I want to apologize for the delay with this decision; it was very difficult to find reviewers. The preprint has now been read by four reviewers and myself. All of us found merit in the study but there are several issues that should be addressed before publication.

By far the biggest issue raised is that the authors used different species for the island vs. mainland. This means that species itself is confounded with island vs. mainland and it seems that the phylogenetic distribution is skewed as well. This issue needs to be addressed up front.

Furthermore, several methodological questions need to be clarified such as sampling strategy and choice of "control genes". Finally, the reviewers had a number of minor comments will improve readability that should be addressed as well. In addition, I had some comments that follow below. I look forward to reading a revised version of this manuscript.

Best,

Emma Berdan

Dear Dr Berdan,

We are very grateful to the four reviewers and you as editor for the useful comments and suggestions that have led to an improved manuscript.

Please find attached our response to each comment (in blue and starting by AU:). The revised manuscript has been submitted to biorxiv (https://doi.org/10.1101/2021.11.21.469450). We also made available the revised manuscript with the changes highlighted as a separate document.

The revised manuscript contains the following major changes :

1. The reviewer comments trigger some additional analyses and simulations in order to quantify the mode of evolution of TLR and BD genes. Particularly, we made the assumption that TLR and BD polymorphism patterns were shaped mostly by purifying selection, which may not be the case. We now test the effect of positive and balancing selection on expected variation of Ps and Pn/Ps among island and mainland birds. These additional analyses demonstrate that, although BD and TLR do not evolve under the same selection regime as control genes, positive selection or balancing selection are unlikely to explain the variation of Ps and Pn/Ps between island and mainland population. We have added a section "*Population genetics of BD and TLR immune genes*" where we present these results.
2. We have redone all the simulations related to the MHC analyses. These new simulations are closers to the observed Pn/Ps of the mainland species as suggested by the reviewer #2.

3. We now clarified the theory behind our hypotheses, by explaining why a species effect is separated from an island vs. mainland effect. Please, see our reply detailed below.
4. Population structure analyses (PCA) allowed us to detect problematic individuals (three potentially contaminated and one mislabelled individuals, 3 *Ploceus* and 1 *Cyanistes*). These individuals were therefore excluded from the revised manuscript.
5. We performed more in depth analyses to identify potential hidden paralogy in the immunity genes. This led us to exclude a few genes (i.e. TLR21) from the revised manuscript. The points 4 and 5 do not lead to changes in significance or effects in the statistical analyses.

First, one of the main concerns, raised by you and Reviewer #2, was about the use of different bird species from the islands and the mainland. We now addressed this issue in the revised version of the ms (lines 143-145). We would like to point out that one of the aims of our paper was to determine whether changes in genetic diversity of immunity genes was part of the insularity syndrome, which refers to parallel changes/evolution in island species. Thus, we do need island species to be compared to mainland ones. Moreover, island populations are relatively difficult to find beyond the recent human-induced colonization.

Second, we would like to point out that we have at least **seven independent island colonizations** ensuring a repeated and testable analysis. We do understand your concern about a potential bias in the phylogenetic distribution. That is why we took the **phylogeny into account in two ways** : (i) by a taxonomical effect in the linear mixed model and (ii) by the Phylogenetic Generalized Least Squares model that takes explicitly the phylogenetic relationships into account (Freckleton et al., 2015; Grafen, 1989). In addition, our statistics are based on polymorphism. The level of polymorphism of a given species is dynamically controlled by drift, mutation rate and natural selection. As soon as two species stop sharing a significant fraction of their polymorphism (i.e. most of their genes are monophyletic), then the level of Pn/Ps is determined by the evolutionary forces acting in this species, irrespective of the Pn/Ps level in their ancestor.

Third, reviewer #2 is concerned that a “species effect” could not be separated from a “island effect”. We are confident that this is not an issue because the Pn/Ps of the control genes is only dependent on the variation of *Ne*, as selection is expected to be similar among species. These control genes are randomly selected and their functions and, the selection pressure acting on them, are likely to be similar in those environments. We also tested that their Pn/Ps reflect the Pn/Ps of the whole transcriptome (see Supp Fig S2). The Pn/Ps of the control genes is determined by the distribution of fitness effect (DFE) of mutation. The DFE depends on effective population size (*Ne*) and selection coefficient of mutation (*s*). **The only possibility for the species to be a confounding factor of the “island effect” is that the distribution of *s* varies among bird species independently of Ne**. There is no evidence for that since birds have a conserved genomic architecture in terms of chromosome numbers, genome size and recombination rate (Backström et al., 2008; Ellegren, 2010; Singhal et al., 2015; Vijay et al., 2017). The best predictor of Pn/Ps variation among species is *Ne* ; Pn/Ps is

correlated to range size or direct census size estimation (Leroy et al., 2021). If the DFE was different among birds because of *s,* then the correlation between Pn/Ps and *Ne* will not be as strong as the one reported in Leroy et al. (2021). In this context, it is safe to assume that **the variation of Pn/Ps of control genes are only a consequence of a variation in *Ne* among species**. In contrast, there is a possibility that the distribution of *s* for the immunity genes changes between islands and mainland. In sum, the species effect is only due to an effect of *Ne* and not of *s*. We now modified the conceptual diagram to include this information (Figure R1).

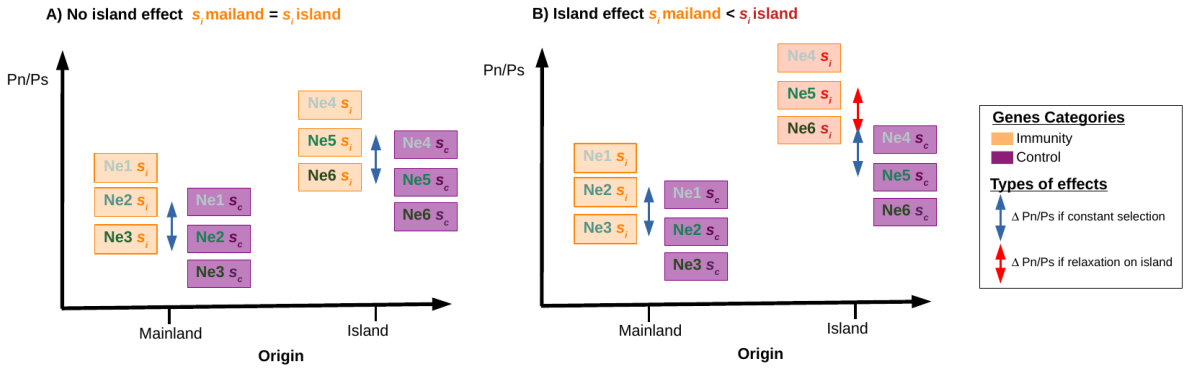

Figure R1: Conceptual diagram showing the expected results under the hypothesis of A) non relaxation in the selection pressure and B) relaxation in the selection pressure on the immunity genes in island species. The diagram includes variation of Ne among species (on average lower for island species) with Ne*i* being the Ne of species *i* (*i* = 1 to 6). The selection coefficient of mutation (s) for control genes and immunity genes are shared among species except for a potential difference between island and mainland for the immunity genes.

Please find below our replies to every reviewer's comments. We hope that you will find that this revised manuscript has been improved, and we greatly thank to all the constructive comments and suggestions.

Best regards,

Mathilde Barthe on the behalf of all co-authors.

Comments:

Species names should always be in italics.

Lines 128-150 - This section is somewhat convoluted and it is hard to understand what your predictions are.

AU: We rephrase this section to make it, hopefully, clearer.

Figure 2 - The same color (purple) should not refer to different groupings in the two figures (A and B). This is confusing for the reader.

AU: We now use a different color for control and simulation grouping in the two panels.

I found table 3 very difficult to read. Please re-do the formatting.

AU: Table 3 has been reformatted.

Reviewer #1 : anonymous reviewer, 14 Feb 2022 12:49

Preprint “Reduction in population size and not a shift in parasite community affect evolution of immune genes in island birds” by Barthe et al. challenges the hypothesis of relaxed selection pressure on immune genes due to reduced diversity of pathogens. Instead, they come to the conclusion that the apparent lack of genomic signatures of selection is, in other than MHC class II genes, due to the decrease in population size and thus stronger genetic drift. The authors study 20 island and 14 mainland bird species. They compare the synonymous and non-synonymous diversity of various innate and acquired immune genes and use a set of non-immune function genes as a control group.

In general, the preprint is well and clearly written. The analyses and conclusions are justified and well explained. The introduction goes through relevant theory and introduces the hypotheses. A minor detail: both immune gene and immunity gene are used in the text, I suggest using only "immune gene" throughout the paper.

AU: Thank you very much for your positive comment. We made the changes according to the reviewer.

RE: In the preprint, the immune genes are handled in three different classes instead of individual examination. I wonder if one would expect all the immune genes to respond in a similar way under relaxed pressure? Would you expect some genes to be more sensitive to selection? Is it possible that some interesting aspects are lost when only the means of the three classes of genes are included in the analyses? Did the authors look at the variation in Pn/Ps among the genes within a species? It is a challenging task to compare mainland and island populations that have differences both in population size, demographic history, time of speciation event and parasite load. Nevertheless, the authors have done a thorough job trying to control for different sources of variation.

AU: We agree that we may lose interesting variation however, a single gene typically has a handful of mutations segregating in the polymorphism preventing accurate estimation of the Pn/Ps ratio. Therefore, we must consider several genes per category.

RE: Another general note: How did the authors take into account gene families? E.g. MHC genes often have multiple repeats in an individual genome and it may be very demanding to pick only one of those genes while alleles are likely to be also recombining into different loci and the copy number changes even between individuals. How many MHC genes were found per species and were there any problems in selecting the correct counterparts from different species?

AU: For old paralogues (i.e, gene duplication ancestral to species divergence), we infer orthology using the approach of the profile hidden Markov models (program HMMER, see "Homology identification" in Supplementary Text). This method has been used successfully to identify orthology in the ORTHOMAM V10 database (Scornavacca et al., 2019).

The reviewer comment stimulates an extra analysis to check for hidden gene duplication. We compute the statistic Fis = 1-H0/He where H0 is the average number of heterozygous individual observed (H0 = #heterozygous / n ; where n is the sample size) and He is the expected number of heterozygous individuals at Hardy-Weinberg (HW) equilibrium (He = (n/(n-1) 2 * p * (1-p))*n where n is the sample size and p the allele frequency of a randomly chosen allele). Fit varies between -1 and 1 with positive value representing excess of homozygous individuals and negative value representing excess of heterozygous individuals compared to the HW expectation. Gene with high value of nucleotide diversity (Pi) and negative value of Fis could represent potential cases where hidden paralogous sequences have not been separated and where all the individuals present heterozygous sites in the positions where a substitution occurred between the paralogous copies. In the figure R2 below, we plotted Fis against Pi and several TLR21 genes appear problematic (Pi > 0.01 and Fis < -0.5). We decided to exclude those genes from the analyses and this did not impact our results.

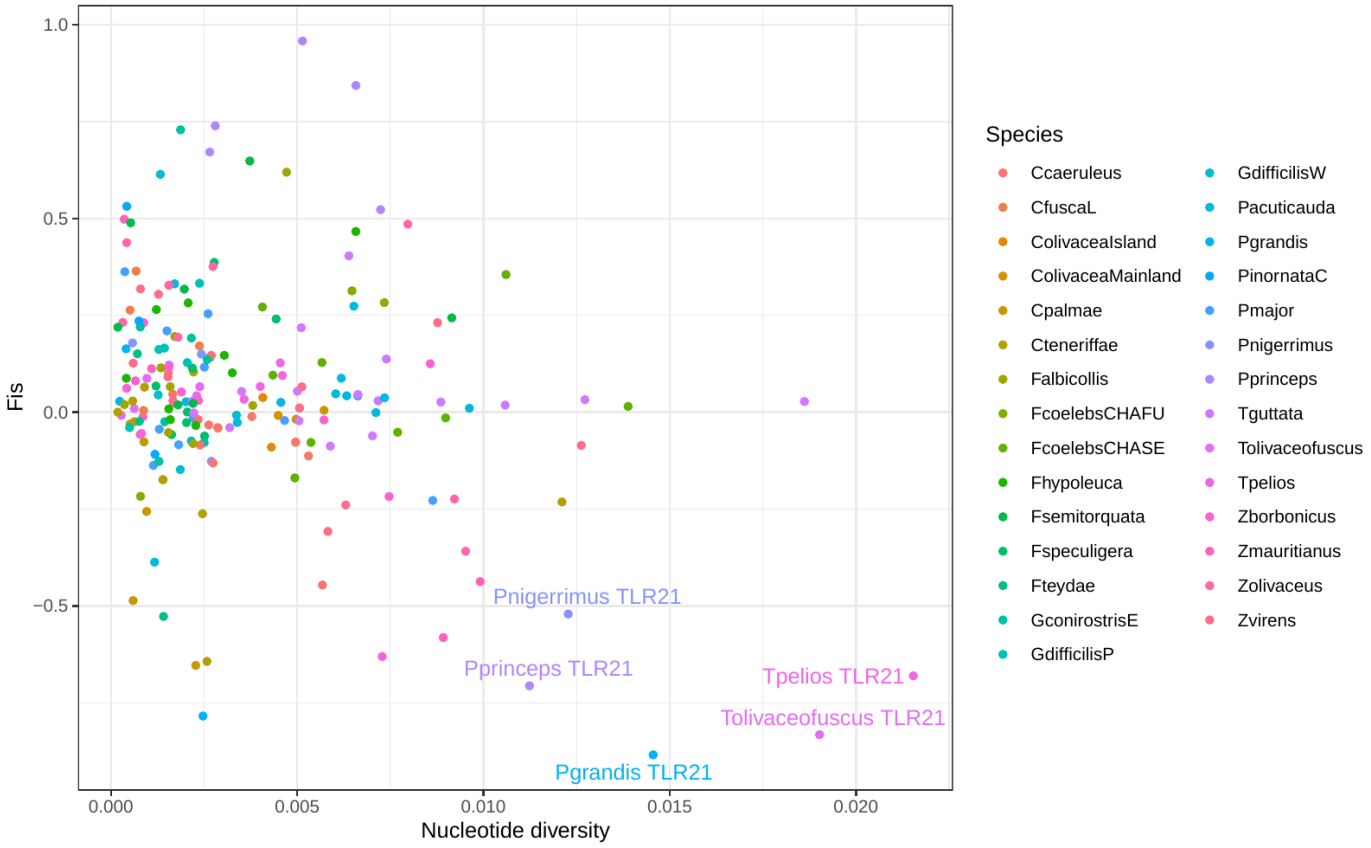

Figure R2 : Deviation of heterozygosity from Hardy-Weinberg proportion (Fis) according to nucleotide diversity (Pi). A negative Fis indicates more heterozygous individuals than expected.

The MHC genes are more difficult to analyze. Indeed, heterozygosity could be comparable to divergence under balancing selection. This makes the identification of orthologs very difficult. In this study, we used the MHC gene I and II copy annotated in Minias et al. (2018). Then, we identify putative homologous sequences only based on sequence homology. This led to the identification of a variable number of genes among species (from 1 to 10 genes for MHC class I and MHC class II). We checked the sequence similarity for the 10 copies of the MHC class II in *F. albicollis* and the 7 copies of the MHC classI genes in *C. caeruleus* using cd-hit (Fu et al., 2012). For MHC class II, sequence divergences are always higher than 15% indicating that reads will likely be correctly assigned to their corresponding gene copy. For MHC class I, sequence identities could be as high as 95%. In this case, we rely on the fact that the reads from very similar paralogous copies will not be confidently assigned to a gene copy sequence by the mapping software. This will lead to a low mapping score quality and are likely to be discarded during the genotype calling procedure. For example, 3 out of 7 genes of the *Cyanistes* MHC class I genes could not have been correctly genotyped and are missing from our final dataset. We have added a section on this potential issue in the revised manuscript and emphasized that the MHC results should be taken with caution.

RE: Detailed comments.

126-127: In this context, the ratio Pn/Ps is typically interpreted as the result of a change in natural selection.→ Why would Pn/Ps indicate change in natural selection, and not a long-lasting evolutionary state of natural selection?

AU: We agree that the sentence can be misleading as the ratio Pn/Ps should reflect both the recent and long-lasting state of natural selection. We rephrased the sentence : “the ratio Pn/Ps is typically interpreted as the result of natural selection”.

RE: 129-132: However, the fixation probability depends on the product Ne*s, and a variation in Ne is also expected to impact the efficacy of selection and thus the ratio Pn/Ps across the entire transcriptome, particularly in the presence of slightly deleterious mutations (Ohta 1992; Charlesworth and Eyre-Walker 2008; Loire et al. 2013; Leroy et al. 2021). → When talking about immune gene diversity, advantageous mutations are also important and likely lost due to drift.

AU: This comment (and others below) stimulated additional analyses regarding the impact of positive, as well as balancing selection, on the relationship between Pn/Ps, Ps and Ne. Indeed, we originally assumed that the polymorphism of TLRs and BDs genes was shaped by purifying selection disregarding positive and balancing selection. We now consider all types of selection and evaluate if the polymorphism of TLRs and BDs could be influenced by positive or balancing selection (please see the section “*Population genetics of BD and TLR Immune*”).

Regarding the specific effect of adaptive selection, genes affected by selective sweeps will experience a lower Ps and higher Pn/Ps compared to genes evolving strictly under purifying selection (Castellano et al., 2018; Chen et al., 2020). This is confirmed by our simulation analyses (see main text for the detail of the simulation and Figure R3). In the presence of recurrent positive selection, Ps is stable or decreases with Ne (Figure R3). Moreover, TLRs and BDs show both higher Ps and higher Pn/Ps than control genes, and island birds have a lower Ps than mainland birds. All these results do not support that positive selection has had a major impact in shaping the polymorphism patterns of TLRs and BDs genes. Importantly, it should be noted that this result does not exclude that positive selection have impacted immune genes evolution in the past (Velová et al., 2018) but positive selection is not frequent enough to significantly impact the current polymorphism patterns.

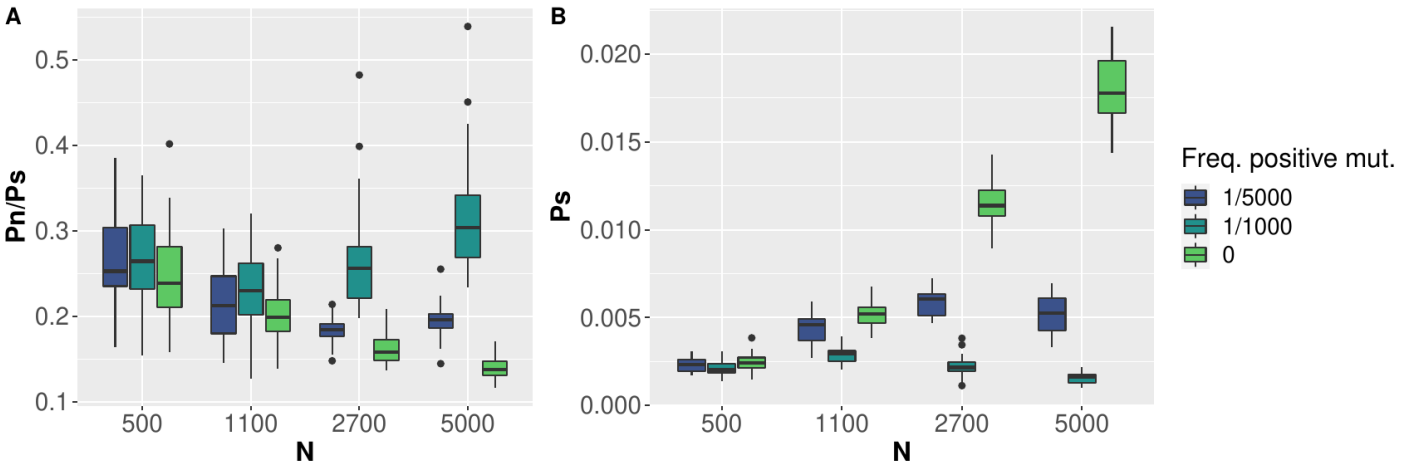

Figure R3: A) PnPs and B) Ps as a function of population size (N). Results of simulation using SLiM where mutation rate, recombination rate and selection coefficient have been multiplied by 100 in order to divide N by 100 in order to speed up the simulation. The relative frequency of positive selection compare to negative selection is set to 0, 1/1000 and 1/5000

RE: 133: Therefore, we predict a significant effect of drift on island species leading to a genome-wide reduction in genetic diversity and efficacy of selection, as reported by previous studies. In addition, due to their lower population sizes, island birds compared to continental species exhibit a genome-wide reduction in genetic diversity and efficacy of selection (Kutschera et al. 2020; Leroy et al. 2021).

→ Here the authors seem to state the same point twice: the effect of drift and population size. Population size is directly linked to the strength of genetic drift, so I suggest either combining these sentences or modifying to make more clear what is the difference that the authors mean here.

AU: We rephrased the sentence as "In addition, due to their lower population sizes, island birds compared to continental species exhibit a genome-wide reduction in genetic diversity and efficacy of selection (Kutschera et al., 2020; Leroy et al., 2021)." line 135 to 138.

RE: 141-143: we randomly selected protein-coding genes (i.e., control genes) implied in various biological functions (Fijarczyk et al. 2016; Leroy et al. 2021).

→ Is "implied" the correct word here?

AU: We changed "implied" by "involved" line 142.

RE: 145-147: More specifically, for genes under purifying selection, non-synonymous weekly deleterious mutations, normally eliminated under strong selection, would be maintained, leading to an increase of genetic diversity.

→ I wonder if it is possible to draw such a direct conclusion about the fate of the alleles of genes under purifying genes.

AU: The assumption that there is a class of mutation that will behave as neutral mutation in small populations and will be selected against in large populations is at the heart of the nearly-neutral theory (Ohta, 1992). It is also the interpretation behind the increase of Pn/Ps (and Dn/Ds) with Ne (Figuet et al., 2016; Leroy et al., 2021; Romiguier et al., 2013).

RE: Anyhow, I assume that for many immune genes, the negative selection pressure also rises from autoimmunity reactivity, so the parasite pressure is not the only driving force.

AU: This is possible, but we expect autoimmunity to be similar between island and mainland.

RE: Figure 2: This figure is good for visualizing the main hypotheses of the manuscript. However, wouldn't we expect genes evolving under purifying selection to follow a similar low ratio of Pn/Ps as random protein-coding genes that often also evolve under purifying selection?

AU: This is a good point that we now discuss more thoroughly in the section "*Population genetics of BD and TLR Immune genes*" line 401. Using simulations and comparing the Ps and Pn/Ps between TLRs, BDs and control genes, we were able to exclude that the higher Pn/Ps of immune genes is caused by positive selection or balancing selection. We also excluded that the higher Pn/Ps is caused by an effect of the lower sample size of the immune genes. The only explanation we can think of is a relaxed selection of immune genes compared to control genes. The exact value of Pn/Ps could vary according to the exact parameter of the distribution of the fitness effect of mutations. This distribution could change among gene categories according to the function of the genes and other aspects such as their expression levels (Drummond and Wilke, 2008).

RE: Table 1: It is not clear which data is from which reference. Also the settings of the table could be more clear, especially for the two last columns.

AU: We now modified the table to make it clearer.

RE: Table 1: Correct "Ploceus cuculatus" to "Ploceus cucullatus".

AU: Corrected.

RE: 268: The island effective population size of 110,000 sounds very large assuming that the census size is much larger. Do you have estimates of the census sizes of the study populations and gene flow between the different populations? I assume gene flow between different island populations and/or between island and mainland populations can have a large effect on both the selection pressures and amount of drift/effective population size.

AU: The exact significance of effective population is an active research question and is beyond the scope of our work (Buffalo, 2021; Corbett-Detig et al., 2015; Galtier and Rousselle, 2020; Leffler et al., 2012; Romiguier et al., 2014). There are many potential explanations for an apparent discrepancy between census size and effective population size. For example, current census size could be lower than effective population size for many island species because of very recent population decline due to human disturbances. Alternatively, continental populations could have been more impacted by the last glaciation. Moreover, the census size of island species could be relatively high for some island species such as the Reunion Grey White-eye (N = 100,000-500,000 ; IUCN Red List).

Gene flow among islands is possible, although population structure is often quite strong (e.g. in the chaffinch from Canaries; Recuerda et al., 2021). We could not find evidence for gene flow between island and mainland in the literature for our species. This is expected as species are different between island and mainland and the islands are often very remote and isolated (e.g., Mascarene and Galapagos archipelago).

RE: 283-285: Unlike for all other species (e.g. Fringilla coelebs, Figure S3), synonymous polymorphism level was very dependent on the number missing data tolerated in P. trochilus alignments (Figure S3).
→ Is there something missing from this sentence?

AU: We rephrased the sentence as : "... synonymous polymorphism level was correlated to the amount missing data in *P. trochilus* alignments (Figure S7)." lines 336 and 337.

RE: 298-301: A linear mixed model was performed, using the Pn/Ps ratio as dependant variable and, as explanatory variables, the mainland or insular origin of species as well as the category of genes (packages lme4 and lmerTest (Bates et al. 2012; Kuznetsova et al. 2017))
→ In general, the regression models in the preprint seem to be well performed and justified. However, the Pn/Ps ratio does not immediately seem like a normally distributed variable. How well does the linear model fit your data with Ps/Pn bound to be positive (and relatively close to zero)? Did you try to fit other error structures and check for the distribution of residuals in the linear model?

AU: Thanks to your suggestion we used a generalized linear mixed model (using the function glmer of the package lme4) with the family "Gamma(link="log")" (Table S15 to S22). The residuals are normally distributed and this did not change our results.

RE: 337: an average of 3.3 millions paired-ends reads per individual was generated
→ "millions" → "million"

AU: Corrected at line 390.

RE: Table 2 & Table 4: The p-values of ANOVA test between simpler models are not reported if a more complex model is significant.

→ I suggest changing this sentence into: “The p-values of ANOVA test between simpler models are not reported if a more complex model explains a larger proportion of the variance”.

AU: Corrected for Tables 2 and 4.

RE: In addition, please clarify in the tables what are models 1-4 in the ANOVA part of the table.

AU: Corrected.

RE: 486-487: “MHC class I genes are primarily involved in the recognition intracellular pathogens (Kappes and Strominger 1988)”
→ Add “of” to: “in the recognition of intracellular pathogens”

AU: Corrected at line 597.

RE: Supplementary material. Figure S2: Correlation between Pn/Ps (a) and Ps (b) calculated on the control genes in this study's dataset and those calculated by (Leroy et al. 2021) Leroy et al. (2021).
→ Remove the extra reference.

AU: Corrected.

RE: Figure S4: Boxplot of Pn/Ps according to population size for simulated sequences under overdomiance with SliM.
→ Correct the word “overdominance”.

AU: Corrected.

RE: Table S3: This is missing, there was no Table S3 in the supplementary material. And I could not find any list of genes that were studied. This was referred as Table S3 in the results on line 341.

AU: The table S3 was added. You can find the list of genes studied in the supplementary folder entitled “**List of genes”.** We corrected referred Tables as “Table S4”

Reviewer #2 : anonymous reviewer, 23 Jan 2022 18:36

This manuscript tests the hypothesis that immune genes of birds living on islands experience reduced selective pressure from parasites, compared to the immune genes of birds living on the mainland. The authors analyze 35 species and compare genetic diversity in synonymous

(Pn) and nonsynonymous sites (Ps) between species from mainland vs. island in control (randomly selected) genes and genes linked to immune functions. To test the hypothesis authors calculate Pn/Ps (or πN/πS), a statistic meant to detect selection within a population, where Pn, in brief, is influenced by effective population size and selection (and mutation rate), and Ps is influenced by effective population size (and mutation rate). The authors want to distinguish between the effects of population size differences (smaller on the island), and selection, (expected to be also smaller on the island). Although the idea is quite simple and attractive, I found major problems with this approach and its implementation. The major problem is the use of a different set of species for scoring Pn/Ps on the mainland and island. I'm afraid this leads to uninterpretable results. I think it is not possible to distinguish with this system effects of relaxed selection caused by reduced Ne, and reduced selective pressure from parasites because both of these properties can vary significantly between species.

Major comments:

1) The statistical test considers two measures, one of them is delta Pn/Ps which compares the difference in selection between immune and control genes, and another is a difference between Pn/Ps between mainland and island species. Whereas delta Pn/Ps is calculated for the same species, the difference between mainland and island is calculated between groups of different species. Delta Pn/Ps will correspond to the average difference in selective pressure between control and immune genes, but the difference mainland-island here corresponds not to the average difference in Ne/selection between origins, but the average difference in Ne/selection between several species. In other words, one can't untangle the effects of origin from the effects of species. I think the only potential solution to test this hypothesis would be to use simulations to find parameters of Ne and selection which correspond to genetic diversity observed in immune and control genes, separately in mainland and island populations, and then compare if estimated selection coefficients are the same or lower in island compared to the mainland.

AU: First, we would like to thank the reviewer for her/his suggestions and comments that led to an improved manuscript.

The first concern raised, also expressed by the editor, was about the "species effect" that potentially could not be separated from the "island effect". We are confident that this is not an issue because the Pn/Ps of the control genes is only dependent on the variation of *Ne* as selection is expected to be similar among species. These control genes are randomly selected and their functions and, the selection pressure acting on them, are likely to be similar in all environments. We also tested that their Pn/Ps reflect the Pn/Ps of the whole transcriptome demonstrating that they represent a good approximation of the genome-wide selection pressure (see Supp Fig S6). The Pn/Ps of the control genes is determined by the distribution of fitness effect (DFE) of mutation. The DFE depends on effective population size (*Ne*) and selection coefficient of mutation (*s*). **The only possibility for the species to be a confounding factor of the "island effect" is that the distribution of *s* varies among bird**

species independently of Ne. **There is no evidence for that** since (i) the control gene represents the genome-wide selection pressure and (ii) birds have a conserved genomic architecture in terms of the number of chromosomes, genome size and recombination rate and, therefore, all the factors that could influence the efficacy of selection are conserved across species (Backström et al., 2008; Ellegren, 2010; Singhal et al., 2015; Vijay et al., 2017).The best predictor of Pn/Ps variation among species is *Ne* and Pn/Ps is correlated to range size or direct census size estimation (Leroy et al., 2021). If the DFE was different among birds because of *s* then the correlation between Pn/Ps and Ne will not be as good as the one reported in Leroy et al. (2021). Therefore, it is safe to assume that **the variation of Pn/Ps of control genes are only a consequence of a variation in *Ne* among species**. In contrast, there is a possibility that the distribution of *s* for the immunity genes changes between island and mainland. Therefore, the species effect is only an effect of Ne and not a *s*. We modified the conceptual diagram to include this information (Figure R1).

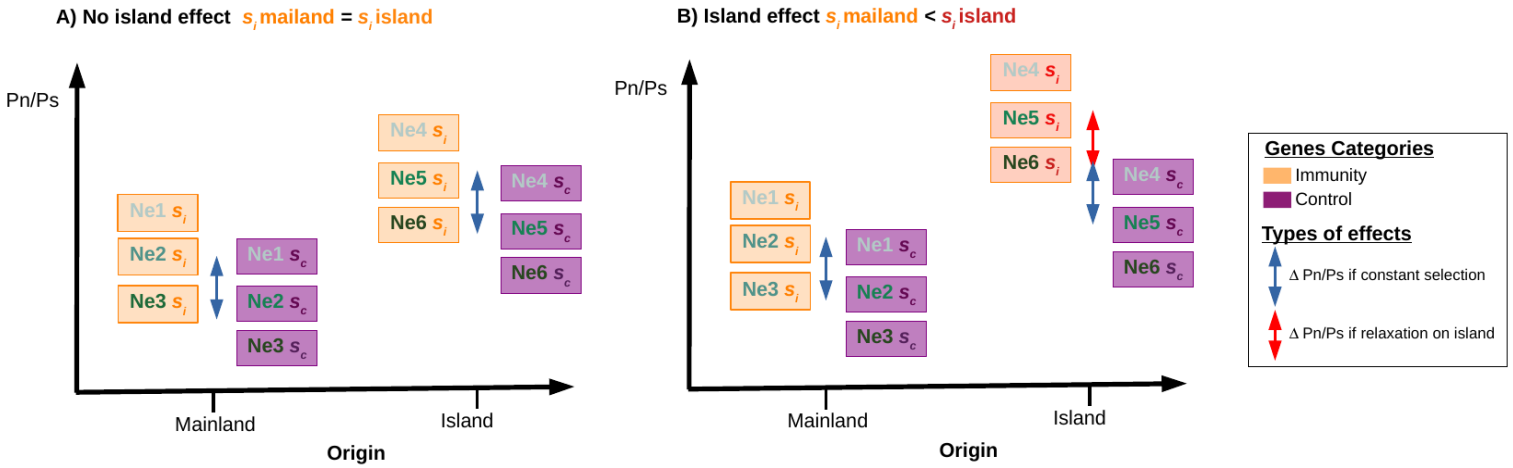

Figure R1: Conceptual diagram showing the expected results under the hypothesis of A) non relaxation in the selection pressure and B) relaxation in the selection pressure on the immunity genes in island species. The diagram includes variation of Ne among species (on average lower for island species) with *Nei* being the *Ne* of species *i* (*i* = 1 to 6). The selection coefficient of mutation (s) for control : genes and immunity genes are shared among species except for a potential difference between island and mainland for the immunity genes.

RE: 2) Not much is mentioned about populations. Pn/Ps is a statistic that only makes sense in populations, eg population structure can impact nucleotide diversity and therefore Pn/Ps ratio. From the text and tables, it's not clear if sampled individuals come from single populations or are just randomly selected from a species. The authors should clarify this or show some results indicating that all individuals from each species come from the same population.

AU: Thank you for this relevant comment. The species for which we have full genome sequenced were obtained from Leroy et al. (2021). For those species, a principal components

analysis (PCA) based on a random sampling of 1000 SNPs was performed to ensure that no population structure or an unfortunate misnaming of individuals was present (Leroy et al., 2021). We did the same for the newly sequenced species by performing PCA on control genes SNP and generate new figures (Sup Fig S2-S4). This control analysis reveals that one individual was actually mislabelled (one Príncipe weaver *Ploceus princeps* was actually a Soa Tomé Kingfisher *Corythornis cristatus*), this individual (P6-015) was removed from our analysis. We also excluded 4 individuals (*Ploceus princeps* P6-174; *P. grandis* ST10_094; *P. nigerrimus* G3_016; *C. teneriffae* TF57) for which we suspected contamination. Otherwise, no extra population structure was detected (See Supp Fig S2-S4).

RE: 3) In MHC simulations, apart from the above-mentioned problem with different species, it seems from simulations of balancing selection that the selection coefficient does not have any effect on the difference between mainland and island genes for a set of chosen Ne sizes (Figure S6A). In this case, testing for the origin effect is not supported because there is no power to distinguish between different selection coefficients. Another thing is that simulated mainland Pn/Ps is different from the observed, so naturally, the category will turn out as significant. I think the results of this test are not interpretable unless simulation parameters are chosen such that simulated Pn/Ps in the mainland will match observed.

AU: We modified our simulation parameters to have comparable Pn/Ps with observed Pn/Ps for mainland species. The new analyses lead to the same results, with a significant interaction effect between origin (island/mainland) and gene category (simulation/MHC class I/MHC class II).

RE: 4) Pn/Ps depends on Ps and this will differ across species. Authors should show the distribution of Ps for each species and group, to make sure that differences in Pn/Ps are not driven by Ps only.

AU: Under the nearly-neutral model of molecular evolution, a link between Pn/Ps and Ps is expected because *Ne* will determine both neutral diversity and the efficacy of selection (see below for more detailed explanation and Welch et al., (2008). This relationship has been supported empirically (Chen et al., 2017; Leroy et al., 2021). As a consequence, Ps should not be used as an explanatory variable of Pn/Ps as it captures the parameter that controls the variation of Pn/Ps in control genes between island and mainland (see also Leroy and Nabholz, (2022)).

RE: Minor comments: line 80: natural selection ”on” low-effect mutations

AU: Corrected at line 79.

RE: line 128: This parapgraph is not clear, it would be help the reader to explain what is expected with lower Ne, what is expected with lower s, in terms of Pn, Ps, and Pn/Ps

AU: We have added a paragraph, at line 118, to better explain the expectations : “Following population genetic theory, in a diploid population, Ps = $4 Ne \mu$ *and* Pn = $4 Ne \mu f$, where *Ne* is the effective population size, μ is the mutation rate and *f* is a function that integrates the

probability of an allele to segregate at a given frequency. *f* depends on the distribution of the fitness effect (DFE) of mutations (Eyre-Walker and Keightley, 2007). This distribution scales with *Ne* as the fitness effect is dependent on *Ne* multiplied by the coefficient of selection *s* (Kimura, 1962). The nearly-neutral theory predicts that the DFE includes a large proportion of mutations with a *Ne.s* close to 0 (Ohta, 1992). 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. This will reduce Pn relatively to the Ps, leading to a negative correlation between Pn/Ps and Ps (through *Ne*) (Welch et al., 2008). The presence of linked positively selected mutations does not change qualitatively this relationship (Castellano et al., 2018; Chen et al., 2020)"

RE: line 129: Please explain why lower s would lead to higher Pn/Ps

AU: We explained it in our response above. The frequency of non-synonymous mutations depends on the DFE that varies with *s*. If it is closer to zero (i.e., the non-synonymous mutations are less deleterious, on average) then non-synonymous mutations will increase in frequencies leading to an increase of Pn/Ps (Welch et al., 2008).

RE: line 131: Is there a reason why do you specifically use "transcriptome" here? If it's a general statement, "genome" would fit better.

AU: Yes, we refer to the coding region of the genome and not to the entire genome.

RE: line 133: The two sentences ("Therefore, we predict..." and "In addition, ...") refer to the same thing: effect of drift (Ne) on selection. I think authors meant to refer in one to lower parasite pressure and in another one to population size decrease.

AU: Yes, the reviewer is right. We modified the sentences at line 135 to 138.

RE: line 144: Genetic diversity of immunity genes is used interchangeably with pn/ps but it's not the same thing, I would stick to pn/ps.

AU: Corrected.

RE: line 147: Please use "nonsynonymous genetic variation" or pn/ps instead of " genetic variation" .

AU: Corrected at line 150.

RE: line 149: Mutations could be eliminated or fixed.

AU: Corrected at line 152.

RE: line 259: This sentence is confusing. Control genes preferentially evolve under purifying selection because this is commonly found across genes in the genome but not because balancing selection is rare.

AU: We rephrased the sentence as “We assumed that control genes mostly evolve under purifying selection”

RE: line 296: Should be nucleotide diversity instead of genetic diversity

AU: Corrected at line 348.

RE: line 298: Is there a reason why PGLS model was not used for the first analysis (immune genes)? Please explain why two different tests were applied to these two similar analyses?

AU: The PGLS method, as implemented in “nlme” packages, can have only one point per species. However, in the first model, we have two points (one Pn/Ps for control genes and one Pn/Ps for immunity genes). In order to use the PGLS method, we chose to use the delta statistic (Pn/Ps immunity - Pn/Ps control).

RE: line 337: It's not clear what N in parentheses referes to. Please explain.

AU: We corrected the sentence by “For the 150 individuals (10 species with 15 individuals each)” at line 389.

Re: line 341: couldn't find Table S3 anywhere

AU: Corrected.

RE: line 338: Please add information about the total length of target regions and average coverage per sample.

AU: We have added Table S4 which shows the total length of locus, the proportion of missing data. We also have added the mean coverage by individuals in Table S3.

RE: line 341: It’s not clear how many species were sequenced with targeted sequencing and how many with WGS. Please add information.

AU: Thank you for your comment, the information was in the abstract, but we added it now in the main text and in Table 1 column “Data Type”

RE: line 386: Some additional results are shown in supplementary material but not commented at all in the main manuscript, eg variation in selection coefficient. Please mention them in the results.

AU: We now mention all supplementary analyses in the main text.

RE: Supplementary methods:

1) Some methods which are discribed in supplements are not metioned in the main paper methods section, eg. test of contaminants. Please add some information in the main methods that is was done.

AU: The methods are now in the main manuscript.

RE: 2) In supplementary methods some analyses are not clear: eg homology detection, please precise at which stage of the data analysis it was performed and what was the goal.

AU: We have clarified the homology detection method by adding : “After defining the genes composition of our differents group, see Selection and identification of immune and control genes, we obtained their reference sequences from the Ensembl database for core groupe and control group and from the annotated genome of *F. albicollis* for Database groupe and Sma3s groupe. We also used the TLRs alignments from (Velová et al., 2018).

Then, in order to identify corresponding sequences in our 11 reference genomes, we search for homologous sequences. To do so, we first translated sequences with transeq (vers. 6.6.0.0; Rice et al., 2000), then we used…”

RE: 3) Why tests for balancing selection (line 42) were performed only for a single species?

AU: This test is time-consuming and computationally intensive. Moreover, it should be done on species with genome-wide data as the method uses linkage information. We therefore chose a species with good genome, annotations and reads sequences, and we assumed that a gene evolving under balancing selection in one species is likely to also evolve under balancing selection in a closely related passerine species.

RE: 4) Supplementary tables are not sorted, and in different formats. Table S3 is missing. Several tables include sequencing information for species, but it’s not clear what the refer to (eg Table S4).

AU: The Table S4 is now available in Table S3 and we tried to make information clearer.

RE: Table 1: It is hard to see for which species reference correpsonds to. Please mark clearly the rows with reference species.

AU: Corrected.

RE: Table 2: it's very unclear what columns 1,2,3 and 4 refer to. Probably you should add model numbers before "Model" column.

AU: Corrected.

RE: Table 3 : Here data is given for one or two best models? It seems like two models, not only the best statistical model.

AU: For Table 3, two best models are shown because the best model based on corrected Akaike Information Criterion (ΔAICc) is different from the best model identified with model simplification using an ANOVA (see Table 2). For this reason, we present results of both of these models in Table 3.

Reviewer #3 : Steven Fiddaman, 14 Jan 2022 16:26

In this study, Barthe and colleagues use genomic and targeted-capture sequencing data from 20 island-dwelling and 14 mainland bird species to test the hypothesis that immune genes experience relaxed selection pressure in island species due to reduced pathogen diversity. They use Toll-like receptors, beta-defensins and MHC classes I and II as genetic proxies for immune function. They identify that neutral processes, such as demographic effects (i.e. bottlenecks), need to be disentangled from the extrinsic effects of a relaxed selection pressure due to changes in pathogen profiles. The authors find no evidence of relaxed selection due to reduced pathogen pressure in the innate immune genes studies, although there is some evidence of this in the MHC class II locus.

Overall, the study has significant merits and makes a good and worthy contribution to the field. It is well written and figures are presented with appropriate clarity. I do, however, have some comments which are detailed below.

AU: Thank you very much for your positive comments.

RE: Line 77. While there is evidence of decrease in MHC diversity leading to a reduction in immune function, as far as I am aware there is no strong evidence for this for TLRs and BDs, and the authors do not cite any (that I can see). This is quite a big assumption and is important for the validity of using genetic proxies for immune function.

AU: We do not expect that a reduction in TLRs and BDs genetic diversity would lead to a reduction of immune function. Rather, we made the assumption that those genes evolved under stabilizing selection (i.e., purifying selection). In the revised manuscript, we now have included a section "*Population genetics of BD and TLR Immune genes*", line 398, in which we demonstrate that TLRs and BDs polymorphism patterns are not strongly influenced by adaptive mutation or balancing selection. If BDs and TLRs polymorphism are influenced by purifying selection, we can assume that a relaxation will lead to an increase in Pn/Ps and not a decrease in genetic diversity.

RE: Line 78. I wanted to chase up information in 'Hale & Briskie 2007' but I cannot find this citation in the reference list.

AU: Corrected.

RE: Line 96. "TLRs and BDs are under purifying selection which usually results in the selective removal of deleterious alleles and stabilizing selection." TLRs, in particular the

extracellular domain, have been found many times to evolve under positive selection. This is demonstrated clearly in the Velova et al. 2018 paper cited within this study. It has also been demonstrated over relatively short timescales (at the subspecies level) so it is certainly relevant here; see for instance Levy et al 2020 (doi.org/10.1093/molbev/msaa040). It is not correct to assume that immune genes, especially Toll-like receptors, are exclusively subject to purifying selection in all cases. See also 10.7554/eLife.41815. Different TLRs are subject to different selective pressures – nucleic acid-recognising TLRs are typically under stronger purifying selection than bacterial-recognising TLRs.

AU: We agree with the reviewer that our sentence was clearly over-simplistic as we overlooked positive selection. We rephrased the sentence and, throughout the manuscript, we added some references to positive selection. Our assumption is that TLRs and BDs evolve mostly under purifying selection, and that a lower parasitic pressure will lead to a relaxation of natural selection. As stated above, the revised manuscript now includes analyses that demonstrate that TLRs and BDs polymorphism patterns are not strongly influenced by adaptative mutation or balancing selection. These results do not exclude that TLRs and/or BDs regularly evolve under positive selection, but only that positive selection is not frequent enough to have left a significant impact at linked selection.

Additionally, we looked more closely at the Pn/Ps of our different TLRs genes to ensure that they all have comparable evolution (Figure R4).

Finally, an interesting follow-up analysis would be to test if positive selection has been reduced in immune genes of insular species compared to continental species using divergence based method (codon models) or contrasting polymorphism and divergence using McDonald-Kreitman type of methods. This analysis will, however, prove to be difficult as the divergence between insular and continental species is small and therefore, the number of substitutions will be low (plus the exact timing of island colonization is usually uncertain).

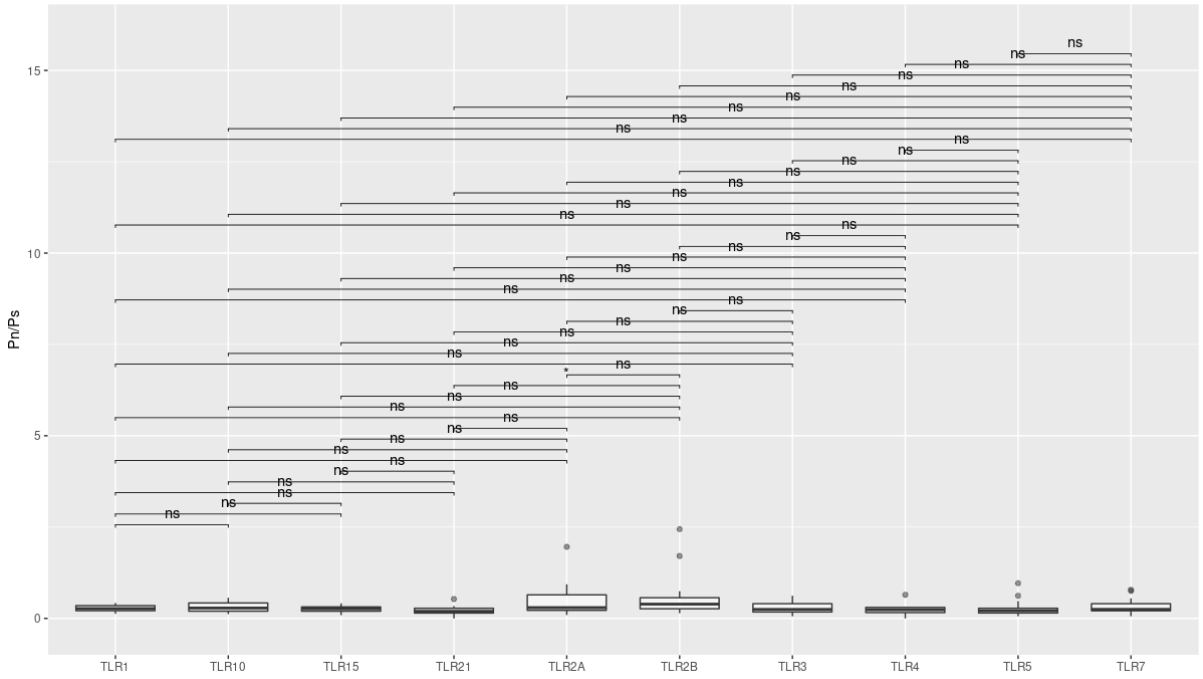

Figure R4: Pn/Ps among different TLRs genes. None pairwise are significant between TLRs families.

RE: Line 146. Typo “weekly”.

AU: Corrected, line 148.

RE: Line 186. Missing information about ethical approval and/or permits required for blood taking.

AU: All the relevant information is stated in the acknowledgement.

RE: Line 191. The MHC locus is notorious for duplications etc (the authors acknowledge this later in the manuscript), which is especially problematic in non-model species. How were loci ensured to be single-copy?

AU: This is clearly an issue that we tried to overcome. Briefly, we analysed different numbers of gene copy among species. We hope that the copies' divergences are high enough for the reads to be correctly mapped to their respective copy. If the copies are very close, then the mapping quality will drop and the reads will be discarded. Reviewer #1 also raised a related comment that we answered as such : “The MHC genes are more difficult to analysed. Indeed, heterozygosity could be comparable to divergence under balancing selection. This makes the identification of orthologs very difficult. In this study, we used the MHC gene I and II copy annotated in Minias et al. (2018). Then, we identify putative homologous sequences only based on sequence homology. This led to the identification of a variable number of genes among species (from 1 to 10 genes for MHC class I and MHC class II). We checked the sequence similarity for the 10 copies of the MHC class II in *F. albicollis* and the 7 copies of

the MHC class I genes in *C. caeruleus* using cd-hit (Fu et al., 2012). For MHC class II, sequences divergences are always higher than 15% indicating that reads will likely be correctly assigned to their corresponding gene copy. For MHC class I, sequence identities could be as high as 95%. In this case, we rely on the fact that the reads from very similar paralogous copies will not be confidently assigned to a gene copy sequence by the mapping software. This will lead to a low mapping score quality and are likely to be discarded during the genotype calling procedure. For example, 3 out of 7 genes of the *Cyanistes* MHC class I genes could not have been correctly genotyped and are missing from our final dataset."

RE: Line 348. 0.12 and 0.2 should be the other way round.

AU: Corrected.

RE: Figure 3. What is the outlier species in the TLR box-plot? A Pn/Ps ratio of >1.5 across the entire coding sequence of all TLRs is implausible – perhaps there is a mistake here, or there is a significant amount of missing data?

AU: The Pn/Ps > 1.5 is indeed implausible. This high estimation is linked to the very low number of SNP recorded in *Platyspiza crassirostris* (a Darwin's finch with a very low genetic diversity). Only 8 SNPs have been called for the TLRs, leading to a very noisy Pn/Ps estimate. Excluding this species does not change the result.

RE: Line 409. Presumably meant to be $P < 0.05$?

AU: Corrected, line 507.

Reviewer #4 : anonymous reviewer, 07 Feb 2022 20:26

This paper is very interesting and adds to the growing evidence that in islands genetic drift is the main force driving the evolution of populations, even in genes that are supposed to be subject to string selection. The paper is generally very well written and easy going. I quite enjoyed reading it. I have only made a few remarks which you will find in the pdf attached.

AU: Thank-you for all the corrections that we have all included in the revised ms except the one below.

RE: Line 430: I wouldn't exactly say that genetic drift reduces the efficacy of selection. Rather, genetic drift is strong and selection cannot counteract it.

AU: The fitness effect of mutation that determines the efficacy of selection is determined by the product of the effective population size *Ne* and the coefficient of selection *s*. So, we think that it is accurate to say that a strong genetic drift (low Ne) reduces the efficacy of selection.

References

Backström, N., Karaiskou, N., Leder, E.H., Gustafsson, L., Primmer, C.R., Qvarnström, A., Ellegren, H., 2008. A Gene-Based Genetic Linkage Map of the Collared Flycatcher (Ficedula albicollis) Reveals Extensive Synteny and Gene-Order Conservation During 100 Million Years of Avian Evolution. Genetics 179, 1479–1495. https://doi.org/10.1534/genetics.108.088195

Buffalo, V., 2021. Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain Lewontin's paradox. eLife 10, e67509. https://doi.org/10.7554/eLife.67509

Castellano, D., James, J., Eyre-Walker, A., 2018. Nearly neutral evolution across the Drosophila melanogaster genome. Mol. Biol. Evol. 35, 2685–2694.

Chen, J., Glémin, S., Lascoux, M., 2020. From drift to draft: how much do beneficial mutations actually contribute to predictions of Ohta's slightly deleterious model of molecular evolution? Genetics 214, 1005–1018.

Chen, J., Glémin, S., Lascoux, M., 2017. Genetic diversity and the efficacy of purifying selection across plant and animal species. Mol. Biol. Evol. 34, 1417–1428.

Corbett-Detig, R.B., Hartl, D.L., Sackton, T.B., 2015. Natural Selection Constrains Neutral Diversity across A Wide Range of Species. PLoS Biol 13, e1002112. https://doi.org/10.1371/journal.pbio.1002112

Drummond, D.A., Wilke, C.O., 2008. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell 134, 341–352. http://dx.doi.org/10.1016/j.cell.2008.05.042

Ellegren, H., 2010. Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol 25, 283–291.

Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8, 610–618. https://doi.org/10.1038/nrg2146

Figuet, E., Nabholz, B., Bonneau, M., Mas Carrio, E., Nadachowska-Brzyska, K., Ellegren, H., Galtier, N., 2016. Life history traits, protein evolution, and the nearly neutral theory in amniotes. Mol. Biol. Evol. 33, 1517–1527.

Freckleton, R.P., Harvey, P.H., Pagel, M., 2015. Phylogenetic analysis and comparative data: a test and review of evidence. Am. Nat.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.

Galtier, N., Rousselle, M., 2020. How much does Ne vary among species? Genetics 216, 559–572.

Grafen, A., 1989. The phylogenetic regression. Sci 326, 119–157.

Kimura, M., 1962. On the Probability of Fixation of Mutant Genes in a Population. Genetics 47, 713–719.

Kutschera, V.E., Poelstra, J.W., Botero-Castro, F., Dussex, N., Gemmell, N., Hunt, G.R., Ritchie, M.G., Rutz, C., Wiberg, R.A.W., Wolf, J.B.W., 2020. Purifying Selection in Corvids Is Less Efficient on Islands. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msz233

Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Ségurel, L., Venkat, A., Andolfatto, P., Przeworski, M., 2012. Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species? PLoS Biol 10, e1001388. https://doi.org/10.1371/journal.pbio.1001388

Leroy, T., Nabholz, B., 2022. Response to Kratochvíl and Rovatsos. Curr. Biol. 32, R30–R31.

Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A.E., Scornavacca, C., Recuerda, M., Fuchs, J., Illera, J.C., De Swardt, D.H., Blanco, G., Thébaud, C., Milá, B., Nabholz, B., 2021. Island songbirds as windows into evolution in small populations. Curr. Biol. 31, 1303-1310.e4. https://doi.org/10.1016/j.cub.2020.12.040

Minias, P., Pikus, E., Whittingham, L.A., Dunn, P.O., 2018. A global analysis of selection at the avian MHC. Evolution 72, 1278–1293.

Ohta, T., 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23,

263–286.
Recuerda, M., Illera, J.C., Blanco, G., Zardoya, R., Milá, B., 2021. Sequential colonization of oceanic archipelagos led to a species-level radiation in the common chaffinch complex (Aves: Fringilla coelebs). Mol. Phylogenet. Evol. 164, 107291.
Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277.
Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., Loire, E., Lourenco, J.M., Nabholz, B., Roux, C., Tsagkogeorga, G., Weber, A. a.-T., Weinert, L.A., Belkhir, K., Bierne, N., Glémin, S., Galtier, N., 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. Nature 515, 261–263. https://doi.org/10.1038/nature13685
Romiguier, J., Ranwez, V., Douzery, E.J.P., Galtier, N., 2013. Genomic Evidence for Large, Long-Lived Ancestors to Placental Mammals. Mol. Biol. Evol. 30, 5–13. https://doi.org/10.1093/molbev/mss211
Scornavacca, C., Belkhir, K., Lopez, J., Dernat, R., Delsuc, F., Douzery, E.J., Ranwez, V., 2019. OrthoMaM v10: scaling-up orthologous coding sequence and exon alignments with more than one hundred mammalian genomes. Mol. Biol. Evol. 36, 861–862.
Singhal, S., Leffler, E.M., Sannareddy, K., Turner, I., Venn, O., Hooper, D.M., Strand, A.I., Li, Q., Raney, B., Balakrishnan, C.N., Griffith, S.C., McVean, G., Przeworski, M., 2015. Data from: Stable recombination hotspots in birds. https://doi.org/10.5061/dryad.fd24j
Velová, H., Gutowska-Ding, M.W., Burt, D.W., Vinkler, M., Yeager, M., 2018. Toll-like receptor evolution in birds: gene duplication, pseudogenisation and diversifying selection. Mol. Biol. Evol.
Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H., Wolf, J.B.W., 2017. Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. Mol Ecol 26, 4284–4295.
Welch, J.J., Eyre-Walker, A., Waxman, D., 2008. Divergence and Polymorphism Under the Nearly Neutral Theory of Molecular Evolution. J. Mol. Evol. 67, 418–426. https://doi.org/10.1007/s00239-008-9146-9