



HAL
open science

Response to Kratochvíl and Rovatsos

Benoit Nabholz, Thibault Leroy

► **To cite this version:**

Benoit Nabholz, Thibault Leroy. Response to Kratochvíl and Rovatsos. *Current Biology - CB*, 2022, 32 (1), pp.R30-R31. 10.1016/j.cub.2021.11.067 . hal-03521111

HAL Id: hal-03521111

<https://hal.science/hal-03521111>

Submitted on 11 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Response to Kratochvíl and Rovatsos

Thibault Leroy^{1,2*} and Benoit Nabholz^{3,4*}

¹ IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071 Beaucouzé, France

² Department of Botany & Biodiversity Research, University of Vienna, Vienna, Austria

³ ISEM, Univ Montpellier, CNRS, IRD, EPHE, Montpellier, France

⁴ Institut Universitaire de France (IUF), Paris, France

*Corresponding authors: thibault.leroy@univie.ac.at and benoit.nabholz@umontpellier.fr

Lead contact : benoit.nabholz@umontpellier.fr

In a recent work¹, we investigated in detail the population genomics of 25 insular or continentally-distributed Passerida species using whole-genome sequences. As part of our investigations, we included simple ratios of non-synonymous to synonymous mutations, a commonly used measure of the strength of purifying selection on coding regions. Concerns regarding the statistical limits associated with the use of this ratio were recently raised by Kratochvíl & Rovatsos². In this commentary, our objective is to continue the debate, not only by introducing the advantages and the drawbacks of this measure, but also by proposing suitable alternatives. Applied to our datasets allowed us to strengthen our initial conclusions, in particular regarding the support for the nearly neutral theory.

The use of levels of nucleotide diversity at non-synonymous (π_N) and synonymous polymorphisms (π_S) as a way to quantify the proportion of deleterious mutations is a fairly old idea (e.g. ref. 3). Rapid decreasing raw sequencing costs over the last decade have made these large scale investigations possible, for instance by Chen and collaborators⁴. We recently investigated the diversity levels in 25 passerine bird species¹. Based on these investigations, we observed marked differences in nucleotide diversities between the 14 insular and 11 continental bird species, suggesting that the smaller area available on islands constrains the upper bound of the effective population sizes (N_e). We also found that the island species exhibit higher π_N/π_S suggesting that the lower N_e experienced by the island species has affected the ability of natural selection to efficiently remove weakly deleterious mutations segregating in the π_N .

In a recent commentary, Kratochvíl and Rovatsos criticized from a statistical point of view the use of the π_N/π_S ratio. One of the crucial concerns Kratochvíl and Rovatsos raised is associated with the high level of correlation observed between the π_N and π_S . We therefore take this opportunity to briefly reintroduce the rationale behind the use of these statistics.

First, as reported by Kratochvíl and Rovatsos, both π_N and π_S are highly correlated, but this correlation does not refute the neutral theory, but rather explains it, since the π_S and π_N are dependent on the same evolutionary forces, i.e. both effective population size Ne and mutation rate μ . Following population genetic theory, the equations are the following: $\pi_S = 2 c Ne \mu$ and $\pi_N = 2 c Ne \mu f$, where c is the ploidy level and f is a function or a constant depending on the evolutionary models considered (see Supplementary Information). Therefore, the strong correlation of π_S and π_N does not come as a surprise, since they share most of their determinants. Under the strict neutral model, f is a constant and π_N and π_S are correlated, independently of Ne^5 . Under the nearly neutral model, the slope value is expected to change depending on Ne , because of the weakly deleterious mutations (f is a function that changes with Ne)⁵.

Kratochvíl and Rovatsos argued that the use of the π_N/π_S ratio is misleading. Given that π_N and π_S are highly correlated, π_N/π_S could be spuriously correlated with π_S , even if the fraction of weakly deleterious mutations does not actually vary across species. In this case, π_N and π_S are strictly linearly correlated, in such a way that the ratio is expected to be equal to a constant ($\pi_N/\pi_S = k$, i.e. strict neutral model). Concretely, when π_N is plotted against π_S , the relationship will be linear with the value of the slope equal to k . As a consequence, the fact that we empirically observed a better fit for a non-linear model can be used as evidence supporting the nearly neutral theory (Fig. 1A; PGLS, ANOVA LTR ratio = 22.7, $p < 0.0001$; AICc PGLS linear model = -410.1, AICc PGLS polynomial model = -429.99). This nonlinear relationship is consistent with the nearly neutral theory and may also explain why Kratochvíl and Rovatsos² find a non-zero intercept, when testing the correlation between total nucleotide diversity with π_N and π_S .

There are other points to consider when using a ratio that are discussed in greater detail in Supplementary Information. We have taken advantage of this correspondence to perform two new analyses that test the robustness of our results.

First, we performed coalescent simulations in order to test the expected relationship between π_N/π_S and π_S assuming a strict neutral model (i.e., with a constant π_N/π_S , see Supplementary Information and Figure S1). Our results produced linear relationships between π_N/π_S and π_S with R^2 of 6% on average considerably lower than the R^2 observed in our datasets ($R^2 = 0.89$). Our simulations confirm the potential issue associated with the use of π_N/π_S over π_S , but without questioning the relationships we initially recovered.

Second, the distributions of fitness effects of mutation are clearly different between mainland and island songbirds (Figure 1B, see Supplementary Information for details). Island species have a higher proportion of weakly and mildly deleterious mutation than mainland species. This is congruent with a reduction of the efficacy of selection in island species, contributing to an increase in frequency of weakly deleterious mutations.

In conclusion, although we agree that the use of the π_N/π_S ratio is not ideal, the way Kratochvíl and Rovatsos have reanalyzed our data is not suited for testing the effect of the nearly-neutral theory. We took the opportunity of this reply to extend our work by generating some results that demonstrate the robustness of our initial analysis. The new analyses we conducted here are again fully in line with the predictions of the nearly neutral theory of molecular evolution. Nevertheless, we agree with Kratochvíl and Rovatsos' call for caution regarding the use of the π_N/π_S ratio, and even more broadly, ratios in biology. They should be avoided, when possible. Alternatively, the potential biases should be carefully assessed, if the use of ratios is still unavoidable and theoretically justified, as in the case of the π_N/π_S ratio. We hope that future research will identify alternative ways to study the dynamics of weakly deleterious polymorphisms.

References

1. Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A.E., Scornavacca, C., Recuerda, M., Fuchs, J., Illera, J.C., De Swardt, D.H., and Blanco, G. (2021). Island songbirds as windows into evolution in small populations. *Curr. Biol.* **31**, 1303–1310.
2. Kratochvíl, L., and Rovatsos, M. Misleading ratios in population genetics and selection detection: little support for nearly neutral theory of evolution in insular birds. *Curr. Biol.*
3. Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**, 263–286.
4. Chen, J., Glémin, S., and Lascoux, M. (2017). Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol. Biol. Evol.* **34**, 1417–1428.
5. Welch, J.J., Eyre-Walker, A., and Waxman, D. (2008). Divergence and Polymorphism Under the Nearly Neutral Theory of Molecular Evolution. *J. Mol. Evol.* **67**, 418–426.

Author Contributions

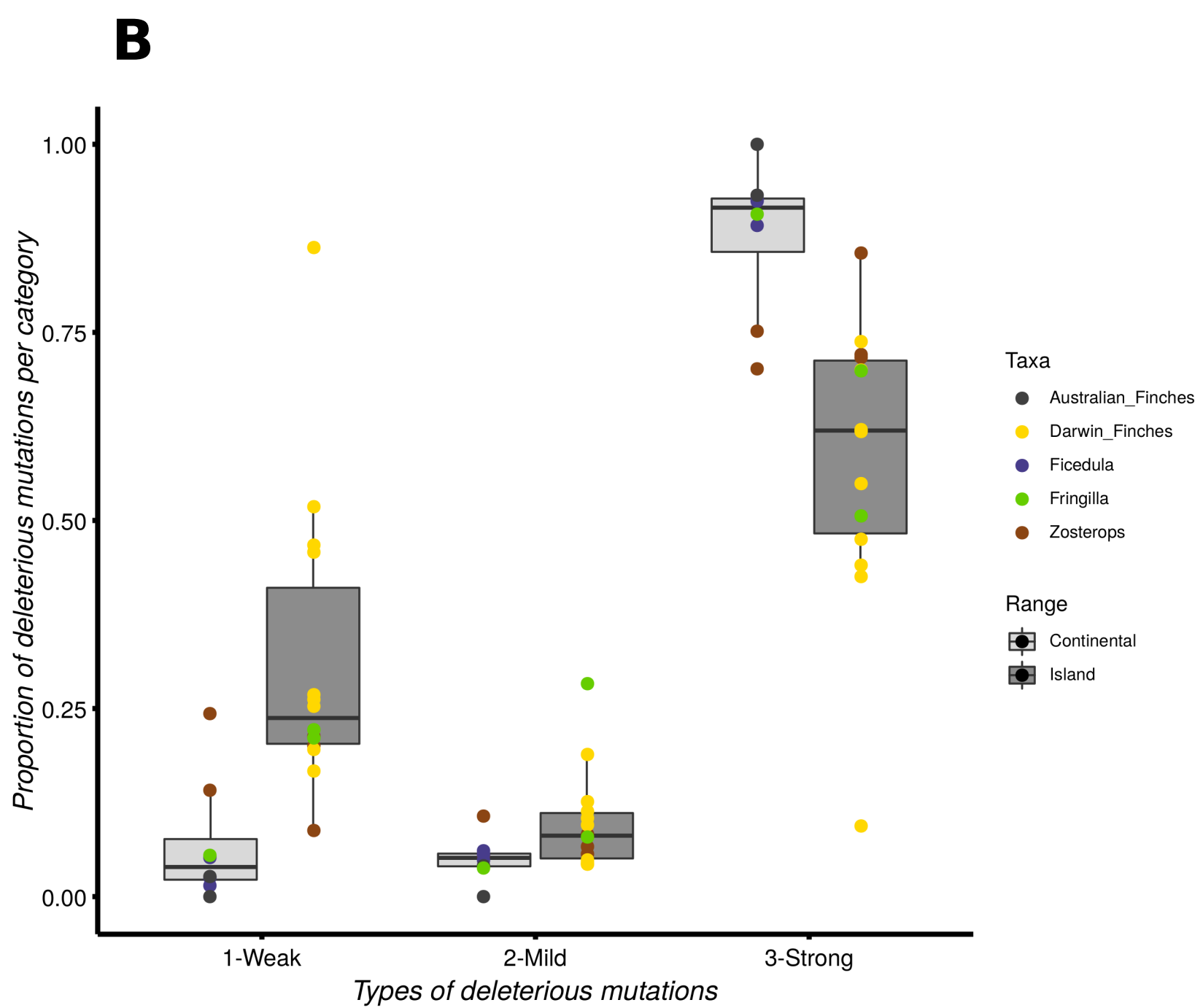
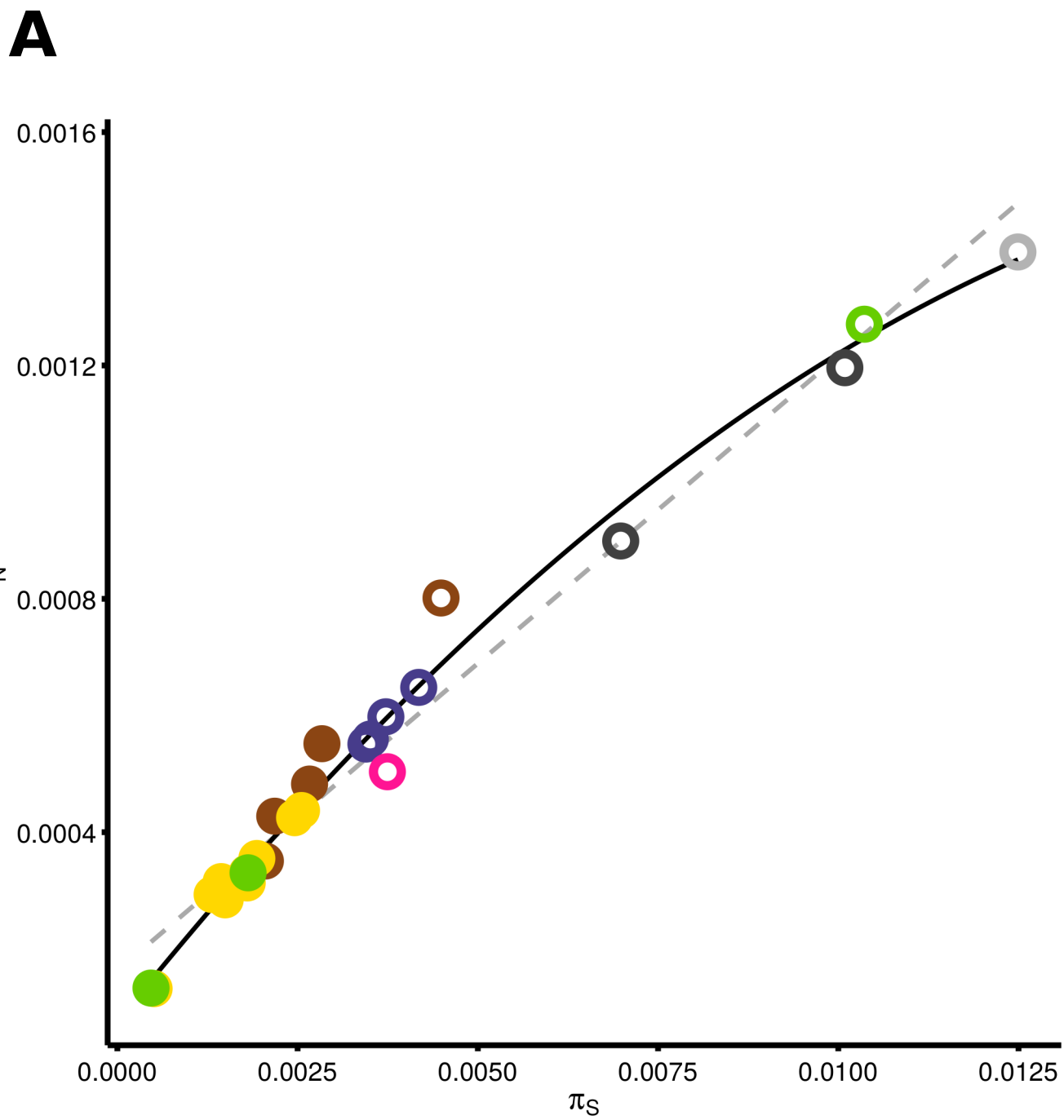
TL and BN, Conceptualization, Statistical analyses and Writing.

Declaration of Interests

The authors declare no competing interests.

Figure Legend:

Figure 1: π_N/π_S ratio-independent support for the results shown by Leroy and collaborators. A) π_N as a function of π_S among all species used in the previous studies. Two best regressions are shown, the best linear (grey dotted line) and the best quadratic regressions (black line). Note that the models tested include a phylogenetic control (not shown here, see main text). B) The distribution of fitness effects of deleterious mutations, as summarized by the proportion of mutations with weak, mild or strong fitness effects among island and mainland songbirds species. Fitness effect is represented as the product $Ne.s$ (i.e. lower than 2, between 2 and 10 and more than 10 for weakly, mildly and strongly deleterious mutations, respectively).



Supplemental Experimental procedures

The pros, cons and alternatives to the use of π_N/π_S ratio in population genomics

Thibault Leroy and Benoit Nabholz

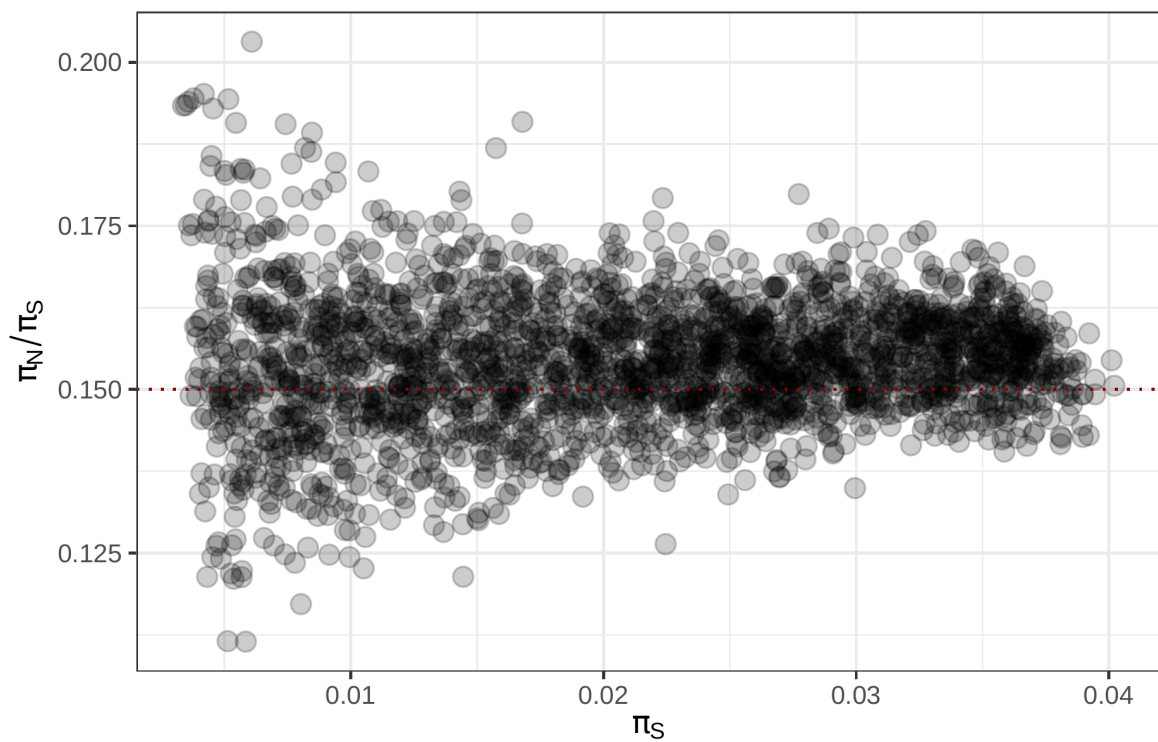


Figure S1: π_N/π_S against π_S for the 2,500 coalescent simulations performed. The red dotted line indicates the expected π_N/π_S of 0.15.

Population genetic theory of the relationship between π_N and π_S

Assuming that $\pi_S = 2 c Ne \mu$ and $\pi_N = 2 c Ne \mu f$, where c is the ploidy level and f is a function that integrates the probability of an allele to segregate at a given frequency, ranging from 0 to 1. f depends on the distribution of the fitness effect (DFE) of mutations. This DFE describes the density probability of the selective advantage of an allele (s). The fixation probability of a non-synonymous allele is dependent on Ne multiplied by s^1 . The nearly-neutral theory predicts that the DFE has a density distribution that includes a large proportion of mutations with a $Ne \cdot s$ close to 0^2 . As a consequence, an increase of Ne will lead to an increase of the fitness effect of weakly deleterious mutations, in such a way that these mutations will be more easily removed from the population by natural selection, therefore decreasing the π_N relative to the π_S (see Welch et al. 2008³ for a detailed theoretical development). Moreover, under both the nearly-neutral and a strict neutral model, there is a fraction of non-synonymous mutations that are effectively neutral and, therefore, behave exactly as synonymous mutations. So, in a nutshell, the two metrics are expected to be strongly correlated. Under the strict neutral model, $\pi_N = 2 c Ne \mu k$, where k is the proportion of neutral mutations independent of Ne^4 . In this case, π_N and π_S are correlated, independently of Ne^3 . Under the nearly neutral model, the slope value is expected to change depending on Ne , because of the weakly deleterious mutations.

Coalescent simulations

We performed coalescent simulations in order to test the expected relationship between π_N/π_S and π_S assuming a strict neutral model (i.e., with a constant π_N/π_S). Our simulation framework was designed in order to use the same range of segregating sites than in our empirical datasets. Simulations were performed assigning population sizes randomly drawn from a uniform distribution (min = 10,000 and max = 100,000). We used a sample of 15 chromosomes with 200,000 nucleotides, a third of these sites with a mutation rate of $1e-7$, $1.5e-8$ otherwise. These categories will be considered as synonymous and non-synonymous mutations, respectively (leading to a constant π_N/π_S of 0.15). Recombination rate was set to $1e-8$ recombination/site/g. Simulations were performed 100 times using msprime (v. 1.0.1;⁵), each including 25 populations. We obtained between ~1,200 and ~13,000 segregating sites per simulation, a number close to what we observed in the empirical dataset. Our results are consistent with a negative correlation between π_N/π_S and π_S (64% of simulations) and an increase in variance of the π_N/π_S toward low values of population size (figure S1). However, as compared to the

empirical relationship, this bias seems weak, producing linear relationships between $\log(\pi_N/\pi_S)$ and $\log(\pi_S)$ with R^2 of 6% on average (max = 28%) considerably lower than the R^2 observed in our datasets ($R^2 = 0.89$).

Distributions of the fitness effect of mutations

Second, we computed the fitness effect of non-synonymous mutations for each species and summarized it as the proportion of mutations falling into three categories, namely weakly deleterious ($0 < Ne.s < 2$), mildly deleterious ($2 < Ne.s < 10$) and strongly deleterious ($Ne.s > 10$). These proportions were obtained from the distribution of fitness effect estimated using the synonymous and non-synonymous site frequency spectrum and the model GammaExpo⁶ focusing on the deleterious mutations, see also "Summary statistics of the divergence data" section of the Method details in Leroy et al.⁷. We recovered markedly different distributions between mainland and island songbirds (Figure 1B). Island species have a higher proportion of weakly and mildly deleterious mutations than mainland species. This is congruent with a reduction of the efficacy of selection in island species, contributing to an increase in the frequency of weakly deleterious mutations.

Acknowledgments:

This research was funded by the French ANR (BirdIslandGenomic project, ANR-14-CE02-0002 and NeGA project, ANR-20-CE02-0008). The analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform services. We would like to thank Lukáš Kratochvíl and Michail Rovatsos for their feedback and time they spent on our study and allow us to consider more deeply the statistical problem inherent with the use of ratio in correlation analyses. Such post-publication feedback is important and should be promoted within the scientific community. In addition, we want to thank the three reviewers who contributed to enhance the content. This is an ISEM publication n° **ISEM 2021-290**.

Author Contributions

TL and BN, Conceptualization, Statistical analyses and Writing.

Supplemental References

- S1. Kimura, M. (1962). On the Probability of Fixation of Mutant Genes in a Population. *Genetics* 47, 713–719.
- S2. Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* 23, 263–286.
- S3. Welch, J.J., Eyre-Walker, A., and Waxman, D. (2008). Divergence and Polymorphism Under the Nearly Neutral Theory of Molecular Evolution. *J. Mol. Evol.* 67, 418–426.
- S4. Kimura, M. (1983). *The neutral theory of molecular evolution.* (Cambridge University Press).
- S5. Kelleher, J., Etheridge, A.M., and McVean, G. (2016). Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.* 12, e1004842.
- S6. Galtier, N. (2016). Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genet.* 12, e1005774.
- S7. Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A.E., Scornavacca, C., Recuerda, M., Fuchs, J., Illera, J.C., De Swardt, D.H., and Blanco, G. (2021). Island songbirds as windows into evolution in small populations. *Curr. Biol.* 31, 1303–1310.