Endemic island songbirds as windows into evolution in small effective population sizes

The nearly neutral theory in light of passerine birds' genomes

Thibault Leroy[1,2], Marjolaine Rousselle[1,3], Marie-Ka Tilak[1], Aude Caizergues[1,4], Céline Scornavacca[1], Maria Recuerda Carrasco[5], Jérôme Fuchs[6], Juan Carlos Illera[7], Dawie H. De Swardt[8], Christophe Thébaud[9], Borja Milà[5] & Benoit Nabholz[1]

[1] *ISEM, Univ. Montpellier, CNRS, IRD, EPHE, Montpellier, France*

[2] *Department of Botany & Biodiversity Research, University of Vienna, Vienna, Austria*

[3] *Bioinformatics Research Centre, Aarhus University,C. F.Møllers Alle 8,DK-8000 Aarhus, Denmark*

[4] *CEFE, Univ Montpellier, CNRS, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France*

[5] *National Museum of Natural Sciences (MNCN), Spanish National Research Council (CSIC), Madrid, Spain*

[6] *Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, CP51, 57 rue Cuvier, 75005 Paris, France*

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

[8] *Department of Ornithology, National Museum, Bloemfontein, South Africa*

[9] *Laboratoire Evolution et Diversité Biologique (EDB), UMR 5174 CNRS - Université Paul Sabatier - IRD, Toulouse, France*

Corresponding authors:
Thibault Leroy, thibault.leroy@univie.ac.at
Benoit Nabholz, benoit.nabholz@umontpellier.fr

Abstract

Due to their limited ranges and inherent isolation, island species provide a unique opportunity to study the impact of non-adaptive forces on molecular evolution, especially how effective population size may influence the accumulation of deleterious mutations. By estimating nucleotide diversity at synonymous and non-synonymous sites for a large set of in-house and published whole genome sequences covering distant phylogenetic taxa (14 endemic insular and 11 mainland passerine bird species), we found support for contrasted patterns of molecular variation between endemic island and mainland species. Even after controlling for the phylogenetic inertia, we observed significantly lower nucleotide diversities, higher mutation loads and lower adaptive substitution rates in the endemic insular species. Overall, our results indicate that the smaller area available on islands likely constrains the upper bound of the effective population size, in such a way that demography has affected the ability of natural selection to efficiently remove weakly deleterious mutations. Endemic insular species' genomes therefore represent excellent windows into evolution in small effective population sizes and we discussed the implications for both evolutionary and conservation biology.

Keywords: *insularity, neutral theory, molecular evolution, purifying selection, adaptive substitutions, background selection, census population sizes*

Introduction

Endemic island species have long been of central interest for a diverse range of biologists. The study of island organisms stimulated the early study of speciation (Darwin 1859; Mayr 1942) and the development of the island biogeography theory (MacArthur & Wilson 1963; Warren et al. 2015; Valente et al. 2020). Oceanic islands also provide emblematic examples of adaptive changes in response to climate fluctuations (e.g. beak size evolution in the Darwin's finches, Grant and Grant 2006), and adaptive evolutionary radiations (Schluter 2000; Losos 2009; Moyle

et al. 2009; Grant and Grant 2019). Islands also represent areas of particular interests for conservation biologists, because of their high species richness (15-20% of the global species diversity for some plant and animal taxa; e.g. Johnson & Stattersfield 1990, Kreft et al. 2008) but also the more severe consequences of recent anthropogenic disturbances (80% of the reported extinctions since the year 1500, Ricketts et al. 2005). From a molecular evolution perspective, less clear is the long-term impact of genetic drift – and more broadly, non-adaptive processes – on island species evolution. Two main reasons explain why genetic drift can play so crucial a role in insular species. First, oceanic islands are often small and remote, thus providing opportunities for extreme founder effects in colonist populations and isolation impeding gene flow with the initial gene pool after colonization. Second, once a founding population is established, its growth will be mechanically limited by the island size – the smaller the island, the smaller the census population size.

The rate of evolutionary changes caused by genetic drift - which is captured by the effective population size (*Ne)* - is expected to be central for endemic island species. These species are expected to be more prone to an increased inbreeding and enhanced effects of genetic drift (Frankham, 2002). Indeed, island species represent groups of species with lower census and, supposedly, lower *Ne* than mainland species. If this hypothesis is true and assuming the nearly neutral theory of molecular evolution (Ohta 1973; 1992), such a lower *Ne* is expected to limit the adaptive potential of the endemic island species, because smaller populations produce lower number of mutations per generation at the species level and therefore carry less alleles that may become beneficial after an environmental change (Lanfear et al. 2014, Nam et al 2017, Gossman et al 2012, Rousselle et al. 2020). This is particularly important, because island species are expected to face higher demographic and environmental stochasticity, in such a way that a low proportion of adaptive mutations may limit the capacity to buffer or trade off these variations. In addition, *Ne* is also expected to influence the efficacy of natural selection to purge deleterious alleles, with higher effectiveness in large populations assuming the nearly neutral models of genome evolution (Ohta 1973, Charlesworth 2009; Eyre-Walker et al. 2002a; Lanfear et al. 2014). To sum up, following the nearly neutral theory, slightly deleterious alleles are supposed to be accumulated at a higher rate in endemic island species as compared to their mainland relatives, leading to an increasing deleterious mutational burden over time. Taxon-specific evidences

supporting this important burden in island species start to accumulate (Loire et al. 2013 for Giant Galápagos tortoises; Rogers and Slatkin 2017 for woolly mammoths; Robinson et al. 2016; 2018 for island foxes and Kutschera et al. 2019 for crows), but the strength of such a general relationship between census and effective population sizes remains debated. For example, James et al. (2016) found no effect of island colonization on mitochondrial molecular evolution. By sequencing the transcriptome of a broad taxonomic sample of animals, Romiguier et al. (2014) found no relationship between range size and the level of genetic diversity. More recently, Díez-del-Molino et al. (2018) found a positive, albeit weak, correlation between census size and genetic diversity in birds and mammals. The absence of a general conclusion was one of our main motivations for this study.

Despite its central importance, estimating *Ne* is challenging in practice. *Ne* is indeed well-known to fluctuate rapidly over short periods of time, in such a way that the current size may not reflect past population size that has shaped present-day genetic diversity (e.g. Atkinson et al. 2008; Schiffels & Durbin, 2014 for humans). As a consequence, depending on the evolutionary time scale under consideration, the *Ne* can be very different. For example, the *Ne* that have determined the dynamics of ancestral alleles detected using between-species (divergence) data might be several orders of magnitude different from recent *Ne* estimates using present-day within-species (polymorphism) data (Eyre-Walker 2002b, Rousselle et al. 2018). In animals or plants, life-history traits such as body mass or longevity are known to provide useful proxies for long-term *Ne* (Popadin et al. 2007, Romiguier et al. 2014, Figuet et al. 2016, Chen et al. 2017). In birds, the same life-history traits also represent proxies of *Ne*, but only after taking into account some confounding effects, suggesting that the relationship between *Ne* and life-history is weaker for birds than for some other clades (Weber et al. 2014, Botero et al. 2017, Mugal et al. 2020). However, there is a fundamental limitation to the use of life-history traits as a proxy of *Ne*. The relationship between these two variables is only poorly described and only holds as a result of the correlations between body size and population density (White et al. 2007). Long-term *Ne* can be more directly estimated using population genomic variables. The level of genetic drift is inversely proportional to *Ne* (Kimura & Crow, 1964; Kimura et al. 1983) and consequently, the observed nucleotide diversity levels at synonymous sites (π_s) can be used as a predictor of the long-term *Ne* (e.g. Romiguier et al. 2014).

The impact of *Ne* on the efficacy of positive selection still remains highly debated among evolutionary biologists (e.g. Kern & Hahn, 2018 and Jensen et al., 2019 for recent publications), despite the significant progress in both the methods and the fundamental knowledge gained in this field over the last decade (e.g. Nikolaev et al. 2007; Keightley and Eyre-Walker 2010; Gossmann et al. 2010; 2012; Galtier 2016; Chen et al. 2017; Rousselle et al. 2020). By combining information at both the within-species (polymorphism) and between-species (divergence) levels, the proportion of adaptive substitutions can be estimated based on the comparisons of the ratios of non-synonymous and synonymous mutations in the polymorphism and substitution data (π_N/π_S & d_N/d_S, McDonald Kreitman, 1991). Over the last decade, several methods inspired by the seminal work of McDonald and Kreitman (1991) have been developed to take into account short-term demographic variation and the presence of slightly deleterious mutations. These methods used the Site Frequency Spectra (SFS) at both synonymous and non-synonymous sites to estimate the Distribution of Fitness Effects (DFE) of non-synonymous mutations (Keightley & Eyre-Walker 2007; Eyre-Walker & Keightley 2009; Galtier 2016; Tataru et al. 2017; see also Moutinho et al. 2019 for a review). Using these methods, can we expect different proportions of adaptive mutations between island and mainland species? On the one hand, we can hypothesize that the population with the larger effective population sizes will also exhibit the higher rates of adaptive evolution, because of a greater number of *de novo* mutations produced per generation and because of the greater standing genetic variability available. On the other hand, assuming that the species with small *Ne* exhibits a higher proportion of deleterious mutations, therefore generating proteins that contribute to pulling away from their fitness optima, we can hypothesize the production of more frequent compensatory adaptive mutations in these species.

Passerine birds represent an excellent group of species to investigate the effect of insularity on evolutionary dynamic for a series of reasons. First, a lot of genomic resources are available for passerine birds. Since the first genome in 2010 (Warren et al. 2010), a lot of passerine bird species have recently joined the list (e.g. Ellegren et al. 2012; Zhang et al. 2012; Cornetti et al. 2015; Laine et al. 2016; Lundberg et al. 2017; Leroy et al. 2019). Beyond genome assemblies, several excellent population genomics studies have focused on passerine birds and released the data in the public domain (e.g. Ellegren et al. 2012; Lamichhaney et al. 2015 and Lundberg et al. 2017 among others). Second, the large community of bird-enthusiasts, including ornithologists but also numerous voluntary birdwatchers, provides excellent monitoring of the

147 species presence, distribution and abundance. Third, probably thanks to their good dispersal
148 capability, a quite large diversity of birds is endemic to islands (17% of the world's bird species,
149 Johnson & Stattersfield 1990). It is important to mention that endemic islands birds have
150 however experienced a dramatic species loss over the last five centuries (Ricketts et al. 2005).
151 Islands habitats are indeed more threatened habitats than continents, especially given the
152 deleterious consequences of human activities and disturbances (Steadman 1995). Fourth, both
153 the chromosome architectures and the recombination landscapes are stable in birds (Ellegren
154 2010; Singhal et al. 2015). Unlike mammals, bird genomes lack the *PRDM9* gene (Baker et al.
155 2017). In mammals, this *PRDM9* gene indeed explains the rapid turnover in recombination
156 hotspots (Baudat et al. 2010; Parvanov et al. 2010). As a consequence, the correlation between
157 the recombination rates and the G+C content is stronger than in mammal species for instance.
158 This relationship is due to the GC-biased gene conversion (gBGC), a recombination-associated
159 segregation bias that favors G and C over A and T alleles (for details, see Duret & Galtier, 2009).
160 In birds, the genomic GC content, or more specifically GC-content at third codon position
161 (hereafter GC3) for coding regions, are therefore a good predictor of the local recombination
162 rate. For all these reasons, we have therefore chosen to focus on Passerida -- a clade of
163 passerines -- species to evaluate the molecular features of insularity. Passerida is a clade of
164 songbirds that originated 20 to 30 Million years ago (Baker et al. 2004; Prum et al. 2015, Oliveros
165 et al. 2019) and represents the most species-rich avian clade. In this study, we however
166 considered species with relatively similar body-mass, longevity and clutch-size to take into
167 account some confounding factors (see Materials and Methods).

168 In a nutshell, we combined in-house (n=89) and published (n=206) whole genome
169 sequences of 25 passerine bird species, including 14 endemic insular species, to estimate
170 parameters of molecular evolution. More precisely, we investigated the following questions
171 regarding the use of endemic insular species for evaluating predictions from the nearly neutral
172 theory: (i) Do endemic insular species exhibit genomic signatures consistent with evolution under
173 lower long-term effective sizes as compared to widely-distributed mainland species? (ii) Are high
174 deleterious mutation loads a common feature of insularity? (iii) Is there any covariation between
175 the deleterious mutation loads and heterogeneities in recombination rate? (iv) Is the proportion
176 of adaptive substitutions significantly lower or higher for the insular species? (v) How informative
177 these statistics are in a conservation biology context?

Results

Genomic signatures of lower long-term *Ne* in island species

Using a combination of in-house (n=89) and published (n=206) whole genome sequences of a set of 25 focal species consisting of 14 insular and 11 mainland species (Table 1, Fig. S1), we estimated the nucleotide diversity at synonymous (π_S) and at non-synonymous sites (π_N) for each species at 6,499 orthologous genes on average (range: 5,018-7,514, among 8,253 orthogroups). π_S varies from 0.07% in the Tenerife blue chaffinch (*Fringilla teydea*) to 1.26% in the willow warbler (*Phylloscopus trochilus*) representing a 17-fold difference (Table S1). These species also have the highest and lowest π_N/π_S with 0.285 and 0.120 for the Tenerife blue chaffinch and the willow warbler respectively (Table S1).

We recovered a strong negative correlation between the ratio of non-synonymous to synonymous nucleotide diversities (π_N/π_S) and π_S (Fig.1A, Spearman's ρ = -0.879, p-value=2.2×10^{-06}, see also Figs. S2-S4). Endemic insular species exhibit significantly less mean π_S than mainland species with only one mainland species (*Zosterops pallidus*) overlapping with the island species (mean π_S = 0.59% and 0.18% or mainland and island species respectively, Fig.1 B; Welch t-test = -3.82; p-value=0.003). Similarly, species endemic to island have a π_N/π_S 40% higher on average than mainland species (Fig. 1 C; mean π_N/π_S = 0.145 and 0.201 for mainland and island species respectively, Welch t-test = 4.86; p-value=6.8×10^{-5}).

Table 1: Sequencing data used in this study. The abbreviation in parenthesis following Darwin's finches names indicate the island of origin (C=Coco, E=Española, L=San Cristobal, P=Pinta, S=Santiago, W=Wolf, Z=Santa Cruz).

	Species	Clade	Range	Individuals	Data
1	*Certhidea olivacea (S)*	Darwin's finches	Island	5	Lamichhaney *et al.* 2015
2	*Certhidea fusca (E)*	Darwin's finches	Island	10	Lamichhaney *et al.* 2015
3	*Certhidea fusca (L)*	Darwin's finches	Island	10	Lamichhaney *et al.* 2015
4	*Platyspiza crassirostris (Z)*	Darwin's finches	Island	5	Lamichhaney *et al.* 2015
5	*Camarhynchus pallidus (Z)*	Darwin's finches	Island	5	Lamichhaney *et al.* 2015
6	*Pinaroloxias inornata (C)*	Darwin's finches	Island	8	Lamichhaney *et al.* 2015
7	*Geospiza difficilis (P)*	Darwin's finches	Island	10	Lamichhaney *et al.* 2015
8	*Geospiza septentrionalis (W)*	Darwin's finches	Island	8	Lamichhaney *et al.* 2015
9	*Geospiza conirostris (E)*	Darwin's finches	Island	10	Lamichhaney *et al.* 2015
10	*Ficedula albicollis*	Ficedula flycatchers	Mainland	20	Ellegren *et al.* 2012
11	*Ficedula hypoleuca*	Ficedula flycatchers	Mainland	20	Ellegren *et al.* 2012
12	*Ficedula speculigera*	Ficedula flycatchers	Mainland	20	Ellegren *et al.* 2012
13	*Ficedula semitorquata*	Ficedula flycatchers	Mainland	20	Ellegren *et al.* 2012
				6	Bourgeois *et al.* 2017
14	*Zosterops borbonicus*	White-eyes	Island	1	Leroy *et al.* 2019
				18	This study
15	*Zosterops olivaceus*	White-eyes	Island	15	This study
16	*Zosterops mauritianus*	White-eyes	Island	9	This study
17	*Zosterops pallidus*	White-eyes	Mainland	2	This study
18	*Zosterops virens*	White-eyes	Mainland	11	This study
19	*Fringilla coelebs*	Chaffinches	Mainland	15	This study
20	*Fringilla coelebs palmae*	Chaffinches	Island	9	This study
21	*Fringilla teydea*	Chaffinches	Island	10	This study
22	*Taeniopygia guttata castanotis*	Estrildidae	Mainland	19	Singhal *et al.* 2015
23	*Poephila acuticauda acuticauda*	Estrildidae	Mainland	10	Singhal *et al.* 2015
24	*Parus major*	Paridae	Mainland	10	Corcoran *et al.* 2017
25	*Phylloscopus trochilus*	Phylloscopidae	Mainland	9	Lundberg *et al.* 2017

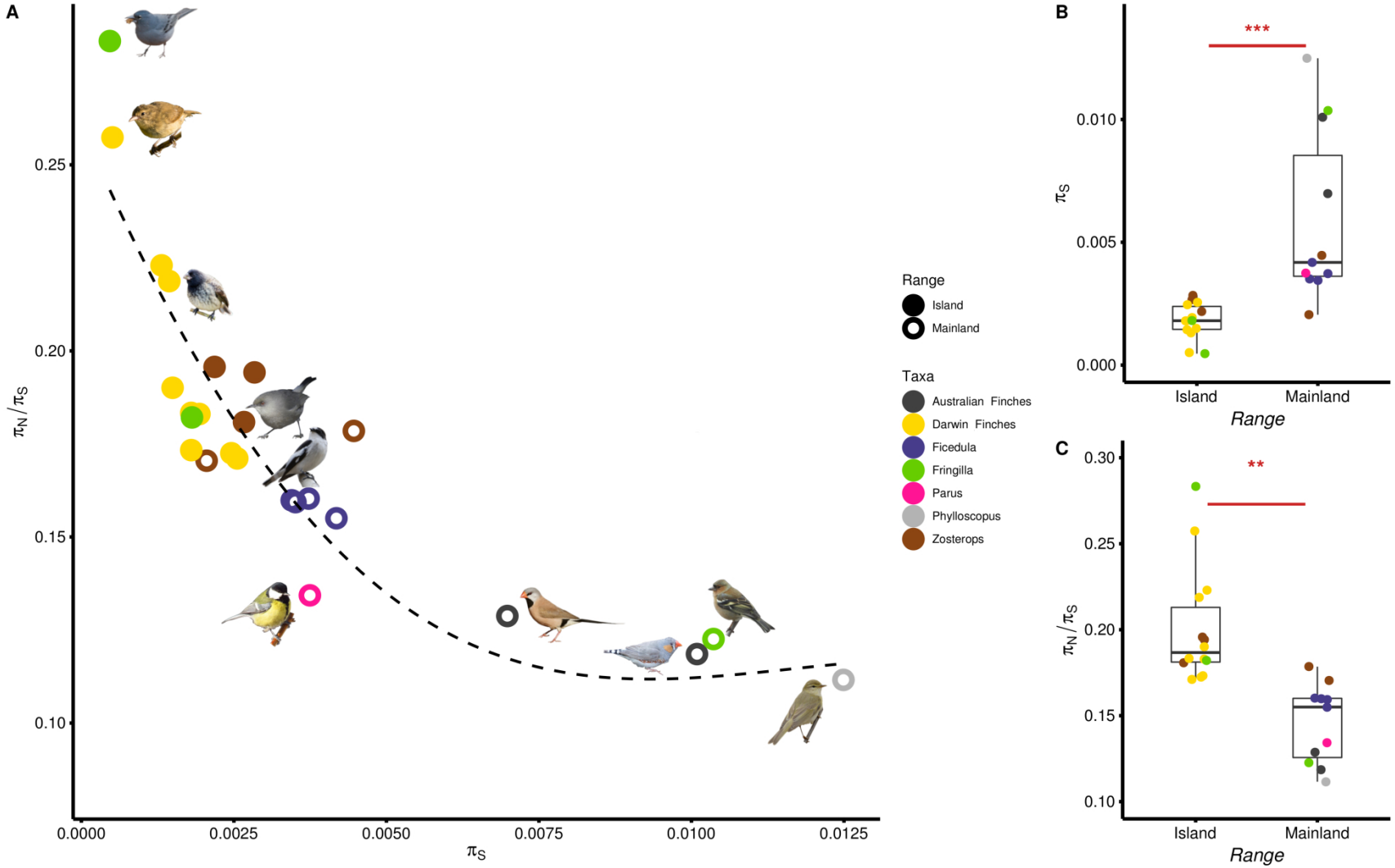

Fig.1: Island species as models for evolution in small effective population sizes. *A. Local polynomial regression (LOESS with span=1.25) between the ratio of nonsynonymous to synonymous nucleotide diversity (π_N/π_S) and the levels of nucleotide diversity (π_S), used as an indicator of effective population sizes. See also Fig. S3 for a log-transformed regression. B & C. Variation in nucleotide diversity (π_S, B) and π_N/π_S between the island endemic and the mainland species (C).Photo credits: F. Desmoulins, E. Giacone, G. Lasley, Lianaj, Y. Lyubchenko, B. Nabholz, J.D. Reynolds, K.Samodurov, A.Sarkisyan (iNaturalist.org); M. Gabrielli (personal communication).*

Linking census, effective population size and conservation status

Current census population size and the size of the geographical range size in square kilometers also significantly correlate with both π_S and π_N/π_S (Fig.2). Median estimates for the current census population sizes (only available for 13 of our 25 focal species) are positively correlated with π_S and negatively with π_N/π_S (R^2 = 0.63, p-value=0.001 and R^2 = 0.66, p-value=6.9 x 10^{-04}, respectively; Fig. 2 A & B). Similarly, geographic range size is also significantly correlated with $\log_{10}$-transformed π_S and π_N/π_S (Fig.2 C & D, table 2, R^2 = 0.54, p-value=3.1x10^{-05} and R^2 = 0.51, p-value=6.1x10^{-05}, respectively). In contrast, IUCN red list assessment has a weak effect on π_S and π_N/π_S (Table 2, R^2 = 0.20, p-value=0.03 and R^2 = 0.18, p-value=0.05, respectively). Census population size, geographical range, IUCN assessment and insularity (*i.e.* island/mainland status) are all cross-correlated. We use linear model analyses to evaluate the relative contribution of insularity in comparison to the other explanatory variables to π_S and π_N/π_S variations (Table 2). Following the minimum AIC criteria, the best model is census size or range size for explaining π_S but models considering insularity exhibit close AIC values (ΔAIC<2). Insularity is, however, the best model to explain π_N/π_S (Table 2). Linear models combining insularity and range size or census size always have lower AIC than models with one explanatory variable (Table 2) demonstrating that the effect of insularity is indeed linked to population size variation and that there is no intersect effect of insularity *per se*. In contrast, IUCN assessment has a weak effect, and this effect completely disappears in combination with insularity (Table 2). Even after explicitly controlling our statistical analyses for the phylogenetic relationship between species using a Phylogenetic Generalized Least Squared (PGLS) method with a Brownian motion model of character evolution (Grafen 1989; Paradis 2012), the significant effect of the island/mainland status, range size and census size on π_S and π_N/π_S remain highly significant (Table 2). Our dataset indeed contains species with at least 4 independent mainland-island transitions (Fig. S1), thus ensuring sufficient statistical power to support our interpretations.

Table 2: Linear models and* Phylogenetic Generalized Least Squared (*PGLS) method-based tests performed.
To perform accurate comparisons, model selections were performed using the maximum number of species (N) for which all information was available (i.e. 13 species with census size estimates, 21 species with IUCN status, and 25 with both insularity status & range sizes). IUCN status was considered as a binary variable "Threatened", when the status is "Vulnerable", or "Non-threatened" otherwise (i.e. "Least Concerned or Near-threatened). For each comparison, the best model is indicated in bold. For the models with PGLS, the p-value of the model was obtained from a Likelihood ratio test using the anova.lme function (nlme package, Pinheiro et al. 2020).

Type	*Model*	*N*	*R^2*	*p-value model*	*p-value variable 1*	*p-value variable 2*	*AIC*	*ΔAIC*
Linear	*$log_{10}(\pi_S)$ ~ Insularity*	*13*	*0.577*	*0.00259*			*5.01*	*1.72*
Linear	***$log_{10}(\pi_S)$ ~ log_{10}(census_size)***	*13*	0.630	***0.00121***			***3.29***	***0***
Linear	*$log_{10}(\pi_S)$ ~ log_{10}(census_size)+Insularity*	*13*	***0.656***	*0.00485*	*0.162*	*0.405*	*4.34*	*1.05*
Linear	*$log_{10}(\pi_S)$ ~ IUCN*	*21*	*0.133*	*0.104*			*16.33*	*13.62*
Linear	***$log_{10}(\pi_S)$ ~ Insularity***	*21*	*0.547*	***0.000128***			***2.71***	***0***
Linear	*$log_{10}(\pi_S)$ ~ IUCN+Insularity*	*21*	***0.548***	*0.000787*	*0.83*	*0.000727*	*4.66*	*1.95*
Linear	*$log_{10}(\pi_S)$ ~ Insularity*	*25*	*0.528*	*0.000039*			*4.45*	*0.52*
Linear	***$log_{10}(\pi_S)$ ~ $log_{10}(range\_km^2)$***	*25*	*0.537*	***0.000031***			***3.93***	***0***
Linear	*$log(\pi_S)$ ~ Insularity + $log(range\_km^2)$*	*25*	***0.552***	*0.000145*	*0.398*	*0.282*	*5.1*	*1.17*
Linear	*$log_{10}(\pi_N/\pi_S)$ ~ Insularity*	*13*	*0.577*	*0.00258*			*-27.68*	*3.01*
Linear	***$log_{10}(\pi_N/\pi_S)$ ~ log_{10}(census_size)***	*13*	*0.665*	***0.00068***			***-30.69***	***0***
Linear	*$log_{10}(\pi_N/\pi_S)$ ~ log_{10}(census_size)+Insularity*	*13*	***0.681***	*0.00331*	*0.102*	*0.496*	*-29.33*	*1.36*
Linear	***$log_{10}(\pi_N/\pi_S)$ ~ Insularity***	***21***	***0.552***	***0.000113***			***-49.67***	***0***
Linear	*$log_{10}(\pi_N/\pi_S)$ ~ IUCN*	*21*	*0.067*	*0.256*			*-34.26*	*15.41*
Linear	*$log_{10}(\pi_N/\pi_S)$ ~ IUCN+Insularity*	*21*	***0.578***	*0.000421*	*0.305*	*0.00019*	*-48.93*	*0.74*
Linear	***$log_{10}(\pi_N/\pi_S)$ ~Insularity***	*25*	*0.529*	***0.000038***			***-59.4***	***0***

Linear	*$log_{10}(\pi_N/\pi_S) \sim log_{10}(range\_km^2)$*	*25*	*0.513*	*0.000057*			*-58.36*	*1.04*
Linear	*$log_{10}(\pi_N/\pi_S) \sim Insularity + log_{10}(range\_km^2)$*	*25*	***0.541***	*0.00019*	*0.254*	*0.461*	*-58.03*	*1.37*
PGLS	***$log_{10}(\pi_S) \sim Insularity$***	***13***		***0.0001***	***0.0005***		***8***	***0***
PGLS	*$log_{10}(\pi_S) \sim log_{10}(census\_size)$*	*13*		*0.0153*	*0.029*		*17.25*	*9.25*
PGLS	*$log_{10}(\pi_S) \sim log_{10}(census\_size) + Insularity$*	*13*		*0.0003*	*0.365*	*0.0058*	*8.87*	*0.87*
PGLS	*$log_{10}(\pi_S) \sim Insularity$*	*25*		*0.0053*	*0.0081*		*20.38*	*2.16*
PGLS	***$log_{10}(\pi_S) \sim log_{10}(range\_km^2)$***	***25***		***0.0016***	***0.0028***		***18.22***	***0***
PGLS	*$log_{10}(\pi_S) \sim log_{10}(range\_km^2) + Insularity$*	*25*		*0.0049*	*0.117*	*0.458*	*19.51*	*1.29*
PGLS	***$log_{10}(\pi_S) \sim Insularity$***	*21*		***0.0066***	***0.0107***		***9.42***	***0***
PGLS	*$log_{10}(\pi_S) \sim IUCN$*	*21*		*0.942*	*0.946*		*16.8*	*7.38*
PGLS	*$log_{10}(\pi_S) \sim IUCN + Insularity$*	*21*		*0.024*	*0.804*	*0.0127*	*11.35*	*1.93*
PGLS	***$log_{10}(\pi_N/\pi_S) \sim$ Insularity***	*13*		***0.0006***	***0.0019***		***-22.93***	***0***
PGLS	*$log_{10}(\pi_N/\pi_S) \sim log_{10}(census\_size)$*	*13*		*0.021*	*0.0378*		*-16.38*	*6.55*
PGLS	*$log_{10}(\pi_N/\pi_S) \sim log_{10}(census\_size) + Insularity$*	*13*		*0.0021*	*0.557*	*0.0233*	*-21.4*	*1.53*
PGLS	***$log_{10}(\pi_N/\pi_S) \sim Insularity$***	*25*		***0.0046***	***0.0071***		***-46.96***	***0***
PGLS	*$log_{10}(\pi_N/\pi_S) \sim log_{10}(range\_km^2)$*	*25*		*0.0089*	*0.0131*		*-45.76*	*1.2*
PGLS	*$log_{10}(\pi_N/\pi_S) \sim log_{10}(range\_km^2) +$ Insularity*	*25*		*0.0121*	*0.41*	*0.191*	*-45.75*	*1.21*
PGLS	***$log_{10}(\pi_N/\pi_S) \sim Insularity$***	*21*		***0.0027***	***0.0049***		***-45.98***	***0***
PGLS	*$log_{10}(\pi_N/\pi_S) \sim IUCN$*	*21*		*0.667*	*0.686*		*-37.17*	*8.81*
PGLS	*$log_{10}(\pi_N/\pi_S) \sim IUCN + Insularity$*	*21*		*0.0085*	*0.496*	*0.0052*	*-44.53*	*1.45*

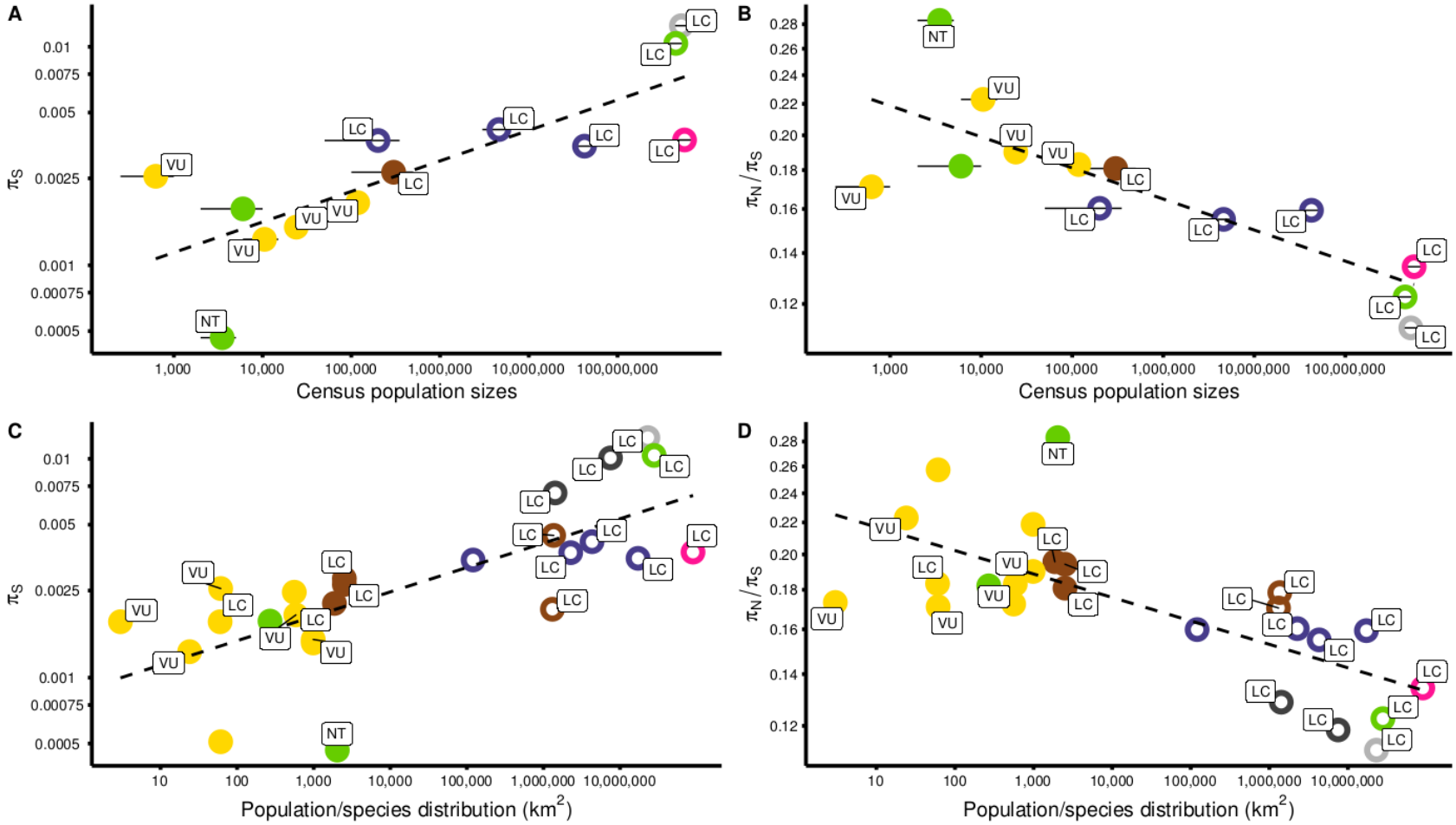

Fig. 2: Ecological-evolutionary correlations based on the variables investigated in this study. *π_S and π_N/π_S are used as proxies of Ne and the efficacy of natural selection to remove deleterious variants and are correlated with both the median estimates for the current census population sizes (A & B) and the geographical range sizes (C & D). Both ecological and evolutionary parameters are log-transformed. Filled and open dots represent the island and the mainland species, respectively (Fig. 1 for details). Only the 13 species with estimates of the current census population sizes are included for the panels A & B (with ranges shown with a thin black line). Where known, the IUCN conservation status of the investigated species is indicated (LC = least concerned, NT = near threatened, VU = vulnerable).*

Within-genome variation in the efficacy of purifying selection

In birds, the GC-content at third codon position (GC3) is known to be highly correlated with recombination rates (Hillier et al. 2004; Backström et al. 2010; Singhal et al. 2015). After ordering genes according to GC3 (see Materials and Methods), we observed negative slopes between (π_N/π_S) and GC3 (Fig. 3.A, B). These results were equally supported when using the estimates of π_N/π_S based on the method described in Rousselle and collaborators (2020) which considers the site frequency spectra at both non-synonymous and synonymous (method 2 in Fig. S5). The relationship between π_N/π_S and GC3 is therefore supported, and consistent with an increased efficiency of natural selection with higher local recombination rate.

By comparing sets of genes exhibiting the lowest and highest GC3 (2Mb of total concatenated sequences used for the computations), we reported significant differences in mean π_N/π_S between island and mainland species in both GC-rich (t=3.316 ,p=0.003) and GC-poor (t=7.300, p=2.1x10^{-07}) regions, but with more marked π_N/π_S differences in genes exhibiting low GC3 ($\Delta mean_{insular\ vs.\ mainland}$= 0.107, 95%CI: 0.077-0.194) than in those exhibiting high GC3 (Δmean= 0.029, 95%CI: 0.011-0.048) (Fig. 3C). The genetic advantage of recombination, which is captured by the slope between the π_N/π_S and the GC3 is itself highly correlated with the levels of nucleotide diversity (Fig. 3E), consistent with a growing importance of the local recombination rate for the efficiency of purifying selection with respect to the effective population size, which is especially important for the endemic island species (filled dots, Fig. 3E). The stronger effect of recombination for island species is is captured by the significant interaction between GC3 and insularity in the linear model : π_N/π_S ~ GC3+insularity+GC3:insularity (R^2 = 0.72, *p model* <2.2x10^{-16}, *p GC3* <2.2x10^{-16}=, *p insularity* = 2.8x10^{-10}, *p interaction* = 1.33x10^{-05}).

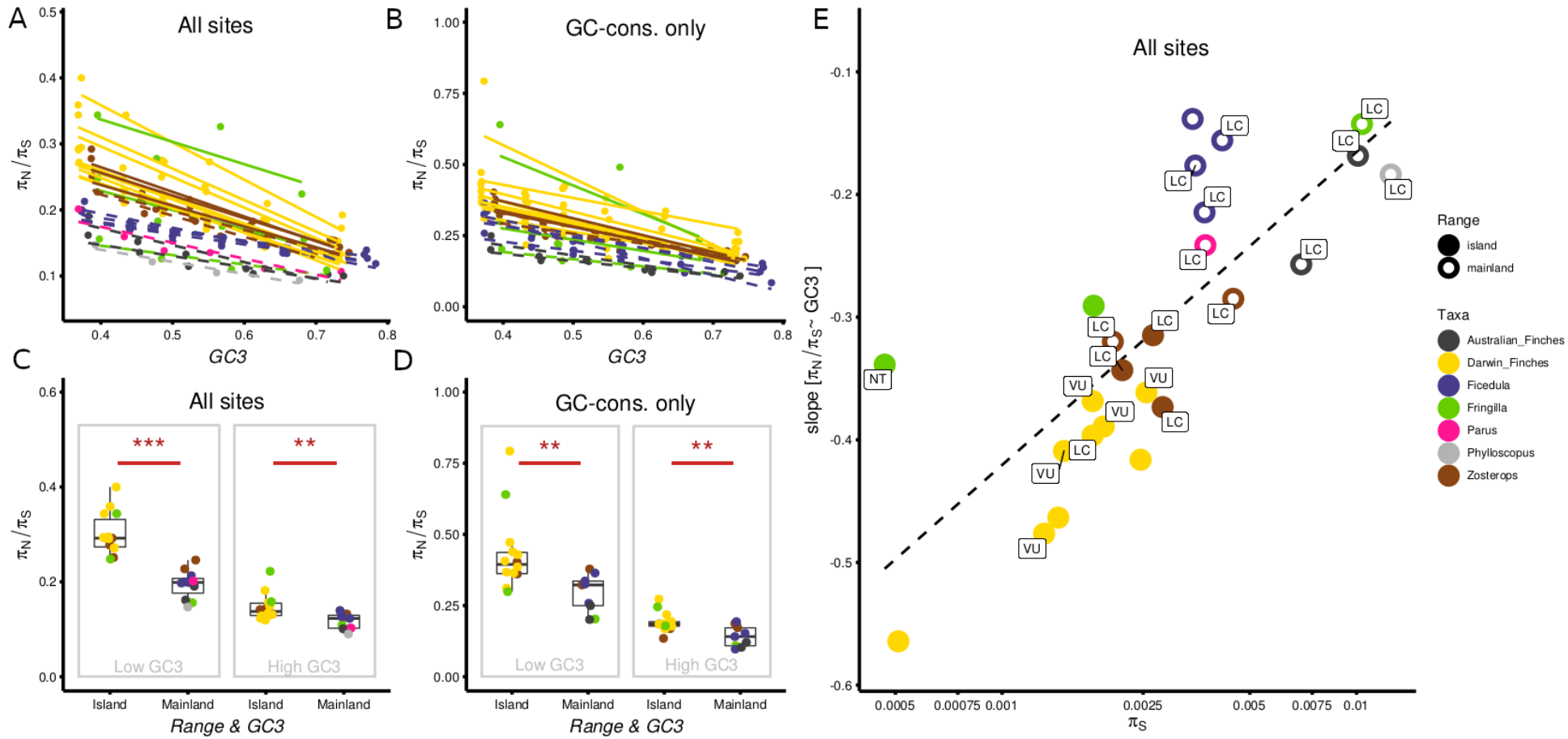

Fig.3: Efficiency of natural selection of insular and mainland passerine bird species to purge potentially deleterious mutations depending on the local recombination rates. *A & B. Linear regressions between π_N/π_S ratios and the third codon position (GC3), a proxy of local recombination rates for all sites (A) and GC-conservative sites only (B). C & D. Observed variation in π_N/π_S ratios at GC-rich and GC-poor genes between island and mainland species for all sites (C) and for GC-conservative sites only (D). E. Slopes of the linear regressions between πN/πS and GC3 (i.e. as shown in the panel A) depending on the synonymous nucleotide diversity. The dotted line indicates the best-fitting linear model.*

To take into account the potential impact of gBGC on the estimations of π_S and π_N, we used the same approach than in Rousselle et al. 2018 and computed π_N/π_S based on the Site Frequency Spectrum (SFS) computed for GC-conservative sites only (i.e., A↔T and G↔C mutations), a part of the total variants which is unaffected by gBGC. Interestingly, we also recovered significant differences in π_N/π_S between island and mainland species at GC-conservative sites (Fig. 3D, see also Fig. S6), albeit slightly less significant (GC-poor: Δmean=0.138 with 95%CI: 0.050-0.225, t=3.284, p=0.004; GC-rich: Δmean=0.050 with 95%CI: 0.018-0.082, t=3.284, p=0.005). The evolutionary advantage of genetic recombination is higher in the island (GC-conservative π_N/π_S, $\Delta mean_{GC\text{-}rich\ vs.\ GC\text{-}poor}$=0.239) than in the mainland species (Δmean=0.151), which seems to imply that recombination more efficiently limits the mutation burden of insular species. However, this advantage is still insufficient to entirely compensate for the higher fixation of deleterious alleles due to the increased drift effects in species in small populations, as typically represented by the island passerine species used in our study.

The double penalty of insularity: fewer advantageous & more weakly deleterious substitutions

For each species, we derived the expected ratio of non-adaptive non-synonymous to synonymous substitutions between species (ie, ω_{NA}: the rate of non-adaptive substitution; Galtier 2016) using the unfolded SFS at both synonymous and nonsynonymous sites. We then compared this expectation to the observed d_N/d_S (ω) as estimated based on the sequence divergence with an outgroup species (for a list, see Table S3), obtaining the proportion of amino-acid substitutions that result from positive selection (α). After scaling this proportion by the observed number of synonymous substitutions, we estimated ω_A, the rate of adaptive substitutions relative to neutral divergence. In other words, we used the following equations: $d_N/d_S = \omega_A+\omega_{NA}$ and $\alpha = d_A/d_N$, where d_A is the adaptive non-synonymous divergence.

Interestingly, we observed significant differences in the rate of adaptive (ω_A) or non-adaptive (ω_{NA}) substitutions between insular and mainland species as well as significant correlations between both variables and π_S, the predictor of long-term *Ne* (Fig. 4; see Sup Fig S7 for α estimates). In detail, ω_A is significantly positively correlated with the log-transformed π_S (slope=0.097, R^2=0.300, p-value=0.004; Fig. 4A) and negatively with the log-transformed π_N/π_S

(slope=-0.403, R^2=0.389, p-value=9x10^{-4}; Fig. S8). Reciprocally, ω_{NA} is significantly negatively correlated with π_S (slope=-0.102, R^2=0.414, p-value=6x10^{-4}; Fig. 4B) and positively with π_N/π_S (slope=0.385, R^2=0.437, p-value=4x10^{-4}; Fig. S8). Mainland and island species exhibits significant differences for both ω_A (Δmean=0.057, t=-2.8407, p-value=0.010) and ω_{NA} (Δmean=0.063, t=4.16, p-value=5x10^{-4}).

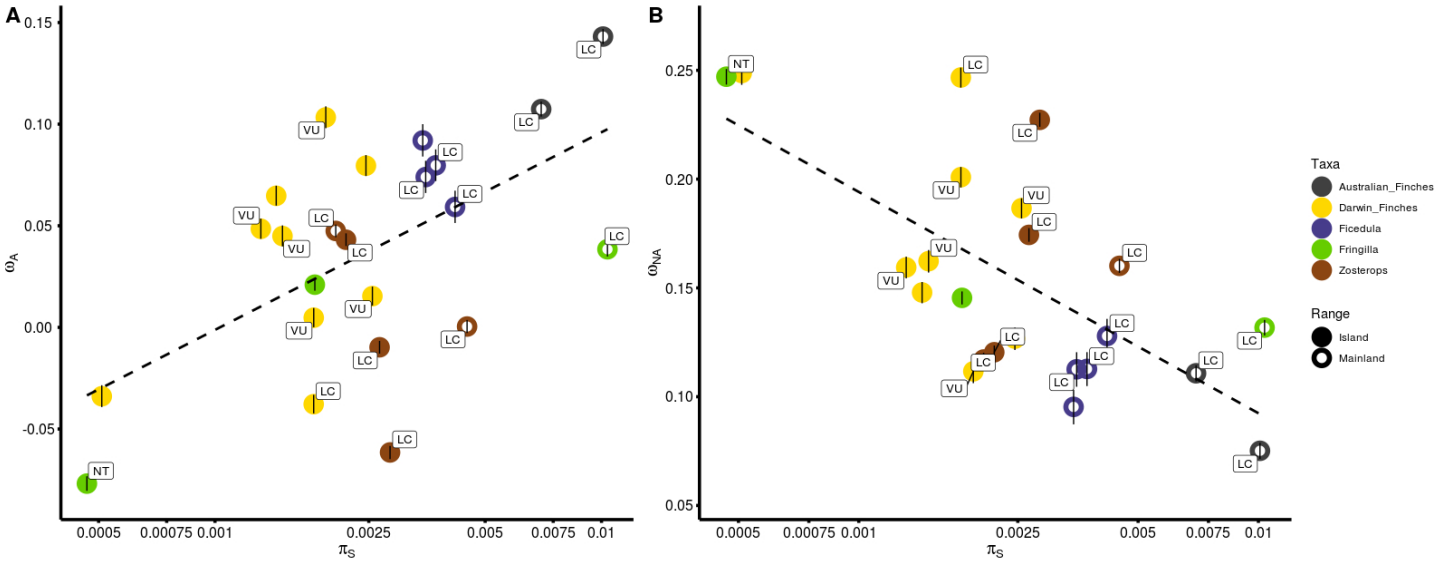

Fig. 4: Proportion of adaptive (A) and non-adaptive (B) substitutions along the neutral genetic diversity gradient (π_S) *as estimated after comparing the observed and the expected d_N/d_S under near neutrality assuming the polymorphism data using the DFE-α method (α shown Fig. S7; with $\omega_A = \alpha(d_N/d_S)$ & $\omega_{NA} = (1-\alpha)(d_N/d_S)$). Estimates were performed using all sites and the GammaExpo model. Error bars (purple line) represent the 95% confidence intervals of each estimate under this model. Where known, the IUCN conservation status of the investigated species is indicated (LC = least concerned, NT = near threatened, VU = vulnerable).*

Discussion

The analysis or reanalysis of full-genome resequencing data allowed us to find consistent results of lower nucleotide diversities, higher mutation loads and lower adaptive substitution rates in endemic insular species as compared to mainland ones. These results provide important insights for both evolutionary and conservation biology, and we will develop these two lines of reasoning one after the other.

Lower *Ne* in endemic island species

The smaller land area available on islands was supposed to constrain the upper bound of both census and effective population sizes of endemic insular species, in such a way that this demography has affected the ability of purifying selection to remove weakly deleterious mutations. Overall, our results are consistent with this general hypothesis. For all parameters, including πS and π_N/π_S, our results were consistent with significant differences associated with insularity (mainland/island), even after explicitly taking into account the phylogenetic signal. The evidence of low *Ne* in insular species is important for evolutionary biology for a series of reasons, (i) even if π_S and π_N/π_S are convenient and informative summary statistics in population genomics, there is a fundamental statistical problem of non-independence of using π_S as both a proxy of *Ne* and as the denominator of π_N/π_S for estimating the efficacy of selection (Piganeau and Eyre-Walker 2009), (ii) leaving aside this statistical issue, there is a conceptual problem of circularity in using population genomics variables to estimate *Ne* for testing the impact of this *Ne* on population genomics variables, (iii) very different scenarios can led to a similar level of nucleotide diversity at synonymous sites, in particular assuming that selected and neutral alleles have different dynamics (Mugal et al. 2013, 2020), (iv) finally, the use of life-history traits is unsatisfactory as well because these proxies of *Ne* are indirect and only rely on the correlations between body size and population density (White et al. 2007).

Our results supporting a lower *Ne* in the insular species is consistent with several previous taxon-specific investigations (Loire et al. 2013 for Giant Galápagos tortoises; Rogers and Slatkin 2017 for woolly mammoths; Robinson et al. 2016; 2018 for island foxes). In contrast, James et al. (2016) found that island and mainland species have a similar level of genetic diversity. James and collaborators' study was however almost exclusively focused on organelle DNA sequences. The difference can probably be explained by the importance of combining information from several independent loci (*i.e.* thousands of orthologous genes in our study). More recently, Kutschera et al. 2019 found consistent results in corvids. Using whole-genome sequencing data of seven species within the genus *Corvus*, they found support for higher mutation loads in the insular species, as compared to their mainland relatives. However, the low sample size (only two island species), as well as the emphasis on a single taxon, limits its degree of generalization (Kutschera et al. 2019). Using a larger set of island and mainland species from several distant taxa across the passerida phylogeny, our conclusion reached another level of significance. Indeed, the difference

between mainland and island species remains highly significant, even after a conservative control for the phylogenetic relationship among the species, thus allowing us to fully exclude the hypothesis of an impact of the shared ancestry on the conclusions (Kutschera et al. 2019). In plants, a four-taxa comparison with both endemic island and widespread relatives for each taxa found significant lower nucleotide diversities and higher mutation loads on the island endemic species, and therefore provide support for the endemic island status in plants (Hamabata et al. 2019). Therefore, it is very likely that the endemic island status provides a good proxy of *Ne*, not only for specific animal clades such as birds or mammals, but for terrestrial species as a whole. More broadly, this result opens up new opportunities for using island species as models to understand the impact of *Ne* on genome evolution, including genome size, or natural selection on non-coding genomic regions.

Bird population genomics data has ridden to the rescue of the nearly neutral theory

Fifty years after the introduction of the neutral theory of molecular evolution by Motoo Kimura (1968) and by Jack Lester King and Thomas Hughes Jukes (1969), the neutralist–selectionist debate has recently been reinvigorated (Kern & Hahn, 2018). In their standpoint, the latter authors reaffirmed the importance of the linked positive and negative selection in such a way that they reached the conclusion that "*the neutral theory has been overwhelmingly rejected*". Kern & Hahn's point of view generated a series of counter-claims supporting the central role of the neutral theory for explaining empirical patterns of molecular evolution, without denying the substantial contribution of linked selection (e.g. Chen et al. 2020; Jensen et al. 2019). Following the latter authors, our investigations are consistent with the high explanatory power of the neutral theory, or more precisely the nearly theory (Ohta, 1992). As expected, we indeed found that the proportion of deleterious mutations scaled negatively with the nucleotide diversity. Similarly, we found support for lower proportion of adaptive substitutions and higher proportion of non-adaptive substitutions, in species with lower effective population sizes. Interestingly, for all the species we investigated, the proportion of adaptive amino-acid replacement substitutions never reached 50% nor 25%, but only 14.3% for the species exhibiting the highest ω_A, which is somewhat incongruent with a pervasive effect of natural selection in the

evolution of protein sequences, as discussed by Kern & Hahn based on the patterns observed in fruit flies (Kern & Hahn 2018; see also Jensen et al. 2019). Similar low values for ω_A (<20%) were observed by Rousselle *et al.* 2020 based on the analysis of 50 species from ten distant groups of animals, including two bird clades: fowls and passerines. To sum up, using our panel of songbird species and based on the empirical patterns of molecular evolution we were able to recover, the nearly neutral theory holds true. It must however be noted that, through GC3 - a proxy of the local recombination rates in birds (e.g. Bolívar et al. 2016) - we also found unambiguous evidence for the contribution of linked selection, influencing both the local levels of nucleotide diversity and loads of deleterious mutation, and consistent with heterogeneous landscapes of *Ne* throughout genomes (Rousselle et al. 2019). Interestingly, our analysis reveals an intriguing interaction between recombination rate and the efficacy of selection. Indeed, the effect of GC3 on π_N/π_S significantly differs between species exhibiting a low *Ne* and species with a high *Ne*, with a stronger effect of recombination in insular (low *Ne*) species. We have however shown that this interaction is independent of the process of gBGC (Fig. 3B & D). This effect could be the result of a different distribution of the fitness effects of mutations between low *Ne* and high *Ne* species. The effect of linked selection is known to be particularly strong for mildly selected mutations (Comeron et al. 2008). We can therefore hypothesize that low *Ne* species contribute to a DFE shifted toward more mildly selected mutations as compared to the DFE of high *Ne* species, and consequently produce more numerous footprints of natural selection in their genomes.

Ecological-evolutionary ties and perspectives

The importance of population genomics for ecological research has long remained quite limited (Hugues et al. 2008), but there is now a growing interest for conservation biology studies on the interplay between ecology and evolution. At a macroevolutionary scale, it has become clear that the determinants of genetic diversity are mostly ecologically-driven, as shown by the strong correlations between some key life-history traits and both polymorphisms levels and mutation loads, namely propagule sizes and longevity, respectively. These ecological-evolutionary correlations have been independently recovered in animals and plants (Romiguier et al. 2014; Chen et al. 2017; Plomion et al. 2018). Interestingly, Romiguier et al. 2014 failed to find an effect of geographical range that might be due to the limitation of investigating these patterns in a limited number of species per taxon over a large diversity of taxa. Using a dataset of

639 mammalian species, James & Eyre-Walker (2020) however found support for significant correlations between the levels of mitochondrial diversity and the range size, in addition to some life-history traits such as the mass-specific metabolic rates and the age of sexual maturity, a trait which is itself strongly correlated with longevity. At a microevolutive scale, the intensity of these ecological-evolutionary links remains more debated. Importantly, our analysis was designed to focus on a group of species exhibiting roughly similar body mass and longevity to allow us to investigate some other ecological factors, because the avian genetic diversity is known to be associated with body mass (Eo et al. 2011). It is important to mention that we also checked a posteriori the absence of correlation between body mass and the levels of genetic diversity (Spearman's Rho = -0.11, p-val = 0.58). Consistent with our earlier discussion about insularity, we found that the nucleotide diversity scaled positively with the species range, which therefore represents a gradual transition from species restricted to small islands to species widely distributed over continents. In addition, census population sizes were found to be another excellent explanatory ecological variable, explaining both the levels of nucleotide diversity and the mutation loads. However, neither a relationship between levels of nucleotide diversity and census size nor between the levels of diversity and range size were recovered when endemic island or mainland species were analyzed separately (p-value > 0.10 but see Díez-del-Molino et al. 2018 and Brüniche-Olsen et al. 2019). At this smaller scale, recent demographic variations may have affected the expected relationship between *Ne* and census size and similarly, *Ne* and range size. It should be noted however that our dataset was not ideal for testing the variation among island or mainland species separately, especially given that no mainland species with small census population sizes or range sizes were included in this study.

How informative population genomics variables can be to evaluate the conservative status of the endangered species? Even if our dataset was not initially designed to tackle this issue, with only five threatened (vulnerable) species, all among the 9 Darwin's Finch species, we found no obvious contrast between these species and the other non-threatened endemic insular species, neither for the levels of nucleotide diversity nor for their efficacy of natural selection. This result came as a surprise given that a general outcome is that threatened animal species generally exhibit lower genetic diversity than non-threatened species, including for birds (at least at microsatellite loci, Doyle et al. 2015; Willoughby et al. 2016 for a review). It must however be noted that more recent investigation using whole-genome data of a single individual from 78 mammal species was unable to recover this difference, at least without explicitly taking into

account the animals' diets (Brüniche-Olsen et al. 2018). Similarly, Díez-del-Molino et al. (2018) were unable to recover a significant effect of the IUCN assessment on the levels of nucleotide diversity using a large set of both birds and mammals. Comparing some critically endangered plant species from 4 Eastern Asian taxa and some of their widespread relatives, Hamabata et al. (2019) however found lower genetic diversity and higher mutation loads in the critically endangered plant species, suggesting a long-term evolution in small population size and a long-term reduction in the efficacy of selection. Using divergence data from 76 species of teleostean fishes, Rolland & Romiguier (2019) similarly found that the IUCN red list index provides an equally significant correlation with the non-synonymous over synonymous substitutions than body length, suggesting that the proportion of deleterious substitutions is strongly associated with vulnerability in fishes. In birds, Brüniche-Olsen et al. (2019) recently investigated this patterns on 17 species of tanagers from 27 islands, including a large majority of Darwin's Finches from the Galapagos and the Coco Islands (8 of the 9 Darwin's finches are also used in our study), but also two species from the Barbados. The authors found that the threatened species of tanagers based on the IUCN status (considering all together 'critically endangered', 'endangered' or 'vulnerable') exhibited lower genetic diversity than non-endangered species (*i.e.* 'near-threatened' and 'least concerned') and that both body size and island area are good predictors for the levels of genomic diversity, but the authors did not explicitly evaluate the efficacy of purifying selection. As a consequence, we still have a long way to go towards precisely describing the strengths of these ecological-evolution ties and how this information is in line with the current species conservation status.

Unresolved issue to tackle: long-term survival of highly loaded species

IUCN provides threat categorizations to guide short-term (*i.e.* a century or less) conservation efforts based on the ecological information available, e.g. the evolution of the census population sizes or the current geographical ranges. This is especially important given the large scientific consensus for a climate emergency and the need for an effective action for preserving biodiversity and ecosystems (Ripple et al. 2020). Beyond all these recent human-induced threats, a still open and non-trivial issue is to evaluate if population genomics can provide summary statistics informative enough to extend these objectives over a long-term timeframe. Small island populations indeed exhibit some intrinsic features that may contribute to an increased maladaptation and accelerate, and virtually lead to extinction, alone or in

association with the ecological reasons. We can in particular question how the mutation load evolves over time and if the deleterious burden can result in the extinction of the species. This hypothesis holds true only if these deleterious mutations are not opposed by compensatory or beneficial mutations (Lynch & Gabriel, 1990). This is especially important to mention given that the 4 threatened species (vulnerable status) are not these species exhibiting the lower proportion of adaptive substitutions (mean ω_A=0.053 as compared to 0.036 for the 15 species with a Least Concerned status, respectively). As a consequence, we can question if these high mutation loads can really contribute to extinction alone, following the general hypothesis that these mutations are accumulated faster in genomes than they are removed from the population by selection (Kondrashov 1988). Using a macro-evolutionary approach, Warren et al. (2018) found that the probability of being endangered for an island species increased with the time spent on an island (*i.e.* the time since the island was colonized). Similarly, Valente et al. (2020) found support for a negative relation between island sizes and extinction rates, confirming the general expectation of the island biogeography theory. These results are interpreted as the consequence of age-dependent processes such as ecological specialisation or coevolutionary balance between immigrant and resident species, but the accumulation of deleterious mutations might explain, in whole or in part, this phenomenon as well. Rogers & Slatkin (2017) proposed that, after a tipping point, this mutational meltdown may contribute to the ultimate steps in the road to extinction, as suspected based on one of the fossil remains of one the latest mammoth specimens (Palkopoulou et al. 2015). Such a study also illustrates the importance of paleogenomics in future conservation studies, in order to trace the temporal erosion of nucleotide diversity and accumulation of deleterious mutations, especially for the endangered species (Díez-del-Molino et al. 2018). With the currently available affordable sequencing strategies, the cost-benefit balance is increasingly favorable to get a reference genome assembly and reduced- or whole-genome resequencing data for ancient and/or modern samples of a given species (Brandies et al. 2019). As shown in this study, endemic island species represent species of high interest to evaluate, in a conservation biology context, the long-term consequences of such limited effective population sizes.

Materials and Methods

In this study, we both reanalyzed publicly available and generated our own sequencing data. By generating the data, our objective was to target taxa containing both island and mainland relatives (chaffinches and white-eyes) in order to increase our statistical power. More broadly, our comparison is only based on species with relatively similar body-mass, longevity and clutch-size. This control was introduced to reduce the risk of some confounding factors that could correlate with *Ne* (Romiguier et al. 2014) in order to be able to truly assess the effect of insularity. All scripts and programs used are available at the following Open Science Framework repository: https://osf.io/uw6mb/?view_only=a887417cbc91429dac0bbdc1705e2f2b.

Re-use of publicly available data for passerine birds

We collected publicly available raw sequencing data on SRA from a large range of studies (Table S2). Because the phylogenetic relationship between these species are not fully resolved (Lamichhaney *et al.* 2015; 2016; Zink & Vázquez-Miranda 2019), we first evaluate the net divergence between all pairs of species to delimitate 9 groups of species with a net divergence (D_A) > 0.1% (see Sup Note 1). Within each group, we selected a single population based on the number of individuals sequenced publicly available (Lamichhaney *et al.* 2015).

Newly generated whole genome sequencing data (*Zosterops* and *Fringilla* species)

All *Zosterops* and *Fringilla* individuals were captured using mist nets. With the notable exception of African *Zosterops* species (*Z. pallidus* and *Z. virens*, see below), we collected blood samples for each bird by venipuncture of the brachial vein and then stored in absolute ethanol at -20°c until DNA extraction. For African species, *Z. pallidus* and *Z. virens* individuals, DNA was

extracted from liver, muscle or blood. For these samples, the vouchers are stored at the Museum National d'Histoire Naturelle (MNHN), Paris, France and a tissue duplicate is deposited in the National Museum Bloemfontein (South Africa). For all *Zosterops* and *Fringilla* samples, total genomic DNA were extracted using the DNeasy Blood and Tissue kit (QIAGEN, Valencia, CA) following the manufacturer instructions. Library preparation (1.0 µg DNA used per sample) and Illumina high-throughput sequencing using a paired-end 150 bp (PE150) strategy were performed at Novogene (Cambridge, UK) to a minimum sequencing yield of 18 Gb per sample. Details on samples are available in Table S2.

Newly generated genome assembly (*Fringilla coelebs*)

DNA from muscle tissues of one *Fringilla* individual captured in Madrid, Spain was sequenced to obtain a draft genome of this species. For this individual, two libraries were prepared, a Chicago library and a Dovetail HiC library (Dovetail Genomics, Scotts Valley, CA), as described previously in Putnam et al. (2016) and Lieberman-Aiden et al. (2009), respectively. Both libraries were sequenced on an Illumina HiSeqX platform to produce over 100 million 2x151 bp reads. A *de novo* assembly was constructed using Meraculous (v. 2.2.2.5 diploid_mode 1, Chapman et al. 2011) with a kmer size of 73 and using 397.3 million paired-end reads (totaling 1,204 Gbp). Reads were trimmed for quality, sequencing adapters, and mate-pair adapters using Trimmomatic (Bolger, Lohse & Usadel, 2014). The input *de novo* assembly, shotgun reads, Chicago library reads, and Dovetail HiC library reads were used as input data for HiRise, a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). The HiRise assembly total length was 907.37 Mb obtaining 3,240 scaffolds and a final N50 of 68.85 Mb. The read coverage was 274X.

Species distribution and IUCN red list status

Species range sizes were obtained using from BirdLife (http://datazone.birdlife.org/) or, when not available, estimated based on the information shown on the IUCN-red list webpage using CalcMaps (https://www.calcmaps.com/map-area/). For endemic island species with populations on different islands (*C. fusca*) or on both island and mainland (*F. coelebs*), we

536 considered the total island area as a maximum bound for the population range. The conservation
537 status shown by the IUCN red list are at the species level and not at below-species level. As a
538 consequence, we considered either a missing information for both populations (e.g. *Certhidea*
539 *fusc*a E and *C. fusca* L) or the status only for the most widely distributed species (e.g. the Least
540 concerned (LC) status for the population with a large continental distribution rather than for the
541 island one (e.g. *F. coelebs palmae*, the population only found in the small La Palma island). For
542 *Ficedula speculigera*, a species with a D_A>0.002 (Sup. Note 1) and recognized as a distinct species
543 from *F. hypoleuca*, no information is yet available in the IUCN red list database.

544

545 **Gene models & orthology prediction**

546

547 We used one reference genome for all species belonging to the same clade (Table 1). For
548 Darwin's finches, we use the genome of a medium ground-finch individual (*Geospiza fortis*;
549 assembly GeoFor_1.0; GCF_000277835; Parker et al. 2012). For *Ficedula* flycatcher, we used the
550 genome and gene models of the collared flycatcher (*Ficedula albicollis*, GCF_000247815;
551 assembly FicAlb_1.4; Ellegren et al. 2012). For the Zosterops, we used the genome of the Reunion
552 grey white-eye (*Zosterops borbonicus*; GCA_007252995; assembly ZoBo_15179_v2.0; Leroy et al.
553 2019). For the Estrildidae, we used the assembly of the zebra finch (*Taeniopygia guttata*;
554 GCF_000151805; assembly taeGut3.2.4; Warren et al. 2010). For the chaffinches, we use the
555 newly generated assembly of *Fringilla coelebs* (see above). Finally, we used the genome of the
556 willow warbler (*Phylloscopus trochilus*; GCA_002305835; assembly ASM230583v1; Lundberg et
557 al. 2017) and of the great tit (*Parus major*; GCF_001522545.2; assembly Parus_major1.1; Laine et
558 al. 2016). For the willow warbler (*Phylloscopus trochilus*) and the chaffinch (*Fringilla coelebs*), we
559 performed a protein homology detection and intron resolution using genBlastG (She et al. 2011;
560 http://genome.sfu.ca/genblast/download.html) with the following options "-p genblastg -c 0.8 -r
561 3.0 -gff -e 1e-10". The proteins of the collared flycatcher (*Ficedula albicollis,* assembly FicAlb_1.4;
562 Ellegren et al. 2012) were used as reference.

563 To analyse the same orthologous sequences in all species, we used the set of 8253
564 orthologs identified in Jarvis et al. 2014. Then, we add the sequence of our species to this set of

565 orthogroups using the method described in Scornavacca et al. 2019. Briefly, each orthogroup was
566 used to build an HMM profile using the HMMER toolkit (Eddy 2011). Then, for each new
567 sequence, *hmmscan* was used on the HMM database to get the best hits among the orthogroups.
568 For each orthogroup, the most similar sequences for each species were then detected via
569 *hmmsearch*. Outputs from *hmmsearch* and *hmmscan* were considered as accurate if the first hit
570 score was substantially better than the second best one (in order to limit the risk of paralogy),
571 following a best-reciprocal-hit approach when the results of both programs were compared
572 (Scornavacca et al. 2019).

573

574 ## Variant identification

575 We used Trimmomatic (v.0.33; Bolger et al. 2014) to remove adapters, stringently trim
576 and filter reads using the following set of parameters: LEADING:3 TRAILING:3
577 SLIDINGWINDOW:4:15 MINLEN:50. All trimmed reads were then mapped against the reference
578 genome for each clade (see above) with BWA mem (v. 0.7.12; Li 2013) using default settings.
579 Unmapped reads and mapped reads with a quality (MQ) below 20 were then discarded. Potential
580 PCR duplicates were then removed using MarkDuplicates v. 1.140 (Picard tools). Variant calling
581 was then performed using GATK (v. 3.7; McKenna et al. 2010). First, we used HaplotypeCaller on
582 single samples (gVCF) to call SNPs using default parameters. For each species, we then performed
583 a joint genotyping ("GenotypeGVCFs"). To ensure high quality in our dataset, we filtered out low-
584 quality SNPs using several settings: a quality by depth (QD) < 2.0, a Fisher Strand (FS) bias >60 , a
585 mapping quality (MQ) <40, a MQranksum < -2 or a ReadPosRankSum < -2 or a Raw Mapping
586 Quality (Raw_MQ) < 45,000. SNPs satisfying one of the conditions described above were
587 discarded. For every group of species, we performed principal components analyses (PCA) based
588 on a random sampling of SNPs over the genome (50-200k) to capture additional levels of
589 population structure or an unfortunate misnaming of an individual that could have occurred at
590 one point between the bird sampling campaign and the analysis of the raw sequencing data.

591

592 ## Empirical estimate of the SNP calling error rate

One individual (namely *Zosterops virens* RCHB1917) was sequenced twice at coverage 10x and both replicates went through all the genotyping pipeline blindly. We estimate the error rate by counting the number of differences between the two replicates. All the differences detected correspond to heterozygous sites. We estimated a rate of 2.9×10^{-5}, which is approximately fifty times lower than the expected genome-wide heterozygosity for this focal species (1.5×10^{-3}).

Sequence reconstruction

For each individual, we then reconstructed fasta sequences from VCF files. First, we discarded all indel variations to ensure an exact match between positions in sequences and coordinates in the reference gff file (*i.e.* no frameshift mutation allowed). For each scaffold, we performed a base-by-base reconstruction of two sequences per individual by adding either a reference or an alternate allele based on the genotype information available in the vcf file at the position under investigation. Albeit simplistic, this approach is valid here because our study is only based on the nucleotide levels and does not require haplotype phase information. To exclude sites with abnormally low or high coverage, we compute percentiles of the distribution of coverage over the whole genome for each individual and then calibrate a minimum and maximum coverage. All sites exhibiting very low or very high coverage were masked. A minimum threshold value of 3 was set for all species. We then used a home-made python script to extract CDS from these reconstructed sequences based on the gff files.

Mitochondrial genomes assembly and phylogeny

For the species with mtDNA genomes available on genbank, we used the data (accessions: *Ficedula albicollis* : KF293721; *Fringilla coelebs* : NC_025599; *Fringilla teydea* : KU705740; *Parus major* : NC_026293; *Taeniopygia_guttata* : NC_007897; *Zosterops borbonicus* : MK529728; *Zosterops pallidus* : MK524996. We also add three outgroups : Corvus_brachyrhynchos : NC_026461; Lanius cristatus: NC_028333; Menura novaehollandiae : NC_007883).

For the remaining species, we used MitoFinder (vers. 1.1; https://github.com/RemiAllio/MitoFinder; Allio et al. 2020) to extract all the mitochondrial

protein coding genes. One individual per species was randomly selected to use in MitoFinder with default parameters. Alignments were performed gene by gene using macse (vers. 2; Ranwez et al. 2018). Alignments are available at the following URL: https://osf.io/uw6mb/?view_only=a887417cbc91429dac0bbdc1705e2f2b. Then all genes were concatenated to form a supermatrix (missing data = 20% on average, sd = 14%). Next, a phylogenetic analysis was performed using IQTREE (Nguyen et al. 2015) using the "GTR+G4" substitution model and an ultrafast bootstrap option to have a crude approximation of each node support.

Summary statistics of the polymorphic data

π_N/π_S ratios were computed using *seq_stat_coding*, a publicly available Bio++ script already used by Leroy *et al.* 2019 (https://osf.io/uw6mb/). In addition, we used the π_N/π_S estimates based on the site frequency spectra at both non-synonymous and synonymous sites and described in Rousselle and collaborators (2020) to check the accuracy of these estimates (method 2 in Figs. S4 & S5, see also Table S1). Guanine-Cytosine (GC) content at third-codon positions of protein-coding genes (hereafter GC3), an excellent proxy of the local recombination rate in birds (Bolívar et al. 2016) was also computed under *seq_stat_coding*. To estimate the within-genome variation in the efficacy of selection, we estimated π_N/π_S on sets of genes representing a total concatenated coding alignment of 2 Mb. To do so, genes were sorted in ascending values of GC3 in order to compute π_N/π_S ratios on genes exhibiting roughly the same GC3 values (median GC3 value for the bin was then considered, e.g. in Fig. 3). To ensure reliable estimates, we excluded the last π_N/π_S estimate corresponding to genes exhibiting the highest GC3 values when the number of genes was likely insufficient to get a reliable estimate (<1 Mb of coding sequence used for the computations).

Summary statistics of the divergence data

We used the method implemented by Galtier (2016) (Grapes. v1.0) to estimate α, ω_A and ω_{NA} using the approach introduced by Eyre-Walker & Keightley (2009). Briefly, we fitted both a negative Gamma distribution and an exponential distribution to the synonymous and non-

synonymous SFS (the so-called GammaExpo model of Galtier 2016) to model the DFE. Fitted parameters of the DFE were then used to compute the expected D_N/D_S under near neutrality, which was compared to the observed d_N/d_S to estimate the adaptive substitution rate (ω_A) and the proportion of adaptive substitutions (α) [with $\omega_A = \alpha(d_N/d_S)$ & $\omega_{NA} = (1-\alpha)(d_N/d_S)$]. Potential recent changes in population size that affect the SFS were taken into account via the use of nuisance parameters capturing distortions of the SFS optimized alongside the DFE parameters (Eyre-Walker et al. 2006).

Model comparisons

All statistical analyses were performed using R (R core team 2018). We only considered models with a similar number of observations and compared these models based on the Akaike information criterion (AIC). To test the influence of the explanatory variables on π_S and π_N/π_S, we either used simple linear models or Phylogenetic Generalized Least Square (PGLS) models. For the latter, we used the model implemented in the nlme package (Pinheiro et al. 2020). The mitochondrial phylogeny was considered as the species tree taken into account assuming a Brownian correlation structure (using "corBrownian" from the "ape" package; Paradis and Schliep 2019). P-value and AIC were computed using the anova.gls function. The rationale of the phylogenetic control is to account for the shared polymorphisms (part of species similarity that is explained by the inheritance from a common ancestor). The level of polymorphism of a given species is dynamically controlled by drift, mutation rate and natural selection. As soon as two species do not share a significant fraction of their polymorphism, there is no need to account for their phylogenetic proximity because their polymorphisms evolved independently.

R Plots were performed using a series of R packages: cowplot (Wilke 2016), ggplot2 (Wickham 2016), ggpubr (Kassambara 2018), ggrepel (Slowikowski 2019) and ggtree (Yu et al. 2017).

Acknowledgements:

This research was funded by the French ANR (BirdIslandGenomic project, ANR-14-CE02-0002). The analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform services, the genotoul bioinformatics platform Toulouse Midi-Pyrenees (Bioinfo Genotoul) and

679 the Biogenouest BiRD core facility (Université de Nantes). We are grateful to Quentin Rougemont
680 for providing feedback on a previous version of the manuscript, and colleagues from the
681 phylogeny and molecular evolution team at ISEM Montpellier for fruitful discussions throughout
682 this project. The authors wish to thank all the farmers and private nature reserves where
683 fieldworks were conducted. More specifically, we want to thank the Comité Cuvier (68-055 to JF)
684 for approving the handling and sampling protocols of the individuals. We are grateful to the
685 provincial authorities in the Eastern Cape and Free State provinces of South Africa, and Eastern
686 Cape Parks (Alan Southwood, Cathy Dreyer, Gavin Shaw, Sizwe Mkhulise) for granting permission
687 to collect samples (permit numbers RA-190, CRO144/14CR and 01-24158). We would also like to
688 acknowledge R.C.K. Bowie, P.-H. Fabre, E. Kolarova and G. Oatley for help during field work and
689 the Percy FitzPatrick Institute (University of Cape Town) for invaluable support with logistics. The
690 National Museum Bloemfontein is also acknowledged for logistic support to JF during fieldwork
691 in the Free State. We also thank the Reunion National Park for granting us permission to conduct
692 fieldwork in Pas de Bellecombe, Reunion Island, France. We are grateful for the logistic support
693 provided by the field station of Marelongue, funded by the P.O.E., Reunion National Park and
694 OSU Reunion. The Canary government gave permission to perform the sampling work to JCI
695 (permit 01-24158). JCI was funded by the Spanish Ministry of Science, Innovation and Universities
696 (Ref.: PGC2018-097575-B-I00) and by a GRUPIN research grant from the Regional Government of
697 Asturias (Ref.: IDI/2018/000151). BM and MR were partly funded by the Spanish Ministry of
698 Science (grant PGC2018-098897-B-I00), and MR was supported by a doctoral fellowship from the
699 Spanish Ministry of Education, Culture, and Sport (FPU16/05724).

700

References:

Allio R, Schomaker-Bastos A, Romiguier J, Prosdocimi F, Nabholz B, Delsuc F. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol Ecol Resour*. in press

Atkinson QD, Gray RD, Drummond AJ. mtDNA Variation Predicts Population Size in Humans and Reveals a Major Southern Asian Chapter in Human Prehistory. *Molecular Biology and Evolution*. 18 déc 2007;25(2):468-74.

Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, et al. The recombination landscape of the zebra finch Taeniopygia guttata genome. *Genome Res*. 2010/03/31. avr 2010;20(4):485-95.

Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, et al. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *eLife*. 2017;6:e24133.

Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J. Phylogeny and diversification of the largest avian radiation. *Proc Natl Acad Sci U S A.* 2004;101(30):11040.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, et al. PRDM9 Is a Major Determinant of Meiotic Recombination Hotspots in Humans and Mice. *Science*. 2010;327(5967):836.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20.

Bolívar P, Guéguen L, Duret L, Ellegren H, Mugal CF. GC-biased gene conversion conceals the prediction of the nearly neutral theory in avian genomes. *Genome Biology*. 2019;20(1):5.

Bolívar P, Mugal CF, Nater A, Ellegren H. Recombination Rate Variation Modulates Gene Sequence Evolution Mainly via GC-Biased Gene Conversion, Not Hill-Robertson Interference, in an Avian System. *Mol Biol Evol*. 2016;33(1):216-27.

Botero-Castro F, Figuet E, Tilak M-K, Nabholz B, Galtier N. Avian Genomes Revisited: Hidden Genes Uncovered and the Rates versus Traits Paradox in Birds. *Molecular Biology and Evolution*. 2017;34(12):3123-31.

Bourgeois YXC, Delahaie B, Gautier M, Lhuillier E, Malé P-JG, Bertrand JAM, et al. A novel locus on chromosome 1 underlies the evolution of a melanic plumage polymorphism in a wild songbird. *Royal Society Open Science*. 4(2):160805.

Brüniche-Olsen A, Kellner KF, Anderson CJ, DeWoody JA. Runs of homozygosity have utility in mammalian conservation and evolutionary studies. *Conservation Genetics*. 2018;19(6):1295-307.

Brüniche-Olsen A, Kellner KF, DeWoody JA. Island area, body size and demographic history shape genomic diversity in Darwin's finches and related tanagers. *Molecular Ecology*. 2019;28(22):4914-25.

Chapman JA, Ho I, Sunkara S, Luo S, Schroth GP, Rokhsar DS. Meraculous: De Novo Genome Assembly with Short Paired-End Reads. *PLOS ONE*. 2011;6(8):e23501.

Charlesworth B. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics.* 2009;10(3):195-205.

Chen J, Glémin S, Lascoux M. Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. *Molecular Biology and Evolution*. 2017;34(6):1417-28.

Chen J, Glémin S, Lascoux M. From Drift to Draft: How Much Do Beneficial Mutations Actually Contribute to Predictions of Ohta's Slightly Deleterious Model of Molecular Evolution? *Genetics.* 2020 genetics.302869.2019.

Comeron JM, Williford A, Kliman RM. The Hill–Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. *Heredity.* 2008;100(1):19-31.

Corcoran P, Gossmann TI, Barton HJ, Great Tit HapMap Consortium, Slate J, Zeng K. Determinants of the Efficacy of Natural Selection on Coding and Noncoding Variability in Two Passerine Species. *Genome Biol Evol.* 2017;9(11):2987-3007.

Cornetti L, Valente LM, Dunning LT, Quan X, Black RA, Hébert O, et al. The Genome of the "Great Speciator" Provides Insights into Bird Diversification. *Genome Biology and Evolution.* 2015;7(9):2680-91.

Darwin C 1809-1882. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London : John Murray, 1859; 1859.

Díez-del-Molino D, Sánchez-Barreiro F, Barnes I, Gilbert MTP, Dalén L. Quantifying Temporal Genomic Erosion in Endangered Species. *Trends in Ecology & Evolution*. 2018;33(3):176-85.

Doyle JM, Hacking CC, Willoughby JR, Sundaram M, DeWoody JA. Mammalian genetic diversity as a function of habitat, body size, trophic class, and conservation status. *Journal of Mammalogy*. 2015;96(3):564-72.

Duret L, Galtier N. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. *Annu Rev Genom Hum Genet*. 2009;10(1):285-311.

Eddy SR. Accelerated Profile HMM Searches. *PLoS Comput Biol*. 2011;7(10):e1002195-e1002195.

Ellegren H. Evolutionary stasis: the stable chromosomes of birds. *Trends in Ecology & Evolution*. 2010;25(5):283-91.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, et al. The genomic landscape of species divergence in Ficedula flycatchers. *Nature.* 2012;491:756.

Eo SH, Doyle JM, DeWoody JA. Genetic diversity in birds is associated with body mass and habitat type. *Journal of Zoology.* 2011;283(3):220-6.

Eyre-Walker A. Changing effective population size and the McDonald-Kreitman test. *Genetics.* 2002;162(4):2017-24.

Eyre-Walker A, Keightley PD. Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. *Molecular Biology and Evolution*. 2009;26(9):2097-108.

Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. Quantifying the Slightly Deleterious Mutation Model of Molecular Evolution. *Molecular Biology and Evolution.* 2002;19(12):2142-9.

Eyre-Walker A, Woolfit M, Phelps T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics.* 2006;173(2):891-900.

Figuet E, Nabholz B, Bonneau M, Mas Carrio E, Nadachowska-Brzyska K, Ellegren H, et al. Life History Traits, Protein Evolution, and the Nearly Neutral Theory in Amniotes. *Molecular Biology and Evolution.* 2016;33(6):1517-27.

Frankham R, Ballou JD, Briscoe DA. Introduction to Conservation Genetics. Cambridge: Cambridge University Press; 2002.

Galtier N. Adaptive Protein Evolution in Animals and the Effective Population Size Hypothesis. *PLOS Genetics.* 2016;12(1):e1005774.

Gossmann TI, Keightley PD, Eyre-Walker A. The Effect of Variation in the Effective Population Size on the Rate of Adaptive Molecular Evolution in Eukaryotes. *Genome Biology and Evolution.* 2012;4(5):658-67.

Gossmann TI, Song B-H, Windsor AJ, Mitchell-Olds T, Dixon CJ, Kapralov MV, et al. Genome Wide Analyses Reveal Little Evidence for Adaptive Evolution in Many Plant Species. *Molecular Biology and Evolution.* 2010;27(8):1822-32.

Grafen A, Hamilton WD. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London B, Biological Sciences.* 1989;326(1233):119-57.

Grant PR, Grant BR. Evolution of Character Displacement in Darwin's Finches. *Science*. 2006;313(5784):224.

Grant PR, Grant BR. Hybridization increases population variation during adaptive radiation. *Proc Natl Acad Sci USA.* 2019;116(46):23216.

839 Hamabata T, Kinoshita G, Kurita K, Cao P-L, Ito M, Murata J, et al. Endangered island endemic
840 plants have vulnerable genomes. *Communications Biology.* 2019;2(1):244.
841
842 Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, et al. Sequence and
843 comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.
844 *Nature.* 2004;432(7018):695-716.
845
846 Hughes AR, Inouye BD, Johnson MTJ, Underwood N, Vellend M. Ecological consequences of
847 genetic diversity. *Ecology Letters*. 2008;11(6):609-23.
848
849 James JE, Lanfear R, Eyre-Walker A. Molecular Evolutionary Consequences of Island Colonization.
850 *Genome Biology and Evolution.* 2016;8(6):1876-88.
851
852 James J, Eyre-Walker A. Mitochondrial DNA Sequence Diversity in Mammals: a Correlation
853 Between the Effective and Census Population Sizes. *bioRxiv.* 2020;2020.02.28.969592.
854
855 Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, et al. Whole-genome analyses resolve early
856 branches in the tree of life of modern birds. *Science.* 2014;346(6215):1320.
857
858 Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, et al. The importance of
859 the Neutral Theory in 1968 and 50 years on: A response to Kern and Hahn 2018. *Evolution.*
860 2019;73(1):111-4.
861
862 Johnson TH, Stattersfield AJ. A global review of island endemic birds. *Ibis*. 1990;132(2):167-80.
863
864 Kassambara A. « ggplot2 » Based Publication Ready Plots. 2018. https://cran.r-
865 project.org/web/packages/ggpubr/index.html
866
867 Keightley PD, Eyre-Walker A. Joint inference of the distribution of fitness effects of deleterious
868 mutations and population demography based on nucleotide polymorphism frequencies. *Genetics.*
869 2007;177(4):2251-61.
870
871 Keightley PD, Eyre-Walker A. What can we learn about the distribution of fitness effects of new
872 mutations from DNA sequence data? *Philosophical Transactions of the Royal Society B: Biological*
873 *Sciences.* 2010;365(1544):1187-93.

Kern AD, Hahn MW. The Neutral Theory in Light of Natural Selection. *Molecular Biology and Evolution.* 2018;35(6):1366-71.

Kimura M. Evolutionary Rate at the Molecular Level. *Nature.* 1968;217(5129):624-6.

Kimura M. The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press; 1983.

Kimura M, Crow JF. The number of alleles that can be maintained in a finite population. *Genetics*. 1964;49(4):725.

Kondrashov AS. Deleterious mutations and the evolution of sexual reproduction. *Nature*. 1988;336(6198):435-40.

Kreft H, Jetz W, Mutke J, Kier G, Barthlott W. Global diversity of island floras from a macroecological perspective. *Ecology Letters.* 2008;11(2):116-27.

Kutschera VE, Poelstra JW, Botero-Castro F, Dussex N, Gemmell NJ, Hunt GR, et al. Purifying Selection in Corvids Is Less Efficient on Islands. *Molecular Biology and Evolution*. 2019;37(2):469-74.

Laine VN, Gossmann TI, Schachtschneider KM, Garroway CJ, Madsen O, Verhoeven KJF, et al. Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications.* 2016;7(1):10474.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature.* 2015;518(7539):371-5.

Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, et al. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science.* 2016;352(6284):470.

908 Lanfear R, Kokko H, Eyre-Walker A. Population size and the rate of evolution. *Trends in Ecology &*
909 *Evolution.* 2014;29(1):33-41.
910
911 Leroy T, Anselmetti Y, Tilak M-K, Bérard S, Csukonyi L, Gabrielli M, et al. A bird's white-eye view on
912 neosex chromosome evolution. *BioRxiv.* 2019;505610, ver. 4 peer-reviewed and recommended by
913 Peer Community in Evolutionary Biology.
914
915 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.
916 2013;(1303.3997v1).
917
918 Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al.
919 Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human
920 Genome. *Science.* 2009;326(5950):289.
921
922 Loire E, Chiari Y, Bernard A, Cahais V, Romiguier J, Nabholz B, et al. Population genomics of the
923 endangered giant Galápagos tortoise. *Genome Biology.* 2013;14(12):R136.
924
925 Losos J. Lizards in an Evolutionary Tree: Ecology and Adaptive Radiation of Anoles. University of
926 California Press; 2009. (Organisms and Environments).
927
928 Lundberg M, Liedvogel M, Larson K, Sigeman H, Grahn M, Wright A, et al. Genetic differences
929 between willow warbler migratory phenotypes are few and cluster in large haplotype blocks.
930 *Evolution Letters.* 2017;1(3):155-68.
931
932 Lynch M, Gabriel W. Mutation load and the survival of small populations. *Evolution.*
933 1990;44(7):1725-37.
934
935 MacArthur RH, Wilson EO. An equilibrium theory of insular zoogeography. *Evolution.*
936 1963;17(4):373-87.
937
938 Mayr E. Systematics and the Origin of Species, from the Viewpoint of a Zoologist. Harvard
939 University Press; 1999.
940
941 McDonald JH, Kreitman M. Adaptive protein evolution at the Adh locus in Drosophila. *Nature.*
942 1991;351(6328):652-4.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-303.

Moutinho AF, Bataillon T, Dutheil JY. Variation of the adaptive substitution rate between species and within genomes. *Evolutionary Ecology*. 2019, in press

Moyle RG, Filardi CE, Smith CE, Diamond J. Explosive Pleistocene diversification and hemispheric expansion of a "great speciator". *Proc Natl Acad Sci USA.* 2009;106(6):1863.

Mugal CF, Kutschera VE, Botero-Castro F, Wolf JBW, Kaj I. Polymorphism Data Assist Estimation of the Nonsynonymous over Synonymous Fixation Rate Ratio ω for Closely Related Species. *Molecular Biology and Evolution.* 2019;37(1):260-79.

Mugal CF, Arndt PF, Ellegren H. Twisted Signatures of GC-Biased Gene Conversion Embedded in an Evolutionary Stable Karyotype. *Molecular Biology and Evolution.* 2013;30(7):1700-12.

Nam K, Munch K, Mailund T, Nater A, Greminger MP, Krützen M, et al. Evidence that the rate of strong selective sweeps increases with population size in the great apes. *Proc Natl Acad Sci USA*. 2017;114(7):1613.

Nikolaev SI, Montoya-Burgos JI, Popadin K, Parand L, Margulies EH, Antonarakis SE. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. *Proc Natl Acad Sci USA*. 2007;104(51):20443.

Ohta T. Slightly Deleterious Mutant Substitutions in Evolution. *Nature.* 1973;246(5428):96-8.

Ohta T. The Nearly Neutral Theory of Molecular Evolution. *Annu Rev Ecol Syst*. 1992;23(1):263-86.

Oliveros CH, Field DJ, Ksepka DT, Barker FK, Aleixo A, Andersen MJ, et al. Earth history and the passerine superradiation. *Proc Natl Acad Sci USA.* 2019;116(16):7916.

976 Palkopoulou E, Mallick S, Skoglund P, Enk J, Rohland N, Li H, et al. Complete genomes reveal
977 signatures of demographic and genetic declines in the woolly mammoth. *Curr Biol.*
978 2015;25(10):1395-400.
979
980 Paradis E. Analysis of Phylogenetics and Evolution with R. Springer-Verlag New York; 2012. 386 p.
981
982 Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses
983 in R. *Bioinformatics.* 2018;35(3):526-8.
984
985 Parker P, Li B, Li H, Wang J. The genome of Darwin's Finch (Geospiza fortis). *GigaScience.* 2012;
986
987 Parvanov ED, Petkov PM, Paigen K. Prdm9 Controls Activation of Mammalian Recombination
988 Hotspots. *Science.* 2010;327(5967):835.
989
990 Piganeau G, Eyre-Walker A. Evidence for variation in the effective population size of animal
991 mitochondrial DNA. *PLoS One.* 2009;4(2):e4396-e4396.
992
993 Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects
994 Models. 2020. https://CRAN.R-project.org/package=nlme
995
996 Plomion C, Aury J-M, Amselem J, Leroy T, Murat F, Duplessis S, et al. Oak genome reveals facets of
997 long lifespan. *Nature Plants*. 2018;4(7):440-52.
998
999 Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. Accumulation of slightly deleterious
1000 mutations in mitochondrial protein-coding genes of large versus small mammals. *Proc Natl Acad*
1001 *Sci USA.* 2007;104(33):13390.
1002
1003 Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, et al. A comprehensive
1004 phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature.*
1005 2015;526(7574):569-73.
1006
1007 Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, et al. Chromosome-scale
1008 shotgun assembly using an in vitro method for long-range linkage. *Genome Research*.
1009 2016;26(3):342-50.
1010

R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://www.R-project.org

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE v2: Toolkit for the Alignment of Coding Sequences Accounting for Frameshifts and Stop Codons. *Molecular Biology and Evolution.* 2018;35(10):2582-4.

Ricketts TH, Dinerstein E, Boucher T, Brooks TM, Butchart SHM, Hoffmann M, et al. Pinpointing and preventing imminent extinctions. *Proc Natl Acad Sci U S A.* 2005;102(51):18497.

Ripple WJ, Wolf C, Newsome TM, Barnard P, Moomaw WR. World Scientists' Warning of a Climate Emergency. *BioScience.* 2019;70(1):8-12.

Robinson JA, Brown C, Kim BY, Lohmueller KE, Wayne RK. Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology.* 2018;28(21):3487-3494.e4.

Robinson JA, Ortega-Del Vecchyo D, Fan Z, Kim BY, vonHoldt BM, Marsden CD, et al. Genomic Flatlining in the Endangered Island Fox. *Current Biology.* 2016;26(9):1183-9.

Rogers RL, Slatkin M. Excess of genomic defects in a woolly mammoth on Wrangel island. *PLOS Genetics.* 2017;13(3):e1006601.

Rolland J, Romiguier J. Vulnerability and life-history traits correlate with the load of deleterious mutations in fish. *bioRxiv.* 2019;2019.12.29.890038.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, et al. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature.* 2014;515(7526):261-3.

Rousselle M, Laverré A, Figuet E, Nabholz B, Galtier N. Influence of Recombination and GC-biased Gene Conversion on the Adaptive and Nonadaptive Substitution Rate in Mammals versus Birds. *Molecular Biology and Evolution.* 2019;36(3):458-71.

Rousselle M, Mollion M, Nabholz B, Bataillon T, Galtier N. Overestimation of the adaptive substitution rate in fluctuating populations. *Biology Letters.* 2018;14(5):20180055.

Rousselle M, Simion P, Tilak M-K, Figuet E, Nabholz B, Galtier N. Is adaptation limited by mutation? A timescale dependent effect of genetic diversity on the adaptive substitution rate in animals. *Plos Genetics*. 2020,16(4):e1008668

Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nature Genetics.* 2014;46(8):919-25.

Schluter D. The Ecology of Adaptive Radiation OUP Oxford; 2000.

Scornavacca C, Belkhir K, Lopez J, Dernat R, Delsuc F, Douzery EJP, et al. OrthoMaM v10: Scaling-Up Orthologous Coding Sequence and Exon Alignments with More than One Hundred Mammalian Genomes. *Molecular Biology and Evolution.* 2019;36(4):861-2.

She R, Chu JS-C, Uyar B, Wang J, Wang K, Chen N. genBlastG: using BLAST searches to build homologous gene models. *Bioinformatics.* 2011;27(15):2141-3.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, et al. Stable recombination hotspots in birds. *Science.* 2015;350(6263):928.

Slowikowski K. ggrepel: Automatically Position Non-Overlapping Text Labels with « ggplot2 » 2019. https://cran.r-project.org/web/packages/ggrepel/index.html

Steadman DW. Prehistoric Extinctions of Pacific Island Birds: Biodiversity Meets Zooarchaeology. *Science.* 1995;267(5201):1123.

Tataru P, Mollion M, Glémin S, Bataillon T. Inference of Distribution of Fitness Effects and Proportion of Adaptive Substitutions from Polymorphism Data. *Genetics.* 2017;207(3):1103-19.

Valente L, Phillimore AB, Melo M, Warren BH, Clegg SM, Havenstein K, et al. A simple dynamic model explains the diversity of island birds worldwide. *Nature.* 2020, in press

Warren BH, Hagen O, Gerber F, Thébaud C, Paradis E, Conti E. Evaluating alternative explanations for an association of extinction risk and evolutionary uniqueness in multiple insular lineages. *Evolution.* 2018;72(10):2005-24.

Warren BH, Simberloff D, Ricklefs RE, Aguilée R, Condamine FL, Gravel D, et al. Islands as model systems in ecology and evolution: prospects fifty years after MacArthur-Wilson. *Ecology Letters.* 2015;18(2):200-17.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, et al. The genome of a songbird. *Nature.* 2010;464:757.

Weber CC, Nabholz B, Romiguier J, Ellegren H. Kr/Kc but not dN/dS correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol.* 2014;15(12):542-542.

White EP, Ernest SKM, Kerkhoff AJ, Enquist BJ. Relationships between body size and abundance in ecology. *Trends in Ecology & Evolution.* 2007;22(6):323-30.

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. http://ggplot2.org

Wilke C. cowplot: Streamlined Plot Theme and Plot Annotations for « ggplot2 ». 2016. https://CRAN.R-project.org/package=cowplot

Willoughby JR, Sundaram M, Wijayawardena BK, Kimble SJA, Ji Y, Fernandez NB, et al. The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biological Conservation.* 2015;191:495-503.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution.* 2017;8(1):28-36.

Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. 2008;18. http://dx.doi.org/10.1101/gr.074492.107

1114 Zink RM, Vázquez-Miranda H. Species Limits and Phylogenomic Relationships of Darwin's Finches
1115 Remain Unresolved: Potential Consequences of a Volatile Ecological Setting. *Systematic Biology.*
1116 2018;68(2):347-57.