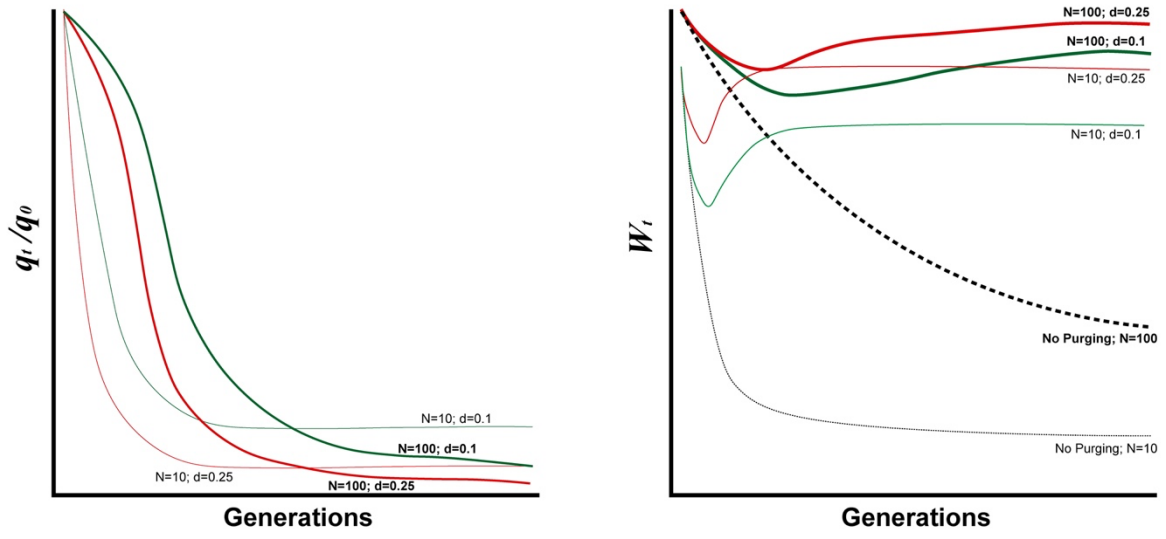


FIGURE 3.1. IP model showing the consequences of purging over 200 generations after the effective population size of an ancestral (*i.e.*, continental) large population with $W_0 = 1$ and $B = 2$ drops to N = 10 (thin lines) or N = 100 (thick lines). For each population size reduction three different deleterious effects are considered (green: $s = 0.2$, $h = 0$, $d = 0.1$ or red: $s = 0.5$, $h = 0$, $d = 0.25$, black: neutral (no purging) predictions. **a)** Average of the frequency of the deleterious alleles relative to the corresponding initial frequency in the initial population (q_t/q_0). The average frequency reaches a larger asymptotic value for smaller populations (thin lines), indicating more efficient purging. Purging is quicker and more efficient for larger d values; **b)** Corresponding expected average fitness through generations showing initial inbreeding depression and later substantial recovery due to purging. Figure modified from Peréz-Pereira *et al.* (2021).

To better understand the characteristics of the mutations (d) that could be purged under contrasting demographic scenarios, Kleinman-Ruiz *et al*. (2022) applied the IP-model for different combinations of homozygous s and h in either a small or a big population (Figure 3.2).

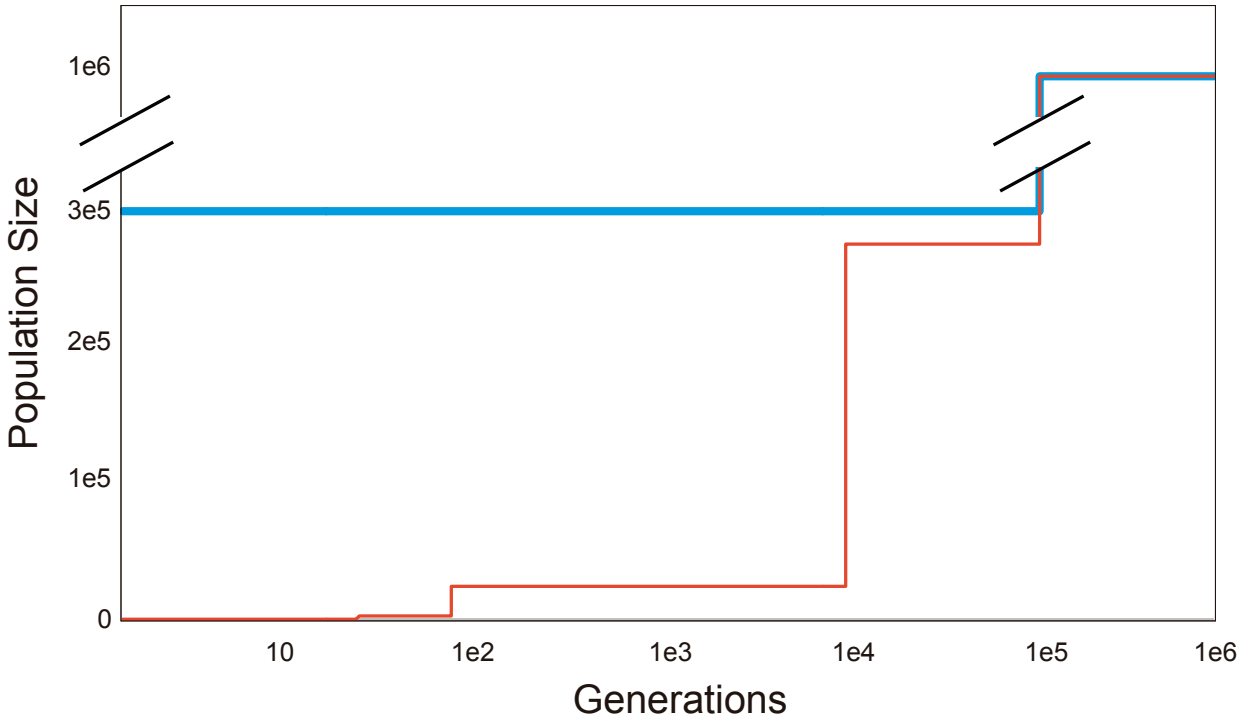


FIGURE 3.2. Long-term demography of the idealized small (thin red line) and big (thick blue line) populations studied by Kleinman-Ruiz *et al.* (2022). Both populations are derived from a common ancestor population close to the mutation-selection balance (N_e=10^6). Figure modified from Kleinman-Ruiz *et al.* (2022).

According to the model, purging only produces a large reduction of the derived count (i.e., the average number of derived alleles per individual) for deleterious, (partially) recessive, mutations with $h \leq 0.25$ in the small population (Figure 3.3). Conversely, for mutations that are roughly additive ($h = 0.45$), reductions in the effective size result in increased burden in the long term. The increase in burden for roughly additive mutations is, however, smaller than the reduction from purging observed for recessive deleterious mutations (note the different scaling of the ordinates axis in Figure 3.3). Thus, the derive count, is mostly determined by recessive deleterious mutations.

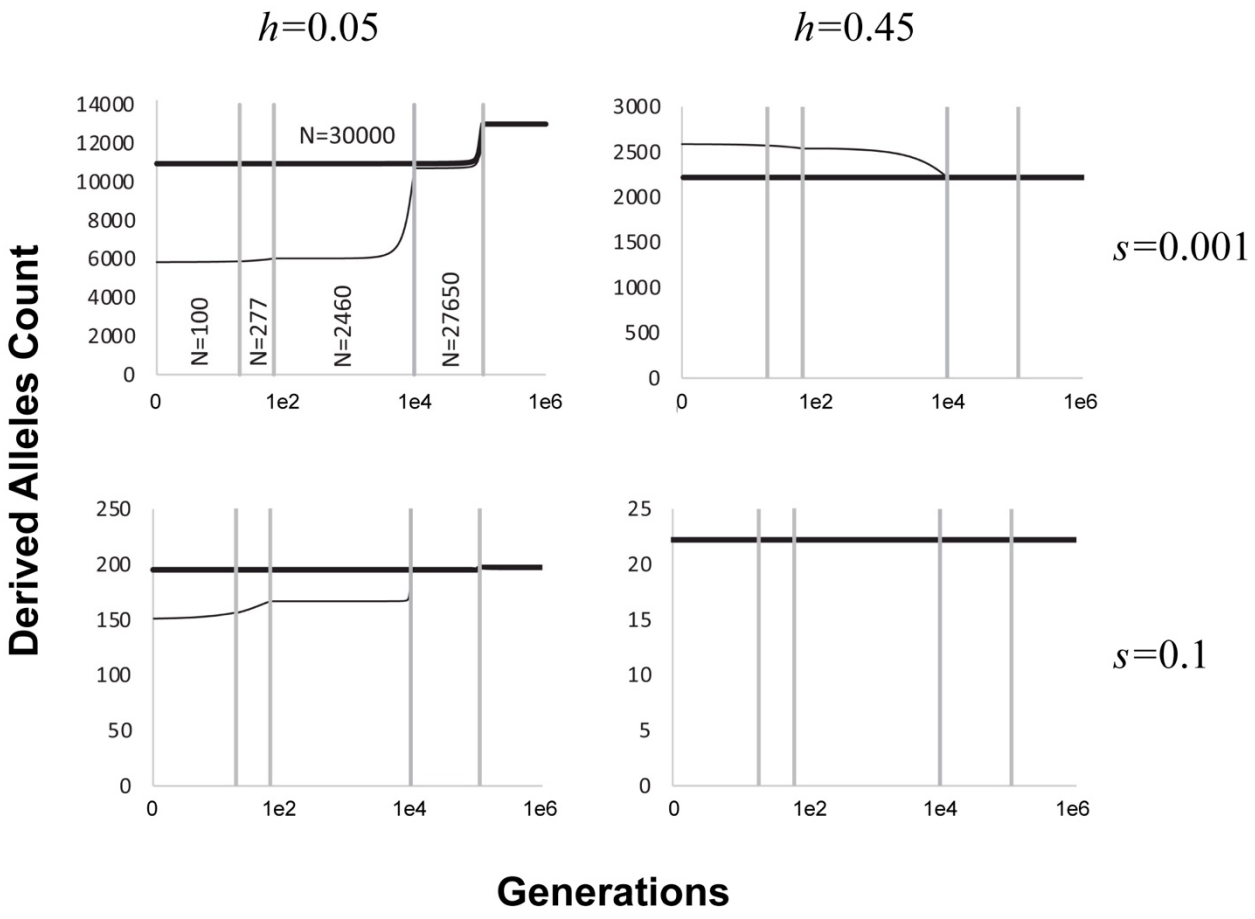


FIGURE 3.3. Predicted evolution of the derived count for two populations (small -thin line- and big -thick line-) derived from a common ancestor. The demography of both populations is represented in Figure 3.2. The effective sizes (N_e) after each bottleneck are shown in the first panel. Panels reflect different values combinations of dominance coefficient (h) and homozygous deleterious effect (s). Up to bottom: small to strong deleterious effects (d). Left to right: dominance gradient form recessive (h=0.05) to additive (h=0.45). These analytical results should be interpreted as a qualitative illustration of the dynamics of the derived count after a bottleneck rather than as precise quantitative predictions. Figure modified from Kleinman-Ruiz *et al.* (2022); see it for a detailed explanation.

The purging-driven differences between the two populations are a consequence of the higher rate of selection against homozygotes following a more pronounced reduction in effective size of the small population. As shown in Figure 3.3, purging can occur in a relatively wide range of population sizes. However, since the ultimate fixation rate of both additive and recessive deleterious alleles drops with $N_e s$, for small enough populations ($\sim N_e < 5/s$), the continuous fixation of deleterious mutations overcomes the reduction of burden from purging (Kleinman-Ruiz, 2022).

In general, the above IP-model predictions are also supported by forward-in-time simulations in which a mainland population gives rise to an island (Figure 3.4, Robinson *et al.,* 2018). Under a wide range of island demographic scenarios, simulations indicate that the predominant factor driving levels of genetic variation is the long-term small N of the island. As predicted by the IP-model, the total number of deleterious alleles per individual on the islands relative to the mainland varied according to selection and dominance coefficients. When mutations were additive, island genomes contained, on average, ~10% more deleterious alleles primarily due to the accumulation of weakly and moderately deleterious alleles. However, the number of strongly deleterious additive alleles were equivalent between the islands and the mainland. At the same time, island genomes contain far fewer moderately or strongly deleterious recessive alleles compared to the mainland (Robinson *et al.,* 2018).

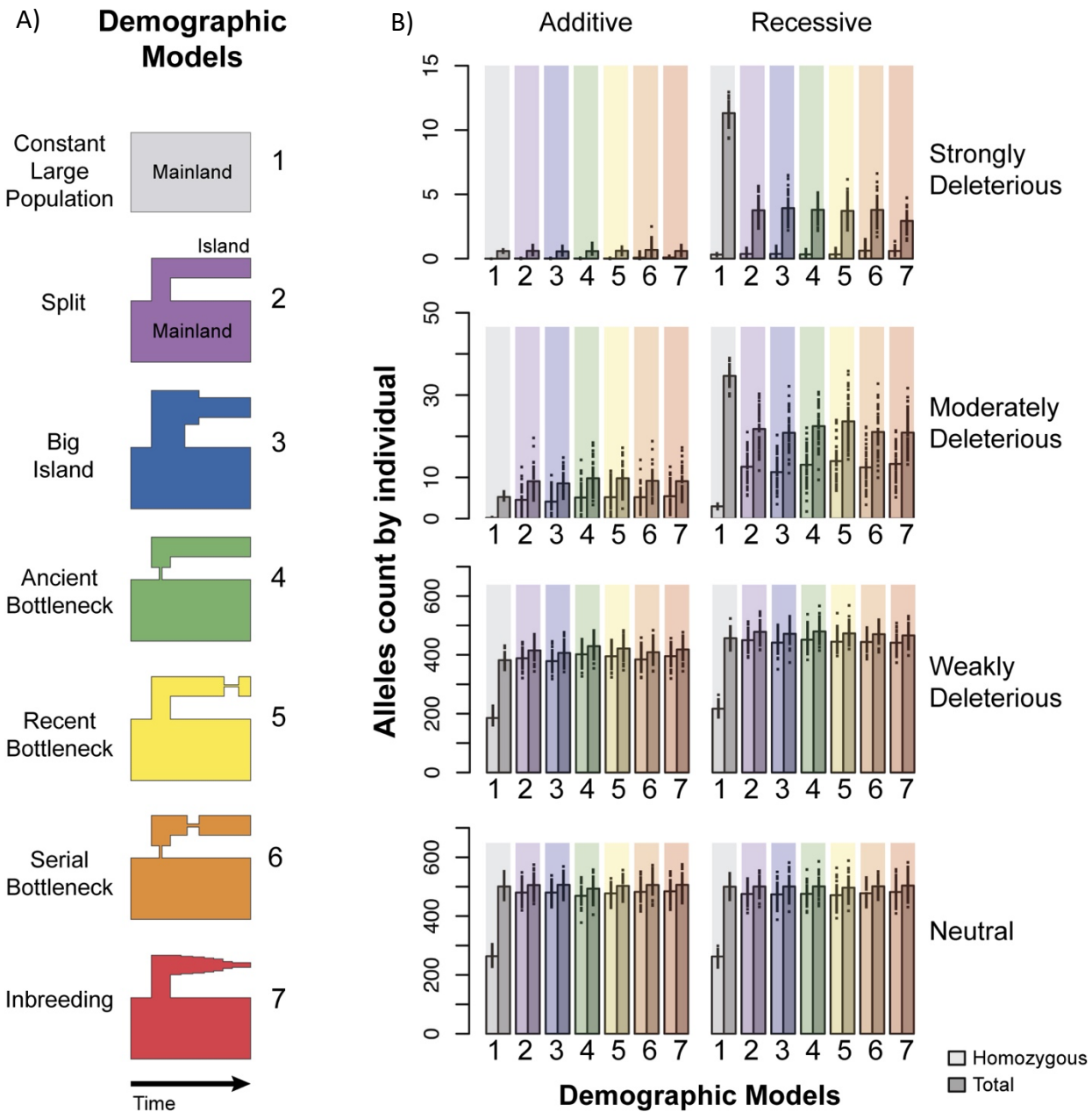


FIGURE 3.4. (A) Depiction of island demographic models (not to scale, $N_{mainland}$ = 10,000 diploids, N_{island} = 1,000 diploids. (B) Total derived count (dark) and the number of homozygous derived alleles (light) per individual. Dots represent the mean per individual for each of the simulation runs; bar height represents the arithmetic mean across simulations (N=: strongly deleterious, $1 \leq s < 0.01$; moderately deleterious, $0.01 \leq s < 0.001$; weakly deleterious, $0.001 \leq s <$

0; and neutral, $s = 0$. Simulations include a mixture of neutral and deleterious alleles in which deleterious alleles are either additive ($h= 0.5$; additive regime) or entirely recessive (recessive regime). Figure modified from Robinson *et al.,* 2018.

While useful, the IP framework described thus far only considers a single migration pulse (e.g., the initial colonization of a single island), and does not take into account the effects of gene flow, which might be important in archipelagos and for those islands located close to a continental coastline. To my knowledge, the effects of repeated migration on the genetic variability and fitness of island populations have not been explicitly addressed in previous studies. However, such effects should be analogous to those of genetic rescue programs, where individuals are translocated into small, isolated, and frequently inbred populations (reviewed by Pérez-Pereira et al., 2021).

On the one hand, the introduction of new migrants into small populations reduce the chance of inbreeding, because they carry beneficial alleles at some of the sites where the individuals in the population are homozygous for deleterious alleles. On the other hand, migrants bear their own inbreeding load due to partially recessive deleterious alleles hidden in heterozygosis, which may fuel future inbreeding depression. Thus, the effects of gene flow will depend on the purging occurring on both the donor and the recipient population. Initial inbreeding decreases as the number of colonizing individuals that arrive on the island increases. Consequently, the lower the initial inbreeding, the slower but more effective the purge will be (Figure 3.1; Eq. 1 and 3). This implies that populations on islands that experienced a slower process of inbreeding to reach their current level are expected to have lower levels of inbreeding depression, reduced hidden genetic load, and smaller fitness gains after subsequent migrations. If the island population is very small

(and remains very small), then the simulations predict a decrease in fitness because the load introduced by migrant individuals is not efficiently purged.

The functionally extinct wolf population of Isle Royale (MI, USA) illustrates the potentially negative effects of migration of colonizing individuals from a large, outbred, mainland population. The island was colonized by two to three mainland wolves that crossed frozen Lake Superior in the 1940s and established a new population that gradually increased to approximately 50 individuals (Peterson *et al.* 2014). After a parvovirus outbreak in the 1980s, this population was dramatically reduced from ~50 to 14. The size of this population remained relatively small until the mid-late 1990s when the population exhibited a substantial but short-lived increase in population size because of the migration of a single wolf from the mainland in 1997. Then, the population size declined significantly by 2010. Recent genomic studies suggest that the collapse of this population may have been prompted by the hidden inbreeding load of the migrant individual (Kyriazis 2020). Although the very small wolf population size of Isle Royale makes this example of limited practical interest in conservation genetics, it may reflect the colonization of remote islands, where we expect a very small number of breeding founders.

By interpreting Pérez-Pereira *et al.* (2021) simulations (Figure 3.5) as scenarios of island colonization, it becomes evident that the fitness of the island populations increases immediately after migration, but at the same time, there is an increase in the genetic load (B). However, while in the sporadic migration models (one or two events), B declines over time approaching similar equilibrium values to those seen in isolated islands, in periodic migration or OMPG models, B oscillates around equilibrium values larger than those observed when there is no migration (Figure 3.6).

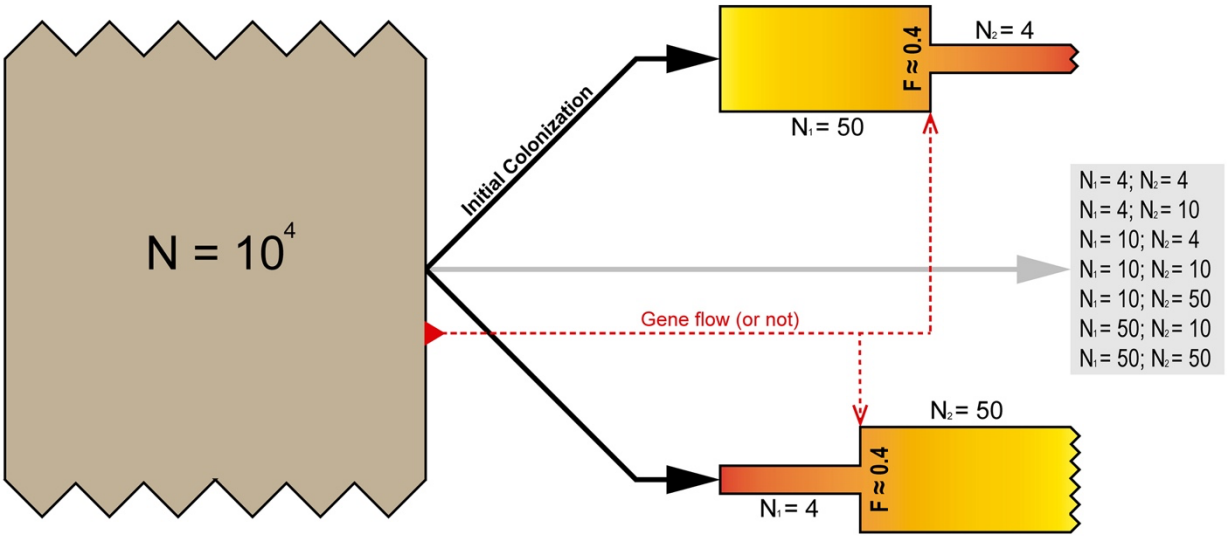


FIGURE 3.5. Pérez-Pereira *et al.* (2021) simulations reinterpreted as different island colonization scenarios: A small number of individuals (N_1) is sampled from a continental population to establish a new island population. The continental population has a long history of large effective population size, so that it can be considered genetically healthy (*i.e.*, it shows little reduction of mean fitness from segregating and fixed deleterious alleles), but it has not been purged (*i.e.*, it hides a large inbreeding load). After $t = N_1$ generations (Phase I) the population size can either be maintained ($N_2 = N_1$) or changed (Phase II, grey box), and the population can receive, or not, an increasing number of migrants from the continent. (*i*) a single event; (*ii*) two events with an interval of five generations; (*iii*) periodic migration every five generations; and (*iv*) one migrant per generation (OMPG). Time progresses downwards, all migrants are assumed to reproduce. Jagged edges indicate that the population size was constant before or after the time represented in the figure.

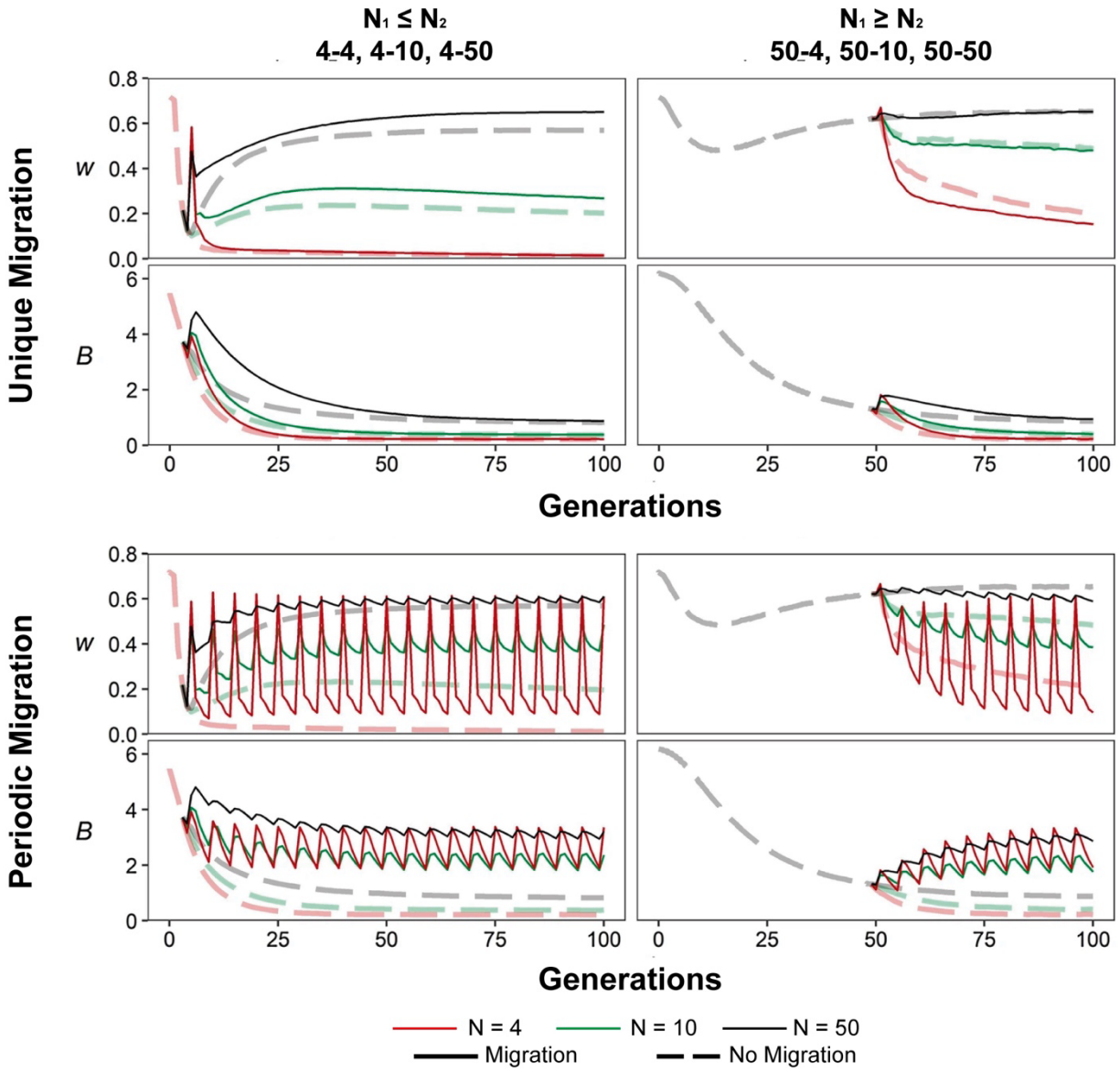


FIGURE 3.6. Evolution of average fitness (*W*) and inbreeding load (*B*) for small (island) populations under different demographic and migration scenarios (coded as *N*1-*N*2). *N*1 is the population size during phase 1, and *N*2 during phase II (red, green, and black lines for population sizes 4, 10 or 50, regardless the phase). Light dashed lines represent isolated populations (no migration). Dark solid lines represent migration starting at generation $t = N1$. (a) One unique migration of 5 males in lines $N2 = 50$, and of 1 male otherwise. (b) Periodic migrations of 5 males

every five generations in lines $N2 = 50$, and of 1 individual otherwise. Figure modified from Pérez-Pereira *et al.* (2021).

Exploring the demographic history and the accumulation or purging of deleterious alleles of species with a common origin, inhabiting oceanic archipelagos, provides an opportunity for understanding the role of gene-flow and purging in the mitigation of inbreeding depression. Advances in whole-genome sequencing enable the reconstruction of demographic history and migration. Moreover, unbiassed estimates of inbreeding and genome-wide assessment of mutation loads can be obtained through genomic data (Welles & Dlugosch, 2018).

This chapter focuses on a pair of sister species of mockingbirds found in the Galapagos islands, the Floreana mockingbird (*Mimus trifasciatus*) and the Galapagos mockingbird (*M. parvulus*). Together with *M. melanotis* and *M.macdonaldi*, these two species of mockingbirds form a monophyletic group. Mitochondrial analyses suggest that the common ancestor of mockingbirds began diversification ~3.5 MYA ago (1.6 – 5.5 MYA), having a similar age as the formation of the oldest islands of the current archipelago configuration (Alborgast *et al.*, 2006; Geist *et al.*, 2014; Sari & Bollmer, 2018). While *M. parvulus* has a broad distribution within the archipelago, *M. trifasciatus* became extinct on Floreana Island around 1890 and, today, this critically endangered species comprises only two small populations located on two satellite islets; Champion with 20-50 individuals and Gardner-by-Floreana with only 200-500 individuals (Grant *et al.* 2000; Hoeck *et al.* 2010).

Mockingbird populations have undergone an evolutionary journey that has been shaped by several factors. The characteristics of the islands they occupy and the isolation from other islands have played a significant role in their evolutionary history (Marlen, 2014; Shaw & Gillespie,

2016). Due to their generalist behaviour, mockingbirds have been able to occupy a considerable part of the islands in which each population resides. Although these birds are not known for their flying ability, recent discoveries suggest that fluctuations in the size of the Galapagos islands caused by geological events, including volcanic eruptions, erosion, and glaciation, have allowed for migration between islands (Ali & Aitchison, 2014; Geist *et al.*, 2014: Woolfit, & Bromham, 2005). The central Galapagos islands were notably more extensive during the Last Glacial Maximum (LGM), and Santa Cruz, Santiago, Isabela, and Fernandina may have joined to form a single large island due to an extremely low sea level stand around 255 ka (Ali & Aitchison 2014; Garg *et al.*, 2018). Human colonization in the past few centuries has caused significant changes to bird populations in the archipelago, leading to drastic population reductions due to introduced predators, parasites, and habitat alterations (Jiménez-Uzcátegui *et al.*, 2019; McNew *et al.*, 2021; Snell *et al.*, 2002; Wikelski *et al.*, 2004).

Mimus melanotis genome (Chapter 2) would enable the examination and comparison of the evolutionary history and demographic patterns of eight populations of mockingbird in the archipelago, which includes two species, *M. trifasciatus* (with two populations) and *M. parvulus* (with six populations), with contrasting sizes and representing a significant distribution range in the archipelago. Deep whole-genome resequencing would reveal previously unknown phylogenetic relationships and demographic histories of the populations, and the patterns of genetic diversity and inbreeding can be explored. Therefore, my objective for this chapter was to clarify the phylogenetic relationships of the populations of *M. trifasciatus* and *M. parvulus* in the archipelago, and to explore the patterns of genetic diversity and inbreeding. The ultimate goal was to provide valuable insights into the origin, plasticity/resilience, and conservation of mockingbird populations. These insights will be valuable for the management and conservation of mockingbird

populations, particularly for the two remaining populations of *M. trifasciatus*, and will provide evidence for the feasibility of establishing a third population on the main island of Floreana.

3.2. METHODS

3.2.1. *Population sampling*

Eight allopatric populations of *Mimus* (i.e., on different islands) were selected to determine which are the population genomic patterns given by different and independent demographic histories. It is assumed that each population of *Mimus* is isolated from other populations/islands and their demography has been shaped mainly by island size. The sampling was focused on populations of the most widespread species *M. parvulus* and its sister species *M. trifasciatus* (Nietlisbach *et al.*, 2013; Štefka et al., 2011). Specifically, the population/island sampled were: a) Isabela (~458,812 Ha) and Santa Cruz (~98,555 Ha) as large population size examples (*M. p. parvulus* specimens), b) Pinta (~5,940 Ha) and Marchena (~12,996 Ha) as intermediate population sizes (*M. p. personatus* specimens), c) Darwin (~110 Ha) and Wolf (~130 Ha) as small population sizes (specimens of *M. p. hulli* and *M. p. wenmani* respectively), and Champion (~9.5 Ha) and Gardner (~81 Ha), which are the two only extant populations of *M. trifasciatus*, with extremely small population sizes. Blood samples from individual birds were collected *in situ* (and *in vivo*) by venipuncture of trapped birds on several expeditions to Galápagos islands between 2009 – 2014 by Hoeck *et al.* (2010b) under permission number PC-48-10 (Disclaimer: the blood samples used in this chapter were collected as a part of a previous study, and were not sampled by me, the author). From these samples, after testing their relatedness using the microsatellite panel of Hoeck *et al.* (2010b), three unrelated representative individuals of each population were selected for

genomic re-sequencing. In this way, any individual relatedness bias in the further population genetic assessments was avoided.

3.2.2. *Whole-genome re-sequencing and variant calling*

The extraction of genomic DNA was carried out using a MasterPure DNA Purification Kit (Lucigen). PCR-free sequencing libraries were prepared by the Norwegian Sequencing Centre and sequenced on six lanes of Illumina HiSeqX10 in a pair-end mode. Later, the adapters removal and trimming of low-quality base pairs of the obtained pair-ended reads was performed using Trimmomatic (v. 0.36, Bolger *et al*. 2014). All reads longer than 120 bp and with a minimum base quality Phred score of 15 were mapped to the *M. melanotis* genome (chapter 2) using BWA – mem v. 0.7.15 (Li & Durbin 2009). Duplicated reads were removed using Picard (v. 2.8.1) (Broad Institute 2018).

Variant calling (SNP's collection) was carried out using GATK v. 3.7 (McKenna *et al.* 2010) with a series of stringent filtering steps to minimize the presence of false genotypes. Reliable biallelic SNPs were obtained following the GATK best practices for hard filtering as described in DePristo *et al*. (2011). Variants failing the following filters were excluded: Quality-by-Depth (QD) < 8; Fisher Strand (FS) > 15; Root Mean Square of the mapping quality (MQ) < 50; Mann-Whitney Rank Sum Test for mapping qualities (MQRankSum) < −5; ReadPosRankSumTest (RPRSlow) < −8.0 and (RPRShigh) > 3; StrandOddsRatio (SOR) > 3.0. Also excluded were sites that exceeded average coverage depth per individual by more than two standard deviations, showed excess heterozygosity ($p < 0.05$) as determined by VCFTools (v. 0.1.16; Danecek *et al.* 2011), had more than 25% missing data, and sites found within repetitive regions. Finally, invariant sites with low quality (QUAL < 15) were removed from the analyses.

3.2.3. *Phylogenetic and population structure analyses*

Several methodologies were applied to infer the phylogenetic relationships between populations. Firstly, a maximum likelihood tree was constructed using SNPphylo (Lee *et al.*, 2014) and 7,245 neutral LD-pruned SNPs. In addition, three *M. polyglottos* individual genomes were included as an outgroup. To avoid potential accuracy issues due to data distribution and randomness, SNPphylo pruning was performed. This was necessary as including too many SNPs in the analysis could lead to biased estimates, since Maximum Likelihood analysis assumes the data is randomly and equally distributed throughout the genome. Therefore, out of the initial set of more than 16 thousand neutral SNPs, only 7,245 were retained for the analysis using SNPhylo. Confidence of the generated branches was calculated based on 1000 bootstrap iterations. Then, a split network was constructed, based on a comparison of the Nei's genetic distance values, in order to explore and discriminate reticular interconnections that are common connected and/or recently split populations. The split network was created using StAMPP (Pembleton *et al.* 2013), and the Neighbour-Net algorithm of Splitstree (Huson & Bryant 2006). Finally, to quantify how the relationships among islands vary across the nuclear genome, an estimation of relative contributions of all possible topologies was carried out using TWISST (Martin and Belleghem, 2017). For this analysis, maximum likelihood trees were estimated in 100 SNPs windows (overlapped or non-overlapped) using PhyML (Guidon *et al.*, 2005). Then, a topology weighting estimation was performed using TWISST with "complete" option, to compute the proportion of subtrees matching each possible topology. TWISST analyses were implemented according to the author's guidelines.

Population structure was estimated using a principal component analysis (PCA) and a clustering analysis. The population structure inference was based on 16,632 neutral LD-pruned

SNPs. The PCA was carried out in the program SNPRelate (Zheng *et al.,* 2012). For clustering analysis, this was performed in ADMIXTURE V. 1.3.0 (Alexander *et al.*, 2009) assuming 1 to 9 clusters (K). Cross-validation error rate was used to identify the number of clusters that best matches genetic structure in the analysed populations.

3.2.4. *Demographic inferences*

To infer the demographic history, four different approaches were used. First, I performed a Multiple Sequentially Markovian Coalescent (MSMC2) analysis to calculate the coalescent rate and infer past population sizes (Schiffels and Durbin, 2014). To minimise the potential bias caused by non-neutral molecular processes, only the neutralome was used. The neutralome was defined as the non-coding regions of the genome that did not include 10kb margins of coding sites, as well as non-coding conserved sites and their 100bp margins. The non-coding conserved regions were obtained by lifting over from the white-collared flycatcher genome (Craig *et al.*, 2018). MSMC2 was run using time segments 1*2+15*1+1*2 (2 segments are joined forcing the coalescence rate to be the same in both segments, then 15 segments each with their own rate, and then again two groups of 2), and confidence of time estimates was calculated based on 20 bootstrap iterations.

Secondly, the program Relate (Speidel *et al*., 2019) was used to estimate both population size and the relative pairwise coalescence rates between populations. Relate (--mode All) was run to estimate genome wide genealogies with population size -N 1000 and mutation rate -m $4.6x10^{-9}$ (estimated for *F. albicollis*; Smeds *et al.*, 2016). Coalescence rates between and within populations were estimated using nine iterations and generation time was set to 4.5 years (Grant *et al.* 2000). Initially, Relate analysis was performed for each chromosome separately, and subsequently, joint estimates were obtained for all autosomal chromosomes excluding chromosome 4A and shorter

scaffolds because of the lack of reliable estimates of the ancestral states. Relative pairwise coalescence values were calculated as the proportion of the cross-coalescence rate between populations and the intra-coalescence rate for the larger population of each pair.

SMC++ analyses (v1.15.2; Terhorst *et al.*, 2017) were carried out to better resolve recent demographic events in the history of the studied populations. SMC++ uses LD (linkage disequilibrium) analysis and HMM (Hidden Markov Models) to reconstruct the demography history of a population departing from unphased whole-genome variants. In this way, the unphased neutralome SNPs data were used to perform this analysis. Initially, the command *vcf2smc* was used to format the input files, with such command each chromosome by each population were converted to SMC++ input file. Then, an individual inference analysis for each population was carried out via the command *estimate* and using the whole population chromosome set and the mutation rate reported for *Ficedula albicollis* (Smeds *et al.*, 2016). Additionally, because the recent bottleneck of *M. trifasciatus*, both remanent populations (Campion and Gardner) were merged to serve as a proxy demographic inference of the extinct *M. trifasciatus* population of Floreana.

Complementary demographic analyses were carried out using G-PhoCS (Generalized Phylogenetic Coalescent Sampler; Gronau *et al.*, 2011) to determine ancestral population sizes, divergence times and post-divergence migration. To generate G-PhoCS input, genomic regions with the least constraint from non-neutral genomic processes (selection, linked selection and GC-biased gene conversion) were used to avoid bias in the inference (Gronau *et al.* 2011). Then, coding regions were masked together with their surrounding 10kb, and also conserved non-coding regions were masked with their surrounding 100bp. Coordinates of conserved non-coding regions were generated by lifting over conserved non-coding regions found in *Ficedula albicollis* (Craig *et al.*

2017) to *M. melanotis* genome. After filtering, more than 80,000 neutral loci (>1000 bp) were recovered. To avoid artifacts because of recombination, I filtered the loci by size, selecting only loci with lengths between 1,300 and 1,500 bp. This resulted in a total of 8,429 neutral loci for input to G-PhoCS. A G-PhoCS MCMC analysis was conducted to test the relative support for the most supported phylogenetic topology that was generated from the previous phylogenetic analysis (refer to Figure 3.10). To explore post-divergence migration events between populations, the migration bands were tested bidirectionally pairwise between all *M. parvulus* populations (Figure 3.7a-h). Additionally, two extra models were tested to avoid any gene flow bias given before the current population diversification model, where one tested an ancestral migration band between two main clades within populations of *M. parvulus* (Figure 3.7i), and the other tested a model without any migration band. Only populations of *M. parvulus* were considered for migration models since these populations were essentially the same population in the near past and there was no evidence of gene flow between *M. trifasciatus* and the sampled populations of *M. parvulus* for this study (see: post-divergence gene flow in the pairwise coalescence rate analysis results). Since G-PhoCS does not have a statistical test for output parameters, nine independent runs were performed based on nine subsets of migrations models (Figure 3.7). The default parameters described by Gronau *et al.* (2011) were used to run the analysis with α= 1.0 and a β = 10,000 for mutation-scaled population sizes and divergence times, and α = 0.002 and β = 0.00001 for mutation-scaled migration rates. The finetunes argument was set on TRUE in the control file, and the MCMC was set to run 200,000 iterations, the first 100,000 were burn-in iterations and leaving the last 100,000 iterations for analysis, the analysis samples were taken each ten iterations. Additionally, *tau-theta-print*=10,000 was used to scale up the raw estimates because the smallest estimates were below the resolution threshold. The procedure described by Freedman *et al*. (2014) was adapted to calibrate G-PhoCS

raw estimates, with two constants; the mutation rate for *Ficedula albicollis* was set as μ=4.6E-9 (Smeds *et al*., 2016) and generation time (g) of 4.5 years for the Galapagos mockingbirds (Grant *et al*., 2000).

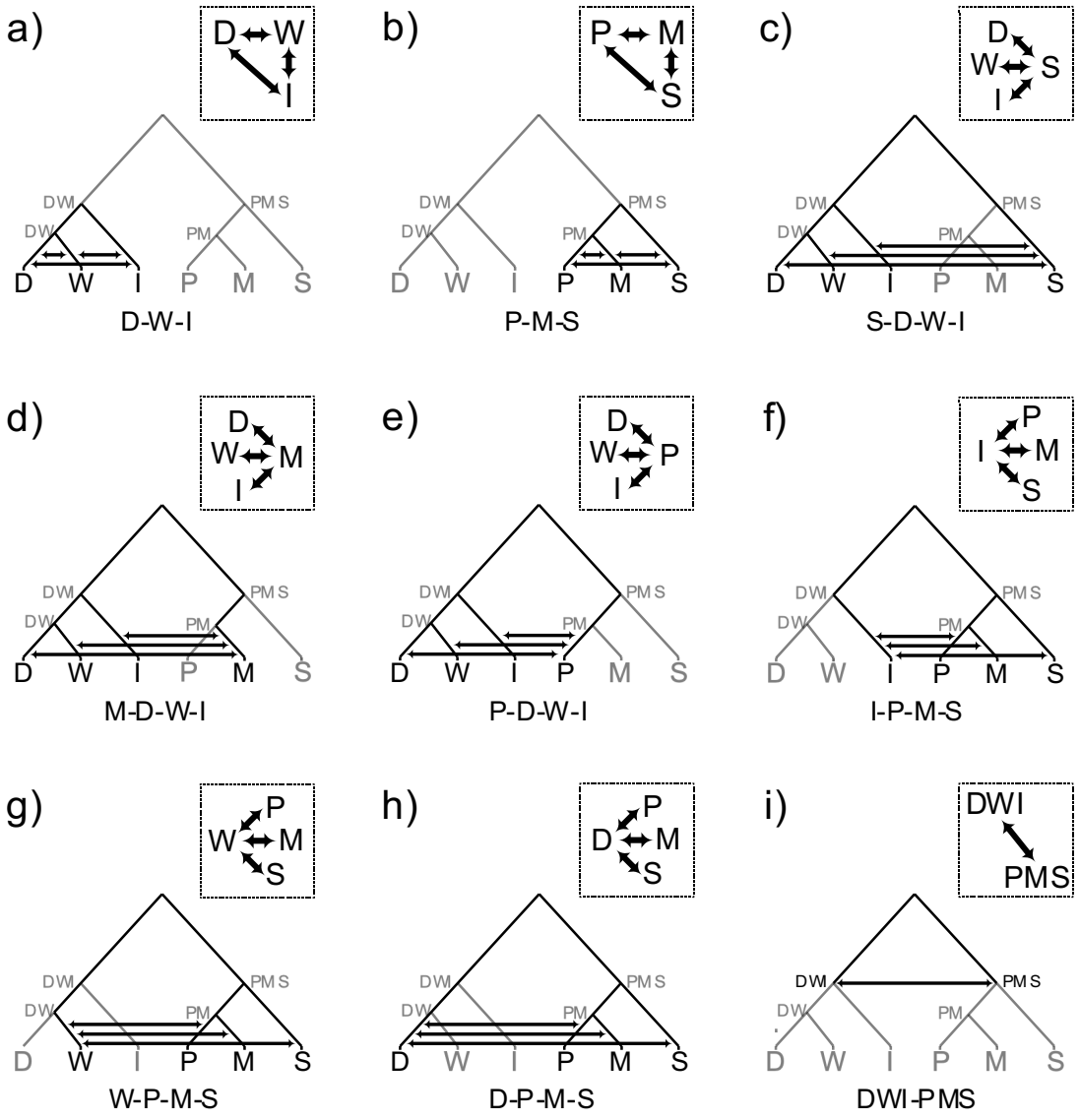


FIGURE 3.7. Independent bidirectional migration band models tested in G-PhoCS between *M. parvulus* populations. Populations are Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M)

and Santa Cruz (S). Models a) and b) tested migration between all three population within the main clades of *M. parvulus*. Models c) to h) tested one population of each main clade of *M. parvulus* as a source/target population against to all three populations in the other main clade. Finally, model i) tested the migration between the ancestral populations of *M. parvulus* main clades. Under each model description is the code name of the model.

3.2.5. *Population size effects in intra-population genetic diversity, inbreeding and genetic load*

Regression analyses were conducted to evaluate the effects of population size on the population genomic indices. In one hand, island size is the main constraint factor that determines the population size in generalist species (as Galapagos mockingbirds). Therefore, island size was used as a proxy for population size. Population genomic indices (i.e., genetic diversity, inbreeding and genetic load) were estimated for each population using the previously filtered whole-genome neutralome.

Firstly, the genome-wide heterozygosity (*H*) and inbreeding coefficient (*F*) were calculated for each individual genome using the argument --het in VCFTools v. 0.1.17 (Danecek *et al.*, 2011). Secondly, nucleotide diversity (π) and heterozygosity (*H*) along the genome were estimated using non-overlapping sliding windows of 10 Kb. To estimate nucleotide diversity along the genome, the individuals were grouped by population and the sliding-window nucleotide diversity estimation were performed using the argument --window-pi 10000 in VCFTools. To estimate population heterozygosity across the genome, a custom script was used. For each non-overlapping 10Kb window, the script calculated the average frequency of heterozygous SNPs, as a proportion of the total number of SNPs in the same window, for each population. The calculation was done by dividing the number of heterozygous SNPs by the product of the number of individuals and the

number of survey SNPs for that population (PopHetWin = nHet/number-of-individuals/number-of-survey-SNPs).

Secondly, homozygosity (ROH) analyses were used to assess the degree of genomic homozygosity. ROHs were identified using the BCFTools/ROH method from BCFTools v. 1.9 (Narasimhan *et al.*, 2016). The categorization thresholds used for ROH were 1) less than 1Mb (ancient relatedness), 2) between 1 and 2 Mb (past bottleneck), and 3) more than 2 Mb (recent or current bottleneck). Furthermore, the linkage disequilibrium was calculated for each population using PopLDdecay v3.41 (Zhang *et al.*, 2019) with default parameters as a different approach to determine genomic homozygosity.

Finally, the genetic load for each population was calculated using two different sources, total count of derived non-synonymous alleles per individual (= the additive load) and the number of homozygous genotypes with derived non-synonymous alleles per individual (= the recessive load). For derived or ancestral allelic state, the inference was carried out with the program Est-sfs v2.03 (Keightley and Jackson, 2018) using the genomic information from common starling (*Sturnus vulgaris*) and northern mockingbird (*M. polyglottos*) as outgroups in the analyses. The effect of derived alleles was analysed using Snpeff 4.3 (Cingolani *et al.* 2012) and the counts of derived alleles were normalized by the number of callable sites to account for variance in sequencing coverage. Lastly, a sliding window nucleotide diversity analysis for 0- and 4-fold positions in coding sites was carried out for each population. In this case, the SNPs in positions 0- and 4-fold were filtered and analysed separately following the previously described methodology for sliding window nucleotide diversity analysis. For this case, the window resulting nucleotide

diversity was corrected using the actual number of all 0- or 4-fold positions contained on each window.

3.3. RESULTS

3.3.1. *Genome re-sequencing and variant calling*

The mean sequence reads per individual was 222 M, with approximately 79% of reads successfully mapped onto the reference genome of *M. melanotis*. These reads covered 88% of the reference genome. The variant calling and filtering resulted in a Variant Call Format (VCF) file containing ~8.6 M bi-allelic SNPs and ~879 M invariant sites. The obtained biallelic SNPs reached a mean effective coverage of 24x of the site in the reference genome of *M. melanotis*.

3.3.2. *Phylogenetic relationships and population structure*

The maximum likelihood tree produced from the genome-wide phylogenetic analyses showed the Galapagos *Mimus* as a clear, strongly supported monophyletic group (bootstrap = 100). Thus, both the Floreana mockingbird (*M. trifasciatus*, blue clade) and the Galápagos mockingbird (*M. parvulus*, dark red clade) each formed a monophyletic group with clades well supported towards each group of individuals as a population (Figure 3.8). The analysis showed a relatively weaker value for the divergence between the Champion (C) and Gardner (F) individuals within the *M. trifasciatus* clade, but it was still relatively high with a bootstrap value of 82%. The relatively lower support for the Champion (C) and Gardner (F) individuals could be explained by their recent forced isolation, which has prevented genetic flow between them, despite previously being part of a single panmictic population (= the extinct Floreana population). As for *M. parvulus*, the individual genomes were divided into two clades with full statistical support (i.e., 100%

bootstrap values). The first clade includes individuals from the central islands of the archipelago (red), Pinta (P), Marchena (M), and Santa Cruz (S); while the second clade comprises individuals from the northwestern islands (orange clade), Isabela (I), Darwin (D) and Wolf (W). While relationships between populations in the central islands clade ((M,S) P) are highly supported (≥ 97% bootstrap values), the topology of the northwestern islands clade is not well resolved ((I,W) D), particularly for the definition between Isabela and Wolf populations (79%). Finally, it should be noted that all individual genomes were clustered based on their islands/populations with full bootstrap support. Similarly, the genetic distance network analysis (Figure 3.9) showed the same population assemblages as the maximum likelihood tree and likewise revealed small basal crosslinks restricted to the DWI clade as a sign of low genetic differentiation between these populations.

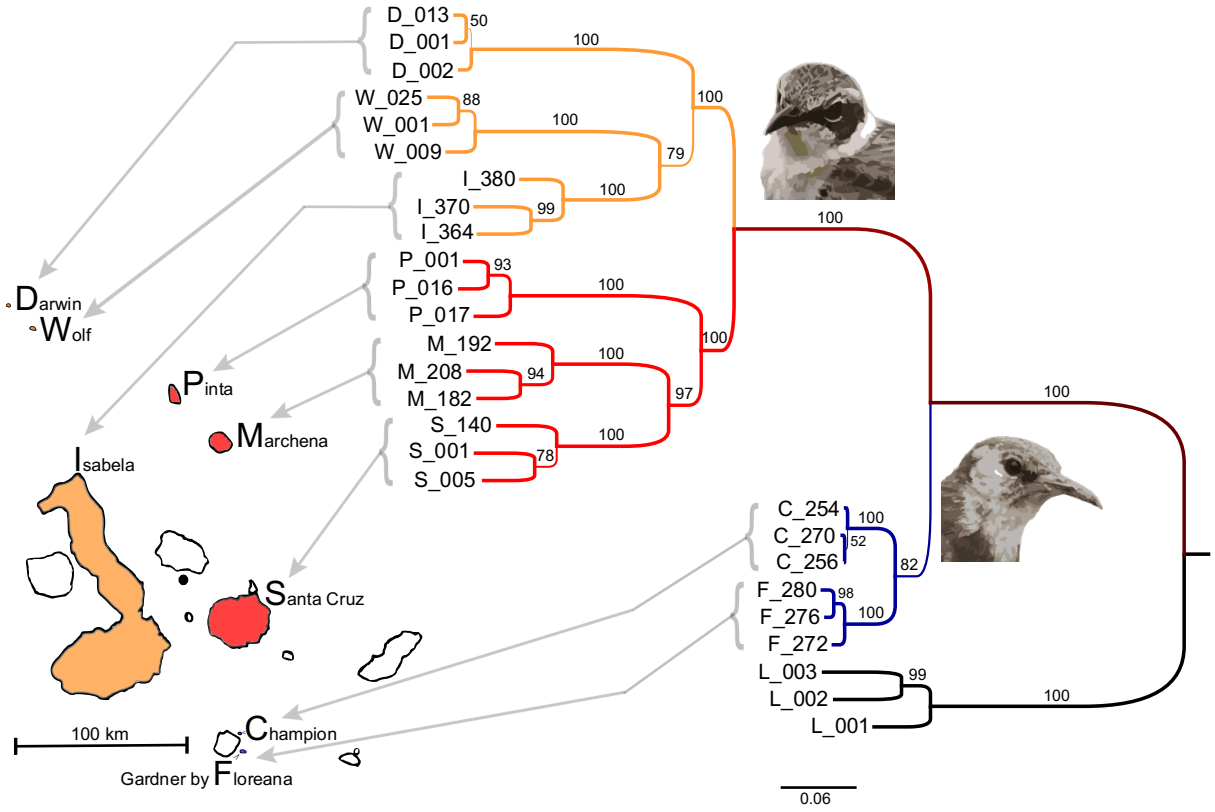


FIGURE 3.8. Maximum likelihood tree with population geographical reference in the archipelago. Bootstrap values above each branch are based on 1,000 iterations. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). Three mayor clades are evident, the first clade containing *M. parvulus* of the north-western populations (orange), the second clade consisting of *M. parvulus* of the eastern populations (red), and the third clade represented by individuals of *M. trifasciatus* (blue).

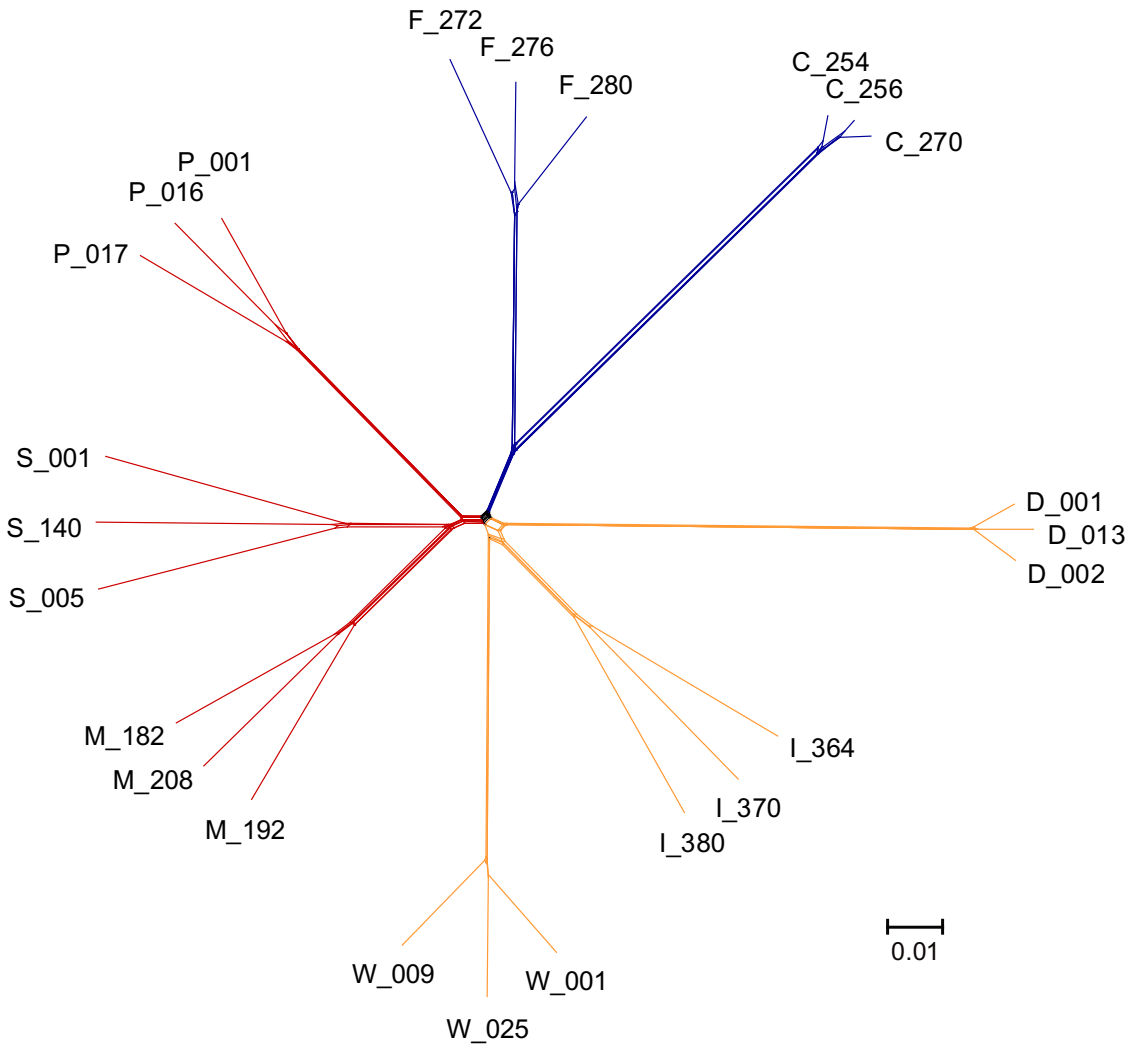


FIGURE 3.9. Split network of mockingbird individuals based on Nei's genetic distances build using the Neighbour-Net algorithm. Three main clusters were recognized; north-western populations of *M. parvulus* (orange), the eastern populations of *M. parvulus* (red), and *M. trifasciatus* (blue).

TWISST analysis showed a persistent phylogenetic differentiation between *M. parvulus* and *M. trifasciatus* along the genome. However, in the two main clades within *M. parvulus* (DWI and PMS), there was not a clear dominance of a single topology along the genome leaving high

uncertainty on their phylogenetic relationships. Specifically, in both clades there were three topologies that showed almost equal proportion of representativeness along the genome, for DWI the conflictive topologies were ((D,W) I): 34%, ((D,I) W): 33% and ((W, I) D): 33%, while for PMS were ((M,P) S): 35% , ((M,S) P): 34% and ((S,P) M): 32% (Figure 3.10).

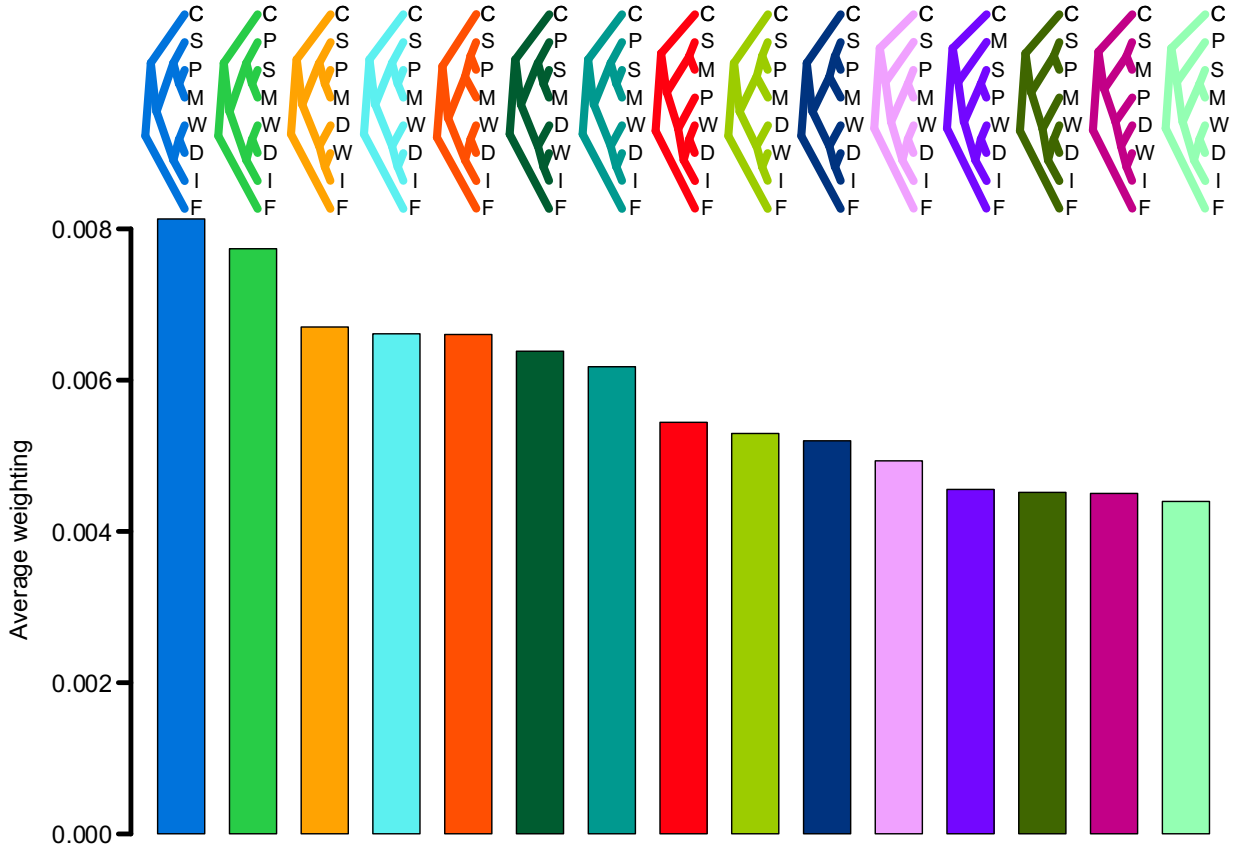


FIGURE 3.10. Average weighting for the best 15 phylogenetic topologies obtained from TWISST analysis. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S).

Finally, the population genetic structure analyses revealed that each population has its own genetic structure, even despite the phylogenetic incongruencies found in DWI and PMS clades. In this way, the PCA showed that the individuals from the same island tend to cluster all together and

there was not a random assortment between individuals (Figure 3.11). Additionally, ADMIXTURE analysis revealed that each population forms a unique cluster and (also) demonstrated that each population was isolated from the others. The model with the lowest cross validation error (0.65), K=7, showed each population as a group, except for Santa Cruz and Marchena which were merged. However, the model in which each of the islands represented a unique genetic cluster (K=8) also had a similar low error rate (0.70) compared to any other K values (Figure 3.12).

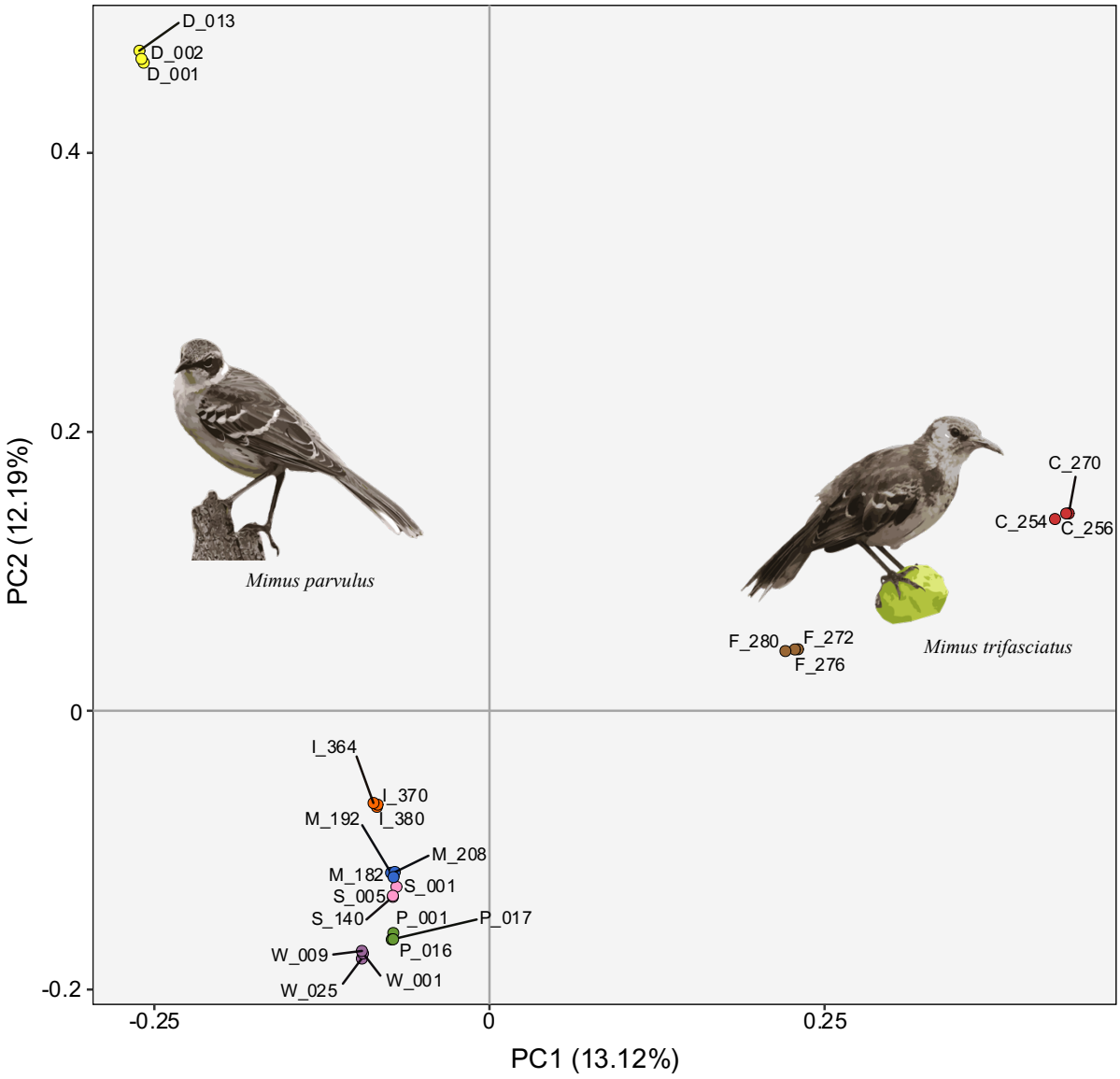


FIGURE 3.11. PCA analysis based on allelic frequencies of 16,632 neutral unlinked SNPs for all 24 individuals sampled. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). In all cases, the three individuals from each population (island) were found forming defined clusters.

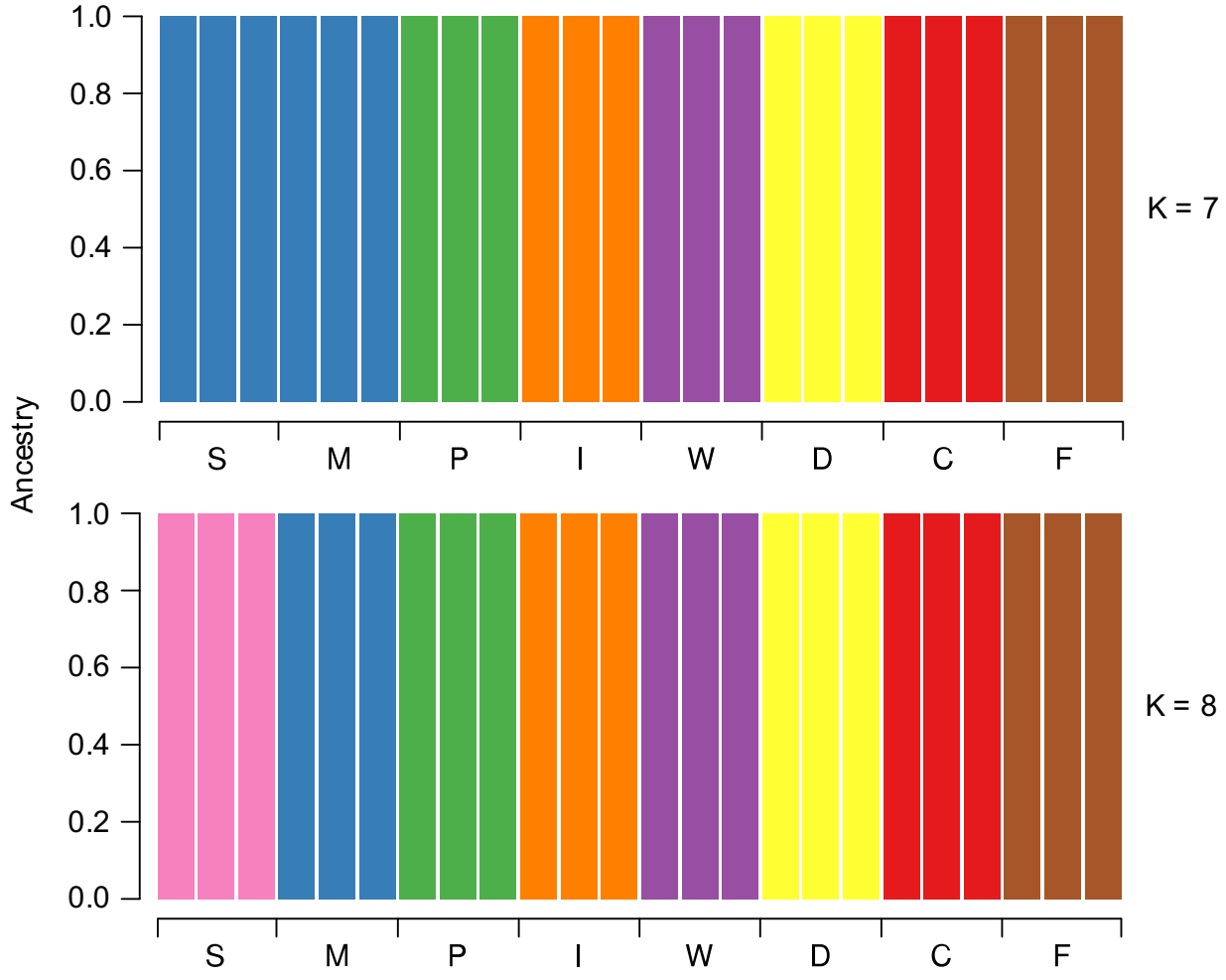
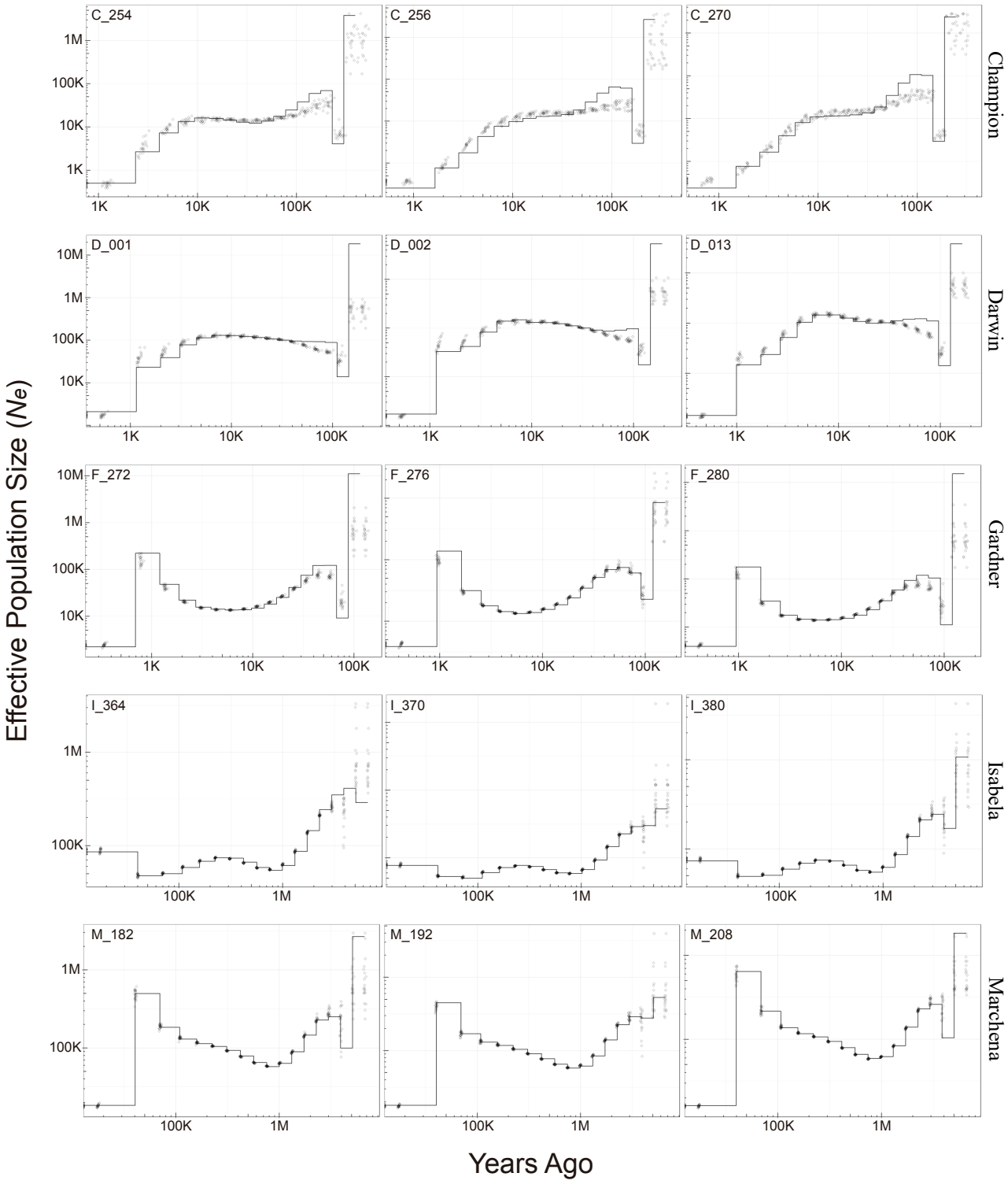


FIGURE 3.12. Best models obtained from ADMIXTURE analysis based on 16,632 neutral unlinked SNPs for all 24 individuals sampled. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). Upper panel shows the model with the lowest error rate (0.65) which is based on 7 lineages (K=7), where Santa Cruz and Marchena shared a common genetic pool while the other populations were isolated from each other. Lower panel shows the model with the second lowest error rate (0.7)

which is based on 8 lineages (K=8), where all populations are isolated from each other. Neither of the models shows genetic flow between individuals nor populations.

3.3.3. *Historical Demography*

Both MSMC2 and Relate analyses pointed out a gradual constriction of population sizes since 10Mya (Figures 3.13 and 3.14). Between 10 – 1 Mya, all populations followed an almost identical population size trajectory with a potential colonisation bottleneck (founder effect) around 1 Mya (Figure 3.14). Shortly after that, populations started diverging in their population size trajectories with small islands showing the most severe reduction in size followed by medium and large size islands (Figure 3.15). The relative cross-coalescence rate between populations shows one panmictic population before 1Mya where the cross-coalescent rate equals the population coalescent rate. But after 1Mya the cross-coalescent rate rapidly decreases indicating decreasing connectivity between islands. The latest split occurred between the two populations of Floreana mockingbird starting 100Kya (Figure 3.16). Finally, the analysis of cross-coalescent rate for each population pair revealed apparent events of post-divergence genetic flow between some *M. parvulus* populations. Specifically, these events started ~100,000 years ago and finished ~10,000 years ago, and it happened between Isabela and Darwin, Santa Cruz and Darwin, Santa Cruz and Marchena, Wolf and Isabela, and Wolf and Santa Cruz (Figure 3.16).


C_254
C_256
C_270
Champion
D_001
D_002
D_013
Darwin
F_272
F_276
F_280
Gardner
I_364
I_370
I_380
Isabela
M_182
M_192
M_208
Marchena
Effective Population Size (Ne)
Years Ago

Continue…

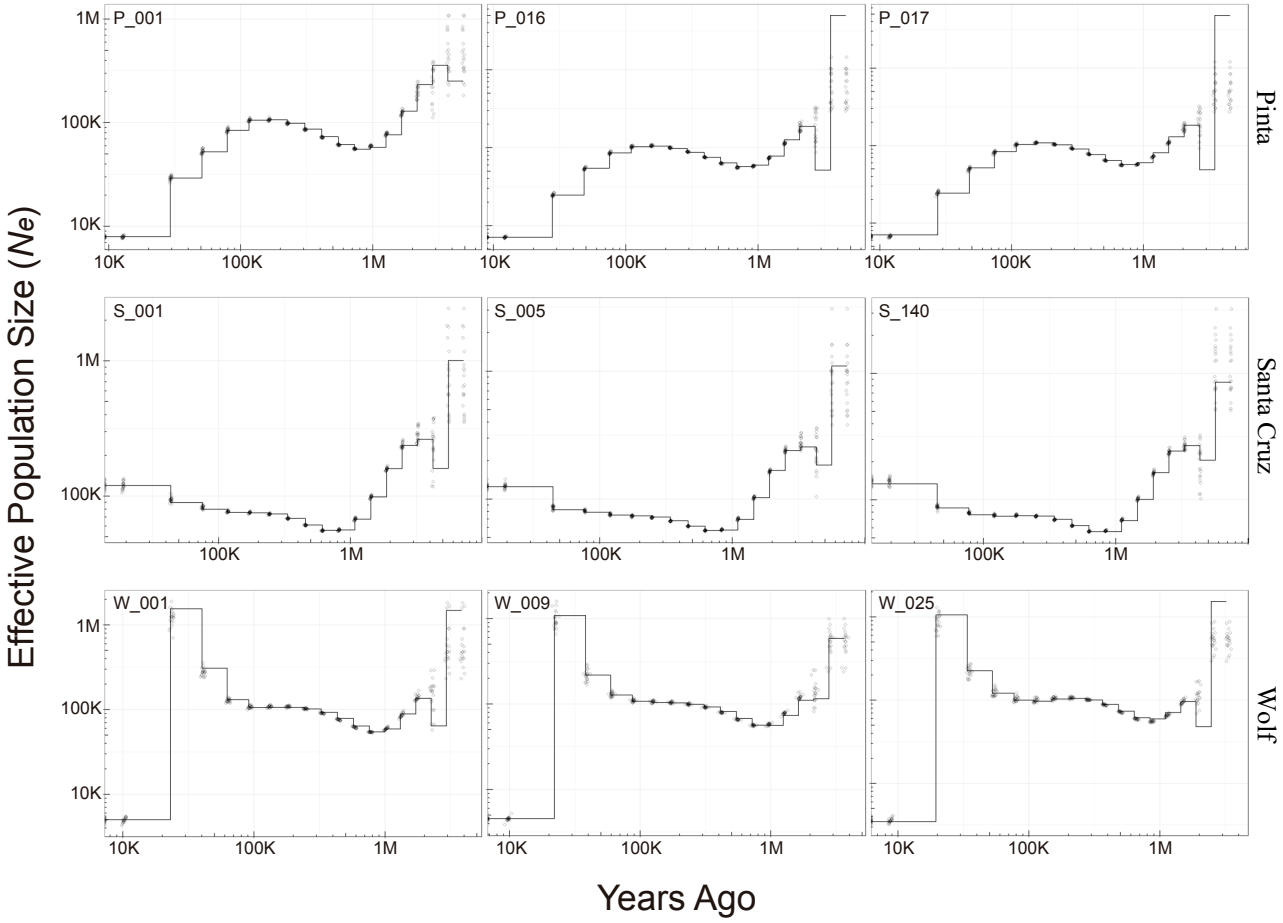


FIGURE 3.13. Individual runs of Multiple Sequentially Markovian Coalescent (MSMC2) analysis for wide-genome demography inference. Plots show the estimation of effective population sizes (*Ne*) of each island/population throughout the time from present towards to past. Each panel line represents a population.

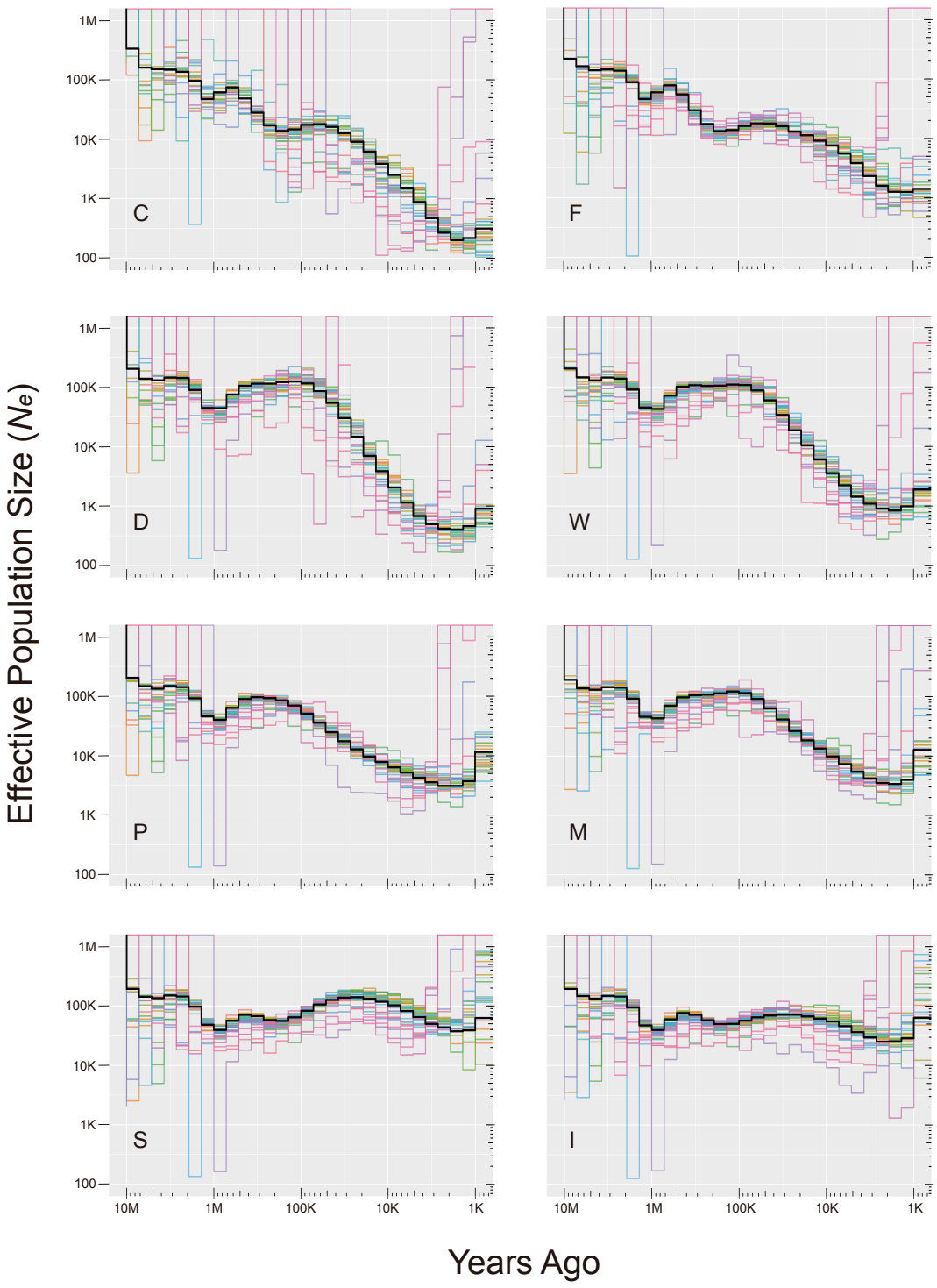


FIGURE 3.14. Demographic history inference for each population based on the joint information from all autosomal chromosomes performed by Relate. Plots show the estimation of effective population sizes (*Ne*) of each island/population throughout the time from past to present.

Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S).

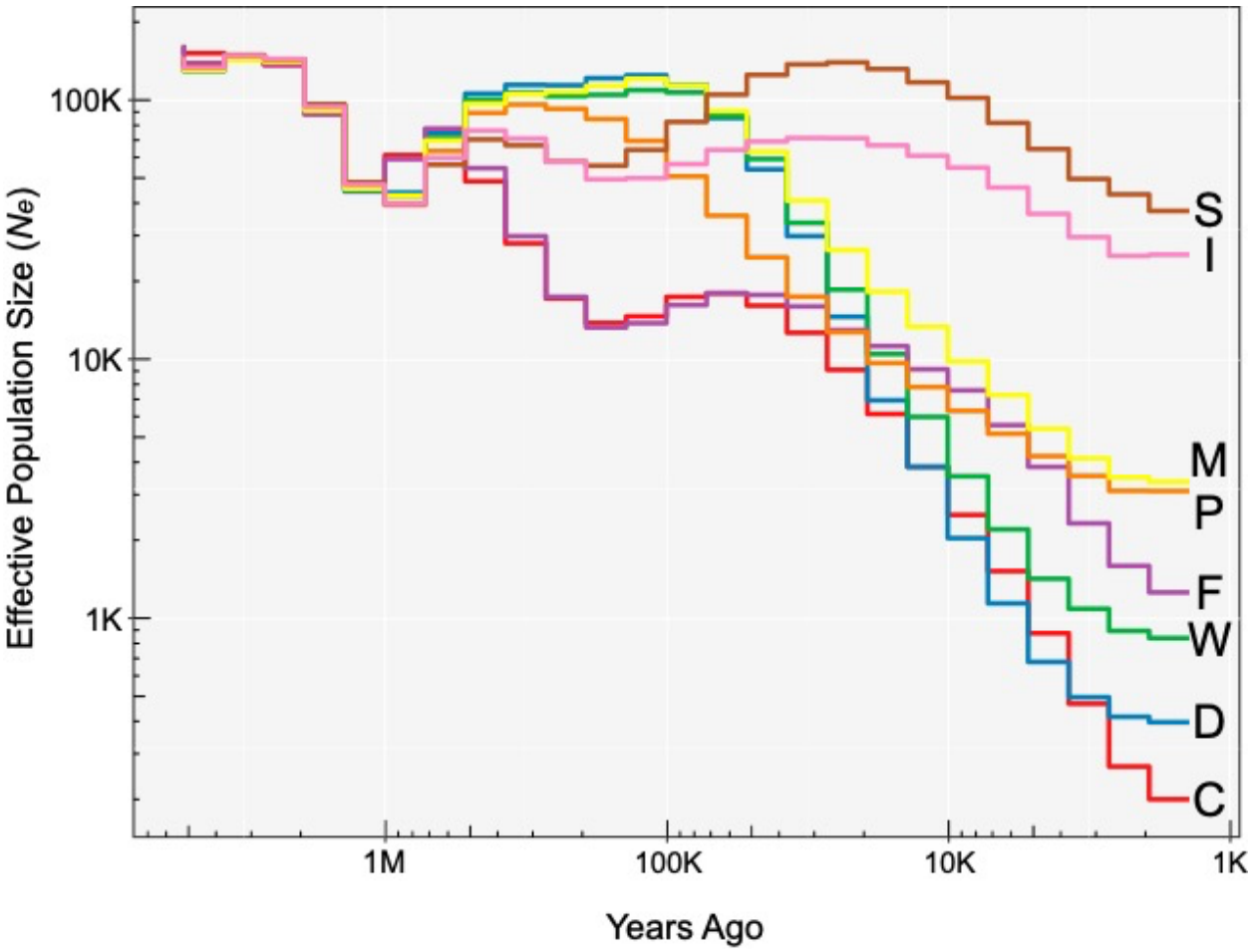
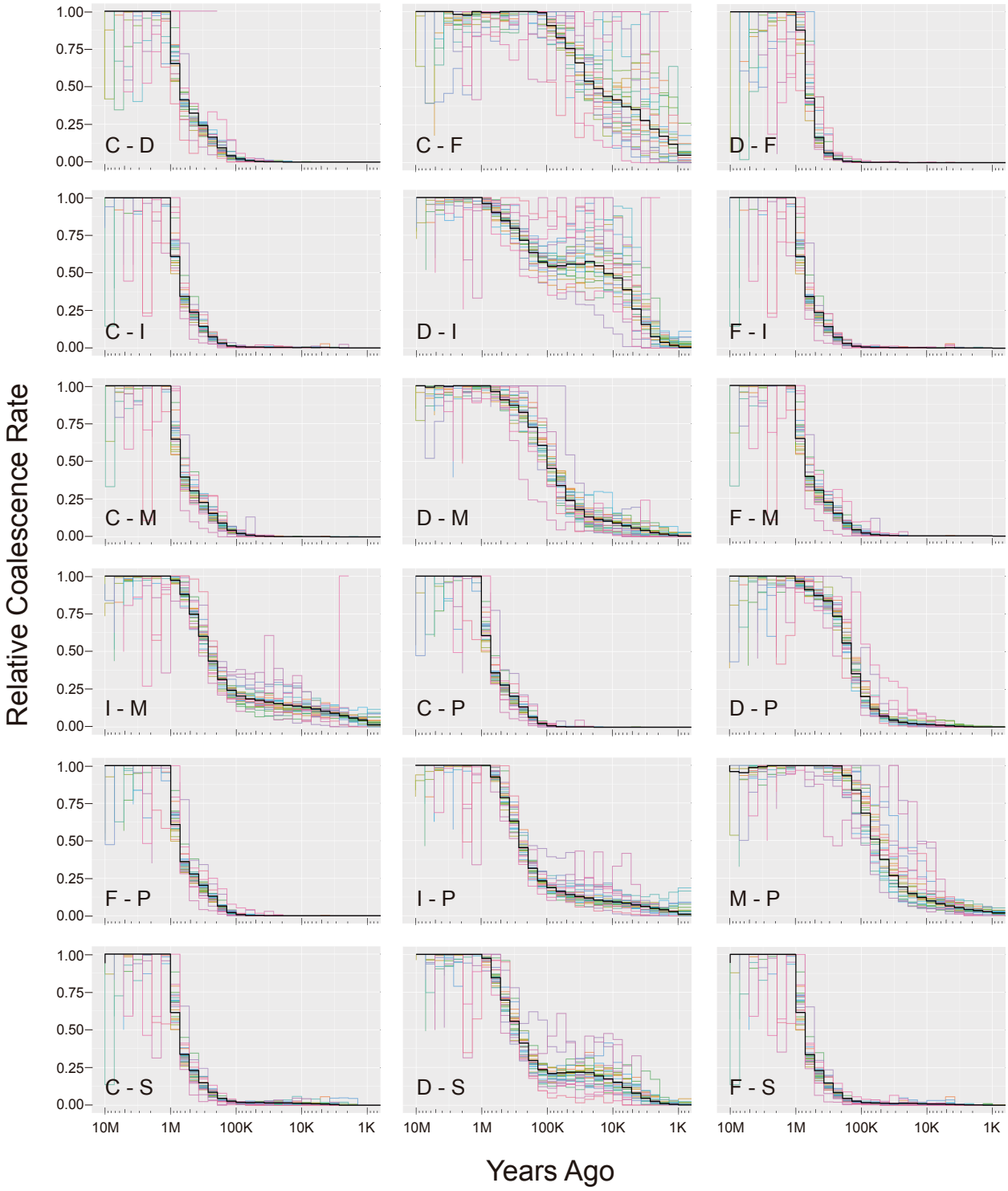


FIGURE 3.15. Comparative summary of demographic history for each island/population inferred using the program Relate. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). The population sizes have been shaped by the constrains of each island size.


Relative Coalescence Rate
1.00
0.75
0.50
0.25
0.00
C - D
C - F
D - F
C - I
D - I
F - I
C - M
D - M
F - M
I - M
C - P
D - P
F - P
I - P
M - P
C - S
D - S
F - S
10M
1M
100K
10K
1K
Years Ago

Continue…

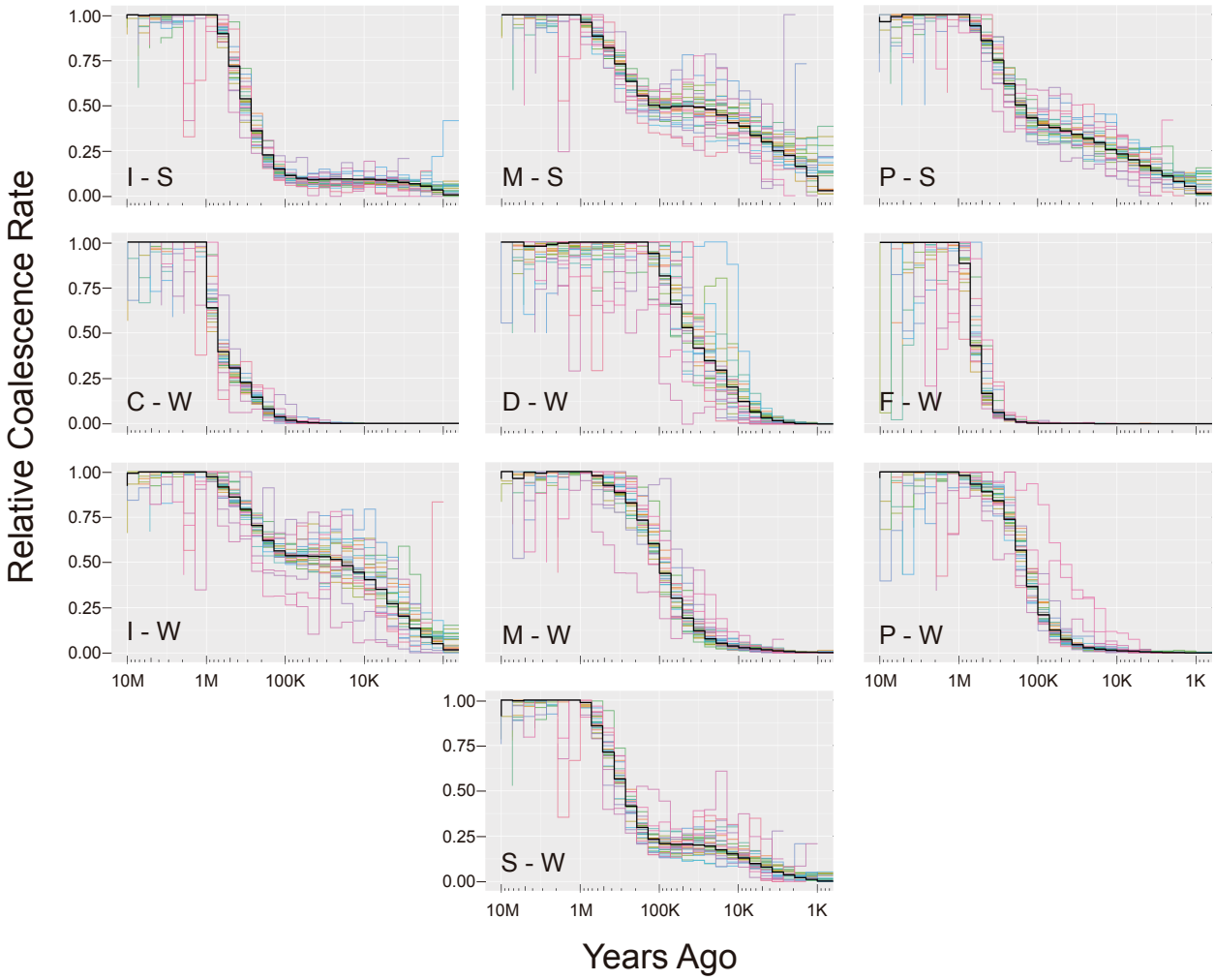


FIGURE 3.16. Pairwise comparisons of the relative coalescence rates through the time (from past to present) for all populations performed by Relate. Coalescence decay revealed a progressive divergence between populations over time, while periods of coalescence stabilization indicated events of gene flow or interconnection between populations. The interconnection is represented by the ratio of within-population coalescent rate over between-population coalescent rate. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). Equity between these rates before 1 Mya indicates that the pair formed a panmictic population while the decrease towards the present indicates increasing separation.

The SMC++ analyses, as in the MSMC2 and Relate analyses, showed an overall decreasing pattern for all populations from 1 Mya and onward. However, SMC++ showed that the demographic histories were much more dynamic over time revealing some clear events of growth and decrease of the population size (Figure 3.17a). For *M. parvulus*, population demographic histories apparently started to split between 500 to 300 Kya. In addition, the demographic history of the two largest islands (Isabela and Santa Cruz) differed remarkably from that of the smaller islands (Marchena, Pinta, Wolf and Darwin), since the population size of the largest islands remained almost constant until the recent past, while the smaller islands displayed important variations that began with a first important decrease ~30 Kya (Figure 3.17b). In particular, the four smaller islands of *M. parvulus* showed an evident population size growth ~20 Kya, and immediately after this abrupt growth their population sizes started to continuously decrease towards to current times. (Figure 3.17c). On the other hand, the two largest populations of *M. parvulus* exhibited similar growth patterns, which, although less pronounced, persisted until approximately 10,000 years ago (Figure 3.17d). Comparing the demographic history of *M. trifasciatus* with the two largest populations of *M. parvulus*, it is recognizable that both species have an early unique history. Their split apparently is found earlier than 1 Mya and since then the population sizes of both species vary independently. However, they reach a stability point at the same time (~150 Kya) and from there, the populations do not show significant changes until the recurrent break point ~20 Kya (Figure 3.17d). Finally, only considering *M. trifasciatus*, the species and its remanent populations alone describe a significant population growth ~5 Kya and then abruptly decayed towards the current times (Figure 3.17e). Additionally, the demographic histories of Champion and Gardner showed significant punctual differences, however, these differences may be caused by a data bias given the high level of inbreeding in the Champion population.

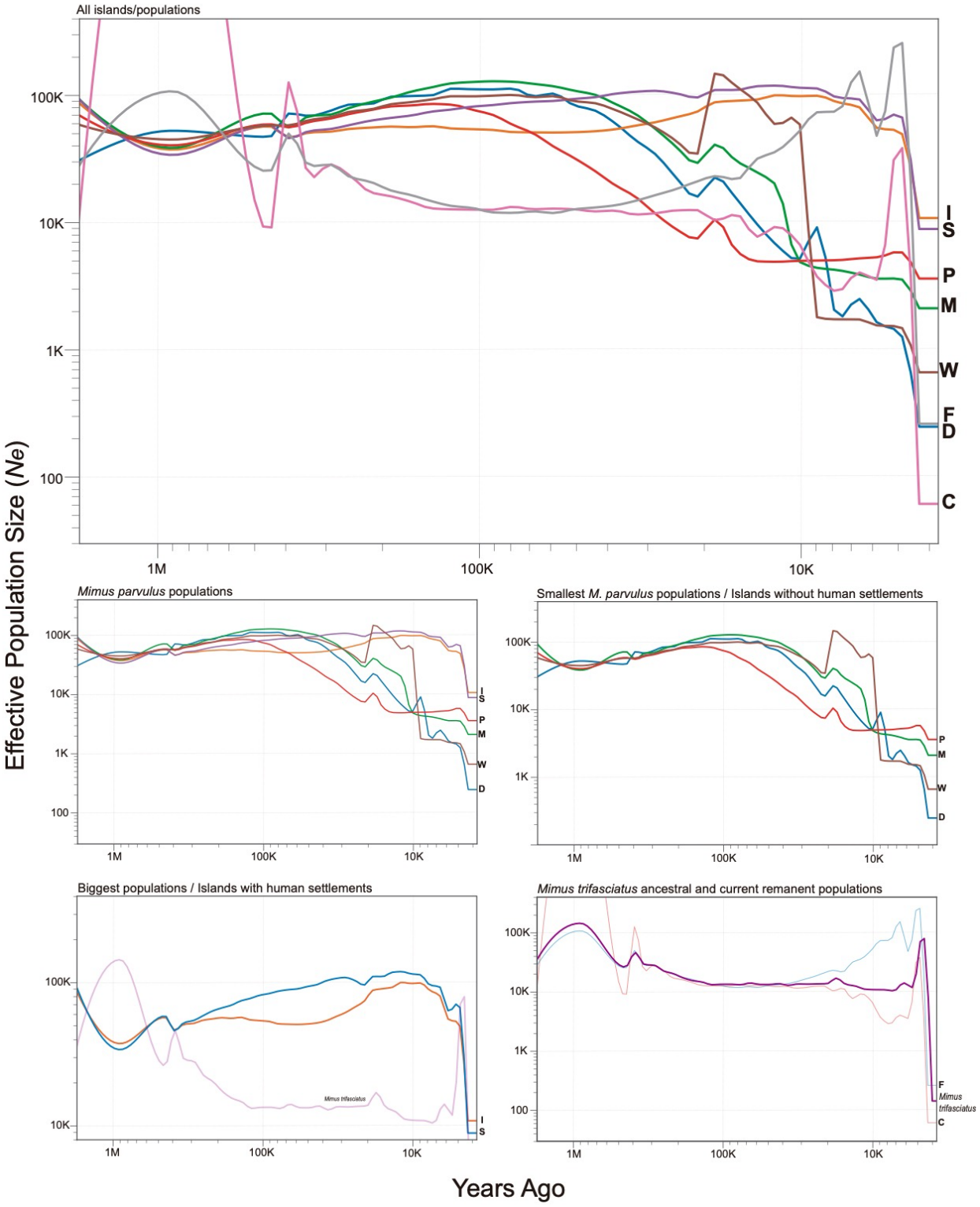


FIGURE 3.17. SMC++ demographic history inferences for Galapagos mockingbirds. In y-axis is the effective population size (*Ne*) and in X-axis is a time scale in years from the past to present.

Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). These plots show the demographic inferences of: a) all current islands/populations; b) *Mimus parvulus* populations; c) the *M. parvulus* four smallest populations, and also these islands have no human settlements; d) the *M. parvulus* two largest and *M. trifasciatus* (Champion + Gardner) populations (the biggest islands), and also these islands have human settlements; and d) *M. trifasciatus* population (C+F) and for each of its remanent sub-populations (C and F).

Given that SMC++ was able to reveal a significant decay event in all the populations analyzed from the recent past to the present, the human presence was analyzed as a triggering factor for this event. This extreme population size decay event started after five thousand years ago in all populations. Therefore, the decay magnitude was estimated based on 1) the variation coefficient and 2) the change percentage during the decay event (from 5 Kya to last *Ne* value estimated) for each population (Table 3.1). The variance analysis was carried out using a fit linear model and the tested factors were: 1) the presence of human settlements on the island, 2) the size of the islands, and 3) the interaction between these two factors. For variation coefficient, the complete model was significant (p=0.041), where the most relevant effect is given by the presence of human settlements (p=0.0784; Table 3.2). Regarding the percentage of change, the complete model was also significant (p=0.033; Table 3.3), where the most important effects are given by the presence of human settlements and also by the size of the island. Therefore, for the two response variables, the decay effect is greater in islands with human settlements versus uninhabited ones. While for change percentage only, it is observed that the decay effect is greater towards smaller islands.

TABLE 3.1. SMC++ results of the last demographic decay event and its relationship with human settlements and the island size. The last decay event magnitude was estimated based on the variation coefficient and the change percentage from 5 Kya to present (last *Ne* value estimated).

Population	*VarCoeff. 5Kya*	*% Change 5Kya*	*Human Settlements*	*Island Size (ha)*
Trifas	1.258596253	98.19818437	Yes	17,300
D	0.726815747	62.57140144	No	106
W	0.38818378	38.73010767	No	134
P	0.231636393	25.7243472	No	5,940
M	0.257908104	27.38872173	No	12,996
S	0.809029043	71.33828785	Yes	98,555
I	0.672212702	60.52645404	Yes	45,8812

**Trifas as Floreana ancestral population → Demographic inference obtained from C + F genomes.*

TABLE 3.2. Factor effect statistics in the last demographic decay based on the population size variation coefficient of the last 5 thousand years (from SMC++ analysis) as the response variable.

	Estimate	**Std. Error**	**t-value**	**Pr(>|t|)**	
(Intercept)	0.90704	0.25182	3.602	0.0367	*
Human Settlements	2.07090	0.78792	2.628	0.0784	·
Island Size	-0.16808	0.07984	-2.105	0.1260	
HumanS:IslandS	-0.24781	0.16937	-1.463	0.2396	

p: ≤ 0.001 (***); ≤ 0.01 (**); ≤ 0.05 (*); ≤ 0.1 (.)

TABLE 3.3. Factor effect statistics in the last demographic decay based on the population size change percentage of the last 5 thousand years (from SMC++ analysis) as the response variable.

	Estimate	**Std. Error**	**t-value**	**Pr(>|t|)**	
(Intercept)	77.536	16.833	4.606	0.0192	*
Human Settlements	131.505	52.669	2.497	0.0880	·
Island Size	-12.934	5.337	-2.424	0.0939	·
HumanS:IslandS	-13.726	11.321	-1.212	0.3121	

p: ≤ 0.001 (***); ≤ 0.01 (**); ≤ 0.05 (*); ≤ 0.1 (.)

The G-PhoCS estimates for theta (θ) and tau (τ) were consistent for all migration tested models giving high reliability in the obtained results. After the raw estimates calibration for theta (θ) in the ancestral populations, the most ancestral population that gave rise to *M. parvulus* and *M. trifasciatus* (PARVTRIF) had an effective population size (Ne) of ~48 thousands. Then, after the first split, the most ancestral population of *M. parvulus* (PARV) increased its effective population size to ~68 thousand individuals, while the *M. trifasciatus* population (TRIF) was constrained to the island of Floreana decreasing its population size to ~15 thousand individuals. Within the *M. parvulus* (PARV) clade, populations decreased in size from PARV ancestral population to central (PMS) and western (DWI) ancestral populations, and then the last two ancestral populations (DW and PM) expanded in size to ~168 and ~189 thousand individuals respectively (Figure 3.18a). On the other hand, the effective population size for current populations showed an evident positive correlation with the island size, where the smallest values were estimated for Champion and Gardner with ~14 and ~62 individuals respectively, which was consistent with the actual population size of these two Floreana mockingbird remanent populations. While the highest population size values were estimated for the largest islands, the islands of Santa Cruz and Isabela had ~50 thousand and ~29 thousand individuals, respectively (Figure 3.18b). Regarding the divergence of the populations, the estimates showed that the split between both species (*M. parvulus* and *M. trifasciatus*) occurred ~427 thousand years ago, then the first divergence within *M. parvulus* clade occurred ~64 thousand years ago splitting the populations of the center of the archipelago (PMS: Pinta, Marchena and Santa Cruz) and the western populations (DWI: Darwin, Wolf and Isabela). Approximately 62 thousand years ago, the central islands clade (PMS) diverged resulting in the current population of Santa Cruz and in the common ancestral population (PM) that would later diverge into the current populations of Pinta (P) and Marchena (M). Then, ~20

thousand years later (43 thousand years ago), the clade of western populations (DWI) diverged giving rise to the current population of Isabela and the ancestral population of Darwin and Wolf (DW). The splits that gave rise to Pinta and Marchena and to Darwin and Marchena occurred ~15 thousand and ~5 thousand years ago respectively (Figure 3.18c). Finally, the most recent split occurred between the remanent populations of *M. trifasciatus* (i.e., on Champion and Gardner) just ~247 years ago, which is consistent with the reported extinction of this species in the main island, Floreana. Detailed population size and divergence estimations of each tested model are found in the Appendix G-PhoCS supplementary tables (SUPPLEMENTARY TABLE 3.1-7).

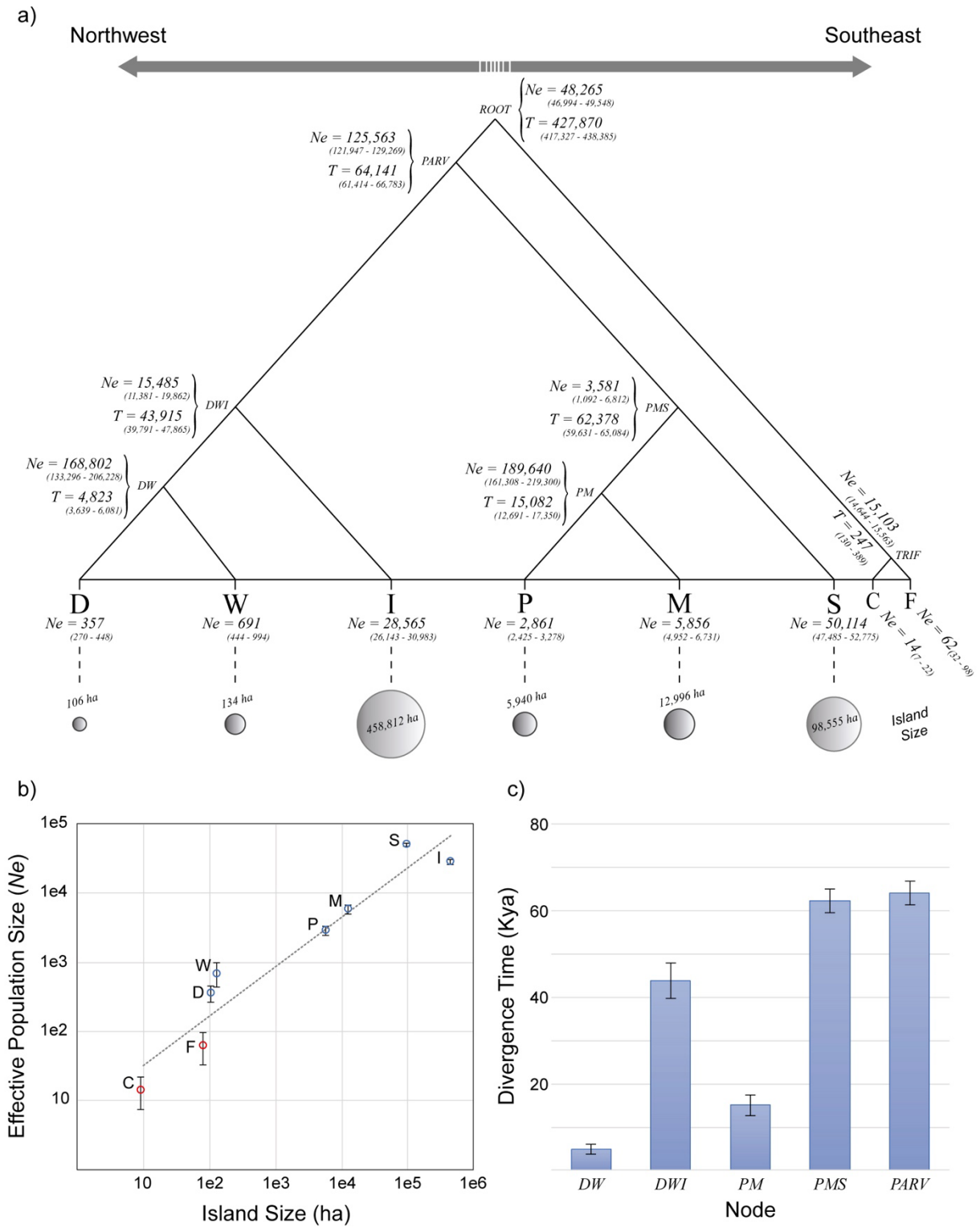


FIGURE 3.18. Overall estimates from all G-PhoCS tested models for population size and divergence. In figure a) are reported all raw and calibrated estimates for population size and

divergence times within the tested phylogenetic topology. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). Figure b) shows the correlation between the effective population size of the current populations and island size, which Spearman correlation coefficient is 0.953. Figure c) shows the divergence times within *M. parvulus* clade, where the more recent divergence events occurred in the northwestern populations clade (DWI) compared with southeastern populations clade (PMS).

The G-PhoCS estimates for total migration rate (m^{tot}) and instantaneous migration (M) were extremely low ($<10^{-3}$ and $<10^{-6}$, respectively) with overlapped confidence intervals for all migration bands in all models tested (Table 3.4, detailed migration estimations of each tested model are found in the supplementary table G-PhoCS). The total migration rate, as cumulative migration probabilities across a given migration band from the most recent point of divergence, is negligible. Similarly, the instantaneous migration (M), which is the probability of migrating through a given migration band in a single generation (adopted as the receptor island colonization probability through a given migration band), is also negligible (Freedman *et al.*, 2014). Therefore, these would show that the probability of post-colonization migratory events is extremely low, as well as the very probability of island colonization.

TABLE 3.4. Summary of the G-PhoCS migration estimates for each migration band and its main variation factors. For each band are reported G-Phocs raw estimates (m; average value for duplicated bands), calculated total migration rate ($m^{tot} = m_{A>B} \cdot \tau_{AB}$) and calculated instantaneous migration rate ($M = m_{A>B} \cdot \mu$).

Migration band	Scaled migration (m)	Total migration rate (m^{tot})	Instantaneous migration rate (M)	Source island size (ha)	Receptor island size (ha)	S-R island distance (Km)	Movement direction
D>I	210.88	8.39E-04	9.70E-07	106	458,812	181.08	NW>SE
D>M	251.23	1.89E-03	1.16E-06	106	12,996	216.01	NW>SE
D>P	228.23	1.66E-03	1.05E-06	106	5,940	176.42	NW>SE
D>S	219.24	1.30E-03	1.01E-06	106	98,555	294.82	NW>SE
D>W	192.05	7.64E-04	8.83E-07	106	134	36.75	NW>SE
I>M	208.45	3.56E-03	9.59E-07	458,812	12,996	82.17	NW>SE
I>P	209.74	3.77E-03	9.65E-07	458,812	5,940	75.78	NW>SE
I>S	200.06	8.32E-03	9.20E-07	458,812	98,555	28.65	NW>SE
M>S	215.04	3.19E-03	9.89E-07	12,996	98,555	85.95	NW>SE
P>M	215.68	3.20E-03	9.92E-07	5,940	12,996	28.15	NW>SE
P>S	218.26	3.23E-03	1.00E-06	5,940	98,555	121.02	NW>SE
W>I	196.07	7.80E-04	9.02E-07	134	458,812	142.97	NW>SE
W>M	219.75	1.49E-03	1.01E-06	134	12,996	181.02	NW>SE
W>P	212.67	1.37E-03	9.78E-07	134	5,940	139.83	NW>SE
W>S	190.47	9.15E-04	8.76E-07	134	98,555	255.90	NW>SE
I>D	186.62	7.43E-04	8.58E-07	458,812	106	181.08	NW<SE
M>D	252.57	1.90E-03	1.16E-06	12,996	106	216.01	NW<SE
P>D	237.23	1.72E-03	1.09E-06	5,940	106	176.42	NW<SE
S>D	232.41	1.36E-03	1.07E-06	98,555	106	294.82	NW<SE
W>D	212.84	8.47E-04	9.79E-07	134	106	36.75	NW<SE
M>I	206.29	3.50E-03	9.49E-07	12,996	458,812	82.17	NW<SE
P>I	197.10	3.55E-03	9.07E-07	5,940	458,812	75.78	NW<SE
S>I	175.22	7.28E-03	8.06E-07	98,555	458,812	28.65	NW<SE
S>M	207.87	3.08E-03	9.56E-07	98,555	12,996	85.95	NW<SE
M>P	184.27	2.73E-03	8.48E-07	12,996	5,940	28.15	NW<SE
S>P	221.14	3.28E-03	1.02E-06	98,555	5,940	121.02	NW<SE
I>W	193.69	7.71E-04	8.91E-07	458,812	134	142.97	NW<SE
M>W	227.01	1.54E-03	1.04E-06	12,996	134	181.02	NW<SE
P>W	203.69	1.31E-03	9.37E-07	5,940	134	139.83	NW<SE
S>W	198.08	9.40E-04	9.11E-07	98,555	134	255.90	NW<SE

3.3.4. *Population genetics and island size associations*

In general, all population genetics indices analysed showed an evident correlation with island size. Individual heterozygosity estimation showed that all individuals formed a cluster based on population location, and they were positively correlated with island size (Figure 3.19a). Conversely, the inbreeding analyses revealed a negative correlation with island size (Figure 3.19b). Additionally, the amount of derived non-synonymous mutations (alleles) in a homozygous condition (recessive load) was negatively correlated with island size (Figure 3.19c). However, the total amount of derived non-synonymous mutations (additive load) just showed a positive correlation for *M. parvulus* populations, while both populations of *M. trifascitus* showed similarly high values, which, when averaged, were higher than the highest value estimated for *M. parvulus* (Figure 3.19d). For detailed values see SUPPLEMENTARY TABLE 3.8.

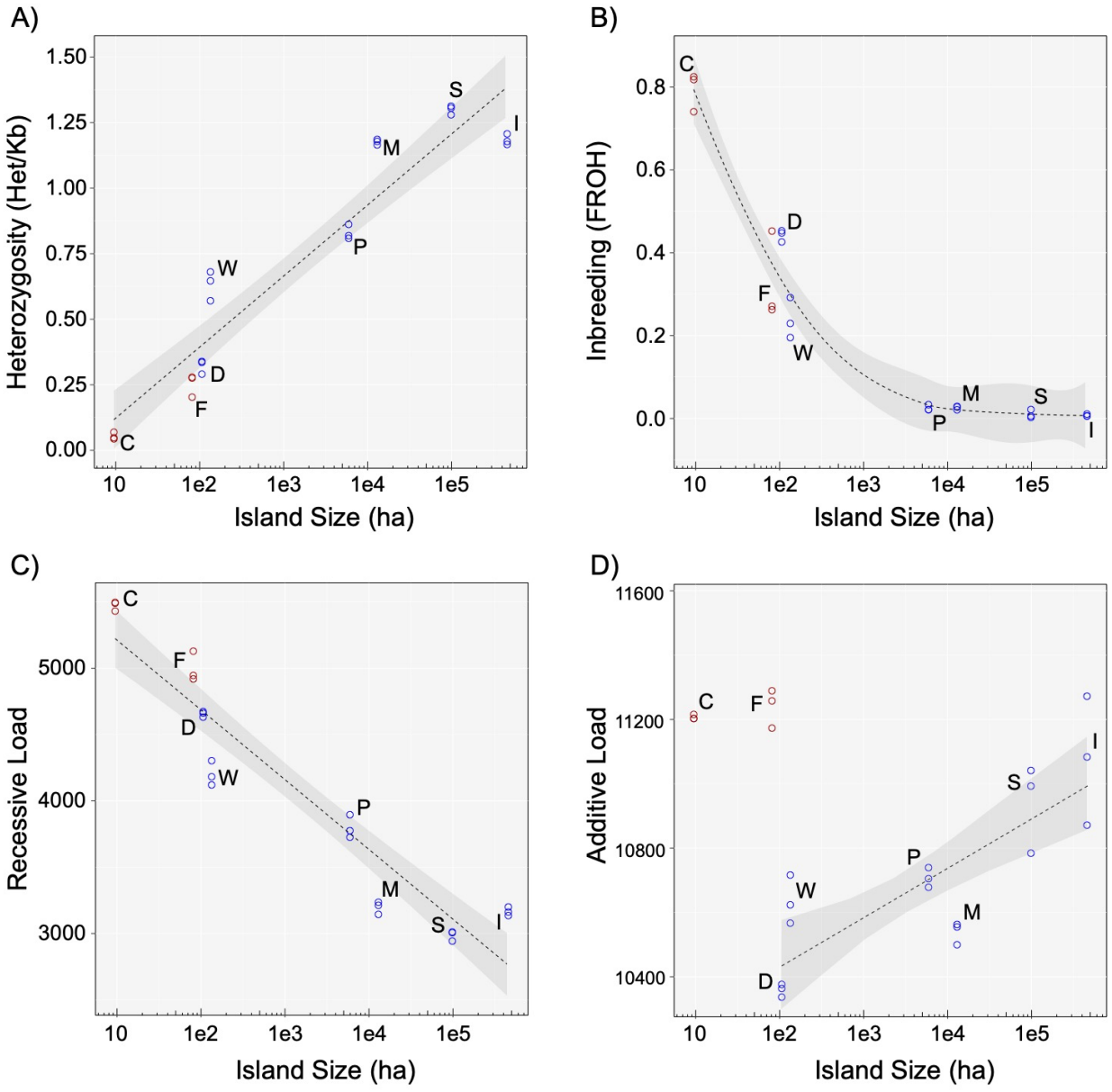


FIGURE 3.19. Regression models between individual genetic indices and island size. Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). Panel shows: A) Mean number of heterozygous SNPs per kilobase, which has a positive logarithmic correlation (R^2=0.892, p<0.000). B) Inbreeding coefficient based on the frequency of long runs of homozygosity in the genome, which has a

negative potential correlation (R^2=0.982, p<0.000). C) Recessive load, which is the total number of homozygous genotypes with a derived non-synonymous allele and has a negative logarithmic correlation (R^2=0.894, p<0.000). D) Additive load, which is the total number of derived non-synonymous alleles. Note that the *M. trifasciatus* populations were excluded from the additive load regression model, only *M. parvulus* populations has a positive logarithmic correlation (R^2=0.718, p<0.000). Each species is color-coded for reference, with red representing *M. trifasciatus* and blue representing *M. parvulus*.

For whole-genome sliding window nucleotide diversity (π) analysis, a quality threshold was imposed based on the quantification of the proportion of covered callable sites for each window to avoid any estimation bias given by low-covered windows. Although no significant correlation between low-coverage and π estimations was found for any population, windows with the lowest values based on a 2.5% quantile cut-off were excluded for the downstream analyses (Figure 3.20). The filtered data set revealed that the nucleotide diversity was highly influenced by the island size. Firstly, the genome-wide nucleotide diversity showed a clear positive correlation with island size, where the nucleotide diversity increased towards the larger islands. Specifically, population π mean values had a significant correlation with island size ($p < 0.01$) based on a Spearman correlation coefficient of 0.929 (Figure 3.21a). In coding regions, the ratio of nucleotide diversity in non-synonymous sites and synonymous sites from coding regions (π_N/π_S), showed different results for both species. While in *M. parvulus* the non-synonymous sites nucleotide diversity proportion showed a negative correlation with island size, whereas the opposite occurred for *M. trifasciatus* (Figure 3.21b). Finally, regarding the distribution of nucleotide diversity (π) and heterozygosity (Het/Kb) throughout the whole genome, the decrease of both parameters from larger to smaller islands is noticeable, but what is equally remarkable is how consistent the patterns

of both parameters are for all populations (Figure 3.22). For detailed values see SUPPLEMENTARY TABLE 3.8.

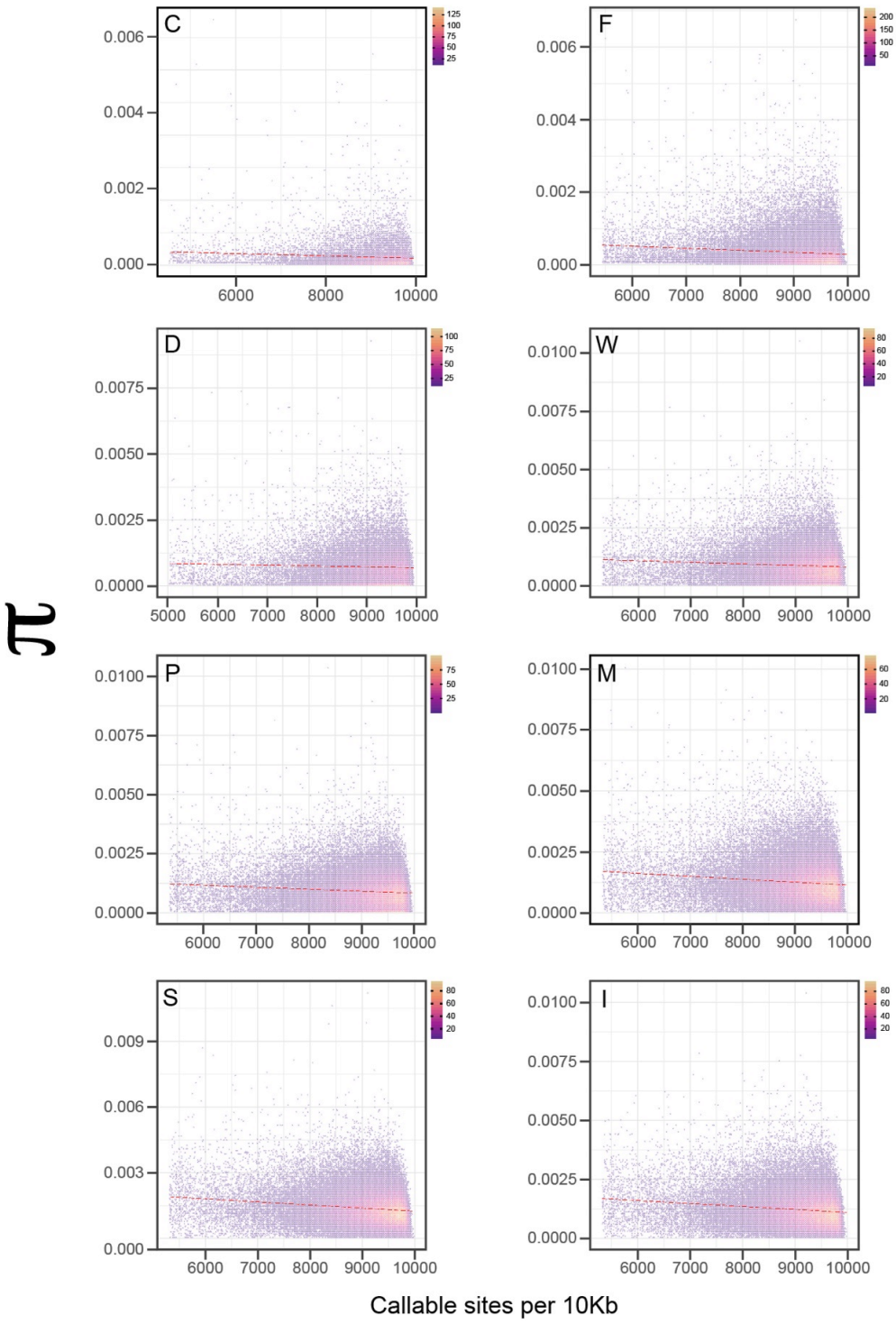


FIGURE 3.20. Scatter plots for each island contrasting the nucleotide diversity (π) and coverage ($n_{callable\text{-}sites}$) in 10 Kb, non-overlapping, windows. After the removal of poorly covered windows ($n_{callable\text{-}sites}$ < quantile 2.5%) no systematic diversity bias of nucleotide diversity (π) is detected.

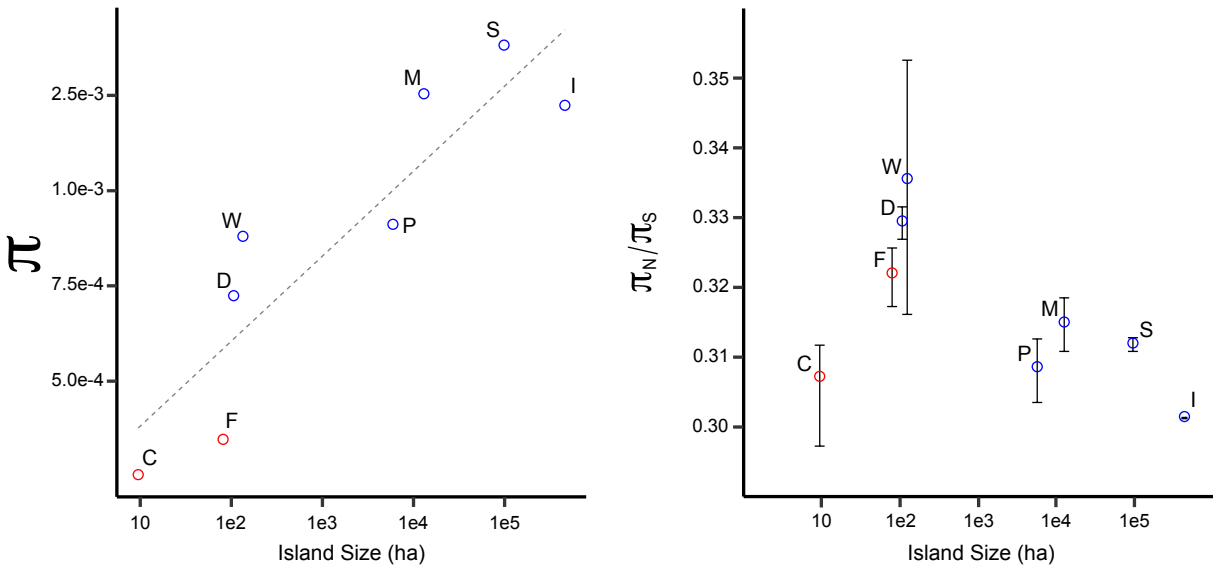


FIGURE 3.21. Sliding-windows analyses for nucleotide diversity (π) by population (red: *M. trifasciatus* populations, blue: *M. parvulus* populations). Populations are Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Isabela (I), Pinta (P), Marchena (M) and Santa Cruz (S). a) Whole-genome sliding-window mean values by population. The results showed a supported positive logarithmical correlation between nucleotide diversity (π) and island size ($\rho_{Spearman}$=0.929; R^2=0.804, p<0.000). b) Ratio of nucleotide diversity (π) between non-synonyms sites and synonyms sites (π_N/π_S) by population. The results did not show a correlation base on all populations, however, only *M. parvulus* populations showed a negative correlation π_N/π_S ratio and island size ($\rho_{Spearman}$=-0.771; R^2=0.852, p<0.000).

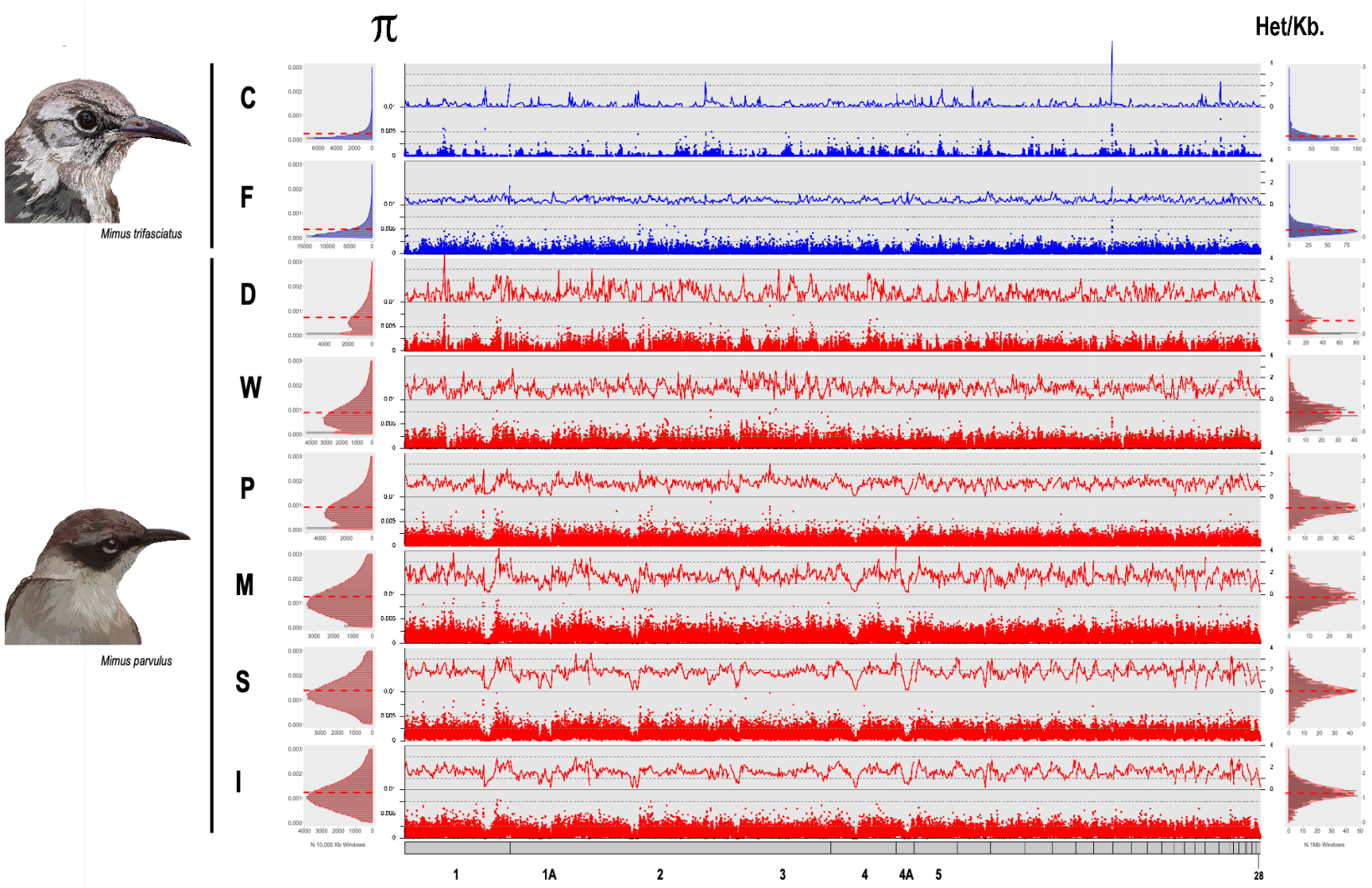


FIGURE 3.22. Nucleotide diversity (π) and heterozygosity (Het/Kb) distribution along whole-genome. Both were estimated by non-overlapping sliding window analysis (sliding-widow size = 10 Kb). Own creation species illustrations from own material.

In contrast, the population inbreeding analyses showed that endogamy is high on small islands, and decreases with increasing island size. Specifically, the lengths distribution of Runs of Homozygosity (ROHs) revealed that the distribution widens with decreasing island size (Figure 3.23a). Additionally, the total sum of ROHs showed striking differences among species and islands. On Champion, the total length of all ROHs reaches up to 880 Mbp (which cover the 99.3% of the genome) while on Santa Cruz it drops down only to 51.9Mbp (5.9% of the genome). The only populations with ROHs over 3 Mb are those of the small islands (Champion, Gardner, Darwin, and Wolf). The individuals on Champion contained between ~75 and 100 long ROH (>2 Mb) (Figure 3.23b). For *M. parvulus*, long ROHs over 2 Mbp are present only on Darwin and Wolf, but they are less frequent than on Champion (~26 ROH/individual). Individuals from Darwin and Wolf showed high frequency of shorter ROHs (0.1-0.5 Kb), indicating that not all small islands had a common inbreeding history. For detailed values see SUPPLEMENTARY TABLE 3.8.

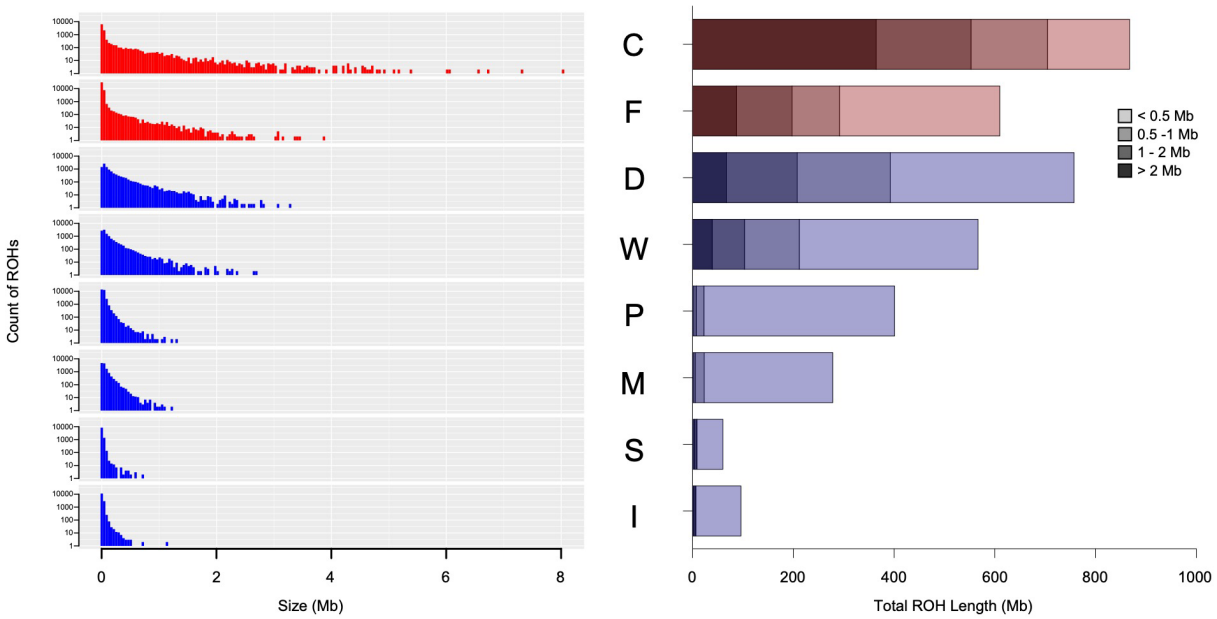


FIGURE 3.23. Runs of Homozygosity (ROHs) distribution by population (red: *M. trifasciatus*, blue: *M. parvulus*). Populations are arranged by island size from the top to the bottom: Champion (C), Gardner-by-Floreana (F), Darwin (D), Wolf (W), Pinta (P), Marchena (M), Santa Cruz (S) and Isabela (I). a) Quantity of ROHs by size for each population, where is recognizable wider distributions the small islands and they become narrow towards to larger islands. b) Total ROHs stacked bars divided in four length categories (light to dark: <0.5, 0.5-1, 1-2, >2 Mb) by population, where the longest ROHs category (>2 Mb) is considerably present in the smaller islands and the most representative example (Champion population) has ~40% of whole genome (~1 Gb) covered by long ROHs (>2 Mb).

Finally, the results of the population LD decay (LDD) analysis also differed between populations and the resulting decay curves could be grouped according to the size of the island ($\rho_{Spearman}$= -0.905**, $p < 0.01$, Figure 3.24). On the big islands of Isabela and Santa Cruz, LDD decayed rapidly and reached the background level (r^2_I= 0.3057, r^2_S= 0.3031) at an average distance

of about ~30 kb between markers. Individuals on medium size islands had similar LD background levels (average: r^2_M= 0.3536, r^2_P= 0.3715) but the decay was not as steep (~150 Kb). In contrast, the small islands of Wolf and Darwin had much higher LD background levels (r^2_D= 0.7333, r^2_W= 0.5342) and much slower decays (Darwin: ~134 Kb, Wolf: ~126 Kb). The two extant populations of *M. trifasciatus* showed contrasting background levels of LD. On Champion, LD extended into unusually large portions of the genome, resulting in an extremely high background of LD (r^2_C= 0.8520). In contrast, Gardner (81 Ha) had background levels of LD (r^2_F= 0.4875) more similar to those found in the small islands of Wolf and Darwin. For detailed values see SUPPLEMENTARY TABLE 3.8.

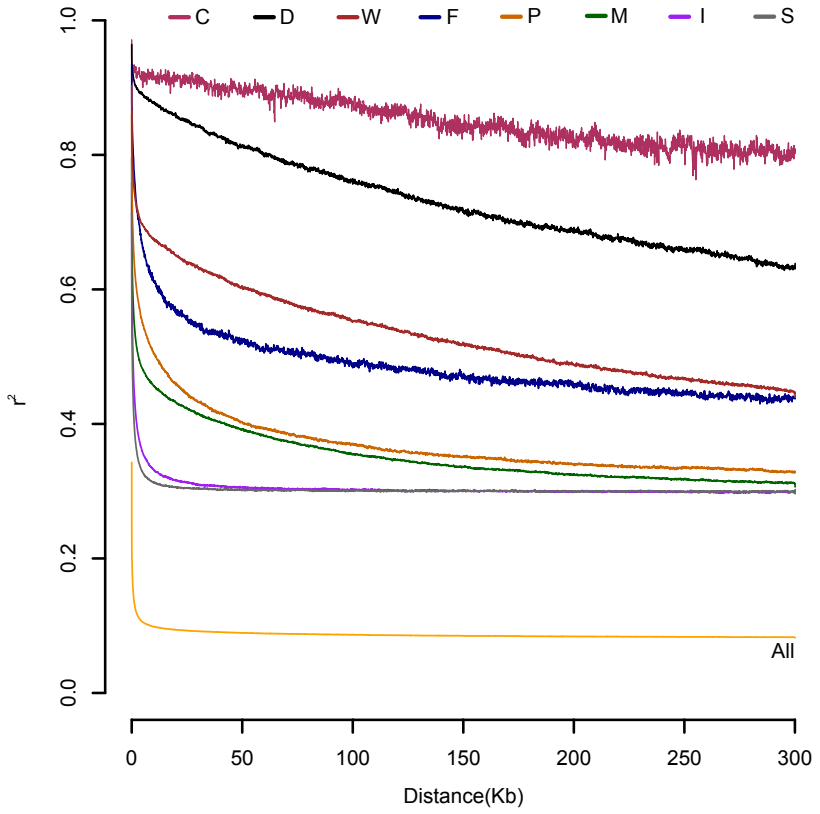


FIGURE 3.24. Linkage disequilibrium decay by population given as the pairwise linkage probability (r^2) by nucleotide distance (Kb). Population are Champion (C), Gardner-by-Floreana

(F), Darwin (D), Wolf (W), Pinta (P), Marchena (M), Isabela (I) and Santa Cruz (S). The r^2 by distance was lower in big islands showing a pronounced decay while towards smaller islands r^2 by distance was increasing showing slower decays.

3.4. DISCUSSION

3.4.1. *Pattern of island colonization and allopatric evolution*

The colonization and distribution of mockingbirds have been shaped by the quite recent origin of the Galapagos archipelago and its continuous genesis of islands. It has been proposed that the volcanic activity of the Galapagos hot-spot began ~56 million years ago; however, the first islands emerged ~5-10 million years ago to ~600 Km away from mainland. Since then, the configuration of the archipelago has been highly dynamic where several islands have been formed and have been eroded to vanish below the ocean (Pindell & Kennan, 2009; Merlen, 2014). The oldest islands of Española and San Cristobal emerged ~3.5 million years ago (Geist *et al.*, 2014). Similarly, the origin of the oldest island terrestrial vertebrate species does not exceed the mean estimation of 3.5 million years ago (Parent *et al.*, 2008). The origin of the Galapagos mockingbirds, dates back to ~3.5 million years ago based on previous phylogenetic analyses of different mitochondrial genes. However, this estimate is relatively broad, between 1.6 to 5 million years ago, suggesting that mockingbirds may have colonized the Galapagos archipelago before any of the current islands existed (Arbogast *et al.*, 2006; Hoeck *et al.*, 2010; Lovette *et al.*, 2012; Sari & Bollmer, 2018). To determine the origin and divergence of the Galapagos mockingbirds it is necessary to include all four species of mockingbird that occur in the archipelago. Although the present study does not include all four species, leaving apart those are expected to be the most evolutionarily basal (i.e., *M. melanotis* and *M. macdonaldi*), my demographic inferences (and

cross-coalescence analysis) were able to determine that the divergence between *M. parvulus* and *M. trifasciatus* occurred between ~0.5 – 1 million years ago. Taking this into account, after that time interval most of the current islands of the archipelago were already formed, with two possible exceptions, Isabela and Darwin (Geist *et al.*, 2014). Even so, the split between the two major clades of *M. parvulus* (DWI and PMS) happened only ~65 thousand years ago, which means that the populations of Isabela, Wolf and Darwin did not exist before this time although these islands were already present for more than 400 thousand years (Geist *et al.*, 2014; Merlen, 2014). Therefore, this leads to the suggestion that the colonization of a new island did not happen as soon as the new island formed, suggesting that the wide-archipelago island colonization was a slow (unlikely) and gradual process for Galapagos mockingbirds.

Mockingbird populations may have evolved in isolation after the colonization of their respective islands. Galapagos mockingbirds are primarily terrestrial birds with limited flight capabilities, which may explain the rarity of inter-island migration events (Curry & Grant, 1990; Grant, 2000). In addition, despite intensive monitoring, no migration events would have been reported since the human history began in the archipelago, with one recent exception. Ortiz-Catedral *et al.* (2021) reported that in 2012 a mockingbird was observed building a nest on Gardner-by-Floreana, an island where only one species, *M. trifasciatus*, is native. The bird was captured and identified as *M. melanotis*, which typically lives on San Cristobal Island. Despite searches, no other individuals of this species were found on Gardner-by-Floreana, and the bird was last seen in October 2012. The relative isolation of a population on an island since establishment would give rise to an independent evolution lineage (Hoeck, 2010b). Based on my analysis, the populations have formed independent groups and have not undergone significant migration events after colonization. This led to a progressive population genetic structuring, consistent with the

island colonization process. My analyses, including maximum likelihood, PCA, and Admixture, revealed genetic structuring among all modern populations studied. The results of demographic and relative cross-coalescence inferences showed that differentiation between populations was a continuous process, with a consistent trend in most cases after their divergence. Furthermore, the G-PhoCS analyses suggested that post-colonization migration rates were extremely low, but not null, indicating that such events may have been unlikely. Consequently, the divergence between populations corresponds to the very moment of their isolation, which has given rise to a unique history of genetic diversity, genetic drift, and inbreeding that were shaped by the environmental conditions of each island. Thus, the Galapagos mockingbirds can be recognized as a remarkable example of allopatric evolution on islands without migration, thereby each of their populations would be linked to the unique variations (environmental, ecological, and geographical) given on each island (Ali & Aitchison, 2014; Bollmer & Nims, 2018; Heads & Grehan, 2021; Woolfit & Bromham, 2005).

Analyses of the phylogenetic relationships and divergence of the different mockingbird populations revealed a directional pattern of colonization of the archipelago. The analyses showed that there was a directionality in a southeast-northwest sense from the most coalescent (or basal) to the most derived populations. Likewise, the estimation of the divergence between populations revealed that the oldest populations were to the southeast, while the newest were located to the northwest. Specifically, *M. trifasciatus* and its populations, being the most southeastern populations of this study, were the most basal phylogenetically and the oldest. Then, for *M. parvulus*, populations were divided in two main clades; the first clade comprised populations at the center of the archipelago (i.e., on Santa Cruz, Marchena and Pinta) being the least derived and representing an older clade than the second (newer clade) that comprised populations to the north-

west (Isabela, Wolf and Darwin). This directionality would be explained by the Galapagos plateau movement direction. Such movement has been ruling the direction from creation to vanish-by-erosion of the islands (NW to SE), and as a consequence has triggered a progressive colonization of the archipelago in the opposite direction (SE to NW) (Ali & Aitchison, 2014; Geist *et al.*, 2014; Merlen, 2014; Head & Grehan, 2021; Shaw & Gillespie, 2016). The colonization progression rule refers to a spatio-temporal progressive colonization model of an oceanic archipelago (hotspot source), where the most basal evolutionary lineages are found on the oldest islands and the divergence of the lineages progressively increases towards the newer islands (Wagner & Funk, 1995). That is, the oldest islands were colonized first (because they were available first) and newer islands that formed were progressively colonized (Shaw & Gillespie, 2016). This pattern has been demonstrated for several organisms in other island archipelago such as the Hawaiian flycatchers (*Chasiempis* spp.; VanderWerf *et al.*, 2010), and Polynesian monarch flycatchers (*Pomarea* spp.) of the Marquesas islands (Polynesia) (Cibois *et al.*, 2004). In the Galapagos archipelago, organisms such as the Galapagos lava lizards (*Microlophus* spp., Benavides *et al.*, 2009), the Galapagos giant tortoises (*Geochelone nigra*; Parent *et al.*, 2008), the Galapagos bulimulid land snails (Parent & Crespi, 2009; Philips *et al.*, 2020), and marine iguanas (*Amblyrhynchus cristatus*; Steinfartz *et al.*, 2009), also follow the same rule of colonization. Although this rule does not apply to all island organisms, it appears to fit remarkably well in organisms that come from a single colonization event and have an allopatric evolutionary model (Parent *et al.*, 2008; Shaw & Gillespie, 2016). The results of the present study showed that for two species of Galapagos mockingbirds, each population revealed a consistently directional pattern that supported the previous findings based on mtDNA analyses (Arbogast *et al.*, 2006; Hoeck *et al.*, 2010b; Štefka *et al.*, 2011).

3.4.2. *Galapagos mockingbird historical demography*

Each island of the Galapagos archipelago has a unique geological history. Since its creation, each island has undergone significant changes in shape and size until its total erosion and disappearance. These changes have taken place mainly due to volcanic activity (as growth factor), erosion (as reduction factor) and global climatic changes (inter-/glaciations as reduction or growth factors respectively). Such dynamic transformations of the islands have direct effects on the effective population size of species that are established on them (Ali & Aitchison, 2014; Geist *et al.*, 2014; Heads & Grehan, 2021; Percy *et al.*, 2016; Woolfit & Bromham, 2005). Consequently, these effects would leave signatures in the demographic history of the organisms. In this way, the significant size changes of islands would give rise to the growth or reduction of the effective population size in those specific moments in time. In the same way, interconnections between islands caused by sea level decrease during glaciations could give rise to genetic flow between ancestral populations and consequently to the growth of subsequent populations (Garg *et al.*, 2018; Gronau *et al.*, 2011; McManus *et al.*, 2015; Spurgin *et al.*, 2014; Szűcs *et al.*, 2014). This assumption would be particularly true in organisms such as the Galapagos mockingbirds, as they exhibit clear population isolation and their population size is positively correlated with the island size they inhabit.

The demographic inferences of the analyzed populations showed highly variable histories for all the populations. Due to their adaptability as generalist species, mockingbirds can exploit a wide range of niches present throughout the island they inhabit, resulting in their widespread distribution (Curry & Grant, 1990; Grant, 2000; Jiménez-Uzcátegui, 2011). Therefore, the variations in the effective population size would have a direct relationship with island size changes.

Accordingly, it can be highlighted that, despite the unique genetic variations among individuals, each population shows a decreasing population size pattern towards the recent past (Figure 3.17). This would be consistent with the gradual erosion of the islands over time and boosted due to anthropogenic effects towards to the recent times (Ali & Aitchison, 2014; Geist *et al.*, 2014; Percy *et al.*, 2016). On the other hand, the change in magnitude of the effective population size is inversely related to island size since the genetic diversity increases logarithmically with population size (Frankham, 1996; Reynolds, 2011). Consequently, the effective population size change is likely to be exacerbated toward smaller populations (islands). Under this assumption, the smaller the island, the more abrupt the population growths and declines would be. Thereby, considering only *M. parvulus*, since the populations of *M. trifasciatus* are the product of a recent bottleneck, it is evident that the four smallest islands of Darwin, Wolf, Pinta and Marchena have much more dynamic demographic histories than those of the larger islands of Santa Cruz and Isabela (Figure 3.17). Therefore, the variation magnitude in the effective population size is given by the magnitude of island-size-changing event, and the magnitude of these events is inversely correlated with the actual size of the island.

The last important glaciation period (Last Maximum Glacial, LMG) would have had highly significant effects in the demographic history of the Galapagos mockingbirds. The LMG started ~20 thousand years ago and declined until it stabilized ~5 thousand years ago. During this period, the ocean level decreased between 140 to 210 meters (Ali & Aitchison, 2014). Consequently, the archipelago increased its land mass by up to ~33% (Ali & Aitchison, 2014; Geist *et al.*, 2014). The effects of these changes are evident in all analyzed populations, for which an important demographic growth is recognized by this period (Figure 3.17). While in populations of small islands this demographic growth is extreme, an abrupt increase. On the large islands of Isabela and

Santa Cruz, as well as (even) for *M. triasciatus* (Champion + Gardner, as the ancestral Floreana population), growth is slight during this same period. Once again, this difference can be explained by the differential impact of the effects depending on the size of the island. In this way, the populations of the large islands originally already had a considerable effective size (~90 thousand individuals for Isabela and Santa Cruz, and ~20 thousand for Floreana) and therefore the addition of a few individuals does not result in significant changes. In contrast, small islands with their limited populations were highly sensitive to any changes, and the introduction of even a small number of individuals would result in significant population fluctuations.

Mimus parvulus and *M. trifasciatus* have contrasting demographic histories. Demographic inferences show that all *M. parvulus* populations follow a similar variation patterns until the recent past, while the *M trifascitus* populations showed independent demographic patterns since its early break with *M. parvulus* (~1MYA). In addition, *M. parvulus* populations show quite dynamic histories, while *M. trifasciatus* populations show a predominantly stable history. The explanation for these differences could lie both in the age of the populations and in their isolation. Based on demographic inferences, the divergence between *M. trifasciatus* and *M. parvulus* is estimated to have happened about 430 thousand years ago. However, the first divergence of *M. parvulus* populations occurred only about 65 thousand years ago, which means that *M. trifasciatus* became isolated as a population much earlier than any other *M. parvulus* population that was examined in this study. Additionally, since its formation (~1.9 MYA), Floreana was considerably isolated from the rest of the islands and its historical geological variations (growth/decrease) remained only local without leading to significant interconnection with other islands in a way that would enable gene flow (Geist *et al.*, 2014). This assumes that once *M. trifasciatus* became established on Floreana, its population filled the whole available niche (i.e., the population grew and spread widely

occupying the entire island) in few generations resulting in a constant equilibrium for effective population size early in its evolution. For *M. parvulus*, after the recent divergence of its populations, a clear instability of the effective population size is observed. This high dynamic would show that the time elapsed since its origin has been insufficient to reach a balance in the population size fluctuations as in *M. trifasciatus*. Apparently, after *M. parvulus* populations had diverged, these still were dealing with the colonization consequences (founder effect) when they had to face additional significant variations given by the LMG (Ali & Aitchison, 2014; Szűcs *et al.*, 2014; Welles & Dlugosch, 2018; Woolfit & Bromham, 2005). Finally, all effective population sizes had an extreme decline within the last ~5,000 years. This extreme phenomenon could be attributed to the beginning of the new inter-glacial era, and later its effect would be enhanced by the beginning of the human era in the archipelago. The beginning of the inter-glacial period gave rise to a gradual rise in sea level and consequently the islands were losing ground which in turn reduced the size of the natural populations (Ali & Aitchison, 2014; Merlen, 2014). Later (~500 years ago), the anthropogenic impact began along with the discovery of the archipelago. This impact directly affected the natural populations due to the introduction of invasive species, changes in the landscape and extractivism (Bollmer & Nims, 2018).

3.4.3. *Island size shapes genome-wide variation patterns*

The effective population sizes of the Galapagos mockingbirds were positively correlated with island size. The Galapagos mockingbirds are generalist species with remarkable niche plasticity, suggesting that each island's population is widely distributed and capable of occupying its entire range. In general, seven vegetation zones in an altitudinal gradient have been recognized on the Galapagos Islands. These areas are grouped into two major groups, arid areas (low-land

areas) and humid areas (highland areas). Only eight (the highest) of the 21 islands have all seven vegetation zones (Percy *et al.*, 2016). The omni-presence of mockingbirds has been reported in all the vegetation zones for each of the islands they occupy. However, they have a predilection for inhabiting arid zones, being the coastal, arid and transition zones where a large number of individuals occur (Curry & Grant, 1990; Kleindorfer *et al.*, 2019). Additionally, their generalist diet, which can include small vertebrates, invertebrates, fruits, seeds, carrion and even blood, allows them to thrive without major setbacks throughout the year despite seasonal variations (Kleindorfer *et al.*, 2019). Conclusively, mockingbirds can occupy the whole extension of the island and can maintain a relatively constant population size throughout time, therefore, the island size would be a meaningful proxy of population size (Curry & Grant, 1990; Grant, 2000; Jiménez-Uzcátegui, 2011). Accordingly, island size would be correlated with the population genetic diversity and inbreeding of mockingbirds (James *et al.*, 2017; Reynolds, 2011; Woolfit & Bromham, 2005).

The population size influences the quantity as well as for the distribution of variants in the population gene pool (Hohenlohe *et al.*, 2018; Woolfit & Bromham, 2005). Mutation gives rise to unique variants in the gene pool of each population. Under a given mutation rate, the number of alleles in the population should increase with population size. Therefore, the larger a population is, the higher its genetic diversity will be (Hohenlohe *et al.*, 2018; Mathur & DeWoody, 2021; Peischl & Excoffier, 2015). As expected, the results of this study showed a clear positive correlation of nucleotide diversity with respect to island size, both from an overall genomic estimation and from the progressive analysis along the genome (i.e., the sliding-window analysis). In this way, the nucleotide diversity increases from the smaller to larger islands (Figure 3.19 and 3.21-22).

Heterozygosity analysis is necessary to understand genotypic distribution in the population genetic pool. The larger a population is the greater the probability of random mating between individuals (panmixia). Consequently, in a large population, the ratio of heterozygous-homozygous of the accumulated variants (alleles) tends to be in balanced proportion; while, for smaller populations, a greater chance of genetic drift and inbreeding will lead to an increase in homozygosity and fixation of alleles within the population (Hohenlohe *et al.*, 2018; Woolfit & Bromham, 2005). Therefore, it is expected that populations on larger islands will have greater heterozygosity than those on small islands (James *et al.*, 2017 Reynolds, 2011; Woolfit & Bromham, 2005). As expected, my results supported this assumption for the populations of Galapagos mockingbird examined in the present study. I found a positive correlation between heterozygosity and island size. I observed the lowest heterozygosity on Champion, the smallest islet, and the values increased progressively towards the largest islands.The greatest level of heterozygosity was detected for individuals on the largest islands, Santa Cruz and Isabela. I also expect that there will be higher inbreeding (homozygosity), and higher allele fixation and linkage disequilibrium (as high genetic drift consequences) on smaller islands than on larger islands (Hedrick & Garcia-Dorado, 2016; Mathur & DeWoody, 2021; Robinson *et al.*, 2019). As expected, I found a negative correlation between inbreeding in mockingbirds and island size. However, the population on Gardner-by-Floreana is an exception. Despite the fact that Gardner has the second smallest area, the estimated rate of inbreeding and LD decay was lower than those for the two small-sized islands, Darwin and Wolf, which are larger in size. This result could be explained by a recent bottleneck event occurring for the Gardner population and by its close location to its sister population on Champion. Both Gardner and Champion populations were stemmed from the large *M. trifasciatus* population from Floreana island and became extinct approximately 150 years ago

(Grant *et al.*, 2000, Jiménez-Uzcátegui *et al.*, 2011). Therefore, while the ancestral genetic diversity may still be in the process of being reduced by inbreeding resulting from the bottleneck on Gardner, the extreme bottleneck on Champion quickly wiped out such ancestral diversity in just a few generations. Thus, on Gardner the few generations that have elapsed in the last ≥150 years have not been able to fix variants due to its population size. In other words, the bottleneck process is still active (it has not reached the equilibrium) in this population. Conversely, on Champion, which contains the smallest, and sister, population of *M. trifasciatus,* the mockingbird population is extremely constrained by the carrying capacity of this small islet. In this case, the effects of the bottleneck were extreme and brought the population to equilibrium in very few generations.

3.4.4. *Human impacts, small islands, and conservation considerations*

Human impact effects in the Galapagos archipelago have been undeniable. Since the discovery of the archipelago in the 16th century, human activity has begun to wreak havoc on island ecosystems and their natural populations. Human disturbances started with the hunting and extractivism of native species, as well as with the introduction of foreign species (which later became invasive species). These disturbances have been intensifying to the present day as human settlements and their economic activities increased. Human presence in the archipelago has resulted in the extinction of several endemic species (and populations). These include the Pinta and Floreana populations of giant tortoise and the Galapagos giant rat. Also, other species or populations are of critical conservation concern. These include the Fernandina giant tortoise, and the Galapagos petrel (IUCN, 2022). The first reported extinction of a land bird in the archipelago caused by anthropogenic effects was the San Cristóbal Island Vermilion Flycatcher, *Pyrocephalus*

dubius, which disappeared during the last decade (Carmi *et al.*, 2016). According to the IUCN, 14 of the 28 endemic species of Galapagos land birds are extinction threatened (IUCN, 2022). Among these, some notable examples of this are the mangrove finch (*Camarhynchus heliobates*), the Galapagos martin (*Progne modesta*) and the Floreana mockingbird (*Mimus trifasciatus*). Nonetheless, it must be noted that populations of all native land bird species have diminished in number since the beginning of human presence in the archipelago, and the Galapagos Mockingbirds have not been the exception.

Anthropogenic effects might be evidenced in the demographic history of the Galapagos mockingbirds. The demographic inferences of the analyzed populations showed dynamic histories over time with certain unique variations for each population. Yet all demographic histories lead to a steep decline (bottleneck) in the recent past. Interestingly, no such extreme change was detectable at any other time in their history, not even in an event as relevant as the LMG. This dramatic decay in the recent past for all populations would be consistent with the beginning of the human era in the archipelago (Steadman *et al.*, 1991, Conrad & Gibbs, 2021). Indeed, the effects of human impact can be seen in the difference in the decline of population intensity between the islands with and without human settlements. My results indicate that the population decline in the islands with human settlements (Santa Cruz, Isabela and Floreana) are more abrupt. This finding was supported by the statistical tests (Figure 3.17; Tables 3.2-3), which showed that there was a proportionally greater population decline on islands with human settlements than on uninhabited islands. On Floreana, human inhabitation would be expected due to the proven recent extirpation of the *M. trifasciatus* population, which also left a population genomic trace of an extreme recent bottleneck revealed by the ROHs analysis for the populations of *M. trifasciatus* on Champion and Gardner. In contrast, the populations on Santa Cruz and Isabela, despite being the most stable and largest

populations over time, have significantly declined compared to all other populations of *M. parvulus*. Therefore, as would be expected, the anthropogenic effects in mockingbird populations have been more intense on islands with human settlements (Wikelski *et al.*, 2004).

Mockingbird populations on small islands face significant challenges to their long-term survival due to reduced genetic diversity compared to other birds, especially in smaller populations like Champion and Gardner (Leroy *et al*., 2021). Despite this, some populations have apparently remained stable over time (Curry & Grant, 1990; Grant, 2000; Jiménez-Uzcátegui *et al*., 2011), which is especially remarkable for a population like Champion's that has more than 90% of its genome in homozygosity. However, all populations have been affected by anthropogenic impacts, with populations on islands with human settlements being particularly vulnerable. This poses a significant risk to small island populations that were already small and have been further reduced by human activities (Kyriazis *et al.*, 2020; Mathur & DeWoody, 2021). In this way, the smaller a population, the less genetic diversity within that population, and the greater the chance of inbreeding. Consequently, populations in extremely small islands could undergo extirpation after a few generations due to the progressive increase in inbreeding and genetic drift (James *et al.*, 2016; Robinson *et al.*, 2016). An increase in inbreeding and genetic drift will lead populations to a vortex of extinction (mutational meltdown) due to the progressive reduction of the average fitness of the population given by the expression of deleterious alleles (genetic load) until its eventual extinction. Deleterious alleles are nonsynonymous mutations (usually recessive) that decrease individual fitness when they are expressed (Kyriazis *et al.*, 2020; Mathur & DeWoody, 2021; Robinson *et al.*, 2019; van der Valk *et al.*, 2019). It is expected that populations on small islands (as Champion, Gardner, Darwin, and even Wolf) show a persistent decline to near disappearance. Instead, it was found that these small populations during their decline found a stasis

point in which they have managed to survive for several generations to the present day. These populations would have reached a balance between genetic load purging and maintain their population size (Hedrick & Garcia-Dorado, 2016; Robinson *et al.*, 2018; van der Valk *et al.*, 2019).

An assessment of genetic load would provide insights for the long-term survival of mockingbird populations on islands (Mathur & DeWoody, 2021; Robinson *et al.*, 2018). Given that mutations are a random phenomenon, it would be expected that the larger the population the more the mutations would be contained on its genetic pool (greater diversity). This correlation would be true even for mutations in coding sites, and consequently, also for non-synonymous mutations. Accordingly, there should be a positive correlation between the number of non-synonymous alleles and the island size. However, the non-synonymous allele fixation (population homozygosity) would be negatively correlated with island size due to the high diversity (high heterozygosity) towards larger populations (Leroy *et al*., 2021; Hedrick & Garcia-Dorado, 2016; Mathur & DeWoody, 2021; Peischl & Excoffier, 2015). Although the homozygosity of non-synonymous alleles (recessive load) reported for mockingbirds in the present study clearly showed a negative correlation with respect to the size of the island as expected, the number of non-synonymous alleles (additive load) did not fit to the expected correlation (Figure 3.19). Specifically, for additive load, it was observed that the expected correlation fits only for *M. parvulus* populations and that the great exceptions are the two remnant populations of *M. trifasciatus* on Champion and Gardner. The explanation for these exceptions would lie in the recent extreme bottleneck that these populations faced. These two populations, Champion (*Ne* ~30) and Gardner (*Ne* ~120) come from a much larger population (*Ne* ~15,000) that itself became extinct only ~150 years ago (only ~33 generations apart). Therefore, the outstanding amount of non-synonymous alleles found in these two remanent populations could have been inherited from the

ancestral population genetic diversity. Currently, these non-synonymous alleles are still in the genetic pools of the populations on Champion and Gardner since the few generations that have passed have not yet been able to purge them (Hedrick & Garcia-Dorado, 2016; Mathur & DeWoody, 2021; Peischl & Excoffier, 2015; Robinson *et al.*, 2016). In contrast, small populations of *M. parvulus*, such as those on Darwin and Wolf, showed a lower additive load than larger populations, which shows that these populations have been small and stable for many generations (i.e., or at least 5,000 years). This has given enough time to allow genetic purging actuates by discarding non-synonymous alleles (especially highly deleterious ones) from the gene pool. As a consequence, individuals in small populations have fewer non-synonymous alleles an also have greater homozygosity (Kyriazis *et al.*, 2020; Robinson *et al.*, 2018; van der Valk *et al.*, 2019).

The ratio of non-synonymous to synonymous mutations could reveal the intensity of a recent bottleneck. Interestingly, a similar result to the additive load can be evidenced in the non-synonymous and synonymous nucleotide diversity ratio (π_{ns}/π_s) analysis. In small populations, high genetic drift is expected to randomly fix synonymous and non-synonymous mutations giving rise to a similar ratio between them. While in larger populations the drift effect is reduced, which allows selection to discard non-synonymous (essentially deleterious) mutations. Consequently, this results in a lower proportion of non-synonymous mutations versus synonymous mutations. This pattern would be observed only in populations of *M. parvulus*, whereas populations of *M. trifasciatus* would be exceptions to this (like happened in the additive load results). In general, the difference found for *M. trifasciatus* can be explained once again by the gene pool inherited from a much larger population (Floreana population) that was drastically and recently reduced towards the current populations (Champion and Gardner). Thus, Champion shows a similar π_{ns}/π_s ratio to that found for the largest populations of *M. parvulus*; while Gardner shows a somewhat higher

ratio than Champion, although it is still lower than the ratio found for the smaller populations of *M. parvulus* (Darwin and Wolf). This "unexpected" pattern between both populations of *M. trifasciatus* could be explained by the severity of the recent bottleneck event (remanent population size) and the genetic diversity contained. On the one hand, Champion being an extremely small population (~30 individuals), the probability of new mutations per generation in the population are exceptionally low in such a short time (<150 years) and the high inbreeding has led to a perennial homogeneity in the population gene pool. Consequently, the π_{ns}/π_s ratio would be an "instantaneous snapshot" of the π_{ns}/π_s ratio for the ancestral population *M. trifasciatus*. On the other hand, Gardner, having a larger population than Champion, allowed the retention of a greater nucleotide diversity from the ancestral population. However, the high genetic drift could give rise to the fixation of non-synonymous alleles that were contained in this greater diversity increasing the current π_{ns}/π_s ratio. As a final note, although it has not been explored in the present investigation, the effect of genetic purging could provide an additional alternative or complementary explanation to the general pattern of the π_{ns}/π_s ratios. Whereas in large populations deleterious (non-synonymous) alleles are masked by heterozygosity and discarded by selection, genetic purging in small populations discards deleterious alleles from the gene pool by removing individuals homozygous for these alleles. The smaller a population, the greater the inbreeding, and hence a greater the probability that individuals with two deleterious alleles are purged. Consequently, genetic purging is expected to be more stringent in smaller populations, decreasing the proportion of non-synonymous mutations in the gene pool. Therefore, this phenomenon could explain the general pattern of inverted u, where extremely small populations reduce the amount of deleterious (non-synonymous) alleles reaching ratios similar to that of larger populations. If this was found to be the case in future investigations, then it would demonstrate that: (1) after a

breakpoint for effective population size, the effect of gene purging becomes significant; and (2) the smaller the population, the greater the purging effect (Hedrick & Garcia-Dorado, 2016; Robinson *et al.*, 2018; Robinson *et al.*, 2019; van der Valk *et al.*, 2019).

Human activities, such as habitat disturbance, the extraction of native species, and the introduction of invasive species, have had a significant impact on the Galapagos archipelago, leading to the extinction of several endemic species and populations, including the Galapagos mockingbird. The impact of human settlements has been more intense on all mockingbird populations. Despite smaller populations, such as those on Champion and Gardner islands, facing challenges due to reduced genetic diversity, these populations have remained stable, as shown in this chapter. To aid conservation efforts for Galapagos mockingbirds, population genomics can be used to assist with genetic rescue, monitoring, and management. Population genomics can provide detailed information on the demographic history of each population/species, their current population diversity, and relatedness with other populations, and can precisely track changes over time. This information can be used to develop specific evolutionary models, and different disturbance scenarios can be simulated to inform conservation strategies and establish informed plans for population recovery, such as genetic rescue. Genetic rescue involves introducing new genetic material into a population with low genetic diversity to increase genetic variability and improve the fitness of the population. Translocations, which involve moving individuals from one population to another, are one method of genetic rescue, but must be carefully planned and monitored to avoid negative consequences such as the introduction of new deleterious alleles, diseases, or parasites. Population genomics can also help identify populations that may be best suited for captive breeding or reintroduction efforts. Therefore, sound scientific knowledge of the genetics and ecology of the populations involved is essential for successful conservation efforts,

with population genetics being a crucial tool. Thus, population genetics is necessary for effective monitoring and designing management strategies to conserve not only mockingbirds but other vulnerable island species.

CHAPTER 4: CONCLUSIONS AND FUTURE PERSPECTIVES

The results of my dissertation reveal some key aspects in the evolution, demographic history, ecology, and conservation of Galapagos mockingbirds. Since (re)colonization events are highly unlikely, demography and population divergence are tightly correlated with the spatio-temporal geographical landscape of the Galapagos islands archipelago, and with anthropogenic impacts. These general conclusions set up the baseline for future studies on ecology, evolution and species management and conservation. In addition to the results I have presented in this thesis, I have also determined the DNA sequences of an additional 30 mitochondrial and nuclear genomes of other specimens of *M. trifasciatus* and the two other *Mimus* species (i.e., *M. melanotis* and *M. macdonaldi*) present in the archipelago but that were not included in this dissertation. These data will allow scientists to better understand the origins and phenotypic diversification of this group, to understand the consequences of hybridization, and to fully characterize the genetic load of the extant populations on the island of Floreana.

4.1. ORIGIN AND DIVERSIFICATION OF GALAPAGOS MOCKINGBIRDS (*Mimus* spp.)

To fully understand the evolutionary process that gave rise to a group of related species in a remote landscape (e.g., oceanic archipelago), it is essential to elucidate and reconstruct the history of colonization and its demographic-evolutionary history (Hohenlohe *et al.*, 2018; Ottenburghs *et al.*, 2019; Welles & Dlugosch, 2018). While my dissertation focuses on the phylogenetic relationships between populations of *M. parvulus* and *M. trifasciatus*, the newly obtained genomes of the other taxa, will allow me to determine the timing of colonization and divergence events.

Previous molecular analyzes based on a handful of nuclear and mitochondrial markers suggested that the Galapagos mockingbird system stemmed from a single colonization event that occurred approximately 3.5 million years ago; however, this timing has been estimated by different researchers to range anywhere between 1.5 to 5.5 million years ago (Arbogast *et al.*, 2006; Hoeck *et al.*, 200b; Nietlisbach *et al.*, 2013; Štefka *et al.*, 2011). The incorporation of additional mitochondrial and nuclear genomes will allow me to obtain a more precise estimate not only on the original colonization event but also a timeline of the formation of each species of Galapagos mockingbird.

Preliminary phylogenetic analyses indeed corroborate the existence of a single common ancestor for the Galapagos mockingbird, and consistently indicate that the Genovesa population of *M. parvulus* (i.e., = *M. parvulus bauri*) is a possible mixed lineage as a result of a hybridization event (Figures 4.1 & 4.2). However, the patterns of mitonuclear discordance are not limited to the clade containing *M. parvulus* from Genovesa. For example, although the results of the nuclear SNPs analyses suggest that individuals from the two remote populations on Darwin Island and Wolf Island are sister taxa (i.e., are the closest relatives of one another) (see Chapter 3), the results of the mitochondrial SNPs analyses show a very different pattern. In this case, the population on Wolf Island are most closely related to the population on the island of Isabela, whereas the population on Darwin Island is most closely related to the population on the island of Marchena. As in the case of the evolutionary origin of the population on Genovesa, this genealogical discordance of populations on the islands of Darwin and Wolf likely reflects reticulated divergence related to genetic flow (Hastings & Harrison, 1994; Nadeau & Kawakami, 2018; Toews *et al.*, 2016). Interestingly, in the mitochondrial maximum likelihood tree, there is low support for the basal node of all *Mimus* species in the Galapagos archipelago, which could be defined as a

polytomy. While it is possible that the divergence between species occurred almost simultaneously (Elgvin *et al.*, 2017; Nadeau & Kawakami, 2018; Nietlisbach *et al.*, 2013; Ottenburghs *et al.*, 2019; Toews *et al.*, 2016).

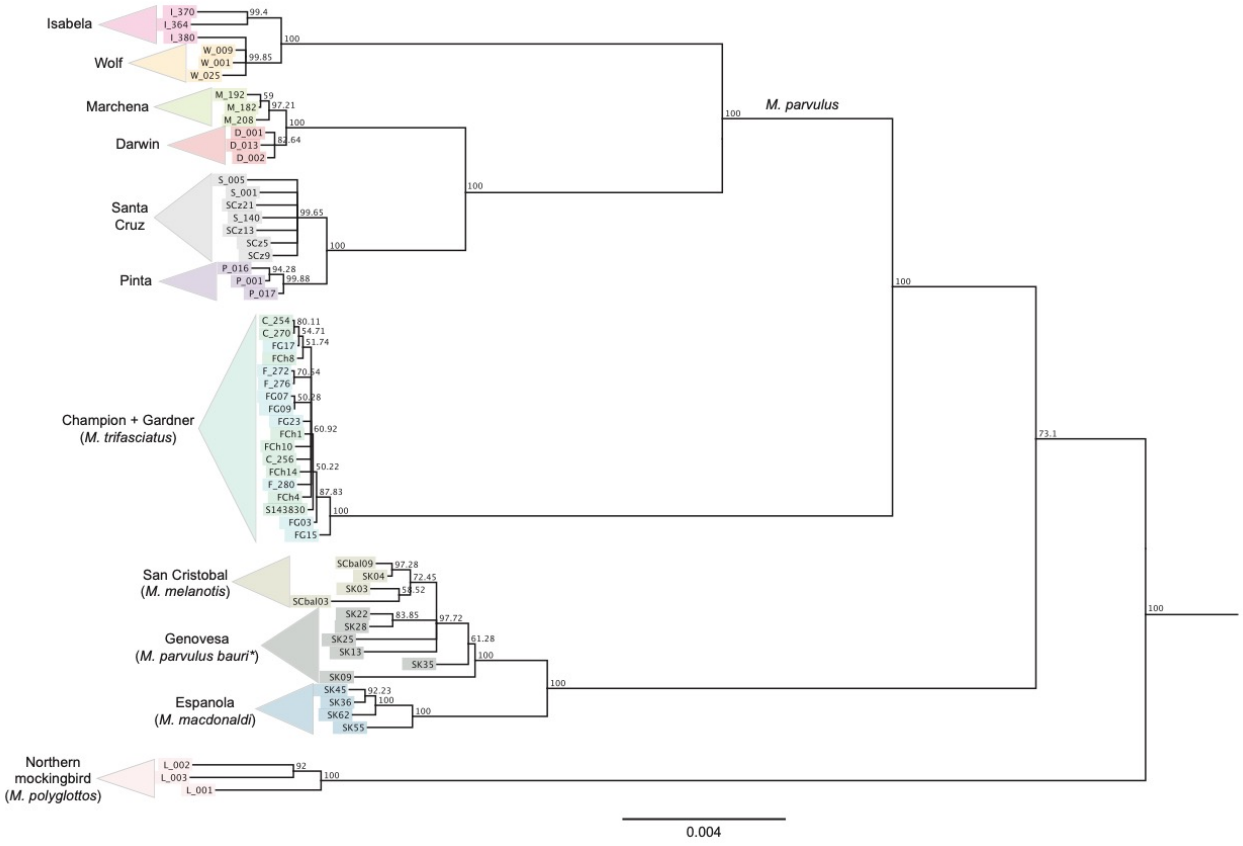


FIGURE 4.1. Maximum-likelihood phylogenetic tree based on whole mitogenome sequences from 57 *Mimus* individuals. The analysis involves 54 Galapagos mockingbird mitogenomes that include the four species (*Mimus parvulus*, *M. trifasciatus*, *M. melanotis*, and *M. macdonaldi*) and eleven population/islands (Darwin, Wolf, Isabela, Pinta, Marchena, Santa Cruz, Genovesa, Champion -Floreana-, Gardner -Floreana-, San Cristobal, Española), and three outgroup mitogenomes from *M. polyglottos*. Bootstrap branch support values were determined after 10,000 iterations.

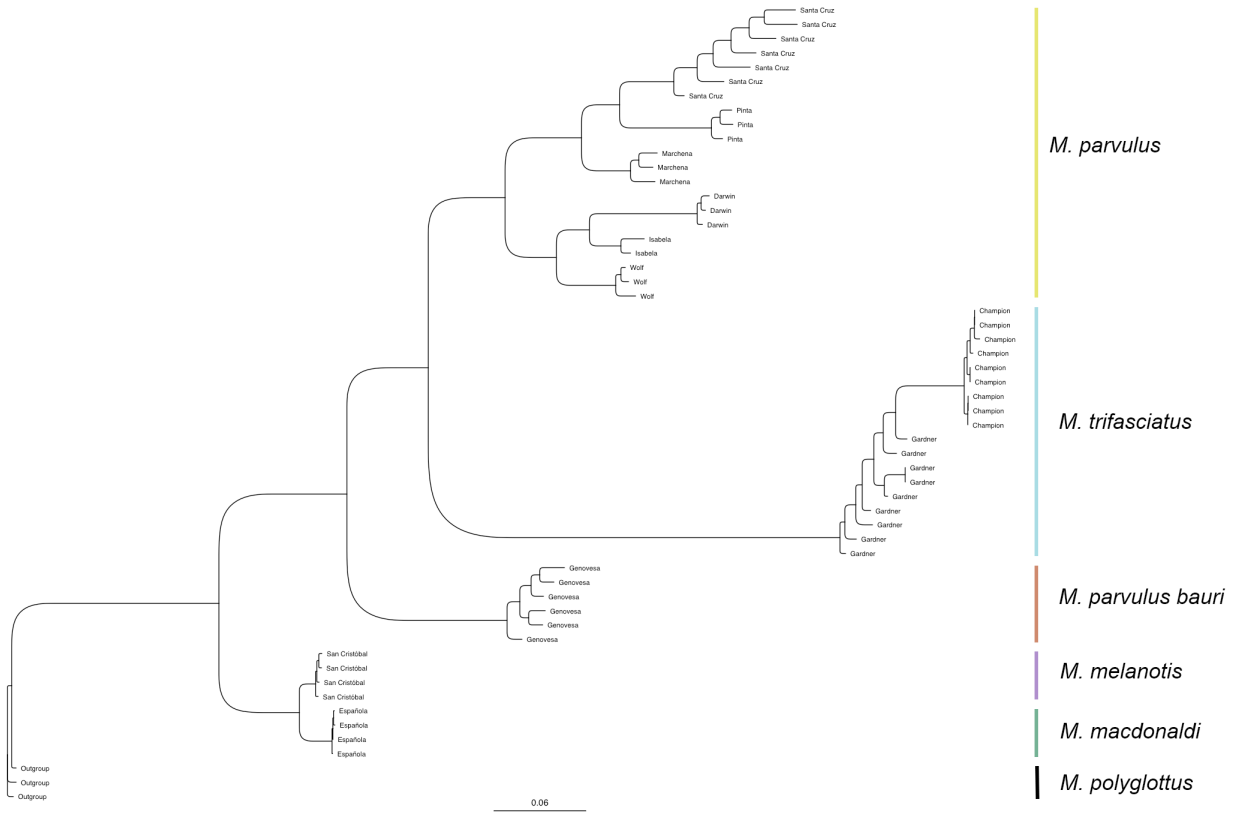


FIGURE 4.2. Maximum-likelihood phylogenetic tree based on whole-genome 10.533 SNPs from 57 *Mimus* individuals. The analysis involves 54 Galapagos mockingbird genomes that include the four species (*M. parvulus*, *M. trifasciatus*, *M. melanotis*, and *M. macdonaldi*) and eleven population/islands (Darwin, Wolf, Isabela, Pinta, Marchena, Santa Cruz, Genovesa, Champion -Floreana-, Gardner -Floreana-, San Cristobal, Española), and three outgroup mitogenomes from M. *polyglottos*.

Establishing a robust framework for determining the dates of divergence will allow me to provide a better explanation for the limited phenotypic differentiation in this group of birds. Although evolution by natural selection is widely regarded as the most important principle of biology, it is still unknown whether phenotypic variations within and between species are mostly

adaptive or neutral. It is possible, that being generalists, mockingbirds have not suffered strong selective pressures to adapt to specific niches. If so, the few relatively small phenotypic variations among individual on different islands and/or different species may have been primarily by genetic drift, and not by selection of functional phenotypic characters driven by the environmental pressures (Hoeck *et al.*, 2010b; Whitlock, 2004). In the future, I plan to use complete dataset of 57 genomes to extensively test this hypothesis.

4.2. HYBRIDIZATION AND HOMOPLOID SPECIATION

Biologists have long recognized that new species may arise because of hybridization between genetically differentiated lineages (Feliner *et al.*, 2017). Hybrid speciation is thought to occur mostly by duplication of a hybrid genome or allopolyploidy. Hybrid sterility and the preservation of co-adapted gene combination as the two main challenges associated with hybrid speciation are ameliorated by the genome duplication. Such duplication doubles a hybrid's chromosomal complement which restores normal pairing and fertility and reduces recombination between homologous chromosomes. However, hybridization can also result in a stable, fertile, and reproductively isolated hybrid lineage where there is no change in ploidy. This process known as, homoploid hybrid speciation (HSS), has been documented in several taxa (e.g., Galapagos finches: *G. fortis* x *G. scandens*, Audubon's Warbler: *Setophaga auduboni*, Italian Sparrow: *Passer italiae*, Golden-crowned Manakin: *Lepidopthrix vilasboasi*) (Lamichhaney *et al.*, 2018; Ottenburghs, 2018).

While homoploid hybridization can result in the formation of a hybrids swarms that are not reproductively isolated from the parental taxa; in the HHS model, reproductive isolation mechanisms develop over time resulting in a continuum from partially reproductively isolated

lineages to independently evolving hybrid species (Ottenburghs, 2018). Although most researchers consider that reproductive isolation to be a direct consequence of past hybridization, some have argued that this criterion is too stringent and that also applies to those cases in which a hybrid population is geographically isolated from its parental species and reproductive isolation subsequently develops as a by-product of random drift or divergent selection (Brelsford, 2011; Hermansen *et al.*, 2011 Feliner *et al.*, 2017).

Nietlisbach *et al*. (2013) considered that the mockingbirds of Genovesa as one of the few cases of incipient homoploid hybrid speciation. According to their study, the Genovesa mockingbird (*M. parvulus bauri*) and the San Cristobal mockingbird (*M. melanotis*) share mitochondrial haplotypes as well as haplotypes at the autosomal locus FIB7. In contrast, at a second autosomal locus (TGF), the mockinbirds on Genovesa belong to a clade with another allopatric (sub)species of *M. parvulus*. However, patterns of genetic variation at only two nuclear loci and ~26 random autosomal microsatellites are insufficient to rule out other hypotheses for the observed genealogical conflict (Ottenburghs *et al.*, 2019).

Whole-genome analyses allow a determination of the species genetic profile and to identify the specific variation that define each species, and with that recognize the introgressive elements in a hybrid population (species) (Elgvin *et al.*, 2017; Nadeau & Kawakami, 2018). The preliminary phylogenomics analyses (Figures 4.1 & 4.2) suggest that this is the case. Beyond just confirming the hybrid origin of the Genovesa mockingbirds, genomic analyses of introgression can reveal variation among loci in the level of incorporation of alleles from one lineage into the other. That information can then use to characterize the extent and timing of gene flow between the two parental species (Nadeau & Kawakami, 2018). Phylogenetic network analyses or modeling

approaches, such as Approximate Bayesian Computation (ABC), can be used to discriminate between incipient hybrid speciation and (recurrent) introgressive hybridization. In addition, studying the patterns of genetic divergence (e.g., *Fst*) between the hybrid and each of the parental lineages will allow us to identify areas of hybrid-specific evolution. Sex chromosomes are known to play a prominent role in the evolution of reproductive isolation and have been suggested to be where genomic incompatibilities, such as hybrid unviability or sterility, first develop (Johnson & Lachance, 2012). Thus, it would be interesting to determine the introgression patterns and evolutionary rates of sex chromosomes in the hybrid population.

4.3. GENETIC LOAD AND CONSERVATION CONSIDERATIONS

Genetic load studies are a valuable tool in species conservation, as they provide important information about the presence and effects of harmful mutations in the genomes of endangered species. Genetic load refers to the deleterious alleles accumulated in a population that decrease individual fitness when expressed. In smaller populations, these variations are more likely to be expressed due to increased inbreeding, which can lead to a higher risk of extinction and a reduced ability to adapt. Therefore, it is crucial to comprehend and manage the effects of genetic load in small populations (Caballero *et al.*, 2017; Hohenlohe *et al.*, 2021; Willi *et al.*, 2022).

Anthropogenic effects, such as habitat disturbance and the introduction of invasive species, have led to the extinction of several endemic species and populations and a decline in the population of native land birds. In the Galapagos archipelago, mockingbird populations have experienced a steep decline due to demographic bottlenecks and reduced genetic diversity caused by inbreeding and genetic drift. Populations on small islands could undergo extirpation after a few generations due to the progressive increase in genetic load, posing a significant risk to small island

populations that were already small and have been further reduced by human activities (Humble *et al.*, 2022; Van Oosterhout, 2020). The relative risks posed by mutation accumulation and demographic stochasticity to a population depend crucially on its size (Chapter 3) (Caballero *et al.*, 2017; Hohenlohe *et al.*, 2021; Supple & Shapiro, 2018; Willi *et al.*, 2022). However, whether purifying natural selection can overcome these negative impacts by "purging" harmful recessive mutations is a topic of active debate with practical consequences reintroduction and rescue conservation efforts. Conservation efforts could be enhanced by maintaining and restoring genetic diversity through translocations and gene flow between populations and monitoring the genetic load and inbreeding to mitigate their nocive effects in the population (Tian *et al.*, 2022; Willi *et al.*, 2022).

The relative risks posed by mutation accumulation and demographic stochasticity to a population depend on its size, and the ability of natural selection to purge harmful recessive mutations is a topic of active debate with practical consequences for conservation efforts. The Floreana Mockingbird (*Mimus trisfasciatus*) became extinct in 1835 on its principal range possibly due to the introduction of black rats (*Rattus rattus*), which acted as nest predators and niche direct competitors (Curry, 1986). However, this species was later rediscovered on two small islets, Champion and Gardner. Despite the small geographic distance separating these two islets (~15 Km) and from the coast of Floreana (~800 m and 8.5 Km, respectively), banding data suggest that these two populations exist in complete isolation (Hoeck *et al*., 2010a; Jiménez-Uzcátegui *et al.*, 2010). While the Gadner-by-Floreana population size is ~137 individuals ($Range_{2003\text{-}2008}$= 65-179 individuals), the population on Champion is comprised of only about 37 individuals ($Range_{2003\text{-}2008}$= 20-52 individuals; Jiménez-Uzcátegui *et al.*, 2010). Any rescue program aiming to restore the original distribution of Floreana mockingbird will benefit from the characterization of the

genetic load at both extant populations and individual levels, by estimating the derived count and fixation rates in coding and non-coding regions of the genome while tracking the frequency of synonymous, missense (tolerated and deleterious), and loss-of-function mutations (Hohenlohe *et al.*, 2021; Humble *et al*., 2022).

Genetic load studies, genetic purging, and genomics are essential tools for understanding and managing the effects of harmful mutations in endangered species. By applying these tools, conservationists can design better strategies for population management, genetic rescue, and captive breeding, and ultimately contribute to the preservation and recovery of endangered species and small populations. Populations of Galapagos mockingbirds on small islands, such as those on Champion, Gardner, Darwin, and even Wolf, show a persistent decline to near disappearance due to inbreeding and genetic drift (Figure 3.13-15 and 3.17). These populations have managed to survive for several generations due to the balance between genetic load purging and maintaining their population size. To mitigate the negative effects of genetic load, conservationists employ various strategies, including maintaining populations in their natural habitats, collecting and storing genetic samples outside of their natural habitats, and using *inter situ* conservation. Another strategy is to let genetic purging take place, which is the reduction of the frequency of a deleterious allele caused by an increased efficiency of natural selection prompted by inbreeding. However, genetic purging also comes with risks such as extinction, fixation of (semi)harmful alleles, and loss of genetic diversity (Caballero *et al.*, 2017; Crnokrak & Barrett, 2002; Hohenlohe *et al.*, 2021; Willi *et al.*, 2022). In addition, to alleviate the effects of human impact, conservation efforts could focus on maintaining and restoring genetic diversity, which could be achieved through translocations and gene flow between populations (Supple & Shapiro, 2018; Willi *et al.*, 2022). The genomics framework presented in this research can be a valuable tool for in-depth analysis of

genetic purging and genetic load in conservation efforts of these species. For example, genomics can reveal the presence of deleterious mutations at the molecular level without potential biases from phenotypic or fitness measures, estimate the mutation load and its effects on population viability and extinction risk, identify genomic regions that are important for maintaining essential functions, and assess how different factors such as selection, drift, or migration (including translocations) influence genetic diversity and adaptation (Caballero *et al.*, 2017; Hohenlohe *et al.*, 2021). Therefore, conservation efforts based on this novel tool could be effectively focus on maintaining and restoring genetic diversity, which could be achieved through translocations and gene flow between populations. Conservationists should also monitor and protect populations that have a high level of inbreeding, which may require immediate intervention to reduce the effects of deleterious alleles on fitness (Crnokrak & Barrett, 2002; Hohenlohe *et al.*, 2021; Tian *et al.*, 2022; Willi *et al.*, 2022). Furthermore, conservation efforts should address the underlying causes of population decline by controlling invasive species and reducing human impact on the archipelago's ecosystems (Allendorf et al., 2010; Hohenlohe *et al.*, 2021).

5. REFERENCES

Abbott, I. & Abbott, LK. (1978). Multivariate study of morphological variation in Galápagos and Ecuadorian mockingbirds. The Condor, 80 (3): 302–308. https://doi.org/10.2307/1368040

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research, 19(9), 1655-1664. http://www.genome.org/cgi/doi/10.1101/gr.094052.109

Ali, J.R. and Aitchison, J.C. (2014). Exploring the combined role of eustasy and oceanic island thermal subsidence in shaping biodiversity on the Galápagos. Journal of Biogeography, 41: 1227-1241. https://doi.org/10.1111/jbi.12313

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. Nature reviews genetics, 11(10), 697-709. https://doi.org/10.1038/nrg2844

Anderson, DL. (1975). Accelerated plate tectonics. Science, 187 (4181): 1077-1079. DOI: 10.1126/science.187.4181.1077.

Andrews, C. B., Mackenzie, S. A., & Gregory, T. R. (2009). Genome size and wing parameters in passerine birds. Proceedings. Biological sciences, 276(1654), 55–61. https://doi.org/10.1098/rspb.2008.1012

Arbogast, B. S., Drovetski, S. V., Curry, R. L., Boag, P. T., Seutin, G., Grant, P. R., Grant, B. R., & Anderson, D. J. (2006). The origin and diversification of Galapagos mockingbirds.

Evolution; international journal of organic evolution, 60(2), 370–382. https://doi.org/10.1554/03-749.1

Arbogast, B. S., Drovetski, S. V., Curry, R. L., Boag, P. T., Seutin, G., Grant, P. R., Grant, B. R., & Anderson, D. J. (2006). The origin and diversification of Galapagos mockingbirds. Evolution; international journal of organic evolution, 60(2), 370–382. https://doi.org/10.1554/03-749.1

Benavides, E., Baum, R., Snell, H.M., Snell, H.L. and Sites, Jr., J.W. (2009). Island biogeography of Galápagos lava lizards (Tropiduridae: Microlophus): species diversity and colonization of the archipelago. Evolution, 63: 1606-1626. https://doi.org/10.1111/j.1558-5646.2009.00617.x

BirdLife International (2017). Spotlight on threatened birds. Presented as part of the BirdLife State of the world's birds website. Available from: http://www.birdlife.org/datazone

Bitter, MC., Kapsenberg, L., Gattuso, JP., & Pfister CA. (2019) Standing genetic variation fuels rapid adaptation to ocean acidification. Nature Communications 10, 5821. https://doi.org/10.1038/s41467-019-13767-1

Black, W. C., 4th, Baer, C. F., Antolin, M. F., & DuTeau, N. M. (2001). Population genomics: genome-wide sampling of insect populations. Annual review of entomology, 46, 441–469. https://doi.org/10.1146/annurev.ento.46.1.441

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

Bollmer, J.L., Nims, B.D. (2018). Genetic diversity in endemic Galápagos birds: Patterns and implications. In: Parker, P. (eds) Disease Ecology. Social and Ecological Interactions in the Galapagos Islands. Springer, Cham. https://doi.org/10.1007/978-3-319-65909-1_4

Botero-Castro, F., Figuet, E., Tilak, M. K., Nabholz, B., & Galtier, N. (2017). Avian genomes revisited: Hidden genes uncovered and the rates versus traits paradox in birds. Molecular biology and evolution, 34(12), 3123–3131. https://doi.org/10.1093/molbev/msx236

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bansal, P., Bridge, A. J., Poux, S., Bougueleret, L., & Xenarios, I. (2016). UniProtKB/Swiss-Prot, the manually annotated section of the UniProt Knowledge Base: How to use the entry view. Methods in molecular biology (Clifton, N.J.), 1374, 23–54. https://doi.org/10.1007/978-1-4939-3167-5_2

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. In: Kraus R. (eds) Avian Genomics in Ecology and Evolution. Springer, Cham. https://doi.org/10.1007/978-3-030-16477-5_6

Brelsford, A. (2011). Hybrid speciation in birds: Allopatry more important than ecology?. Molecular Ecology, 20(18): 3705-3707.

Burns, KJ., Hackett, SJ. and Klein, NJ. (2002). Phylogenetic relationships and morphological diversity in Darwin's finches and their relatives. Evolution 56:1240–1252.

Burt, DW. (2002). Origin and evolution of avian microchromosomes. Cytogenetic and Genome Research, 96: 97–112. DOI: 10.1159/000063018. PMID: 12438785.

Caballero, A., Bravo, I., & Wang, J. (2017). Inbreeding load and purging: implications for the short-term survival and the conservation management of small populations. Heredity, 118(2), 177-185. https://doi.org/10.1038/hdy.2016.80

Cantarel, B. L., Korf, I., Robb, S. M., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., & Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome research, 18(1), 188–196. https://doi.org/10.1101/gr.6743907

Carmi, O., Witt, C., Jaramillo, A., Dumbacher, J. (2016). Phylogeography of the Vermilion Flycatcher species complex: Multiple speciation events, shifts in migratory behavior, and an apparent extinction of a Galápagos-endemic bird species. Molecular Phylogenetics and Evolution, 102, 152-173. https://doi.org/10.1016/j.ympev.2016.05.029

Cibois, A., Thibault, J. C., & Pasquet, E. (2004). Biogeography of eastern Polynesian monarchs (Pomarea): an endemic genus close to extinction. The Condor, 106(4), 837-851. https://doi.org/10.1093/condor/106.4.837

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single

nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2), 80-92. https://doi.org/10.4161/fly.19695

Conrad, C., & Gibbs, J. P. (2021). The era of exploitation: 1535–1959. In Galapagos Giant Tortoises (pp. 63-81). Academic Press. https://doi-org.cyber.usask.ca/10.1016/B978-0-12-817554-5.00002-2

Costanzi, J., & Steifetten, Ø. (2019). Island biogeography theory explains the genetic diversity of a fragmented rock ptarmigan (*Lagopus muta*) population. Ecology and Evolution, 9, 3837 - 3849. https://doi.org/10.1002/ece3.5007

Craig, R. J., Suh, A., Wang, M., & Ellegren, H. (2018). Natural selection beyond genes: identification and analyses of evolutionarily conserved elements in the genome of the collared flycatcher (Ficedula albicollis). Molecular ecology, 27(2), 476-492. https://doi.org/10.1111/mec.14462

Crnokrak, P., & Barrett, S. C. (2002). Perspective: purging the genetic load: a review of the experimental evidence. Evolution, 56(12), 2347-2358. https://doi.org/10.1111/j.0014-3820.2002.tb00160.x

Curry, R. L., and P. R. Grant. 1990. Galapagos mockingbirds: territorial cooperative breeding in a climatically variable environment. Pp. 290–331 in P. B. Stacey and W. D. Koenig, eds. Cooperative breeding in birds: long-term studies of ecology and behavior. Cambridge Univ. Press, Cambridge, U.K.

Damas J., O'Connor R.E., Griffin D.K., Larkin D.M. (2019). Avian chromosomal evolution. In: Kraus R. (eds) Avian genomics in ecology and evolution. Springer, Cham. https://doi.org/10.1007/978-3-030-16477-5_4

Damas, J., Kim, J., Farré, M., Griffin, D. K., & Larkin, D. M. (2018). Reconstruction of avian ancestral karyotypes reveals differences in the evolutionary history of macro- and microchromosomes. Genome biology, 19(1), 155. https://doi.org/10.1186/s13059-018-1544-8

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., & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics (Oxford, England), 27(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics, 43(5), 491–498. https://doi.org/10.1038/ng.806

Dodgson J. B. (2003). Chicken genome sequence: a centennial gift to poultry genetics. Cytogenetic and genome research, 102(1-4), 291–296. https://doi.org/10.1159/000075765

Dvorak, M., Nemeth, E., Wendelin, B., Herrera, P., Mosquera, D., Anchundia, D., Sevilla, C., Tebbich, S. and Fessl, B. (2017). Conservation status of landbirds on Floreana: the smallest inhabited Galapagos island. J. Field Ornithol. 88(2):132–145. DOI: 10.1111/jofo.12197.

Elgvin, TO., Trier, CN., Tørresen, OK., Hagen, IJ., Lien, S., Nederbragt, AJ., Ravinet, M., Jensen, H. and Sætre, GP. (2017). The genomic mosaicism of hybrid speciation. Science Advances, 3(6): e1602996. DOI: 10.1126/sciadv.1602996.

Ellegren H. (2010). Evolutionary stasis: The stable chromosomes of birds. Trends in ecology & evolution, 25(5), 283–291. https://doi.org/10.1016/j.tree.2009.12.004

Ellegren, H., Smeds, L., Burri, R. *et al.* (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature 491, 756–760. https://doi.org/10.1038/nature11584

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. (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature, 491, 756-760. https://doi.org/10.1038/nature11584

Epler, B. (2007). Tourism, the economy, population growth, and conservation in Galapagos. Charles Darwin Foundation. Puerto Ayora, Santa Cruz Island, Galapagos Islands, Ecuador.

Feliner, GN., Álvarez, I., Fuertes-Aguilar, J., Heuertz, M., Marques, I., Moharrek, F., Piñeiro, R., Riina, R., Rosselló, JA., Soltis, PS. and Villa-Machío, I. (2017). Is homoploid hybrid speciation that rare? An empiricist's view. Heredity, 118:513–516.

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A. H., Xie, D., Chen, G., Guo, C., Faircloth, B. C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J. T., Parent, C., Pacheco, G., … 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

Forester, B. R., Beever, E. A., Darst, C., Szymanski, J., & Funk, W. C. (2022). Linking evolutionary potential to extinction risk: applications and future directions. Frontiers in ecology and the environment. https://doi.org/10.1002/fee.2552

Frankham, R. (1996). Relationship of genetic variation to population size in wildlife. Conservation biology, 10(6), 1500–1508. http://www.jstor.org/stable/2387021

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

Freedman, A. H., Gronau, I., Schweizer, R. M., Ortega-Del Vecchyo, D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., Parker, H. G., *et al.*, Wayne, R., & Novembre, J. (2014). Genome sequencing highlights the dynamic early history of dogs. PLoS genetics, 10(1), e1004016. https://doi.org/10.1371/journal.pgen.1004016

Furlan, E., Stoklosa, J., Griffiths, J., Gust, N., Ellis, R., Huggins, R. M., & Weeks, A. R. (2012). Small population size and extremely low levels of genetic diversity in island populations of

the platypus, Ornithorhynchus anatinus. Ecology and evolution, 2(4), 844-857. https://doi.org/10.1002/ece3.195

García-Dorado, A. (2003). Tolerant versus sensitive genomes: The impact of deleterious mutation on fitness and conservation. Conservation Genetics, 4(3), 311-324. https://doi.org/10.1023/A:1024029432658

García-Dorado, A. (2007). Shortcut predictions for fitness properties at the mutation–selection–drift balance and for its buildup after size reduction under different management strategies. Genetics, 176(2), 983-997. https://doi.org/10.1534/genetics.106.065730

Garg, K. M., Chattopadhyay, B., Wilton, P. R., Malia Prawiradilaga, D., & Rheindt, F. E. (2018). Pleistocene land bridges act as semipermeable agents of avian gene flow in Wallacea. Molecular phylogenetics and evolution, 125, 196–203. https://doi.org/10.1016/j.ympev.2018.03.032

Geist, D.J., Snell, H., Snell, H., Goddard, C. and Kurz, M.D. (2014). A paleogeographic model of the galápagos islands and biogeographical and evolutionary implications. In the Galápagos (eds K.S. Harpp, E. Mittelstaedt, N. d'Ozouville and D.W. Graham). https://doi.org/10.1002/9781118852538.ch8

Gignoux-Wolfsohn, S. A., Pinsky, M. L., Kerwin, K., Herzog, C., Hall, M., Bennett, A. B., Fefferman, N. & Maslo, B. (2021). Genomic signatures of selection in bats surviving white-nose syndrome. Molecular ecology, 30(22), 5643-5657. https://doi.org/10.1111/mec.15813

González, J., Montes, C., Rodríguez, J. and Tapia, W. (2008). Rethinking the Galapagos Islands as a complex social-ecological system: implications for conservation and management. Ecology and Society 13(2): 13.

Goodnight, C.J. (2004). Metapopulation quantitative genetics: The quantitative genetics of population differentiation in Hanski, I. & Goggiotti, O.E. (Eds.) Ecology, genetics and evolution of metapopulations, Elsevier Academic Press, Burlington, MA.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature biotechnology, 29(7), 644–652. https://doi.org/10.1038/nbt.1883

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics, 26(9), 1145-5. https://doi.org/10.1093/bioinformatics/btq102

Grant, P. R., Curry, R. L., & Grant, B. R. (2000). A remnant population of the Floreana mockingbird on Champion island, Galápagos. Biological Conservation, 92(3), 285-290. https://doi.org/10.1016/S0006-3207(99)00092-0

Gregory, T. R. (2002). A bird's-eye view of the C-value enigma: Genome size, cell size, and metabolic rate in the class Aves. Evolution, 56(1), 121–130. http://www.jstor.org/stable/3061525

Gregory, T. R. (2005). The evolution of the genome. Elsevier academic, Burlington, MA. https://doi.org/10.1016/B978-0-12-301463-4.X5000-1

Grehan, J. (2001). Biogeography and evolution of the Galapagos: Integration of the biological and geological evidence. Biological Journal of the Linnean Society, 74: 267–287.

Grenier, C. (2000). Conservation contre nature. Les îles Galápagos. Collection Latitude 23, IRD Éditions, Paris, France.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. Nature genetics, 43(10), 1031–1034. https://doi.org/10.1038/ng.937

Guindon, S., Lethiec, F., Duroux, P., & Gascuel, O. (2005). PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic acids research, 33(suppl_2), W557-W559. https://doi.org/10.1093/nar/gki352

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., … Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols, 8(8), 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hastings, A., & Harrison, S. (1994). Metapopulation dynamics and genetics. Annual review of ecology and systematics, 25, 167–188. http://www.jstor.org/stable/2097309

Heads, M., and Grehan, J.R. (2021). The Galápagos islands: Biogeographic patterns and geology. Biol Rev, 96: 1160-1185. https://doi.org/10.1111/brv.12696

Hedrick, P. W., & Garcia-Dorado, A. (2016). Understanding inbreeding depression, purging, and genetic rescue. Trends in ecology & evolution, 31(12), 940–952. https://doi.org/10.1016/j.tree.2016.09.005

Hermansen, JS., Sæther, SA., Elgvin, TO., Borge, T., Hjelle, E. and Saetre, GP. (2011). Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. Molecular ecology, 20, 3812–3822.

Hill, R., Loxterman, J. L., and Aho, K. (2017). Insular biogeography and population genetics of dwarf mistletoe (*Arceuthobium americanum*) in the Central Rocky Mountains. Ecosphere, 8(5):e01810.10.1002/ecs2.1810

Hoeck, P. E., Beaumont, M., James, K., Grant, R. B., Grant, P. R., and Keller L. (2010a). Saving Darwin's muse: evolutionary genetics for the recovery of the Floreana mockingbird. Biol. Lett. 6, 212–215. http://doi.org/10.1098/rsbl.2009.0778

Hoeck, P. E., Bollmer, J. L., Parker, P. G., & Keller, L. F. (2010b). Differentiation with drift: a spatio-temporal genetic analysis of Galapagos mockingbird populations (*Mimus* spp.). Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 365(1543), 1127–1138. https://doi.org/10.1098/rstb.2009.0311

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2021). Population genomics for wildlife conservation and management. *Molecular Ecology*, *30*(1), 62-82. https://doi.org/10.1111/mec.15720

Hohenlohe, P.A., Hand, B.K., Andrews, K.R., Luikart, G. (2018). Population genomics provides key insights in ecology and evolution. In: Rajora, O. (eds) Population genomics. Springer, Cham. https://doi.org/10.1007/13836_2018_20

Hughes, A. L., & Friedman, R. (2008). Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. Molecular biology and evolution, 25(12), 2681–2688. https://doi.org/10.1093/molbev/msn207

Hughes, A. L., & Piontkivska, H. (2005). DNA repeat arrays in chicken and human genomes and the adaptive evolution of avian genome size. BMC evolutionary biology, 5, 12. https://doi.org/10.1186/1471-2148-5-12

Hughes, A., & Hughes, M. (1995). Small genomes for better flyers. Nature 377, 391. https://doi.org/10.1038/377391a0

Humble, E., Stoffel, M. A., Dicks, K., Ball, A. D., Gooley, R. M., Chuven, J., ... & Ogden, R. (2022). Conservation management strategy impacts inbreeding and genetic load in scimitar-horned oryx. bioRxiv, 2022-06. https://doi.org/10.1101/2022.06.19.496717

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution, 23(2), 254-267. https://doi.org/10.1093/molbev/msj030

IUCN. 2021. The IUCN red list of threatened species. Version 2021-3. https://www.iucnredlist.org. Accessed on [08 April 2022].

Jackiw, RN., Mandil, G., and Hager, HA. (2015). A framework to guide the conservation of species hybrids based on ethical and ecological considerations. Conservation biology, 29, 1041-1051.

James, J., 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

Jiménez-Uzcátegui, G., Llerena, W., Milstead, WB., Lomas, E. and Wiedenfeld, D. (2010). Is the population of Floreana Mockingbird *Mimus trifasciatus* declining?. Cotinga 33(1):1-7.

Jiménez-Uzcátegui, G., Llerena, W.F., Milstead, W.B., Lomas, E., & Wiedenfeld, D.A. (2011). Is the population of Floreana Mockingbird *Mimus trifasciatus* declining. Cotinga 33: OL 1–7.

Jiménez-Uzcátegui, G., Wiedenfeld, D., Valle, CA., Vargas, H., Piedrahita, P., Muñoz, LJ. and Alava, JJ. (2019). Threats and vision for the conservation of Galápagos birds. The open ornithology journal, 12: 1-15. DOI: 10.2174/1874453201912010001.

Johnson, NA. and Lachance, J. (2012). The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. Ann N Y Acad Sci., 1256: E1–22. doi:10.1111/j.1749-6632.2012.06748.x.

Keightley, P. D., & Jackson, B. C. (2018). Inferring the probability of the derived vs. The ancestral allelic state at a polymorphic site. Genetics, 209(3), 897–906. https://doi.org/10.1534/genetics.118.301120

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology, 37(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kimura, M. (1983). The neutral theory of molecular evolution. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511623486.

Kislik, E., Mantilla, G., Torres, G. and Borbor-Córdova, M. (2017). Biological hotspots in the Galápagos islands: Exploring seasonal trends of ocean climate drivers to monitor algal blooms. International journal of bioengineering and life sciences, 11(12): 814-824.

Kleindorfer, S., Fessl, B., Peters, K., & Anchundia, D. (2019). Field guide. Resident land birds of Galápagos. Charles Darwin foundation. https://www.darwinfoundation.org/en/publications/identification-guides/field-guide-resident-landbirds-of-galapagos

Kleinman-Ruiz, D., Lucena-Pérez, M., Villanueva, B., Fernández, J., Saveljev, A. P., Ratkiewicz, M., Schmidt, K., Galtier, N., García-Dorado, A., & Godoy, J. A. (2022). Purging of deleterious burden in the endangered Iberian lynx. Proceedings of the national academy of sciences, 119(11), e2110614119. https://doi.org/10.1073/pnas.2110614119

Korf, I. (2004). Gene finding in novel genomes. BMC Bioinformatics 5, 59. https://doi.org/10.1186/1471-2105-5-59

Kyriazis, C. C., Wayne, R. K., & Lohmueller, K. E. (2020). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. Evolution letters, 5(1), 33–47. https://doi.org/10.1002/evl3.209

Lamichhaney, S., Han, F., Webster, MT., Andersson, L., Grant, RB., and Grant, PR. (2018). Rapid hybrid speciation in Darwin's finches. Science, 359, 224–228.

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology, 10(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Latorre, O. (1999). El hombre en las islas encantadas: La historia humana de Galápagos. FUNDACYT, Quito, Ecuador.

Lee, T. H., Guo, H., Wang, X., Kim, C., & Paterson, A. H. (2014). SNPhylo: A pipeline to construct a phylogenetic tree from huge SNP data. BMC genomics, 15, 162. https://doi.org/10.1186/1471-2164-15-162

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

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England), 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics (Oxford, England), 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lovell, P. V., Wirthlin, M., Wilhelm, L., Minx, P., Lazar, N. H., Carbone, L., Warren, W. C., & Mello, C. V. (2014). Conserved syntenic clusters of protein coding genes are missing in birds. Genome biology, 15(12), 565. https://doi.org/10.1186/s13059-014-0565-1

Lovette, I.J., Arbogast, B.S., Curry, R.L., Zink, R.M., Botero, C.A., Sullivan, J.P., Talaba, A.L., Harris, R.B., Rubenstein, D.R., Ricklefs, R.E., & Bermingham, E.P. (2012). Phylogenetic relationships of the mockingbirds and thrashers (Aves: Mimidae). Molecular phylogenetics and evolution, 63 (2), 219-229. https://doi.org/10.1016/j.ympev.2011.07.009

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. Nature reviews. Genetics, 4(12), 981–994. https://doi.org/10.1038/nrg1226

Mapleson, D., Venturini, L., Kaithakottil, G., & Swarbreck, D. (2018). Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. GigaScience, 7(12), giy131. https://doi.org/10.1093/gigascience/giy131

Martin, S. H., & Van Belleghem, S. M. (2017). Exploring evolutionary relationships across the genome using topology weighting. Genetics, 206(1), 429-438. https://doi.org/10.1534/genetics.116.194720

Mathur, S., & DeWoody, J. A. (2021). Genetic load has potential in large populations but is realized in small inbred populations. Evolutionary applications, 14(6), 1540–1557. https://doi.org/10.1111/eva.13216

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 research, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McManus, K. F., Kelley, J. L., Song, S., Veeramah, K. R., Woerner, A. E., Stevison, L. S., Ryder, O. A., Ape Genome Project, G., Kidd, J. M., Wall, J. D., Bustamante, C. D., & Hammer, M. F. (2015). Inference of gorilla demographic and selective history from whole-genome sequence data. Molecular biology and evolution, 32(3), 600–612. https://doi.org/10.1093/molbev/msu394

McNew, S. M., Boquete, M. T., Espinoza-Ulloa, S., Andres, J. A., Wagemaker, N. C., Knutie, S. A., ... & Clayton, D. H. (2021). Epigenetic effects of parasites and pesticides on captive and wild nestling birds. Ecology and Evolution, 11(12), 7713-7729. https://doi.org/10.1002/ece3.7606

Merlen, G. (2014). Plate tectonics, evolution, and the survival of species. In the Galápagos (eds K.S. Harpp, E. Mittelstaedt, N. d'Ozouville and D.W. Graham). https://doi.org/10.1002/9781118852538.ch7

Monceau K, Cézilly F, Moreau J, Motreuil S, Wattier R (2013). Colonisation and diversification of the Zenaida dove (*Zenaida aurita*) in the Antilles: Phylogeography, contemporary gene flow and morphological divergence. PLoS ONE 8(12): e82189. https://doi.org/10.1371/journal.pone.0082189

Morgan, WJ. (1972). Deep mantle convection plumes and plate motions. AAPG Bulletin; 56 (2): 203–213. https://doi.org/10.1306/819A3E50-16C5-11D7-8645000102C1865D

Nadeau, N.J., Kawakami, T. (2018). Population genomics of speciation and admixture. In: Rajora, O. (eds) Population Genomics. Population Genomics. Springer, Cham. https://doi.org/10.1007/13836_2018_24

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics, 32(11), 1749-1751. https://doi.org/10.1093/bioinformatics/btw044

Nielsen, R. and Wakeley, J. (2001). Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158 (2), 885-896.

Nietlisbach, P., Wandeler, P., Parker, P. G., Grant, P. R., Grant, B. R., Keller, L. F., & Hoeck, P. E. (2013). Hybrid ancestry of an island subspecies of Galápagos mockingbird explains

discordant gene trees. Molecular phylogenetics and evolution, 69(3), 581–592. https://doi.org/10.1016/j.ympev.2013.07.020

Organ, C., Shedlock, A., Meade, A. *et al.* (2007). Origin of avian genome size and structure in non-avian dinosaurs. Nature 446, 180–184. https://doi.org/10.1038/nature05621

Ortiz-Catedral, L., Lichtblau, A., Anderson, M. G., Sevilla, C., & Rueda, D. (2021). First report of co-occurrence of two species of mockingbird in the Galapagos islands: A San Cristóbal mockingbird *Mimus melanotis* in a population of Floreana mockingbird M. trifasciatus. Galapagos Research, (70).

Oswald, J. A., Overcast, I., Mauck, W. M., 3rd, Andersen, M. J., & Smith, B. T. (2017). Isolation with asymmetric gene flow during the nonsynchronous divergence of dry forest birds. Molecular ecology, 26(5), 1386–1400. https://doi.org/10.1111/mec.14013

Ottenburghs, J. (2018). Exploring the hybrid speciation continuum in birds. Ecology and Evolution, 8(24):13027-13034.

Ottenburghs, J., Lavretsky, P., Peters, J.L., Kawakami, T., Kraus, R.H.S. (2019). Population genomics and phylogeography. In: Kraus, R. (eds) Avian genomics in ecology and evolution. Springer, Cham. https://doi.org/10.1007/978-3-030-16477-5_8

Padilla, D. P., Spurgin, L. G., Fairfield, E. A., Illera, J. C., & Richardson, D. S. (2015). Population history, gene flow, and bottlenecks in island populations of a secondary seed disperser, the southern grey shrike (*Lanius meridionalis koenigi*). Ecology and evolution, 5(1), 36–45. https://doi.org/10.1002/ece3.1334

Parent, C. E., & Crespi, B. J. (2009). Ecological opportunity in adaptive radiation of Galapagos endemic land snails. The American Naturalist, 174(6), 898-905. https://doi.org/10.1086/646604

Parent, C. E., Caccone, A., & Petren, K. (2008). Colonization and diversification of Galápagos terrestrial fauna: A phylogenetic and biogeographical synthesis. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 363(1508), 3347–3361. https://doi.org/10.1098/rstb.2008.0118

Peischl, S. & Excoffier, L. (2015). Expansion load: recessive mutations and the role of standing genetic variation. Molecular Ecology, 24(9):2084–94. https://doi.org/10.1111/mec.13154

Pembleton, L. W., Cogan, N. O., & Forster, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular ecology resources, 13(5), 946–952. https://doi.org/10.1111/1755-0998.12129

Percy, M.S., Schmitt, S.R., Riveros-Iregui, D.A. and Mirus, B.B. (2016). The Galápagos archipelago: A natural laboratory to examine sharp hydroclimatic, geologic and anthropogenic gradients. WIREs Water, 3: 587-600. https://doi.org/10.1002/wat2.1145

Pérez-Pereira, N., Caballero, A., & García-Dorado, A. (2021). Reviewing the consequences of genetic purging on the success of rescue programs. Conservation genetics, 1-17. https://doi.org/10.1007/s10592-021-01405-7

Pérez-Wohlfeil, E., Diaz-del-Pino, S. & Trelles, O. (2019). Ultra-fast genome comparison for large-scale genomic experiments. Sci Rep 9, 10274. https://doi.org/10.1038/s41598-019-46773-w

Peterson, R. O., Vucetich, J. A., Bump, J. M., & Smith, D. W. (2014). Trophic cascades in a multicausal world: Isle Royale and Yellowstone. Annual review of ecology, evolution, and systematics, 45, 325-345.

Phillips, J. G., Linscott, T. M., Rankin, A. M., Kraemer, A. C., Shoobs, N. F., & Parent, C. E. (2020). Archipelago-Wide Patterns of Colonization and Speciation Among an Endemic Radiation of Galápagos Land Snails. The Journal of heredity, 111(1), 92–102. https://doi.org/10.1093/jhered/esz068

Pindell, J.L. and Kennan, L. (2009). Tectonic Evolution of the Gulf of Mexico, Caribbean and Northern South America in the Mantle Reference Frame: An Update. In: James, K., Antonieta-Lorente, M. and Pindell, J., Eds., The Geology and Evolution of the Region between North and South America, Geological Society of London, London, 227-242.

Prost, S., Armstrong, E. E., Nylander, J., Thomas, G., Suh, A., Petersen, B., Dalen, L., Benz, B. W., Blom, M., Palkopoulou, E., Ericson, P., & Irestedt, M. (2019). Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. GigaScience, 8(5), giz003. https://doi.org/10.1093/gigascience/giz003

Reynolds, R. G. (2011). Islands, metapopulations, and archipelagos: genetic equilibrium and non-equilibrium dynamics of structured populations in the context of conservation. Ph.D. dissertation, University of Tennessee. https://trace.tennessee.edu/utk_graddiss/1016

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. Current biology: CB, 28(21), 3487–3494.e4. https://doi.org/10.1016/j.cub.2018.08.066

Robinson, J. A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B. Y., vonHoldt, B. M., Marsden, C. D., Lohmueller, K. E., & Wayne, R. K. (2016). Genomic flatlining in the endangered island fox. Current biology: CB, 26(9), 1183–1189. https://doi.org/10.1016/j.cub.2016.02.062

Robinson, J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., & Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. Science advances, 5(5), eaau0757. https://doi.org/10.1126/sciadv.aau0757

Rogers, R. L., & Slatkin, M. (2017). Excess of genomic defects in a woolly mammoth on Wrangel island. PLoS genetics, 13(3), e1006601. https://doi.org/10.1371/journal.pgen.1006601

Ronce, O., and Olivieri, I. (2004). Life history evolution in metapopulations in Hanski, I. & Goggiotti, O.E. (Eds.) Ecology, genetics and evolution of metapopulations, Elsevier Academic Press, Burlington, MA.

Santiago, E. and Caballero, A. (2016). Joint prediction of the effective population size and the rate of fixation of deleterious mutations. GENETICS, 204(3): 1267-1279. https://doi.org/10.1534/genetics.116.188250

Sari, E.H.R., Bollmer, J.L. (2018). Colonization of Galápagos birds: Identifying the closest relative and estimating colonization. In: Parker, P. (eds) Disease ecology. social and ecological interactions in the Galapagos islands. Springer, Cham. https://doi.org/10.1007/978-3-319-65909-1_2

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., Tse, T., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. Nucleic acids research, 50(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nature genetics, 46(8), 919-925. https://doi.org/10.1038/ng.3015

Sean, GT., Zimova, M., Drake, DL., and Urban, MC. (2021). Balancing selection and drift in a polymorphic salamander metapopulation. Biology Letters 17: 20200901. http://doi.org/10.1098/rsbl.2020.0901

Sequeira, A.S., Lanteri, A.A., Albelo, L.R., Bhattacharya, S. and Sijapati, M. (2008). Colonization history, ecological shifts and diversification in the evolution of endemic Galápagos weevils. Molecular Ecology, 17: 1089-1107. https://doi.org/10.1111/j.1365-294X.2007.03642.x

Shaw, K.L., & Gillespie, R.G. (2016). Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. Proceedings of the national academy of sciences, 113, 7986 - 7993. https://doi.org/10.1073/pnas.1601078113

Silva, P., Galaverni, M., Ortega-Del Vecchyo, D., Fan, Z., Caniglia, R., Fabbri, E., Randi, E., Wayne, R., & Godinho, R. (2020). Genomic evidence for the old divergence of Southern European wolf populations. Proceedings. Biological sciences, 287(1931), 20201206. https://doi.org/10.1098/rspb.2020.1206

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics (Oxford, England), 31(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

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, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015 <http://www.repeatmasker.org>.

Smit, AFA, Hubley, R. RepeatModeler Open-1.0. 2008-2015 <http://www.repeatmasker.org>.

Smith, E., Shi, L., Drummond, P., Rodriguez, L., Hamilton, R., Powell, E., Nahashon, S., Ramlal, S., Smith, G., & Foster, J. (2000). Development and characterization of expressed sequence tags for the turkey (*Meleagris gallopavo*) genome and comparative sequence analysis with other birds. Animal genetics, 31(1), 62–67. https://doi.org/10.1046/j.1365-2052.2000.00590.x

Snell, H., Tye, A., Causton, C. and Bensted-Smith R. (2002). Current status of and threats to the terrestrial biodiversity of Galapagos. Pages 30-47 in R. Bensted-Smith, editor. (2002). A biodiversity vision for the Galapagos islands. Charles Darwin Foundation and World Wildlife Fund, Puerto Ayora, Ecuador.

Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. Nature genetics, 51(9), 1321-1329. https://doi.org/10.1038/s41588-019-0484-x

Spurgin, L.G., Wright, D.J., van der Velde, M., Collar, N.J., Komdeur, J., Burke, T. and Richardson, D.S. (2014). Museum DNA reveals the demographic history of the endangered Seychelles warbler. Evol Appl, 7: 1134-1143. https://doi.org/10.1111/eva.12191

Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel, Bioinformatics, 19(suppl_2), 215–225. https://doi.org/10.1093/bioinformatics/btg1080

Steadman, D., Stafford, T., Donahue, D., & Jull, A. (1991). Chronology of holocene vertebrate extinction in the Galápagos Islands. Quaternary Research, 36(1), 126-133. https://doi.org/10.1016/0033-5894(91)90021-V

Štefka, J., Hoeck, P. E., Keller, L. F., & Smith, V. S. (2011). A hitchhikers guide to the Galápagos: co-phylogeography of Galápagos mockingbirds and their parasites. BMC evolutionary biology, 11, 284. https://doi.org/10.1186/1471-2148-11-284

Steinfartz, S., Glaberman, S., Lanterbecq, D., Russello, M. A., Rosa, S., Hanley, T. C., ... & Caccone, A. (2009). Progressive colonization and restricted gene flow shape island-dependent population structure in Galápagos marine iguanas (*Amblyrhynchus cristatus*). BMC evolutionary biology, 9(1), 1-18. https://doi.org/10.1186/1471-2148-9-297

Stoletzki, N. and Eyre-Walker, A. (2011). Estimation of the Neutrality Index. Mol. Biol. Evol. 28(1):63–70. doi:10.1093/molbev/msq249.

Supple, M.A., Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biol* **19**, 131. https://doi.org/10.1186/s13059-018-1520-3

Szűcs, M., Melbourne, B.A., Tuff, T., & Hufbauer, R.A. (2014). The roles of demography and genetics in the early stages of colonization. Proceedings of the Royal Society B: Biological Sciences, 281. https://doi.org/10.1098/rspb.2014.1073

Taylor, E., Hardner, J. and Stewart, M. (2006). Ecotourism and economic growth in the Galapagos: An island economy-wide analysis. Agriculture and resource economics working papers, department of agricultural and resource economics, UCD, UC Davis.

Taylor, T. L., Volkening, J. D., DeJesus, E., Simmons, M., Dimitrov, K. M., Tillman, G. E., Suarez, D. L., & Afonso, C. L. (2019). Rapid, multiplexed, whole genome and plasmid sequencing of foodborne pathogens using long-read nanopore technology. Scientific reports, 9(1), 16350. https://doi.org/10.1038/s41598-019-52424-x

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nature genetics, 49(2), 303–309. https://doi.org/10.1038/ng.3748

The UniProt Consortium, (2017). UniProt: the universal protein knowledgebase. Nucleic acids research, 45(D1), D158–D169. https://doi.org/10.1093/nar/gkw1099

The UniProt Consortium, (2018). UniProt: The universal protein knowledgebase. Nucleic acids research, 46(5), 2699. https://doi.org/10.1093/nar/gky092

Tian, D., Patton, A. H., Turner, B. J., & Martin, C. H. (2022). Severe inbreeding, increased mutation load and gene loss-of-function in the critically endangered Devils Hole pupfish. *Proceedings of the Royal Society B*, *289*(1986), 20221561. https://doi.org/10.1098/rspb.2022.1561

Tiersch, T. R., Wachtel, S. S. (1991). On the evolution of genome size of birds. Journal of heredity, 82(5), 363–368. https://doi.org/10.1093/oxfordjournals.jhered.a111105

Toral-Granda, M., Causton, C., Jager, H., Trueman, M., Izurieta, J., Araujo, E., Cruz, M., Zander, K., Izurieta, A. and Garnett, S. (2017). Alien species pathways to the Galapagos islands, Ecuador. PLoS ONE 12(9): e0184379.

Torres-Carvajal, O., Barnes, C.W., Pozo-Andrade, M.J., Tapia, W. and Nicholls, G. (2014), Older than the islands: Origin and diversification of Galápagos leaf-toed geckos (Phyllodactylidae: Phyllodactylus) by multiple colonizations. Journal biogeography., 41: 1883-1894. https://doi.org/10.1111/jbi.12375

van der Valk, T., de Manuel, M, Marques-Bonet T, Guschanski K. (2019). Estimates of genetic load suggest frequent purging of deleterious alleles in small populations. bioRxiv; 2019. DOI: 10.1101/696831

Van Oosterhout, C. (2020). Mutation load is the spectre of species conservation. *Nature Ecology & Evolution*, *4*(8), 1004-1006. https://doi.org/10.1038/s41559-020-1204-8

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

Veeckman, E., Ruttink, T., & Vandepoele, K. (2016). Are we there yet?: Reliably estimating the completeness of plant genome sequences. The plant cell, 28(8), 1759–1768. https://www.jstor.org/stable/plantcell.28.8.1759

Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L., & Swarbreck, D. (2018). Leveraging multiple transcriptome assembly methods for improved gene structure annotation. GigaScience, 7(8), giy093. https://doi.org/10.1093/gigascience/giy093

Vignal A., Eory L. (2019). Avian genomics in animal breeding and the end of the model organism. In: Kraus R. (eds) Avian genomics in ecology and evolution. Springer, Cham. https://doi.org/10.1007/978-3-030-16477-5_3

Wagner, W.L., & Funk V. A. (1995). Hawaiian biogeography: Evolution on a hot spot archipelago. Smithsonian Institution Press, Washington, D.C., U.S.A.

Wallis, J. W., Aerts, J., Groenen, M. A., Crooijmans, R. P., Layman, D., Graves, T. A., Scheer, D. E., Kremitzki, C., Fedele, M. J., Mudd, N. K., Cardenas, M., Higginbotham, J., Carter, J., McGrane, R., Gaige, T., Mead, K., Walker, J., Albracht, D., Davito, J., Yang, S. P., … Warren, W. C. (2004). A physical map of the chicken genome. Nature, 432(7018), 761–764. https://doi.org/10.1038/nature03030

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, B.H., Simberloff, D., Ricklefs, R.E., Aguilée, R., Condamine, F.L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernández-Palacios, J.M., Hengl, T., Norder, S.J., Rijsdijk, K.F., Sanmartín, I., Strasberg, D., Triantis, K.A., Valente, L., Whittaker, R.J., Gillespie, R.G., Emerson, B.C., & Thébaud, C. (2015). Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. Ecology letters, 18 2, 200-17. https://doi.org/10.1111/ele.12398

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., Heger, A., Kong, L., Ponting, C. P., Jarvis, E. D., Mello, C. V., Minx, P., Lovell, P., Velho, T. A., Ferris, M., Balakrishnan, C. N., … Wilson, R. K. (2010). The genome of a songbird. Nature, 464(7289), 757–762. https://doi.org/10.1038/nature08819

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular biology and evolution, 35(3), 543–548. https://doi.org/10.1093/molbev/msx319

Watkins, G. and Cruz, F. (2007). Galapagos at risk: a socioeconomic analysis of the situation in the archipelago. Charles Darwin Foundation, Puerto Ayora, Ecuador.

Watterson, G.A. (1975). On the number of segregating sites in genetical models without recombination. Theoretical Population Biology, 7 (2): 256–276, doi:10.1016/0040-5809(75)90020-9.

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. Genome research, 27(5), 757–767. https://doi.org/10.1101/gr.214874.116

Weissensteiner M.H., Suh A. (2019). Repetitive DNA: The dark matter of avian genomics. In: Kraus R. (eds) Avian genomics in ecology and evolution. Springer, Cham. https://doi.org/10.1007/978-3-030-16477-5_5

Weissensteiner, M. H., Pang, A., Bunikis, I., Höijer, I., Vinnere-Petterson, O., Suh, A., & Wolf, J. (2017). Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. Genome research, 27(5), 697–708. https://doi.org/10.1101/gr.215095.116

Welles, S.R., Dlugosch, K.M. (2018). Population genomics of colonization and invasion. In: Rajora, O. (eds) Population genomics. Springer, Cham. https://doi.org/10.1007/13836_2018_22

Whitlock, MC. (2004). Selection and drift in metapopulations in Hanski, I. & Goggiotti, O.E. (Eds.) Ecology, Genetics and Evolution of Metapopulations, Elsevier Academic Press, Burlington, MA.

Whitlock, MC. (2004). Selection and drift in metapopulations in Hanski, I. & Goggiotti, O.E. (Eds.) Ecology, genetics and evolution of metapopulations, Elsevier Academic Press, Burlington, MA.

Wiedenfeld, DA. and Jiménez-Uzcátegui, GA. (2008). Critical problems for bird conservation in the Galápagos islands. Cotinga 29 (2008): 22–27.

Wikelski, M., Foufopoulos, J., Vargas, H., & Snell, H. (2004). Galápagos birds and diseases: invasive pathogens as threats for island species. Ecology and society, 9(1). http://www.jstor.org/stable/26267654

Willi, Y., Kristensen, T. N., Sgrò, C. M., Weeks, A. R., Ørsted, M., & Hoffmann, A. A. (2022). Conservation genetics as a management tool: The five best-supported paradigms to assist the

management of threatened species. *Proceedings of the National Academy of Sciences*, *119*(1), e2105076119. https://doi.org/10.1073/pnas.2105076119

Wilson, E. O., & MacArthur, R. H. (1967). The theory of island biogeography (Vol. 1). Princeton, NJ: Princeton University Press.

Wilson, JT. (1963). A possible origin of the Hawaiian islands. Canadian journal of physics, 1963, 41(6): 863-870, https://doi.org/10.1139/p63-094

Wink M. (2019). A historical perspective of avian genomics. In: Kraus R. (eds) Avian genomics in ecology and evolution. Springer, Cham. https://doi.org/10.1007/978-3-030-16477-5_2

Wolf, JB. and Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics, 18, 87–100.

Woolfit, M., and Bromham, L. (2005). Population size and molecular evolution on islands. Proc. R. Soc. B. 272, 2277–2282. http://doi.org/10.1098/rspb.2005.3217

Wright, S. (1931). Evolution in mendelian populations. Genetics, 16(2): 97–159.

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., & Yang, T. L. (2019). PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics, 35(10), 1786-1788. https://doi.org/10.1093/bioinformatics/bty875

Zhang, G. (2015). Bird sequencing project takes off. Nature 522, 34. https://doi.org/10.1038/522034d

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

APPENDIX

SUPPLEMENTARY TABLE 2.1. Specimens banding and blood sample collection by venipuncture carried out under permission MAE-DBN-CM-2016-0041 from the Ministry of Environment of Ecuador between 2017 and 2018.

Location	Species	Field Code	Band Code	Collection Date	Alcohol 70%	RNALater
San Cristobal	*Mimus melanotis*	SCbal01		2017-05-25	1	
San Cristobal	*Mimus melanotis*	SCbal02		2017-05-25	1	
San Cristobal	*Mimus melanotis*	SCbal03		2017-05-25	1	
San Cristobal	*Mimus melanotis*	SCbal04		2017-05-25	1	
San Cristobal	*Mimus melanotis*	SCbal05	JP5873	2017-05-26	1	
San Cristobal	*Mimus melanotis*	SCbal06	JP5874	2017-05-26	1	
San Cristobal	*Mimus melanotis*	SCbal07		2017-05-26	1	
San Cristobal	*Mimus melanotis*	SCbal08		2017-05-27	1	
Santa Cruz	*Mimus parvulus*	SCz001		2017-05-30	1	
Santa Cruz	*Mimus parvulus*	SCz002		2017-05-30	1	
Santa Cruz	*Mimus parvulus*	SCz003	97-101	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz004	JH-009	2017-05-31		
Santa Cruz	*Mimus parvulus*	SCz005	97-102	2017-05-31		1
Santa Cruz	*Mimus parvulus*	SCz006		2017-05-31	1	1
Santa Cruz	*Mimus parvulus*	SCz007		2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz008	97-103	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz009	97-104	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz010	97-105	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz011	97-106	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz012	97-107	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz013	97-108	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz014	97-109	2017-05-31	1	
Santa Cruz	*Mimus parvulus*	SCz015		2017-05-31		
Santa Cruz	*Mimus parvulus*	SCz016	97-110	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz017	97-200	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz018	97-199	2017-06-01	1	1
Santa Cruz	*Mimus parvulus*	SCz019	97-198	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz020	97-197	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz021	97-196	2017-06-01	1	

Santa Cruz	*Mimus parvulus*	SCz022	97-195	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz023	97-194	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz024	97-193	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz025	97-192	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz026	97-191	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz027	97-190	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz028	97-189	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz029	97-188	2017-06-01	1	
Santa Cruz	*Mimus parvulus*	SCz030	97-187	2017-06-01	1	
San Cristobal	*Mimus melanotis*	SCbal09	SK1	2017-08-01	1	
San Cristobal	*Mimus melanotis*	SCbal10	SK2	2017-08-02	1	
San Cristobal	*Mimus melanotis*	SCbal11	SK3	2017-08-03	1	
San Cristobal	*Mimus malenotis*	SCbal12	SK4	2017-08-03	1	
San Cristobal	*Mimus melanotis*	SCbal13	SK5	2017-08-03	1	1
Santa Cruz	*Mimus parvulus*	SCz31	SK6	2018-07-09		1
Champion	*Mimus trifasciatus*	FCh01	C97059	2018-01-22	1	1
Champion	*Mimus trifasciatus*	FCh02	S143777	2018-01-22	1	1
Champion	*Mimus trifasciatus*	FCh03	C97035	2018-01-22	1	1
Champion	*Mimus trifasciatus*	FCh04	S143623	2018-01-22	1	1
Champion	*Mimus trifasciatus*	FCh05	S143773	2018-01-22	1	
Champion	*Mimus trifasciatus*	FCh06	C97020	2018-01-22	1	
Champion	*Mimus trifasciatus*	FCh07	S143410	2018-01-22	1	1
Champion	*Mimus trifasciatus*	FCh08	C97025	2018-01-22	1	
Champion	*Mimus trifasciatus*	FCh09	C97026	2018-01-22	1	1
Champion	*Mimus trifasciatus*	FCh10	S143783	2018-01-22	1	
Champion	*Mimus trifasciatus*	FCh11	C97029	2018-01-22	1	
Champion	*Mimus trifasciatus*	FCh12	S143830	2018-01-22	1	
Champion	*Mimus trifasciatus*	FCh13	S143810	2018-01-22	1	1
Champion	*Mimus trifasciatus*	FCh14	C97030	2018-01-22	1	
Gardner	*Mimus trifasciatus*	FG01	C97033	2018-01-25	1	1
Gardner	*Mimus trifasciatus*	FG02	C97034	2018-01-25	1	1
Gardner	*Mimus trifasciatus*	FG03	C97035	2018-01-25	1	1
Gardner	*Mimus trifasciatus*	FG04	C97036	2018-01-25	1	1
Gardner	*Mimus trifasciatus*	FG05	S143597	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG06	C97038	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG07	C97039	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG08	C97040	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG09	C97041	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG10	C97042	2018-01-25	1	

Gardner	*Mimus trifasciatus*	FG11	C97043	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG12	C97044	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG13	C97045	2018-01-25	1	
Gardner	*Mimus trifasciatus*	FG14	C97046	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG15	S143793	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG16	C97047	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG17	S143835	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG18	S143795	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG19	C97048	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG20	C97049	2018-01-26	1	1
Gardner	*Mimus trifasciatus*	FG21	C97068	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG22	C97069	2018-01-26	1	
Gardner	*Mimus trifasciatus*	FG23	C97005	2018-01-26	1	
Genovesa	*Mimus parvulus*	G01	SK7	2018-07-12	1	
Genovesa	*Mimus parvulus*	G02	SK8	2018-07-12	1	1
Genovesa	*Mimus parvulus*	G03	SK9	2018-07-12	1	1
Genovesa	*Mimus parvulus*	G04	SK10	2018-07-12	1	1
Genovesa	*Mimus parvulus*	G05	SK11	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G06	SK101	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G07	SK13	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G08	SK15	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G09	SK16	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G10	SK17	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G11	SK18	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G12	SK19	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G13	SK20	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G14	SK21	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G15	SK22	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G16	SK23	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G17	SK24	2018-07-13	1	1
Genovesa	*Mimus parvulus*	G18	SK25	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G19	SK26	2018-07-14		
Genovesa	*Mimus parvulus*	G20	SK27	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G21	SK28	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G22	SK29	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G23	SK30	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G24	SK31	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G25	SK32	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G26	SK33	2018-07-14	1	1

Genovesa	*Mimus parvulus*	G27	SK34	2018-07-14	1	1
Genovesa	*Mimus parvulus*	G28	SK35	2018-07-14	1	1
Española	*Mimus macdonaldi*	E01	SK36	2018-07-20	1	1
Española	*Mimus macdonaldi*	E02	SK37	2018-07-20	1	1
Española	*Mimus macdonaldi*	E03	SK38	2018-07-20	1	1
Española	*Mimus macdonaldi*	E04	SK39	2018-07-20	1	1
Española	*Mimus macdonaldi*	E05	SK40	2018-07-21	1	1
Española	*Mimus macdonaldi*	E06	SK41	2018-07-21	1	1
Española	*Mimus macdonaldi*	E07	SK42	2018-07-21	1	1
Española	*Mimus macdonaldi*	E08	SK43	2018-07-21	1	1
Española	*Mimus macdonaldi*	E09	SK44	2018-07-21	1	1
Española	*Mimus macdonaldi*	E10	SK45	2018-07-21	1	1
Española	*Mimus macdonaldi*	E11	SK46	2018-07-21	1	1
Española	*Mimus macdonaldi*	E12	SK47	2018-07-21	1	1
Española	*Mimus macdonaldi*	E13	SK48	2018-07-21	1	1
Española	*Mimus macdonaldi*	E14	SK49	2018-07-21	1	1
Española	*Mimus macdonaldi*	E15	SK50	2018-07-21	1	1
Española	*Mimus macdonaldi*	E16	S143314	2018-07-21	1	1
Española	*Mimus macdonaldi*	E17	SK51	2018-07-21	1	1
Española	*Mimus macdonaldi*	E18	SK52	2018-07-21	1	1
Española	*Mimus macdonaldi*	E19	SK53	2018-07-21	1	1
Española	*Mimus macdonaldi*	E20	SK54	2018-07-21	1	1
Española	*Mimus macdonaldi*	E21	SK55	2018-07-21	1	1
Española	*Mimus macdonaldi*	E22	SK56	2018-07-21	1	1
Española	*Mimus macdonaldi*	E23	SK57	2018-07-21	1	1
Española	*Mimus macdonaldi*	E24	SK58	2018-07-21	1	1
Española	*Mimus macdonaldi*	E25	SK59	2018-07-22	1	1
Española	*Mimus macdonaldi*	E26	SK60	2018-07-22	1	1
Española	*Mimus macdonaldi*	E27	SK61	2018-07-22	1	1
Española	*Mimus macdonaldi*	E28	SK62	2018-07-22	1	1
Española	*Mimus macdonaldi*	E29	SK63	2018-07-22	1	1
Española	*Mimus macdonaldi*	E30	SK64	2018-07-22	1	1

SUPPLEMENTARY TABLE 3.1. G-PhoCS raw estimates for theta (θ) for each population on each tested model. The estimates are scaled by a multiplier of 10,000 for better readability and handling. Under each estimate are the confidence interval values at 95%. The Overall column shows the average values of all tested models by each population.

	D-W-I	P-M-S	S-D-W-I	M-D-W-I	P-D-W-I	I-P-M-S	W-P-M-S	D-P-M-S	DWI-PMS	No Migration	Overall
Champion	2.47E-03	1.76E-03	2.40E-03	2.79E-03	2.41E-03	1.85E-03	2.65E-03	3.13E-03	2.96E-03	3.58E-03	2.60E-03
	1.16E-3 - 4.35E-3	*9.9E-4 - 3.64E-3*	*1.39E-3 - 3.23E-3*	*1.46E-3 - 3.85E-3*	*9.2E-4 - 3.69E-3*	*1.03E-3 - 2.68E-3*	*1.51E-3 - 4.48E-3*	*1.64E-3 - 5.63E-3*	*1.93E-3 - 4.03E-3*	*1.66E-3 - 5.46E-3*	*0.001369 - 0.004104*
Gardner (F)	0.0108	7.73E-03	0.0105	0.0123	0.0106	8.11E-03	0.0116	0.0137	0.0129	0.0157	0.01139442
	5.06E-3 - 0.0189	*4.32E-3 - 0.016*	*6.1E-3 - 0.0143*	*6.61E-3 - 0.0172*	*4.04E-3 - 0.0162*	*4.42E-3 - 0.0117*	*6.52E-3 - 0.0197*	*7.19E-3 - 0.0248*	*8.5E-3 - 0.0177*	*7.38E-3 - 0.0239*	*0.006014 - 0.01804*
Darwin	0.0535	0.0536	0.0477	0.0981	0.091	0.0444	0.0803	0.0994	0.0523	0.0376	0.06579
	0.0396 - 0.0689	*0.0389 - 0.0664*	*0.032 - 0.064*	*0.0775 - 0.123*	*0.0734 - 0.1079*	*0.0317 - 0.0585*	*0.0595 - 0.1018*	*0.0826 - 0.1176*	*0.0385 - 0.0649*	*0.0242 - 0.0528*	*0.04979 - 0.08258*
Wolf	0.1033	0.1037	0.0927	0.1904	0.1753	0.0857	0.1526	0.1961	0.1007	0.0725	0.1273
	0.0672 - 0.1615	*0.0717 - 0.1559*	*0.0559 - 0.1386*	*0.1258 - 0.2704*	*0.1211 - 0.2354*	*0.0523 - 0.1255*	*0.0745 - 0.2149*	*0.1421 - 0.262*	*0.0655 - 0.1475*	*0.0411 - 0.1176*	*0.08172 - 0.18293*
Pinta	0.504	0.5042	0.4471	0.5554	0.5961	0.615	0.518	0.5488	0.47	0.5074	0.5266
	0.4152 - 0.5797	*0.4139 - 0.5846*	*0.3664 - 0.5242*	*0.4884 - 0.6205*	*0.5287 - 0.6592*	*0.5277 - 0.7025*	*0.4269 - 0.6178*	*0.4647 - 0.6175*	*0.3945 - 0.5412*	*0.4358 - 0.5845*	*0.44622 - 0.60317*
Marchena	1.0316	1.0262	0.9146	1.1037	1.2597	1.2586	1.0606	1.1244	0.9612	1.036	1.07766
	0.85 - 1.1904	*0.8459 - 1.197*	*0.7475 - 1.0746*	*0.9694 - 1.2384*	*1.118 - 1.4043*	*1.0752 - 1.4373*	*0.8713 - 1.2746*	*0.9486 - 1.2665*	*0.8057 - 1.1124*	*0.8807 - 1.1897*	*0.91123 - 1.23852*
Sta. Cruz	9.178	9.0969	9.5227	9.1356	9.2553	9.3387	9.1929	9.1212	9.1861	9.1829	9.22103
	8.6664 - 9.6857	*8.6304 - 9.5888*	*9.0532 - 10.0126*	*8.6375 - 9.6306*	*8.7647 - 9.742*	*8.8543 - 9.8278*	*8.7195 - 9.6803*	*8.6448 - 9.5852*	*8.718 - 9.6646*	*8.6849 - 9.69*	*8.73737 - 9.71076*
Isabela	5.4291	5.4186	4.8002	5.2982	5.2822	4.8666	5.313	5.2977	5.4149	5.4403	5.25608
	5.0175 - 5.8611	*4.9702 - 5.8526*	*4.3782 - 5.2521*	*4.829 - 5.7504*	*4.8734 - 5.6819*	*4.3303 - 5.3902*	*4.8585 - 5.7563*	*4.8652 - 5.7416*	*4.9592 - 5.8712*	*5.023 - 5.8516*	*4.81045 - 5.7009*
DW	32.271	32.1729	30.0826	30.1031	29.9066	31.9081	30.3384	28.4979	32.3327	32.9841	31.05974
	25.8918 - 39.1266	*25.6272 - 39.2068*	*22.8959 - 37.7093*	*23.6104 - 36.8624*	*23.374 - 36.5844*	*25.0742 - 38.8043*	*23.5589 - 37.4063*	*22.566 - 34.9194*	*26.068 - 39.1676*	*26.5984 - 39.6737*	*24.52648 - 37.94608*
DWI	2.5576	2.6164	3.8994	2.7278	2.5898	3.4299	2.8039	2.7635	2.5508	2.5544	2.84935
	1.8362 - 3.2437	*1.8434 - 3.4882*	*3.1104 - 4.6667*	*1.9083 - 3.5733*	*2.0232 - 3.2382*	*2.5237 - 4.5422*	*2.0173 - 3.6675*	*2.0169 - 3.4407*	*1.7723 - 3.4334*	*1.89 - 3.2529*	*2.09417 - 3.65468*
PM	35.9357	35.7907	39.7274	32.3537	32.104	33.4001	34.7738	33.5124	35.9706	35.371	34.89394
	30.5075 - 41.5924	*30.2633 - 41.6221*	*33.9556 - 45.665*	*27.3238 - 37.401*	*27.3336 - 37.1201*	*28.2455 - 38.7104*	*29.5657 - 40.4396*	*28.5872 - 38.5821*	*30.8253 - 41.6427*	*30.1995 - 40.7369*	*29.6807 - 40.35123*
PMS	0.4902	0.423	0.5543	1.2742	0.45	1.0915	0.6056	0.7355	0.4448	0.5214	0.65905
	0.0655 - 1.1471	*0.1004 - 0.844*	*0.2697 - 0.9764*	*0.3162 - 2.2839*	*0.0622 - 0.9173*	*0.4618 - 1.7717*	*0.232 - 1.1414*	*0.2613 - 1.3207*	*0.1423 - 0.7177*	*0.098 - 1.4151*	*0.20094 - 1.25353*
PARV	23.0966	23.0211	22.6842	23.209	23.3897	23.0206	23.1433	23.3248	23.1106	23.0367	23.10366
	22.4116 - 23.7948	*22.3137 - 23.7617*	*22.1121 - 23.2894*	*22.5515 - 23.9101*	*22.716 - 24.0628*	*22.3176 - 23.717*	*22.4451 - 23.8558*	*22.6811 - 23.9988*	*22.4253 - 23.804*	*22.4096 - 23.6613*	*22.43836 - 23.78557*
TRIFAS	2.7745	2.7866	2.7932	2.7809	2.7798	2.7759	2.7812	2.7664	2.778	2.7738	2.77903
	2.6883 - 2.8608	*2.6974 - 2.8765*	*2.7136 - 2.8743*	*2.698 - 2.8626*	*2.6951 - 2.866*	*2.6927 - 2.8607*	*2.6955 - 2.8679*	*2.6833 - 2.8487*	*2.6917 - 2.8587*	*2.6902 - 2.8603*	*2.69458 - 2.86365*
Root	8.9077	8.8786	8.8673	8.8736	8.8536	8.884	8.87	8.9017	8.8839	8.8879	8.88083
	8.6685 - 9.1544	*8.6325 - 9.121*	*8.6394 - 9.0934*	*8.6531 - 9.1051*	*8.6139 - 9.0911*	*8.6543 - 9.1215*	*8.6276 - 9.11*	*8.6821 - 9.1368*	*8.654 - 9.1183*	*8.6442 - 9.1174*	*8.64696 - 9.1169*

SUPPLEMENTARY TABLE 3.2. Effective population size (*Ne*) estimates in number of individuals for each model and population. The calibration of Ne estimates was obtained as $Ne=\theta/(4\cdot\mu)$, where theta (θ) estimations were taken from G-PhoCS results and the mutation rate (μ) is 4.6E-9 (adopted from *Ficedula albicollis* mutation rate, Smeds et al., 2016). Under each estimate are the confidence interval values at 95%. The Overall column shows the average values of all tested models by each population.

	D-W-I	P-M-S	S-D-W-I	M-D-W-I	P-D-W-I	I-P-M-S	W-P-M-S	D-P-M-S	DWI-PMS	No Migration	Overall
Champion	13.43	9.58	13.02	15.18	13.11	10.05	14.43	17.02	16.08	19.45	14.13
	6.30 - 23.64	*5.38 - 19.78*	*7.55 - 17.55*	*7.93 - 20.92*	*5.00 - 20.05*	*5.60 - 14.57*	*8.21 - 24.35*	*8.91 - 30.60*	*10.49 - 21.90*	*9.02 - 29.67*	*7.44 - 22.30*
Gardner (F)	58.70	42.03	57.07	66.85	57.61	44.08	63.04	74.46	70.11	85.33	61.93
	27.50 - 102.72	*23.48 - 86.96*	*33.15 - 77.72*	*35.92 - 93.48*	*21.96 - 88.04*	*24.02 - 63.59*	*35.43 - 107.07*	*39.08 - 134.78*	*46.20 - 96.20*	*40.11 - 129.89*	*32.68 - 98.04*
Darwin	290.76	291.30	259.24	533.15	494.57	241.30	436.41	540.22	284.24	204.35	357.55
	215.22 - 374.46	*211.41 - 360.87*	*173.91 - 347.83*	*421.20 - 668.48*	*398.91 - 586.41*	*172.28 - 317.93*	*323.37 - 553.26*	*448.91 - 639.13*	*209.24 - 352.72*	*131.52 - 286.96*	*270.60 - 448.80*
Wolf	561.41	563.59	503.80	1034.78	952.72	465.76	829.35	1065.76	547.28	394.02	691.85
	365.22 - 877.72	*389.67 - 847.28*	*303.80 - 753.26*	*683.70 - 1469.57*	*658.15 - 1279.35*	*284.24 - 682.07*	*404.89 - 1167.93*	*772.28 - 1423.91*	*355.98 - 801.63*	*223.37 - 639.13*	*444.13 - 994.18*
Pinta	2739.13	2740.22	2429.89	3018.48	3239.67	3342.39	2815.22	2982.61	2554.35	2757.61	2861.96
	2256.52 - 3150.54	*2249.46 - 3177.17*	*1991.30 - 2848.91*	*2654.35 - 3372.28*	*2873.37 - 3582.61*	*2867.93 - 3817.93*	*2320.11 - 3357.61*	*2525.54 - 3355.98*	*2144.02 - 2941.30*	*2368.48 - 3176.63*	*2425.11 - 3278.10*
Marchena	5606.52	5577.17	4970.65	5998.37	6846.20	6840.22	5764.13	6110.87	5223.91	5630.43	5856.85
	4619.57 - 6469.57	*4597.28 - 6505.43*	*4062.50 - 5840.22*	*5268.48 - 6730.43*	*6076.09 - 7632.07*	*5843.48 - 7811.41*	*4735.33 - 6927.17*	*5155.43 - 6883.15*	*4378.80 - 6045.65*	*4786.41 - 6465.76*	*4952.34 - 6731.09*
Sta. Cruz	49880.43	49439.67	51753.80	49650.00	50300.54	50753.80	49961.41	49571.74	49924.46	49907.07	50114.29
	47100.00 - 52639.67	*46904.35 - 52113.04*	*49202.17 - 54416.30*	*46942.93 - 52340.22*	*47634.24 - 52945.65*	*48121.20 - 53411.96*	*47388.59 - 52610.33*	*46982.61 - 52093.48*	*47380.43 - 52525.00*	*47200.54 - 52663.04*	*47485.71 - 52775.87*
Isabela	29505.98	29448.91	26088.04	28794.57	28707.61	26448.91	28875.00	28791.85	29428.80	29566.85	28565.65
	27269.02 - 31853.80	*27011.96 - 31807.61*	*23794.57 - 28544.02*	*26244.57 - 31252.17*	*26485.87 - 30879.89*	*23534.24 - 29294.57*	*26404.89 - 31284.24*	*26441.30 - 31204.35*	*26952.17 - 31908.70*	*27298.91 - 31802.17*	*26143.75 - 30983.15*
DW	175385.87	174852.72	163492.39	163603.80	162535.87	173413.59	164882.61	154879.89	175721.20	179261.41	168802.93
	140716.30 - 212644.57	*139278.26 - 213080.43*	*124434.24 - 204941.85*	*128317.39 - 200339.13*	*127032.61 - 198828.26*	*136272.83 - 210892.93*	*128037.50 - 203295.11*	*122641.30 - 189779.35*	*141673.91 - 212867.39*	*144556.52 - 215617.93*	*133296.09 - 206228.70*
DWI	13900.00	14219.57	21192.39	14825.00	14075.00	18640.76	15238.59	15019.02	13863.04	13882.61	15485.60
	9979.35 - 17628.80	*10018.48 - 18957.61*	*16904.35 - 25362.50*	*10371.20 - 19420.11*	*10995.65 - 17598.91*	*13715.76 - 24685.87*	*10963.59 - 19932.07*	*10961.41 - 18699.46*	*9632.07 - 18659.78*	*10271.74 - 17678.80*	*11381.36 - 19862.39*
PM	195302.72	194514.67	215909.78	175835.33	174478.26	181522.28	188988.04	182132.61	195492.39	192233.70	189640.98
	165801.63 - 226045.65	*164474.46 - 226207.07*	*184541.30 - 248179.35*	*148498.91 - 203266.30*	*148552.17 - 201739.67*	*153508.15 - 210382.61*	*160683.15 - 219780.43*	*155365.22 - 209685.33*	*167528.80 - 226319.02*	*164127.72 - 221396.20*	*161308.15 - 219300.16*
PMS	2664.13	2298.91	3012.50	6925.00	2445.65	5932.07	3291.30	3997.28	2417.39	2833.70	3581.79
	355.98 - 6234.24	*545.65 - 4586.96*	*1465.76 - 5306.52*	*1718.48 - 12412.50*	*338.04 - 4985.33*	*2509.78 - 9628.80*	*1260.87 - 6203.26*	*1420.11 - 7177.72*	*773.37 - 3900.54*	*532.61 - 7690.76*	*1092.07 - 6812.66*
PARV	125525.00	125114.67	123283.70	126135.87	127117.93	125111.96	125778.80	126765.22	125601.09	125199.46	125563.37
	121802.17 - 129319.57	*121270.11 - 129139.67*	*120174.46 - 126572.83*	*122562.50 - 129946.20*	*123456.52 - 130776.09*	*121291.30 - 128896.74*	*121984.24 - 129651.09*	*123266.85 - 130428.26*	*121876.63 - 129369.57*	*121791.30 - 128594.02*	*121947.61 - 129269.40*
TRIFAS	15078.80	15144.57	15180.43	15113.59	15107.61	15086.41	15115.22	15034.78	15097.83	15075.00	15103.42
	14610.33 - 15547.83	*14659.78 - 15633.15*	*14747.83 - 15621.20*	*14663.04 - 15557.61*	*14647.28 - 15576.09*	*14634.24 - 15547.28*	*14649.46 - 15586.41*	*14583.15 - 15482.07*	*14628.80 - 15536.41*	*14620.65 - 15545.11*	*14644.46 - 15563.32*
Root	48411.41	48253.26	48191.85	48226.09	48117.39	48282.61	48206.52	48378.80	48282.07	48303.80	48265.38
	47111.41 - 49752.17	*46915.76 - 49570.65*	*46953.26 - 49420.65*	*47027.72 - 49484.24*	*46814.67 - 49408.15*	*47034.24 - 49573.37*	*46889.13 - 49510.87*	*47185.33 - 49656.52*	*47032.61 - 49555.98*	*46979.35 - 49551.09*	*46994.35 - 49548.37*

SUPPLEMENTARY TABLE 3.3. G-PhoCS raw estimates for tau (τ) for each node on each tested model. The estimates are scaled by a multiplier of 10,000 for better readability and handling. Under each estimate are the confidence interval values at 95%. The Overall column shows the average values of all tested models by each node.

	D-W-I	P-M-S	S-D-W-I	M-D-W-I	P-D-W-I	I-P-M-S	W-P-M-S	D-P-M-S	DWI-PMS	No Migration	Overall
DW	0.0398	0.0397	0.0355	0.0742	0.0684	0.0328	0.0598	0.0762	0.0388	0.0279	0.04931
	0.0293 - 0.0514	*0.0287 - 0.0492*	*0.0238 - 0.0479*	*0.058 - 0.0937*	*0.0553 - 0.0816*	*0.0233 - 0.0433*	*0.0442 - 0.0761*	*0.063 - 0.0909*	*0.0287 - 0.0484*	*0.0177 - 0.0392*	*0.0372 - 0.06217*
DWI	0.4651	0.4641	0.4058	0.4552	0.4525	0.4261	0.4483	0.4449	0.4627	0.4644	0.44891
	0.4275 - 0.5038	*0.4212 - 0.5023*	*0.37 - 0.452*	*0.4095 - 0.4964*	*0.4143 - 0.4845*	*0.3626 - 0.4703*	*0.407 - 0.4894*	*0.4084 - 0.4859*	*0.4202 - 0.5062*	*0.4269 - 0.5021*	*0.40676 - 0.48929*
PM	0.1464	0.1482	0.1294	0.1614	0.1788	0.1813	0.1515	0.1612	0.1365	0.1471	0.15418
	0.1201 - 0.1694	*0.1209 - 0.1734*	*0.1052 - 0.1527*	*0.1419 - 0.1809*	*0.1575 - 0.1982*	*0.1543 - 0.2083*	*0.1236 - 0.1817*	*0.1348 - 0.1814*	*0.1137 - 0.1575*	*0.1254 - 0.1701*	*0.12974 - 0.17736*
PMS	0.6345	0.6404	0.6734	0.621	0.6299	0.6494	0.6351	0.6262	0.6333	0.6333	0.63765
	0.6038 - 0.6659	*0.6114 - 0.6708*	*0.6471 - 0.6995*	*0.5909 - 0.6515*	*0.6024 - 0.6557*	*0.6213 - 0.6763*	*0.6077 - 0.6599*	*0.5996 - 0.6507*	*0.6068 - 0.6605*	*0.6047 - 0.6623*	*0.60957 - 0.66531*
PARV	0.6474	0.6512	0.6889	0.6564	0.6423	0.6806	0.6515	0.6463	0.645	0.6471	0.65567
	0.6212 - 0.6734	*0.6208 - 0.6807*	*0.6611 - 0.7125*	*0.6247 - 0.6853*	*0.6175 - 0.6658*	*0.6511 - 0.7095*	*0.6216 - 0.6802*	*0.6199 - 0.6727*	*0.6184 - 0.6736*	*0.6216 - 0.6731*	*0.62779 - 0.68268*
TRIFAS	2.39E-03	1.71E-03	2.33E-03	2.72E-03	2.35E-03	1.79E-03	2.57E-03	3.03E-03	2.86E-03	3.48E-03	0.00252343
	1.13E-3 - 4.24E-3	*9.6E-4 - 3.56E-3*	*1.34E-3 - 3.14E-3*	*1.42E-3 - 3.74E-3*	*8.9E-4 - 3.59E-3*	*9.9E-4 - 2.58E-3*	*1.45E-3 - 4.28E-3*	*1.63E-3 - 5.46E-3*	*1.87E-3 - 3.91E-3*	*1.61E-3 - 5.33E-3*	*0.001329 - 0.003983*
Root	4.3578	4.3824	4.393	4.3754	4.377	4.3733	4.3793	4.3563	4.3729	4.3704	4.37378
	4.2361 - 4.4662	*4.2653 - 4.4963*	*4.2974 - 4.4869*	*4.2763 - 4.4742*	*4.2691 - 4.4921*	*4.2648 - 4.4789*	*4.2677 - 4.495*	*4.255 - 4.4585*	*4.2627 - 4.4771*	*4.2657 - 4.4875*	*4.26601 - 4.48127*

SUPPLEMENTARY TABLE 3.4. Divergence time (T) estimates calibrated in years for each node under each model. The calibration of T estimations was obtained as $T=\tau \cdot g/\mu$, where tau (τ) estimations were taken from G-PhoCS results, the generation time (g) is 4.5 years (Grant *et al.*, 2000), and the mutation rate (μ) is 4.6E-9 (adopted from *Ficedula albicollis* mutation rate, Smeds et al., 2016). Under each estimate are the confidence interval values at 95%. The Overall column shows the average values of all tested models by each node.

	D-W-I	**P-M-S**	**S-D-W-I**	**M-D-W-I**	**P-D-W-I**	**I-P-M-S**	**W-P-M-S**	**D-P-M-S**	**DWI-PMS**	**No Migration**	**Overall**
DW	3893.48	3883.70	3472.83	7258.70	6691.30	3208.70	5850.00	7454.35	3795.65	2729.35	4823.80
	2866.30 - 5028.26	*2807.61 - 4813.04*	*2328.26 - 4685.87*	*5673.91 - 9166.30*	*5409.78 - 7982.61*	*2279.35 - 4235.87*	*4323.91 - 7444.57*	*6163.04 - 8892.39*	*2807.61 - 4734.78*	*1731.52 - 3834.78*	*3639.13 - 6081.85*
DWI	45498.91	45401.09	39697.83	44530.43	44266.30	41683.70	43855.43	43522.83	45264.13	45430.43	43915.11
	41820.65 - 49284.78	*41204.35 - 49138.04*	*36195.65 - 44217.39*	*40059.78 - 48560.87*	*40529.35 - 47396.74*	*35471.74 - 46007.61*	*39815.22 - 47876.09*	*39952.17 - 47533.70*	*41106.52 - 49519.57*	*41761.96 - 49118.48*	*39791.74 - 47865.33*
PM	14321.74	14497.83	12658.70	15789.13	17491.30	17735.87	14820.65	15769.57	13353.26	14390.22	15082.83
	11748.91 - 16571.74	*11827.17 - 16963.04*	*10291.30 - 14938.04*	*13881.52 - 17696.74*	*15407.61 - 19389.13*	*15094.57 - 20377.17*	*12091.30 - 17775.00*	*13186.96 - 17745.65*	*11122.83 - 15407.61*	*12267.39 - 16640.22*	*12691.96 - 17350.43*
PMS	62070.65	62647.83	65876.09	60750.00	61620.65	63528.26	62129.35	61258.70	61953.26	61953.26	62378.80
	59067.39 - 65142.39	*59810.87 - 65621.74*	*63303.26 - 68429.35*	*57805.43 - 63733.70*	*58930.43 - 64144.57*	*60779.35 - 66159.78*	*59448.91 - 64555.43*	*58656.52 - 63655.43*	*59360.87 - 64614.13*	*59155.43 - 64790.22*	*59631.85 - 65084.67*
PARV	63332.61	63704.35	67392.39	64213.04	62833.70	66580.43	63733.70	63225.00	63097.83	63303.26	64141.63
	60769.57 - 65876.09	*60730.43 - 66590.22*	*64672.83 - 69701.09*	*61111.96 - 67040.22*	*60407.61 - 65132.61*	*63694.57 - 69407.61*	*60808.70 - 66541.30*	*60642.39 - 65807.61*	*60495.65 - 65895.65*	*60808.70 - 65846.74*	*61414.24 - 66783.91*
TRIFAS	233.80	167.28	227.93	266.09	229.89	175.11	251.41	296.41	279.78	340.86	246.86
	110.54 - 414.78	*93.91 - 348.26*	*131.09 - 307.17*	*138.91 - 365.87*	*87.07 - 351.20*	*96.85 - 252.39*	*141.85 - 418.70*	*159.46 - 534.13*	*182.93 - 382.50*	*157.50 - 521.41*	*130.01 - 389.64*
Root	426306.52	428713.04	429750.00	428028.26	428184.78	427822.83	428409.78	426159.78	427783.70	427539.13	427869.78
	414401.09 - 436910.87	*417257.61 - 439855.43*	*420397.83 - 438935.87*	*418333.70 - 437693.48*	*417629.35 - 439444.57*	*417208.70 - 438153.26*	*417492.39 - 439728.26*	*416250.00 - 436157.61*	*417003.26 - 437977.17*	*417296.74 - 438994.57*	*417327.07 - 438385.11*

SUPPLEMENTARY TABLE 3.5. G-PhoCS raw estimates for migration (*m*) for each migration band. The migration bands are arranged in two groups depending on the sense of the band (SE to NW, or NW to SE). Over each estimate is the model where this value come from. Under each estimate are the confidence interval values at 95%.

SE>NW			NW>SE		
W>D	D-W-I		**D>W**	D-W-I	
	212.8396			192.0544	
	200.4162 - 224.8609			*170.9111 - 215.117*	
I>D	D-W-I		**D>I**	D-W-I	
	186.6227			210.8777	
	171.2695 - 200.5557			*196.1779 - 227.3521*	
I>W	D-W-I		**W>I**	D-W-I	
	193.6929			196.0742	
	182.3413 - 204.6566			*183.9161 - 207.5392*	
S>D	D-P-M-S	S-D-W-I	**D>S**	D-P-M-S	S-D-W-I
	263.855	200.9694		255.924	182.5543
	224.167 - 313.141	*194.0697 - 208.9009*		*217.429 - 303.728*	*176.2868 - 189.759*
S>W	W-P-M-S	S-D-W-I	**W>S**	W-P-M-S	S-D-W-I
	194.9831	201.1848		196.6415	184.3041
	179.633 - 211.7854	*194.2776 - 209.1247*		*187.468 - 207.3775*	*177.9765 - 191.5778*
S>I	S-D-W-I	S-D-W-I	**I>S**	I-P-M-S	S-D-W-I
	169.0644	181.3848		194.849	205.2613
	152.3211 - 188.375	*175.1575 - 188.5434*		*175.5522 - 217.1049*	*198.2142 - 213.3621*
M>D	D-P-M-S	M-D-W-I	**D>M**	D-P-M-S	M-D-W-I
	238.275	266.8663		250.468	252.0001
	202.435 - 282.783	*206.4853 - 387.3826*		*212.794 - 297.253*	*194.9828 - 365.803*
M>W	W-P-M-S	M-D-W-I	**W>M**	W-P-M-S	M-D-W-I
	194.6508	259.363		200.4417	239.0584
	171.1612 - 217.8781	*200.6797 - 376.4909*		*188.0028 - 214.5036*	*184.9692 - 347.0167*
M>I	I-P-M-S	M-D-W-I	**I>M**	I-P-M-S	M-D-W-I
	168.3032	244.2858		191.7845	225.1202
	151.6353 - 187.527	*189.0139 - 354.605*		*172.7912 - 213.6903*	*174.1846 - 326.7841*
P>D	D-P-M-S	P-D-W-I	**D>P**	D-P-M-S	P-D-W-I
	249.954	224.5074		246.547	209.9124
	212.3567 - 296.6424	*208.8815 - 240.3537*		*209.462 - 292.599*	*195.3023 - 224.7286*
P>W	W-P-M-S	P-D-W-I	**W>P**	W-P-M-S	P-D-W-I
	189.5218	217.8643		196.7998	228.5434
	173.0697 - 207.3762	*202.7007 - 233.2417*		*171.3013 - 222.1887*	*212.6366 - 244.6746*
P>I	I-P-M-S	P-D-W-I	**I>P**	I-P-M-S	P-D-W-I
	199.7021	194.5061		188.1088	231.3747
	179.9246 - 222.5122	*180.9683 - 208.2348*		*169.4795 - 209.5948*	*215.2708 - 247.7057*
S>M	P-M-S		**M>S**	P-M-S	
	207.8662			215.0371	
	198.0493 - 217.6661			*204.8815 - 225.175*	
S>P	P-M-S		**P>S**	P-M-S	
	221.1418			218.2595	
	210.6979 - 231.5676			*207.9517 - 228.5494*	
M>P	P-M-S		**P>M**	P-M-S	
	184.2672			215.6826	
	175.5648 - 192.9545			*205.4966 - 225.851*	
PMS>DWI	DWI-PMS		**DWI>PMS**	DWI-PMS	
	195.3818			203.4354	
	181.3157 - 206.7434			*188.7895 - 215.2653*	

SUPPLEMENTARY TABLE 3.6. Total migration rate (m^{tot}) calculated as $m_{A>B} \cdot \tau_{AB}$. The calculated estimates are arranged in two groups depending on the sense of the band (SE to NW, or NW to SE).

SE>NW			NW>SE		
W>D	D-W-I		**D>W**	D-W-I	
	0.0008471			0.00076438	
I>D	D-W-I		**D>I**	D-W-I	
	0.00074276			0.00083929	
I>W	D-W-I		**W>I**	D-W-I	
	0.0007709			0.00078038	
S>D	D-P-M-S	S-D-W-I	**D>S**	D-P-M-S	S-D-W-I
	0.00201058	0.00071344		0.00195014	0.00064807
S>W	W-P-M-S	S-D-W-I	**W>S**	W-P-M-S	S-D-W-I
	0.001166	0.00071421		0.00117592	0.00065428
S>I	I-P-M-S	S-D-W-I	**I>S**	I-P-M-S	S-D-W-I
	0.00720383	0.0073606		0.00830252	0.0083295
M>D	D-P-M-S	M-D-W-I	**D>M**	D-P-M-S	M-D-W-I
	0.00181566	0.00198015		0.00190857	0.00186984
M>W	W-P-M-S	M-D-W-I	**W>M**	W-P-M-S	M-D-W-I
	0.00116401	0.00192447		0.00119864	0.00177381
M>I	I-P-M-S	M-D-W-I	**I>M**	I-P-M-S	M-D-W-I
	0.00305134	0.00394277		0.00347705	0.00363344
P>D	D-P-M-S	P-D-W-I	**D>P**	D-P-M-S	P-D-W-I
	0.00190465	0.00153563		0.00187868	0.0014358
P>W	W-P-M-S	P-D-W-I	**W>P**	W-P-M-S	P-D-W-I
	0.00113334	0.00149019		0.00117686	0.00156324
P>I	I-P-M-S	P-D-W-I	**I>P**	I-P-M-S	P-D-W-I
	0.0036206	0.00133042		0.00341041	0.0015826
S>M	P-M-S		**M>S**	P-M-S	
	0.00308058			0.00318685	
S>P	P-M-S		**P>S**	P-M-S	
	0.00327732			0.00323461	
M>P	P-M-S		**P>M**	P-M-S	
	0.00273084			0.00319642	
PMS>DWI	DWI-PMS		**DWI>PMS**	DWI-PMS	
	0.00904032			0.00941296	

SUPPLEMENTARY TABLE 3.7. Instantaneous migration rate (M) calculated as $m_{A>B} \cdot \mu$. The calculated estimates are arranged in two groups depending on the sense of the band (SE to NW, or NW to SE).

SE>NW			NW>SE		
W>D	D-W-I		**D>W**	D-W-I	
	9.79062E-07			8.8345E-07	
I>D	D-W-I		**D>I**	D-W-I	
	8.58464E-07			9.70037E-07	
I>W	D-W-I		**W>I**	D-W-I	
	8.90987E-07			9.01941E-07	
S>D	D-P-M-S	S-D-W-I	**D>S**	D-P-M-S	S-D-W-I
	1.21373E-06	9.24459E-07		1.17725E-06	8.3975E-07
S>W	W-P-M-S	S-D-W-I	**W>S**	W-P-M-S	S-D-W-I
	8.96922E-07	9.2545E-07		9.04551E-07	8.47799E-07
S>I	I-P-M-S	S-D-W-I	**I>S**	I-P-M-S	S-D-W-I
	7.77696E-07	8.3437E-07		8.96305E-07	9.44202E-07
M>D	D-P-M-S	M-D-W-I	**D>M**	D-P-M-S	M-D-W-I
	1.09607E-06	1.22758E-06		1.15215E-06	1.1592E-06
M>W	W-P-M-S	M-D-W-I	**W>M**	W-P-M-S	M-D-W-I
	8.95394E-07	1.19307E-06		9.22032E-07	1.09967E-06
M>I	I-P-M-S	M-D-W-I	**I>M**	I-P-M-S	M-D-W-I
	7.74195E-07	1.12371E-06		8.82209E-07	1.03555E-06
P>D	D-P-M-S	P-D-W-I	**D>P**	D-P-M-S	P-D-W-I
	1.14979E-06	1.03273E-06		1.13411E-06	9.65597E-07
P>W	W-P-M-S	P-D-W-I	**W>P**	W-P-M-S	P-D-W-I
	8.718E-07	1.00218E-06		9.05279E-07	1.0513E-06
P>I	I-P-M-S	P-D-W-I	**I>P**	I-P-M-S	P-D-W-I
	9.1863E-07	8.94728E-07		8.653E-07	1.06432E-06
S>M	P-M-S		**M>S**	P-M-S	
	9.56185E-07			9.89171E-07	
S>P	P-M-S		**P>S**	P-M-S	
	1.01725E-06			1.00399E-06	
M>P	P-M-S		**P>M**	P-M-S	
	8.47629E-07			9.9214E-07	
PMS>DWI	DWI-PMS		**DWI>PMS**	DWI-PMS	
	8.98756E-07			9.35803E-07	

SUPPLEMENTARY TABLE 3.8. Summary of genomic indices by population.

Parameter	Champion	Darwin	Gardner-by-Floreana	Isabela	Marchena	Pinta	Santa Cruz	Wolf
	C	D	F	I	M	P	S	W
Island Size (ha)	9.5	110	81	45,8812	12,996	5,940	98,555	130
Heterozygosity	0.0224243	0.0732099	0.0313929	0.1186875	0.1235383	0.0915943	0.1347328	0.085778
Het/Kb	0.2249561	0.7298943	0.3150451	1.1979957	1.2556768	0.9247685	1.3629785	0.8576315
ROHs Sum (Mb)	867.50714	757.291331	609.832127	95.9943047	278.362258	401.020858	60.1669993	566.559076
ROHs >0.5 Mb Sum	704.662375	392.315508	291.489571	6.67456733	22.934552	22.3974777	8.82081133	211.760838
Inbreeding (FROH)	0.622	0.234	0.223	0.005	0.006	0.009	0.008	0.116
Recessive Load (Count)	5,472	4,656	4,998	3,166	3,198	3,799	2,987	4,202
Additive Load (Count)	11,207	10,359	11,240	11,076	10,539	10,707	10,940	10,636
LD mean (r^2)	0.852	0.7333	0.4875	0.3057	0.3536	0.3715	0.3031	0.5342
10 kb win π	2.55E-04	7.24E-04	3.46E-04	1.22E-03	1.25E-03	9.12E-04	1.38E-03	8.80E-04
10 kb win π 0-fold	6.1361E-05	6.3498E-05	5.9887E-05	6.5014E-05	7.0936E-05	6.6237E-05	6.9791E-05	6.7484E-05
10 kb win π 4-fold	0.002326246	0.00172736	0.001604296	0.001621716	0.001823584	0.001746063	0.001788354	0.001813174
Ratio π(0-fold/4-fold)	0.307051647	0.32934283	0.321784007	0.301351469	0.31462422	0.308209624	0.311639903	0.33552126