Determinants of mutation load in birds

Fidel Botero-Castro [1], Jochen B. W. Wolf [1*]

[1] Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Germany

* Address for correspondence: Jochen Wolf, Division of Evolutionary Biology, Faculty of Biology, Ludwig Maximilian University of Munich, Großhaderner Straße 2, 82152 Planegg-Martinsried, Germany, Phone: 0049-89-2180-74102, E-mail j.wolf@bio.lmu.de; Fidel Botero-Castro, fidel.boteroc@gmail.com

Running title: Mutation load and recombination

Keywords: Selection, distribution of fitness effects, mutation load, purging, effective population size

Summary. Harmful genetic mutations can build up in populations, affecting individual and population-level health. The authors analyzed population genomic data of 19 bird species to understand the genomic and population determinants of such mutation load. They confirm that small populations tend to experience a heavier mutation burden. Moreover, harmful mutations accumulate more readily in regions of high recombination due to GC-biased gene conversion (gBGC), which is offset by purging efficacy in large, but not small populations. This study suggests that genetic risk factors for mutational meltdown in small populations are fueled by gBGC and cluster in regions of low recombination.

Abstract

Mutations contribute substantially to mortality and disease. The mutation load in a population depends on the efficacy of purifying selection in removing deleterious genetic variation. Here, we estimated the proportion of deleterious mutations segregating in population samples of 19 bird species. Exploiting the conserved avian karyotype with high variation in recombination rate and GC content, we quantified the joint effects of effective population size (N_e), recombination and GC-biased gene-conversion (gBGC). In agreement with the nearly-neutral theory of molecular evolution, mutation load was substantially elevated in populations with small N_e. Purging efficacy increased with recombination rate resulting in more than a two-fold difference of genetic load between large and small chromosomes. GC-biased mutations contributed about one third to the pool of deleterious mutations. Their expected accumulation in regions of high recombination was offset by purging efficacy in large, but not small populations. This study suggests that genetic risk factors for mutational meltdown in small populations are fueled by gBGC and cluster in regions of low recombination.

Introduction

Novel mutations enter a population each generation. Many are neutral, but of those affecting the fitness of its carrier most are deleterious, some are advantageous (Keightley and Lynch 2003). The overall frequency of selective effects is summarized in the genome-wide distribution of fitness effects (DFE, **Figure 1)** reflecting properties of the organism and the distance to its global fitness optimum (Eyre-Walker and Keightley 2007; Lanfear et al. 2014; Chen et al. 2022). Lethal or strongly deleterious mutations are quickly removed from a population by purifying selection, and beneficial mutations are promoted to fixation. Neutral and weakly deleterious mutations, however, can persist for many generations and constitute

the bulk of segregating genetic variation. The time they spend in a population depends on the interaction of their fitness effects (s) with the effective size of the population (N_e), the population selection coefficient $N_e s$ (Kimura 1983; Ohta 1992; Charlesworth 2009; Akashi et al. 2012). As a consequence, deleterious mutations can rise to high frequencies in populations with small N_e (**Figure 1**).

These considerations of the nearly-neutral theory of molecular evolution are of real-world importance (Dussex et al. 2023). The accumulation of deleterious mutations maintained by mutation-selection balance, the mutation load (Muller 1950), can eventually elevate the risk of population extinction (Lande 1994; Frankham 1998; Soulé and Mills 1998; Agrawal and Whitlock 2012). There is growing empirical evidence that mutation load indeed scales negatively with proxies for N_e as has been predicted in theory (Kimura et al. 1963; Lynch et al. 1995). Across a broad taxonomic range, the efficacy of purifying selection tends to be elevated in species and populations characterized by a large number of effectively breeding individuals (Slotte et al. 2010; Deinum et al. 2015; Chen et al. 2017; Castellano et al. 2018; Galtier and Rousselle 2020). As a consequence, mutation load tends to be increased in small, isolated island populations (Kutschera et al. 2020; Leroy et al. 2021).

While the effect of population size on mutation load has been well studied, other factors have received less attention. Recombination, for instance, is known to modulate the efficacy of selection at the intragenomic scale. In regions of low recombination, Hill-Robertson interference acts as to reduce the local effective population size increasing the probability of deleterious alleles going to fixation (Hill and Robertson 1966; Felsenstein 1974; Comeron et al. 2008). Similarly, the purging effect of synergistic epistasis between deleterious mutations is influenced by recombination (Agrawal and Whitlock 2012). As a consequence, mutation

load is expected to differ within the genome by recombination intensity (Comeron et al. 2008). Empirical evidence comes from exploiting variation in recombination rates within and across chromosomes (Gossmann et al. 2011; Webster and Hurst 2012; Corcoran et al. 2017; Chase et al. 2024). Due to obligate crossover small chromosomes have high recombination rates per physical unit (cM/Mb) (Hillier et al, 2004; Kawakami et al. 2014, Backström et al 2010, Bascón-Cardozo et al 2023) and are expected to carry the smallest load, whereas large low-recombining chromosomes will accumulate more deleterious mutations. Non-recombining, uniparentally inherited sex-chromosomes or organellar genomes are predicted hotspots of mutation load and have accordingly received most attention (Charlesworth and Charlesworth 2000; Marais 2007).

Importantly, recombination is also known to be associated with GC-biased gene conversion (gBGC) referring to preferential reparation of double-strand breaks towards G and C alleles which eventually results in an increase of GC content in highly recombining regions (Galtier et al. 2001; Duret and Galtier 2009). Most of these mutations are prone to deleterious effects but are constantly replenished through the mutational bias in regions of high recombination mimicking the effect of positive selection (Berglund et al. 2009). gBGC is thus predicted to counteract efficient purging of weakly deleterious mutation in regions of high recombination (Bolívar et al. 2018). Yet, it is unclear at which rate these opposing forces operate, in particular, as the strength of gBGC also seems to interact with the overall effective population size (Bolívar et al. 2019) (**Figure 1**).

Avian genomes are well suited to study these interacting effects, as they provide the necessary variation in N_e across a stable, yet highly variable, recombination landscape within and between chromosomes (**Figure 2**) (Ellegren 2010). To test the influence of 8/18/2025

6:04:00 AMrecombination, effective population size and GC-biased gene conversion, we collated genome-wide resequencing data across 24 populations from 19 avian species and estimated the DFE to obtain estimates of the proportion of segregating, mildly deleterious mutations ($N_e s \in (-1;0]$) serving as proxy for mutation load. According to theory, we expect the following: mutation load should be: 1) higher in populations with low N_e, 2) reduced in regions of high recombination and 3) elevated for the mutational spectrum subject to gBGC (weak-to-strong mutations: [A,T]->[C,G]). 4) Relative rates of the two recombination-mediated, opposing processes of purging and gBGC may differ and be additionally modified by N_e (**Figure 1**). In addition to these effects, we expect species-specific variance in mutation load reflecting a species' position in the fitness landscape (Lanfear et al. 2014).

Materials and Methods

Sequence Data

We collated publicly available resequencing data for 443 individuals from 47 populations of 30 passerine species and utilized *.vcf* files generated by comparable pipelines described in (Leroy et al. 2021) and (Kutschera et al. 2020) (see **Supplementary Table 1** for accession numbers of NCBI's Sequencing Read Archive). All *.vcf* files were then filtered to remove indels and multi-nucleotide variants. Only sites with a minimum Qval of 20 and belonging to autosomal scaffolds/chromosomes were retained. *.vcf* files were further split by chromosome. Finally, sites were marked as 0-fold, 2-fold, 3-fold and 4-fold degenerate on the basis of the corresponding genome annotation of the closest reference genome (**Supplementary Table 1**) using the script *NewAnnotateRef.py* b Williamson et al. (2014) available at (https://github.com/fabbyrob/science/tree/master/pileup_analyzers). The output file was then converted into bed files for each partition using bedtools. An additional dataset was produced by excluding CpG-prone sites and weak-to-strong mutations following the approach of

(Kutschera et al. 2020). This dataset was used to assess the effect of gBGC (A-to-G, A-to-C, T-to-G, T-to-C substitutions) by comparison to the dataset including all sites.

Population Structure

The presence of population structure and gene flow between populations can distort inference of the DFE, although the effect can partly be absorbed by the inclusion of a demographic parameter of Keightley and Eyre-Walker's (2007) (2e model, see also (Kutschera et al. 2020)). To reduce any effect of population structure, we performed a PCA for each species using the full set of coding autosomal SNPs in *plink* (v1.90b6.21). Individuals appearing as outliers in the first two axes of variation were removed from dataset (**Supplementary Figure 1**). Moreover, the Italian population of collared flycatchers (*Ficedula albicollis*) and the Spanish population of pied flycatchers (*Ficedula hypoleuca*) were analyzed separately from the remaining populations, since they diverged strongly from the remaining populations, which has been previously documented (Burri et al. 2015). Two species of white eyes (*Zosterops borbonicus* and *Z. virens*) were excluded due to significant structure and gene flow between the subpopulations (Bourgeois et al. 2020). Furthermore, as the inference of the SFS requires sites to be sampled for all individuals in the sample (i.e no missing data), individuals with low-coverage and low-quality genotypes resulting in more than 5 % missing data were also excluded (**Supplementary Table 1**). Moreover, SFS required a minimum of eight chromosomes to be sampled, i.e. four individuals with sufficient sequencing coverage (cf. Kutschera et al. 2020). After applying these individual and population filters, the final dataset consisted of 24 populations from 19 species, each represented by a median average of 1,604,1053 total and 46,342 segregating sites (**Supplementary Table 2**).

Inference of the DFE

We inferred the folded site frequency spectrum (SFS) for each population and each of 28 autosomes separately using the scripts *vcfSummarizer.py* and *bootstrap.py* available at (https://github.com/fabbyrob/science/tree/master/pileup_analyzers). The inferred folded SFS was used to model the DFE of each species both per-chromosome and using the full autosomal dataset using DFEalpha v2.16 (Keightley and Eyre-Walker 2007). This approach requires a set of sites under selection, in which we included the 0-fold, 2-fold and 3-fold degenerated sites, and a set of sites evolving under or close to neutrality, for which we used the set of 4-fold degenerated sites. Two demographic models with constant size (1-epoch) and one with change in population size (2-epochs), respectively, were run for comparison and the best fitting model was selected using a likelihood ratio test.

Inference of the best fitting model differed by chromosome (**Supplementary Figure 2**). Inference using the full autosomal dataset was largely dominated by the largest chromosomes. Given this variability and our interest in obtaining DFE estimates per chromosome, we selected the statistically preferred model per chromosome rather than imposing a single model to all genomic regions.

Estimation of the Effective Population Size

Effective populations sizes were obtained from (Kutschera et al. 2020) and (Leroy et al. 2021), inferred as the harmonic mean of N_e estimates derived from the sequential Markovian coalescent (McKenna et al. 2010; Schiffels and Durbin 2014) weighted by the length of each time period. For details see Supplementary Text S1 in Kutschera et al. (2020) and Method detail in Leroy et al. (2021).

Recombination Rate Estimation

The average recombination rate (*r)* per chromosome was approximated in centimorgans per megabase [cM/Mb] as follows: we extracted estimates of the per-base population recombination rate ρ[1/bp] obtained by Vijay et al. (2016) in bins of 50kb windows for 3 populations of hooded and carrion crows (*Corvus (corone) cornix*). We then calculated, for each population, the median across the 50kb windows to obtain per chromosome estimates. Last, we averaged the estimates across the three populations. We then estimated genetic variation along each chromosome and converted the population recombination rate estimates ρ[1/bp] to recombination estimates r[cM/Mb] using the relationship $\rho / \pi = 4N_e r / 4Ne\mu = r / \mu$. Multiplication with an avian mutation rate of 4.6×10^{-9} mutations/site/generation (Smeds et al. 2016) and scaling to cM (10^2) and Mb (10^6) yields r[cM/Mb]. Per-chromosome estimates range between 2.46 and 11.5 cM/Mb (mean weighted by chromosome lenght: 3.57) adding up to a total genetic map length of 3,500 cM which is comparable to previous, pedigree-based estimates of passerine birds (Backström et al. 2010; Kawakami et al. 2014).

Central to our study is the assumption that the karyotype of passerine species considered here is stable enough to reflect comparable recombination rates across syntenic chromosomes. To test the assumption that a single set of recombination rates is representative for the species involved in the analyses, we compared chromosome-level recombination rates of hooded crow (Vijay et al. 2016) with those of zebra finch (*Taeniopygia guttata*) (two estimates from Singhal et al. (2015) and Stapley et al. (2010), collared flycatcher (*Ficedula albicollis*) (Chase et al. 2024), and Eurasian blackcap (*Sylvia atricapilla*)(Bascón-Cardozo et al. 2024). As the estimates of zebra finch in Singhal et al (2015) and the ones of the Eurasian blackcap were provided as estimates of the population recombination rate (rho, per base), we used the values of genetic diversity provided in the respective studies (genomic estimate of 0.082 for

zebra finch; per chromosome estimates in eurasian blackcap as in Table S3 of Bascón-Cardozo et al (2024)) , and the avian mutation rate as in Smeads et al ($4.6x10^{-9}$ mutations/site/generation), to convert them into cM/Mb using the same approach as for hooded crow outlined above.

Chromosome-level recombination rates were high across all comparisons (median R^2=0.668, **Supplementary Figure 3**) supporting that at chromosome-level, recombination rates are comparable across passerine birds. A single recombination estimate, here from crows, appears thus suitable for comparative analysis, although it will introduce noise which may conceal its effect. Moreover, chromosome length strongly correlates with recombination rate, and is highly conserved across avian species (Ellegren 2010, this study; Bascón-Cardozo et al. 2024). We therefore also used chromosome length as a proxy for recombination rate.

In order to ensure that the chromosomes correspond to the names assigned in the different references used here (**Supplementary Table 1**), we performed a verification of synteny using the corresponding annotations. First, all annotated genes for each chromosome were extracted for hooded crow, zebra finch, collared flycatcher, great tit and chaffinch and the lists were compared between each possible pair of references. The annotations were highly consistent and synteny verified, with few exceptions, for all references. Most mismatches occurred between the reference for hooded crow and the other references, mainly due to crow scaffolds belonging to a) chromosomes that were not present in the crow assembly version GCF_000738735.1, to which the *.vcf* files utilized here correspond, being assigned to other chromosomes (e.g. 25, LGE22) or to b) small chromosomes (e.g. 23, 27). The identified mismatches were corrected and the scaffolds included in the correct chromosomes, if they were confirmed to be syntenic across all the other analyzed references or excluded, if the

chromosomes were not present in the hooded crow assembly. A few mismatches happening in the used zebra finch reference (GCF_000151805.1), are already corrected and verified as syntenic in a more recent annotation (GCF_003957565.2).

Statistical Analyses

We used a mixed linear model framework to test the influence of recombination (r), effective population size (N_e) and GC-biased gene conversion (gBGC) on the proportion of mildly deleterious alleles ($N_e s \in (-1;0]$). Recombination and effective population size were as approximated by two alternative measures (r: recombination rate [cM/Mb] or chromosome length (see above); N_e: species restricted to island yes/no or log(N_e)). The main results were largely unaffected by the choice of measure. In the main text we present the results for recombination rate [cM] and island yes/no. Results for the alternative measures can be found in **Supplementary Table 3.** The effect of gBGC was assessed as categorical variable by comparing the estimate of DFE using all polymorphisms (with gBGC) or the subset including only weak-to-weak (A-T), and strong-to-strong mutations (C-G) (without gBGC).

We fitted a mixed effect model including all three parameters and a choice of meaningful interactions as fixed effects using the lmer function of the R package lme4 (Bates et al. 2015). We further incorporated chromosome (chr) and population identity (pop) as random effects including variation in slope for the latter. More than a single population was available for a subset of five species, which led us to also include species (spec) as a nested random effect (population nested in species). Model choice was then performed largely following (Zuur et al. 2009): we started with a model containing the full set of random factors and all relevant fixed factors and their interactions: *DFE ~ r + pop + gBGC + r*pop + r*gBGC + pop*gBGC +(1|chr) + (1+r|pop/spec).* We then selected the best combination of random

parameters using restricted maximum likelihood estimation (REML) minimizing the Akaike Information Criterion (AIC) which eliminated the nested population/species component and reduced the random effects to *(1+r|pop)+(1|chr)*. Next, we formulated a set of 10 biologically meaningful models including the null model (see **Supplementary Table 3**) and selected the model minimizing AIC using full maximum likelihood which provides information about the goodness of fit of the respective model to the observed data and penalizes overly complex models (Johnson and Omland 2004). Models differing by less than two units of AIC were considered to be supported. For parameter estimation, we then re-ran the best model using REML estimation and conducted an analysis of variance as implemented in the R package car (Fox and Weisberg 2019). Finally, we assessed normality of residuals by visual inspection and model performance by plotting observed to predicted values (**Supplementary Figure 4**).

Results

We calculated the DFE for each of the 24 species from segregating sites in coding sequence with reliable genotype information for at least 8 chromosomal copies. The DFE was calculated separately for each of 28 autosomes by contrasting the site frequency spectrum (SFS) of neutral mutations at 4-fold degenerate sites with the frequency of mutations subject to selection at 0-fold degenerate sites controlling for demographic perturbation where indicated (see methods). Recombination rate was approximated in cM/Mb and its covaried to a high degree with the chromosome length and (gBGC-impacted) GC content (log-log correlations: R^2=0.77 and 0.81, respectively; both $p<0.001$) (**Figure 2**).

In order to establish a baseline without the influence of gBGC, we first performed the analyses on the SFS after excluding CpG-prone sites and weak-to-strong mutations [A,T]-

>[C,G]. The median, genome-wide proportion of mildly deleterious mutations differed substantially between populations and species (min/median/max: 0.09/0.18/0.25). This variation in mutation load appears to be well explained by population size, regardless of whether it was approximated by coalescence-based estimates of N_e or by species distribution (island vs. mainland) (**Figure 3A, Supplementary Figure 5**). Species restricted to islands had a median average N_e of approx. 90,000, mainland species approx. 160,000. This 2-fold difference in N_e was reflected by a median increase in the mutation load of approximately 1.6-fold. Mutation load was not only dependent on N_e of the species, but also varied along the genomes of nearly all populations as function of recombination rate: low-recombining chromosomes carried substantially more load than high-recombining genomes for both island and mainland populations (**Figure 3A**).

gBGC increased the mutation load by approximately 1.5-fold on median average (**Figure 3B**). At first glance, the patterns of mutation load in relation to recombination rate and species-level effective population size (N_e) appear similar whether or not biased gene conversion is considered (**Figure 3A** vs. **Figure 3B**). To formally evaluate the effects of population size, recombination, gBGC, and their interactions on mutation load, we employed a linear mixed-effects model (**Supplementary Table 3**). Chromosome and population identity were included as random effects, as they accounted for a significant proportion of additional variance. The best-fitting model (**Table 1, Supplementary Figure 4**) indicates a strong influence of all three factors, summarized as follows: (i) mutation load is substantially higher in small populations, (ii) is further elevated for mutations affected by gBGC, and (iii) is markedly reduced in genomic regions with high recombination rates. The model also predicts that the absolute contribution of gBGC to deleterious variation is greater in small populations (i.e., a significant N_e *x gBGC* interaction), with island populations showing a

9.6% higher proportion of deleterious mutations compared to a 6.4% increase in mainland populations. Nonetheless, the relative increase due to gBGC remains consistent across population sizes, with island and mainland populations exhibiting a similar fold-increase in deleterious mutations (1.44× vs. 1.46×, respectively).

The data further allowed us to assess the relative rates by which the two opposing, recombination-dependent processes affect mutation load. If recombination affected both processes equally, then a doubling in the increase of deleterious mutations by gBGC should be fully offset by a proportional increase in purging efficacy. The relative contribution of both processes may further depend on the species' N_e (**Figure 1**). We find that the purging efficacy of recombination on mutation load was reduced in small populations for the mutational spectrum affected by gBGC. In other words, small populations suffer from a disproportionate introduction of deleterious [A,T]->[C,G] mutations in regions of high recombination that is not fully offset by purging efficacy (**Figure 3C**). This effect (three-way interaction $r \times N_e \times gBGC$) was weak, but statistically supported when using a binary measure of population size (island/mainland) (**Table 1**, ΔAIC to 2nd best model =2.229), but not when parametrizing by coalescent based estimates of *Ne* (**Supplementary Table 3**).

Discussion

In this study, we estimated the distribution of fitness effect for 24 populations from 19 species including representatives from across the major songbird clades (Corvoidea, Muscicapoidea, Passeroidea, Sylvoidea) (Jønsson and Fjeldså 2006). Oscine songbirds are the most speciose avian clade arising broadly around the Cretaceous-Palaeogene boundary some 66 million years ago (Sibley and Ahlquist 1990; Barker et al. 2004; Jarvis et al. 2014). The study thus claims some generality at the higher end of evolutionary timescales compared

to previous comparative work in the area (within species: (Eyre-Walker et al. 2006; Slotte et al. 2010; Deinum et al. 2015; Castellano et al. 2018), genera: (Kutschera et al. 2020), families: (Castellano et al. 2019), and beyond (Chen et al. 2017; Leroy et al. 2021; Chen et al. 2022)). With the exception of few, chosen species are phylogenetically distant enough as not to share ancestral polymorphism (isolation > 9 Ne generations Hudson et al. 2002; Mugal et al. 2020). While some evolutionary dependency may remain, in particular for populations within species, we abstained from additional phylogenetic contrasts in the random part of the model, as the population effect was eliminated during model selection (see methods).

Focusing on the mildly deleterious portion of the DFE, we tested the effects of N_e, recombination rate, gBGC and their interaction on mutation load. The effective population size, as parameterized by species distribution (island/mainland) or coalescent-based estimates, was a strong predictor. In accordance with the central tenet of the nearly neutral theory and previous studies, mutation load was significantly elevated in small effective populations (Ohta 1992; Charlesworth 2009; Agrawal and Whitlock 2012; Kutschera et al. 2020; Leroy et al. 2021) . Island populations had approximately half the coalescent effective population size as on the mainland. Our model predicts this difference to translate to an increase in the genome-wide mutation load of approximately 1.6 on islands (cf. (Castellano et al. 2018) for a similar estimate considering local genomic N_e variation within the *Drosophila* genome). While the difference in mutation load on islands is clearly mediated by small population size, a rather discrete change between mainland and island populations may also suggest additional factors – such as bottlenecks – that are not fully accounted for by the coalescent-based N_e estimates (**Supplementary Figure 5**). Overall, these results add to the evidence that species living on islands accumulate mildly deleterious mutations more readily than their mainland counterparts (Johnson and Seger 2001; Woolfit and Bromham 2005;

Kutschera et al. 2020) making them more vulnerable to mutational meltdown (Frankham 1998).

Our results further suggest that the risk associated with mutation load is unevenly distributed across the genome. In accordance with relatively few previous studies touching upon the subject (Gossmann et al. 2011; Leroy et al. 2021; Chase et al. 2024), we find a clear reduction of mutation load with an increase in recombination. In our model, the consistent ~ 5-fold range of recombination rate across avian chromosomes (Backström et al. 2010; Kawakami et al. 2014; Bascón-Cardozo et al. 2024, this study) translates to a ~1.7-fold change in mutation load. The order of this change is comparable to the effect of N_e-change between island and mainland. Importantly, however, the effect of recombination was independent of N_e (no statistical interaction N_e x r) suggesting the same relative purging efficiency between chromosomes regardless of population size. This result is in conflict with conclusions from Leroy et al. (2021). While they also find reduced mutation load in genomic regions of high recombination, the effect of recombination was stronger in island populations. This disconnect may in part be explained by them using GC3 content as proxy for recombination, a different set of species and presence of population structure which can influence the DFE (Kutschera et al. 2020) and which we removed for this study (**Supplementary Figure 1**).

In addition to N_e and intra-genomic heterogeneity, the quality of mutations also affected mutation load. When including mutations subject to gBGC biased gene conversion the genome-wide proportion of mildly deleterious variation was elevated ~1.5 fold, both in island and mainland populations. An elevation of mutation load in the fraction of these mutations is expected, as they are replenished through the mutational bias mimicking the effect of positive

selection (Berglund et al. 2009). Yet, its interaction with recombination and overall population size remains largely unclear. Since gBGC relies on double-strand breaks associated with elevated levels of recombination, gBGC mediated mutation load is expected to be more pronounced in regions of high recombination (Duret and Galtier 2009; Kostka et al. 2012; Pouyet et al. 2018). We thus expected to see relatively more deleterious mutations in regions of high recombination with gBGC then without (flatter slope, **Figure 1**). We observed the effect only in island populations, while the effect may be too slight to pick up in the mainland with an overall increased efficacy of purifying selection.

Returning to the hypotheses of the introduction we conclude that 1) mutation load is elevated in (island) populations with low N_e, 2) is reduced in regions of high recombination and 3) elevated for the mutational spectrum subject to gBGC. All three effects are of comparable magnitude across the relevant biological range of variation in our sample. 4) Relative rates of the two recombination-mediated, opposing processes of purging and gBGC differ in populations with low N_e. Only few studies have hitherto explored these interactions which require more attention (Jiang et al. 2025). In addition, non-negligible variance for the species random effect is consistent with additional components affected a species' DFE, such as its current position in the fitness landscape. Overall, this study highlights that genetic risk factors for mutational meltdown are non-homogeneously distributed across the genome (Kardos et al. 2021). Emerging medical and conservation genomic studies screening for deleterious mutations and their fitness effects (Bertorelle et al. 2022; Hoffman et al. 2024) may benefit from taking the interaction between recombination, gBGC and N_e into account.

Data Availability

Whole genome resequencing data are available through NCBI. Accession numbers are provided in **Supplementary Table 1**. Analysis scripts are available at https://github.com/EvoBioWolf/CORVID_DFE.

Acknowledgements

Benoit Nabholz and Verena Kutschera provided access to raw data and .vcf files. Funding was provided to J.B.W.W. by the European Research Council (ERCStG-336536 FuncSpecGen), the Swedish Research Council Vetenskapsrådet (621-2013-4510), the Knut and Alice Wallenberg Foundation and LMU Munich. HPC computing was performed on the BioHPC hosted at Leibniz Rechenzentrum Munich funded by the German Research Foundation (grant INST 86/2050-1 FUGG).

Author Contributions

JBWW and FB conceived of the study. FB conducted all bioinformatic analyses. JBWW conducted the statistical analyses and wrote the manuscript with input from FB.

References

Agrawal AF, Whitlock MC. 2012. Mutation Load: The Fitness of Individuals in Populations Where Deleterious Alleles Are Abundant. Annu Rev Ecol Evol Syst. 43(1):115–135. doi:10.1146/annurev-ecolsys-110411-160257. [accessed 2024 Feb 23]. https://www.annualreviews.org/doi/10.1146/annurev-ecolsys-110411-160257.

Akashi H, Osada N, Ohta T. 2012. Weak selection and protein evolution. Genetics. 192(1):15–31. doi:10.1534/genetics.112.140178. [accessed 2012 Oct 29]. http://www.genetics.org/content/192/1/15.

Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, Webster MT, Öst T, Schneider M, Kempenaers B, et al. 2010. The recombination landscape of the zebra

finch Taeniopygia guttata genome. Genome Res. 20(4):485–495. doi:10.1101/gr.101410.109. [accessed 2024 Mar 8]. https://genome.cshlp.org/content/20/4/485.

Barker FK, Cibois A, Schikler P, Feinstein J, Cracraft J. 2004. Phylogeny and diversification of the largest avian radiation. Proc Natl Acad Sci U S A. 101(30):11040–11045. doi:10.1073/pnas.0401892101.

Bascón-Cardozo K, Bours A, Manthey G, Durieux G, Dutheil JY, Pruisscher P, Odenthal-Hesse L, Liedvogel M. 2024. Fine-Scale Map Reveals Highly Variable Recombination Rates Associated with Genomic Features in the Eurasian Blackcap. Genome Biol Evol. 16(1):evad233. doi:10.1093/gbe/evad233. [accessed 2025 Jun 8]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10781513/.

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 67(1). doi:10.18637/jss.v067.i01. [accessed 2016 Aug 22]. http://www.jstatsoft.org/v67/i01/.

Berglund J, Pollard KS, Webster MT. 2009. Hotspots of biased nucleotide substitutions in human genes. Plos Biol. 7:e1000026(1).

Bertorelle G, Raffini F, Bosse M, Bortoluzzi C, Iannucci A, Trucchi E, Morales HE, van Oosterhout C. 2022. Genetic load: genomic estimates and applications in non-model animals. Nat Rev Genet. 23(8):492–503. doi:10.1038/s41576-022-00448-x. [accessed 2024 Apr 12]. https://www.nature.com/articles/s41576-022-00448-x.

Bolívar P, Guéguen L, Duret L, Ellegren H, Mugal CF. 2019. GC-biased gene conversion conceals the prediction of the nearly neutral theory in avian genomes. Genome Biol. 20(1):1–13. doi:10.1186/s13059-018-1613-z. [accessed 2024 Feb 27]. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1613-z.

Bolívar P, Mugal CF, Rossi M, Nater A, Wang M, Dutoit L, Ellegren H. 2018. Biased Inference of Selection Due to GC-Biased Gene Conversion and the Rate of Protein Evolution in Flycatchers When Accounting for It. Mol Biol Evol. 35(10):2475–2486. doi:10.1093/molbev/msy149. [accessed 2025 Jul 25]. https://doi.org/10.1093/molbev/msy149.

Bourgeois YXC, Bertrand JAM, Delahaie B, Holota H, Thébaud C, Milá B. 2020. Differential divergence in autosomes and sex chromosomes is associated with intra-island diversification at a very small spatial scale in a songbird lineage. Mol Ecol. 29(6):1137–1153. doi:10.1111/mec.15396. [accessed 2025 Jun 8]. https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15396.

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S, Garamszegi LZ, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Res. 25:1656–1665. [accessed 2015 Nov 3]. http://genome.cshlp.org/content/early/2015/10/07/gr.196485.115.

Castellano D, James J, Eyre-Walker A. 2018. Nearly Neutral Evolution across the Drosophila melanogaster Genome. Mol Biol Evol. 35(11):2685–2694. doi:10.1093/molbev/msy164. [accessed 2021 Jul 20]. https://doi.org/10.1093/molbev/msy164.

Castellano D, Macià MC, Tataru P, Bataillon T, Munch K. 2019. Comparison of the Full Distribution of Fitness Effects of New Amino Acid Mutations Across Great Apes. Genetics. 213(3):953–966. doi:10.1534/genetics.119.302494. [accessed 2025 Jul 25]. https://doi.org/10.1534/genetics.119.302494.

Charlesworth B. 2009. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 10(3):195–205. http://dx.doi.org/10.1038/nrg2526.

Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. Philos Trans R Soc Lond Ser B-Biol Sci. 355(1403):1563–1572.

Chase MA, Vilcot M, Mugal CF. 2024. The role of recombination dynamics in shaping signatures of direct and indirect selection across the *Ficedula* flycatcher genome. Proc R Soc B Biol Sci. 291(2015). doi:10.1098/rspb.2023.2382.

Chen J, Bataillon T, Glémin S, Lascoux M. 2022. What does the distribution of fitness effects of new mutations reflect? Insights from plants. New Phytol. 233(4):1613–1619. doi:10.1111/nph.17826. [accessed 2025 Jul 25]. https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.17826.

Chen J, Glémin S, Lascoux M. 2017. Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. Mol Biol Evol. 34(6):1417–1428. doi:10.1093/molbev/msx088. [accessed 2018 Oct 19]. http://academic.oup.com/mbe/article/34/6/1417/3049540.

Comeron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations. Heredity. 100(1):19–31. doi:10.1038/sj.hdy.6801059. [accessed 2011 Jan 26]. http://www.ncbi.nlm.nih.gov/pubmed/17878920.

Corcoran P, Gossmann TI, Barton HJ, Great Tit HapMap Consortium, Slate J, Zeng K. 2017. Determinants of the Efficacy of Natural Selection on Coding and Noncoding Variability in Two Passerine Species. Genome Biol Evol. 9(11):2987–3007. doi:10.1093/gbe/evx213.

Deinum EE, Halligan DL, Ness RW, Zhang Y-H, Cong L, Zhang J-X, Keightley PD. 2015. Recent Evolution in Rattus norvegicus Is Shaped by Declining Effective Population Size. Mol Biol Evol. 32(10):2547–2558. doi:10.1093/molbev/msv126. [accessed 2015 Oct 12]. http://mbe.oxfordjournals.org/content/32/10/2547.

Duret L, Galtier N. 2009. Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes. Annu Rev Genomics Hum Genet. 10(1):285–311. doi:10.1146/annurev-genom-082908-150001. [accessed 2020 Mar 13]. http://www.annualreviews.org/doi/10.1146/annurev-genom-082908-150001.

Dussex N, Morales HE, Grossen C, Dalén L, Van Oosterhout C. 2023. Purging and accumulation of genetic load in conservation. Trends Ecol Evol. 38(10):961–969. doi:10.1016/j.tree.2023.05.008. [accessed 2024 Feb 27]. https://linkinghub.elsevier.com/retrieve/pii/S0169534723001313.

Ellegren H. 2010. Evolutionary stasis: the stable chromosomes of birds. Trends Ecol Evol. 25(5):283–291. doi:10.1016/j.tree.2009.12.004.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8(8):610–618. doi:10.1038/nrg2146. [accessed 2014 Sep 16]. http://www.nature.com/nrg/journal/v8/n8/abs/nrg2146.html.

Eyre-Walker A, Woolfit M, Phelps T. 2006. The Distribution of Fitness Effects of New Deleterious Amino Acid Mutations in Humans. Genetics. 173(2):891–900. doi:10.1534/genetics.106.057570. [accessed 2014 Sep 19]. http://www.genetics.org/content/173/2/891.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics. 78(2):737–756.

Fox J, Weisberg S. 2019. An R Companion to Applied Regression. 3rd ed. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Frankham R. 1998. Inbreeding and extinction: island populations. Conserv Biol. 12(3):665–675. doi:10.1111/j.1523-1739.1998.96456.x. [accessed 2014 May 2]. http://onlinelibrary.wiley.com/doi/10.1111/j.1523-1739.1998.96456.x/abstract.

Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-Content Evolution in Mammalian Genomes: The Biased Gene Conversion Hypothesis. Genetics. 159(2):907–911. doi:10.1093/genetics/159.2.907. [accessed 2024 Mar 15]. https://doi.org/10.1093/genetics/159.2.907.

Galtier N, Rousselle M. 2020. How Much Does Ne Vary Among Species? Genetics. 216(2):559–572. doi:10.1534/genetics.120.303622. [accessed 2024 Mar 15]. https://doi.org/10.1534/genetics.120.303622.

Gossmann TI, Woolfit M, Eyre-Walker A. 2011. Quantifying the variation in the effective population size within a genome. Genetics. 189(4):1389–1402. doi:10.1534/genetics.111.132654. [accessed 2012 May 22]. http://www.genetics.org/content/189/4/1389.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8(3):269–294.

Hoffman JI, Vendrami DLJ, Hench K, Chen RS, Stoffel MA, Kardos M, Amos W, Kalinowski J, Rickert D, Köhrer K, et al. 2024. Genomic and fitness consequences of a near-extinction event in the northern elephant seal. Nat Ecol Evol. 8(12):2309–2324. doi:10.1038/s41559-024-02533-2. [accessed 2025 Jun 8]. https://www.nature.com/articles/s41559-024-02533-2.

Hudson RR, Coyne JA, Huelsenbeck J. 2002. Mathematical consequences of the genealogical species concept. Evolution. 56(8):1557–1565. doi:10.1554/0014-3820(2002)056[1557:MCOTGS]2.0.CO;2. [accessed 2016 Jun 1]. http://www.bioone.org/doi/abs/10.1554/0014-3820%282002%29056%5B1557%3AMCOTGS%5D2.0.CO%3B2.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 346(6215):1320–1331. doi:10.1126/science.1253451. [accessed 2015 Mar 18]. http://www.sciencemag.org/content/346/6215/1320.

Jiang J, Chen J-F, Li X-T, Wang L, Mao J-F, Wang B-S, Guo Y-L. 2025. Incorporating genetic load contributes to predicting Arabidopsis thaliana's response to climate change. Nat Commun. 16(1):2752. doi:10.1038/s41467-025-58021-z. [accessed 2025 Jun 8]. https://www.nature.com/articles/s41467-025-58021-z.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. Trends Ecol Evol. 19(2):101–108. doi:10.1016/j.tree.2003.10.013. [accessed 2024 Apr 12]. https://www.sciencedirect.com/science/article/pii/S0169534703003458.

Johnson KP, Seger J. 2001. Elevated Rates of Nonsynonymous Substitution in Island Birds. Mol Biol Evol. 18(5):874–881. doi:10.1093/oxfordjournals.molbev.a003869. [accessed 2019 Feb 15]. http://academic.oup.com/mbe/article/18/5/874/1018679.

Jønsson KA, Fjeldså J. 2006. A phylogenetic supertree of oscine passerine birds (Aves: Passeri). Zool Scr. 35(2):149–186. doi:10.1111/j.1463-6409.2006.00221.x. [accessed 2024 Apr 12]. https://onlinelibrary.wiley.com/doi/10.1111/j.1463-6409.2006.00221.x.

Kardos M, Armstrong EE, Fitzpatrick SW, Hauser S, Hedrick PW, Miller JM, Tallmon DA, Funk WC. 2021. The crucial role of genome-wide genetic variation in conservation. Proc Natl Acad Sci. 118(48):e2104642118. doi:10.1073/pnas.2104642118. [accessed 2025 Jun 8]. https://www.pnas.org/doi/10.1073/pnas.2104642118.

Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, Olason P, Ellegren H. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol Ecol. 23(16):4035–4058. doi:10.1111/mec.12810. [accessed 2016 Nov 18]. http://onlinelibrary.wiley.com.emedien.ub.uni-muenchen.de/doi/10.1111/mec.12810/abstract.

Keightley PD, Eyre-Walker A. 2007. Joint Inference of the Distribution of Fitness Effects of Deleterious Mutations and Population Demography Based on Nucleotide Polymorphism Frequencies. Genetics. 177(4):2251–2261. doi:10.1534/genetics.107.080663. [accessed 2016 Dec 5]. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2219502/.

Keightley PD, Lynch M. 2003. Toward a Realistic Model of Mutations Affecting Fitness. Evolution. 57(3):683–685. doi:10.1111/j.0014-3820.2003.tb01561.x. [accessed 2019 Feb 1]. http://onlinelibrary.wiley.com/doi/abs/10.1111/j.0014-3820.2003.tb01561.x.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge (UK): Cambridge University Press.

Kimura M, Maruyama T, Crow JF. 1963. The mutation load in small populations. Genetics. 48(10):1303–1312. doi:10.1093/genetics/48.10.1303. [accessed 2024 Mar 8]. https://academic.oup.com/genetics/article/48/10/1303/6033817.

Kostka D, Hubisz MJ, Siepel A, Pollard KS. 2012. The Role of GC-Biased Gene Conversion in Shaping the Fastest Evolving Regions of the Human Genome. Mol Biol Evol. 29(3):1047–1057. doi:10.1093/molbev/msr279. [accessed 2025 Jun 8]. https://doi.org/10.1093/molbev/msr279.

Kutschera VE, Poelstra JW, Botero-Castro F, Dussex N, Gemmell NJ, Hunt GR, Ritchie MG, Rutz C, Wiberg RAW, Wolf JBW. 2020. Purifying Selection in Corvids Is Less Efficient on

Islands. Mol Biol Evol. 37(2):469–474. doi:10.1093/molbev/msz233. [accessed 2020 Mar 2]. http://academic.oup.com/mbe/article/37/2/469/5601620.

Lande R. 1994. Risk of population extinction from fixation of new deleterious mutations. Evolution. 48(5):1460–1469. doi:10.2307/2410240. [accessed 2017 Jun 8]. http://www.jstor.org/stable/2410240.

Lanfear R, Kokko H, Eyre-Walker A. 2014. Population size and the rate of evolution. Trends Ecol Evol. 29(1):33–41. doi:10.1016/j.tree.2013.09.009. [accessed 2015 Apr 28]. http://www.sciencedirect.com/science/article/pii/S0169534713002322.

Leroy T, Rousselle M, Tilak M-K, Caizergues AE, Scornavacca C, Recuerda M, Fuchs J, Illera JC, De Swardt DH, Blanco G, et al. 2021. Island songbirds as windows into evolution in small populations. Curr Biol. 31(6):1303-1310.e4. doi:10.1016/j.cub.2020.12.040. [accessed 2025 Jul 25]. https://www.sciencedirect.com/science/article/pii/S0960982220318923.

Lynch M, Conery J, Burger R. 1995. Mutation Accumulation and the Extinction of Small Populations. Am Nat. 146(4):489–518. doi:10.1086/285812. [accessed 2024 Mar 8]. https://www.journals.uchicago.edu/doi/10.1086/285812.

Marais GAB. 2007. The Hill-Robertson effects extend from nucleus to mitochondria. Heredity. 99:357–358.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303. doi:10.1101/gr.107524.110. [accessed 2016 Sep 1]. http://genome.cshlp.org/content/20/9/1297.

Mugal CF, Kutschera VE, Botero-Castro F, Wolf JBW, Kaj I. 2020. Polymorphism Data Assist Estimation of the Nonsynonymous over Synonymous Fixation Rate Ratio ω for Closely Related Species. Mol Biol Evol. 37(1):260–279. doi:10.1093/molbev/msz203. [accessed 2020 Mar 2]. http://academic.oup.com/mbe/article/37/1/260/5556818.

Muller HJ. 1950. Our load of mutations. Am J Hum Genet. 2(2):111–176.

Ohta T. 1992. The Nearly Neutral Theory of Molecular Evolution. Annu Rev Ecol Evol Syst. 23(1):263–286. doi:10.1146/annurev.es.23.110192.001403. http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.es.23.110192.001403.

Pouyet F, Aeschbacher S, Thiéry A, Excoffier L. 2018. Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. Veeramah K, Wittkopp PJ, Gronau I, editors. eLife. 7:e36317. doi:10.7554/eLife.36317. [accessed 2025 Jun 8]. https://doi.org/10.7554/eLife.36317.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. Nat Genet. 46(8):919–925. doi:10.1038/ng.3015. [accessed 2016 Mar 30]. http://www.nature.com/ng/journal/v46/n8/full/ng.3015.html.

Sibley CG, Ahlquist JE. 1990. Phylogeny and Classification of the Birds: A Study in Molecular Evolution. New Haven (CT): Yale University Press.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, et al. 2015. Stable recombination hotspots in birds. Science. 350(6263):928–932. doi:10.1126/science.aad0843. [accessed 2016 Feb 9]. http://science.sciencemag.org/content/350/6263/928.

Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-Wide Evidence for Efficient Positive and Purifying Selection in Capsella grandiflora, a Plant Species with a Large Effective Population Size. Mol Biol Evol. 27(8):1813–1821. doi:10.1093/molbev/msq062. [accessed 2014 Sep 23]. http://mbe.oxfordjournals.org/content/27/8/1813.

Smeds L, Qvarnström A, Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. Genome Res. 26(9):1211–1218. doi:10.1101/gr.204669.116. [accessed 2018 Jul 30]. http://genome.cshlp.org/content/26/9/1211.

Soulé ME, Mills LS. 1998. No need to isolate genetics. Science. 282(5394):1658–1659. doi:10.1126/science.282.5394.1658. [accessed 2019 Mar 27]. http://science.sciencemag.org/content/282/5394/1658.

Stapley J, Birkhead TR, Burke T, Slate J. 2010. Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. Genome Res. 20(4):496–502. doi:10.1101/gr.102095.109.

Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JBW. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat Commun. 7:13195. doi:10.1038/ncomms13195.

Webster MT, Hurst LD. 2012. Direct and indirect consequences of meiotic recombination: implications for genome evolution. Trends Genet. 28(3):101–109. doi:10.1016/j.tig.2011.11.002. [accessed 2025 Jul 25]. https://www.sciencedirect.com/science/article/pii/S0168952511001867.

Williamson RJ, Josephs EB, Platts AE, Hazzouri KM, Haudry A, Blanchette M, Wright SI. 2014. Evidence for Widespread Positive and Negative Selection in Coding and Conserved Noncoding Regions of Capsella grandiflora. PLOS Genet. 10(9):e1004622. doi:10.1371/journal.pgen.1004622. [accessed 2025 Jul 10]. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004622.

Woolfit M, Bromham L. 2005. Population size and molecular evolution on islands. Proc R Soc B Biol Sci. 272(1578):2277–2282. doi:10.1098/rspb.2005.3217. [accessed 2019 Jan 31]. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560185/.

Zuur AF, Ieno EN, Walker N, Saveliev AA, Smith GM. 2009. Mixed effects models and extensions in ecology with R. New York, NY: Springer (Statistics for Biology and Health).

Figures

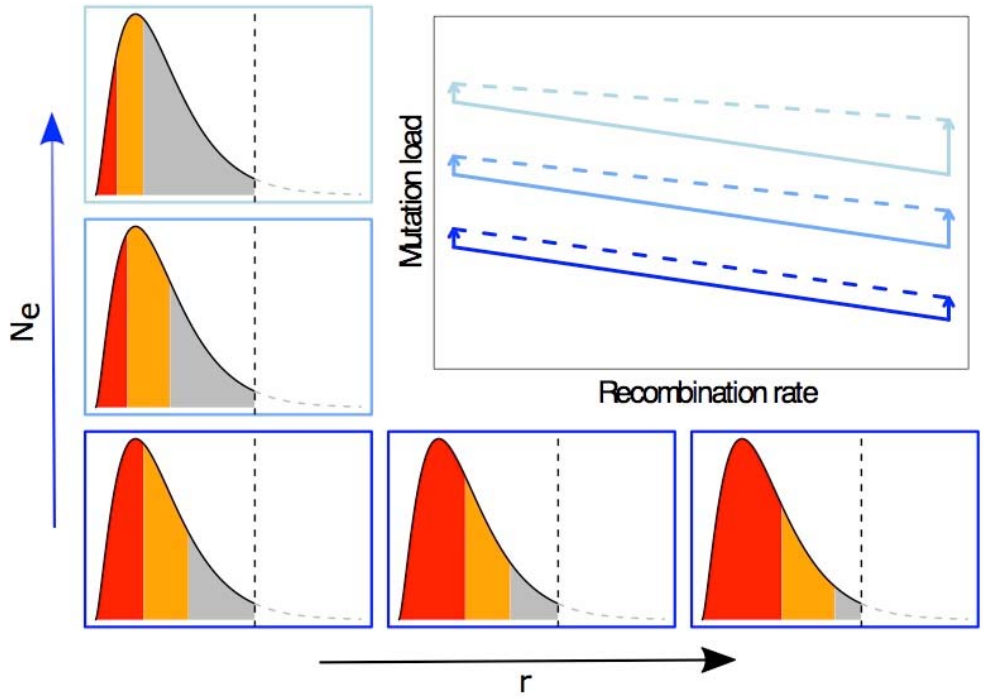

Figure 1 | Theoretical considerations on mutation load. The curves outlined in black illustrate a mock example of a genome-wide distribution of fitness effects (DFE) at a given point in the fitness landscape. We here only consider the fraction of mutations with s<0 reducing fitness (left of the vertical dashed line marking *s*=0). The efficacy of selection in removing those crucially depends on the effective population size N_e (left column) and the recombination rate *r* (bottom row) changing the fraction of mutations behaving as strongly deleterious (red: $N_e s < -10$), deleterious (orange: $N_e s \in (-10;-1]$), or mildly deleterious (grey: $N_e s \in (-1;0]$) which are the focus of this study. *Central figure, solid lines:* The y-axis shows the fraction of all mutations falling into the mildly deleterious category $N_e s \in (-1;0]$. This corresponds to the mutation load which is expected to be highest in populations with small N_e. Mutation load is further expected to decrease in genomic regions of elevated recombination rate (x-axis). Line colours correspond to N_e categories small (light blue), intermediate (blue) and large (dark blue). gBGC is a recombination-associated process

expected to increase mutation load overall, and particularly so in regions of high recombination *(dashed lines*). Difference in slope between solid (without gBGC) and dashed lines (with gBGC) is expected if the relative rate by which recombination introduces mildly deleterious mutations through gBGC differs from the rate by which recombination increases the efficacy of selection. Such difference in relative rates of gBGC and purging efficacy may further interact with the effective population size (length of arrows) mediating the overall and/or recombination-dependent contribution of gBGC to mutation load.

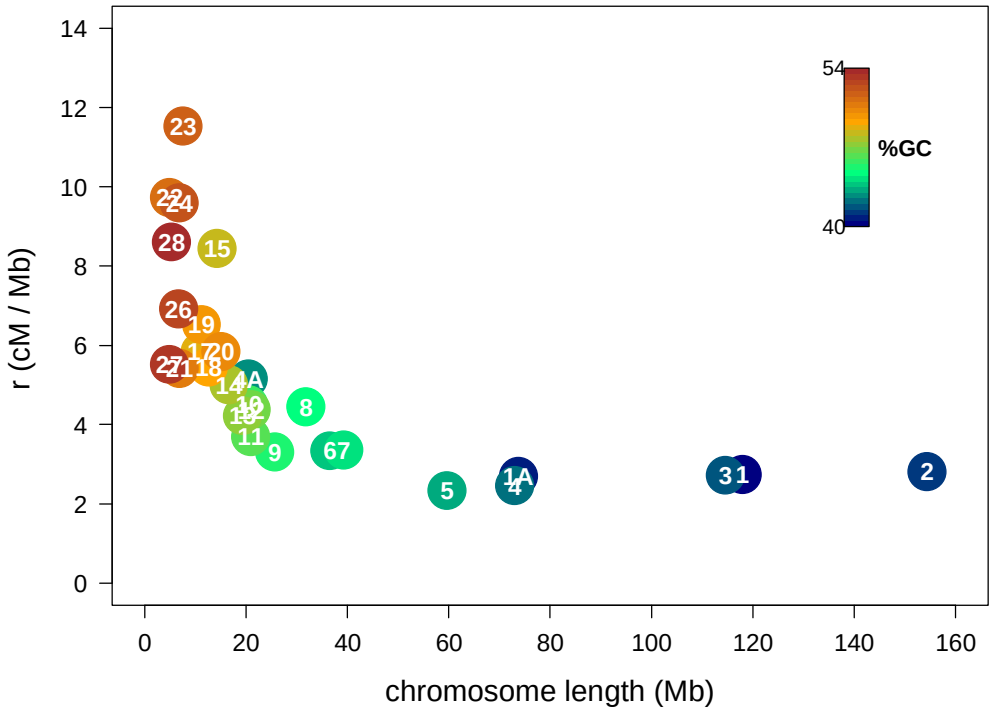

Figure 2 | Relationship between recombination rate [cM/Mb], chromosome length [Mb] and GC content [%] in Eurasian crows, *Corvus (corone) ssp*. A near 5-fold difference in recombination rate is closely associated with chromosome length which strongly covaries with GC content (R^2=0.90, p<0.001). In small, highly recombining chromosomes GC content is elevated by up to 15% reflecting an evolutionary history of gBGC.

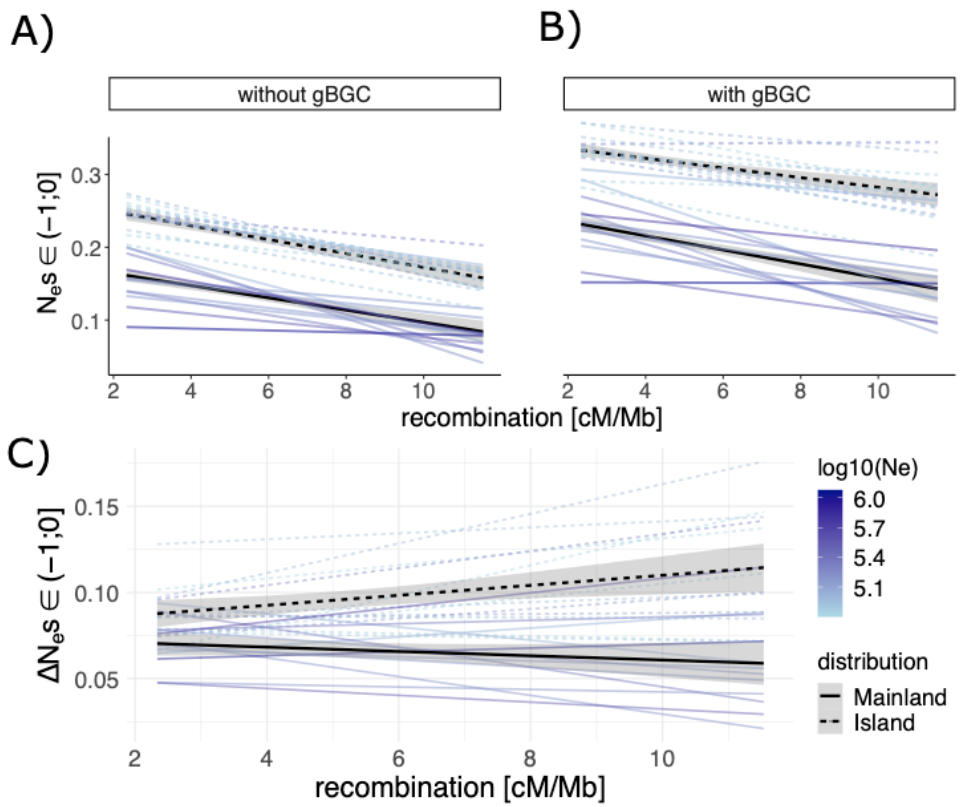

Figure 3. Relationship between genomic variation in recombination rate and mutation load estimated as the fraction of mildly deleterious mutations with $N_e s \in [-1;0)$. Correlation between recombination rate and mutation load A) excluding and B) including mutations subject to gBGC (right). Each regression line represents a population sample of a species differing in effective population sizes (N_e) as is indicated by the blue colour gradient on a logarithmic scale. Dashed and solid lines illustrate island and mainland populations, respectively. C) as above, but the the y-axis displays the difference between estimates of mutation load with and without mutations subject to gBGC. Note that the slopes differ between island and mainland populations indicating that relative rates of the two recombination-mediated processers, gBGC and purging efficacy, may interact with the effective population size of a species.

Table 1: Summary of the four best statistical models and the null model exploring the relationship between the mutation load quantified as the proportion of mildly deleterious alleles (DFE) with recombination rate (r), effective population size (Ne) and GC-biased gene conversion (gBGC). *r* was here approximated by chromosome length, N_e was here approximated by whether a population was restricted to an island (yes/no), gBGC was estimated for all sites (with gBGC) and for a subset of sites excluding weak-to-strong mutations (without gBCG). For other proxies of r and N_e and the full set of models with parameter estimates see **Supplementary Table 3**. Model selection is based on Akaike's Information criterion (AIC) with models by $\Delta AIC_c<2$ are considered supported (best model in bold). In addition, thresholds of Type 1 error probabilities of p<0.001 (***), p<0.01 (**), p<0.05 (*) and p>0.05 (n.s) are indicated for all fixed effects and their interactions included in a specific model.

Model	r [cM/Mb]	Ne [yes/no]	gBGC [with/without]	Ne*gBGC	r*Ne	r*gBGC	r*Ne*gBGC	k	AIC	ΔAIC
10	*****	*****	*****	*****	n.s.	n.s.	*****	**13**	**-4479.5754**	**0**
9	*****	*****	*****	*****	n.s.	n.s.	-	12	-4477.2763	2.299
8	*****	*****	*****	*****	-	-	-	10	-4476.8571	2.718
7	*****	*****	*****	-	-	-	-	11	-4440.2616	39.31
NULL	-	-	-	-	-	-	-	6	-3555.9041	923.671