Purifying Selection in Corvids Is Less Efficient on Islands

Verena E. Kutschera ,*[,1,2] Jelmer W. Poelstra,[3] Fidel Botero-Castro ,[4] Nicolas Dussex,[†,5,6] Neil J. Gemmell,[†,5] Gavin R. Hunt,[†,7] Michael G. Ritchie,[†,8] Christian Rutz,[†,8] R. Axel W. Wiberg,[†,8,9] and Jochen B.W. Wolf *[,1,4]

[1]Department of Evolutionary Biology, Uppsala University, Uppsala, Sweden

[2]Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

[3]Biology Department, Duke University, Durham, NC, USA

[4]Division of Evolutionary Biology, Faculty of Biology, LMU Munich, Planegg-Martinsried, Germany

[5]Department of Anatomy, University of Otago, Dunedin, New Zealand

[6]Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

[7]Unaffiliated

[8]Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, United Kingdom

[9]Department of Environmental Sciences, Evolutionary Biology, University of Basel, Basel, Switzerland

[†]Listed in alphabetical order.

***Corresponding authors:** E-mails: verena.kutschera@scilifelab.se; j.wolf@bio.lmu.de.

Associate editor: Stephen Wright

Abstract

Theory predicts that deleterious mutations accumulate more readily in small populations. As a consequence, mutation load is expected to be elevated in species where life-history strategies and geographic or historical contingencies reduce the number of reproducing individuals. Yet, few studies have empirically tested this prediction using genome-wide data in a comparative framework. We collected whole-genome sequencing data for 147 individuals across seven crow species (*Corvus* spp.). For each species, we estimated the distribution of fitness effects of deleterious mutations and compared it with proxies of the effective population size N_e. Island species with comparatively smaller geographic range sizes had a significantly increased mutation load. These results support the view that small populations have an elevated risk of mutational meltdown, which may contribute to the higher extinction rates observed in island species.

Key words: **molecular evolution, distribution of fitness effects, comparative analysis, avian genomics, mutation load, selection.**

Letter

Introduction

The fate of a mutation entering a population depends on the effect it exerts on the fitness of its carrier. Under the assumption that most changes to the DNA are detrimental, selection is expected to primarily act to remove deleterious alleles (Elena et al. 1998; Keightley and Lynch 2003). Lethal or strongly deleterious mutations have almost no possibility of becoming fixed in a population, but the retention of weakly deleterious mutations depends on the effective population size, N_e (Charlesworth 2009; Akashi et al. 2012). In populations where a large number of individuals contribute genes to the next generation, purifying selection is expected to be efficient relative to genetic drift. In small populations, however, weakly deleterious mutations can rise to high frequencies, elevating the risk of population extinction (Lande 1994; Soulé and Mills 1998). Island species are not only constrained by limited geographic range, and hence relatively small population size, but also by limited opportunities to compensate for local population crashes through immigration from larger source populations (Frankham 2015). Consistent with these theoretical expectations, extinction rates of island populations are elevated in comparison to their mainland counterparts (Frankham 1998).

The distribution of fitness effects of deleterious mutations (hereafter simply referred to as DFE) is a key evolutionary parameter describing the interplay between purifying selection and genetic drift (Eyre-Walker and Keightley 2007). It quantifies the proportion of harmful mutations segregating in a population and thereby provides a useful measure of mutation load. By convention the DFE is conceptualized as discrete classes of mutations scaled by selection coefficients (s) and population size (N). Mutations in the ranges of $Ns \in [1;10)$ and $Ns > 10$ include deleterious and strongly deleterious mutations, respectively. Mutations with $Ns \in [0;1)$ include slightly deleterious mutations ($Ns \sim 1$) as well as the class of neutral and effectively neutral mutations ($Ns \ll 1$); we refer to this class of mutations jointly as "mildly deleterious" (cf. Deinum et al. 2015). Note that, we report the absolute value of population-scaled selection coefficients with higher values representing ever more deleterious effects.

Despite the conceptual importance of the DFE, empirical data on its shape are rare (but see Huber et al. 2017; Castellano et al. 2019), and its biological determinants remain elusive. Insights come from mutagenesis or mutation accumulation experiments (Elena et al. 1998; Keightley and Lynch 2003) and indirect inference from sequencing data (Keightley and Eyre-Walker 2010; Racimo and Schraiber 2014; Deinum et al. 2015). Consistent with a central prediction of the nearly neutral theory, the proportion of mildly deleterious mutations tends to scale negatively with proxies for N_e (Slotte et al. 2010; Deinum et al. 2015; Chen et al. 2017). Because long-term N_e is generally unknown, it has been approximated by geographic range size (Leffler et al. 2012) or a number of life-history traits, including body mass, age of sexual maturity, fecundity, or propagule size (Romiguier et al. 2014; Figuet et al. 2016; Chen et al. 2017).

Here, we estimated the DFE from whole-genome sequencing data of seven avian species within the genus *Corvus* (fig. 1*A* and *C*). The data set includes five widely distributed mainland species and two species that are restricted to small tropical islands. Crows and jackdaws from North America (*C. brachyrhynchos*), Eurasia (*C.* (*corone*) spp., *C. dauuricus*, *C. monedula*), or the Indian subcontinent (*C. splendens*) inhabit large ranges, encompassing on an average $27\times10^6\,km^2$ (range: $5–82\times10^6\,km^2$). In contrast, New Caledonian and white-billed crows (*C. moneduloides*, *C. woodfordi*) evolved on small islands ($0.02–0.05\times10^6\,km^2$) and remained separated from their closest mainland relatives for millions of years (Haring et al. 2012; Jønsson et al. 2012, 2016). The species also vary in body size (length: 36–53 cm, mass: 208–570 g) (Dunning 1992; del Hoyo et al. 2017), which may serve as an additional indicator of long-term N_e (Figuet et al. 2016). We expect from theory that mutation load should be elevated in large-bodied species or in species with small geographic ranges.

Results

We estimated the DFE for population samples of all seven species from segregating sites with reliable genotype information for least eight chromosomal copies per species (fig. 1*A*). In order to control for demographic perturbation, we contrasted 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 (2-epoch model; Keightley and Eyre-Walker 2007). The inferred proportion of mildly deleterious sites $Ns \in [0;1)$ ranged from 0.178 to 0.221 in the species with broad geographic ranges, and was higher in the two island species *C. moneduloides* (0.252) and *C. woodfordi* (0.332) (supplementary table S1, Supplementary Material online). Statistical models including origin (mainland vs. island) were most strongly supported, based on Akaike's information criterion for small sample sizes ($\Delta AIC_c \leq 2$) (table 1 and fig. 1*B*). Summing over the conditional model probabilities, origin had the strongest effect ($wAIC_c = 0.665$), followed by geographic range size ($wAIC_c = 0.142$), and body size with the least support ($wAIC_c = 0.008$). This result suggests that colonization of islands with subsequent persistence in a confined geographic area may indeed limit long-term effective size of a population, and hence reduce the efficacy of selection (Corbett-Detig et al. 2015). Consistent with this expectation, long-term N_e, as independently approximated by the harmonic mean of the cross-coalescence rate through time (Schiffels and Durbin 2014), was higher for individuals of the widely distributed species *C.* (*c.*) *cornix* ($N_e = 88{,}219$) and *C. brachyrhynchos* ($N_e = 122{,}430$) than for the island species *C. moneduloides* ($N_e = 63{,}236$).

The reliance of the DFE on estimates of genetic variation at both selected as well as selectively neutral sites makes it sensitive to perturbation in the "neutral SFS" introduced by factors other than selection. Processes such as demographic change or population structure can accordingly modulate the SFS, mimicking the outcome of selection. For validation of the results above, we therefore explored the effects of sample size, population structure, and demographic population history using a large test data set of 118 individuals from the European crow species complex (Vijay et al. 2016). Without applying any correction for the distribution of the "neutral SFS" (1-epoch model; Keightley and Eyre-Walker 2007) the DFE varied substantially by sample size and population. However, applying model-based correction using an explicit demographic parametrization (2-epoch model) stabilized the DFE across a large range of sample sizes (3–118 individuals), demographic histories, and population stratification, including artificial admixture by pooling across all populations (supplementary text, table S2, and figs. S1 and S2, Supplementary Material online). Moreover, different methods controlling for perturbation in the "neutral SFS" yielded very similar estimates for the proportion of mildly deleterious sites in all seven species (supplementary table S1, Supplementary Material online). Results from the 2-epoch model, presented above, were near-identical to estimates using an unspecified "nuisance parameter" in a diffusion approximation framework (Eyre-Walker et al. 2006), or when simply considering the ratio of nucleotide diversity of 0- and 4-fold degenerate sites (π_0/π_4) (Pearson's $r = 0.95–0.99$; supplementary table S3, Supplementary Material online). As for the test data set, applying model-based correction (2-epoch model) also generally improved the fit to the data for the remaining six species (supplementary table S1, Supplementary Material online). Estimates derived both from π_0/π_4 and the method of Eyre-Walker et al. (2006) (supplementary table S1, Supplementary Material online) likewise identified island origin and small geographic range size as the main factors associated with increased mutation load (supplementary table S4 and fig. S3, Supplementary Material online).

Species samples were not evenly distributed across the phylogeny (fig. 1). Although phylogenetic conservatism for N_e is not expected, shared ancestry may contribute to explaining variation in the DFE across species and lead to overconfident statistical inference. We therefore quantified the degree of shared polymorphism as a proxy for the contribution of shared ancestral population signatures to the DFE of individual species. Proportions were nonzero for the closest species pairs, but overall low suggesting that the species under

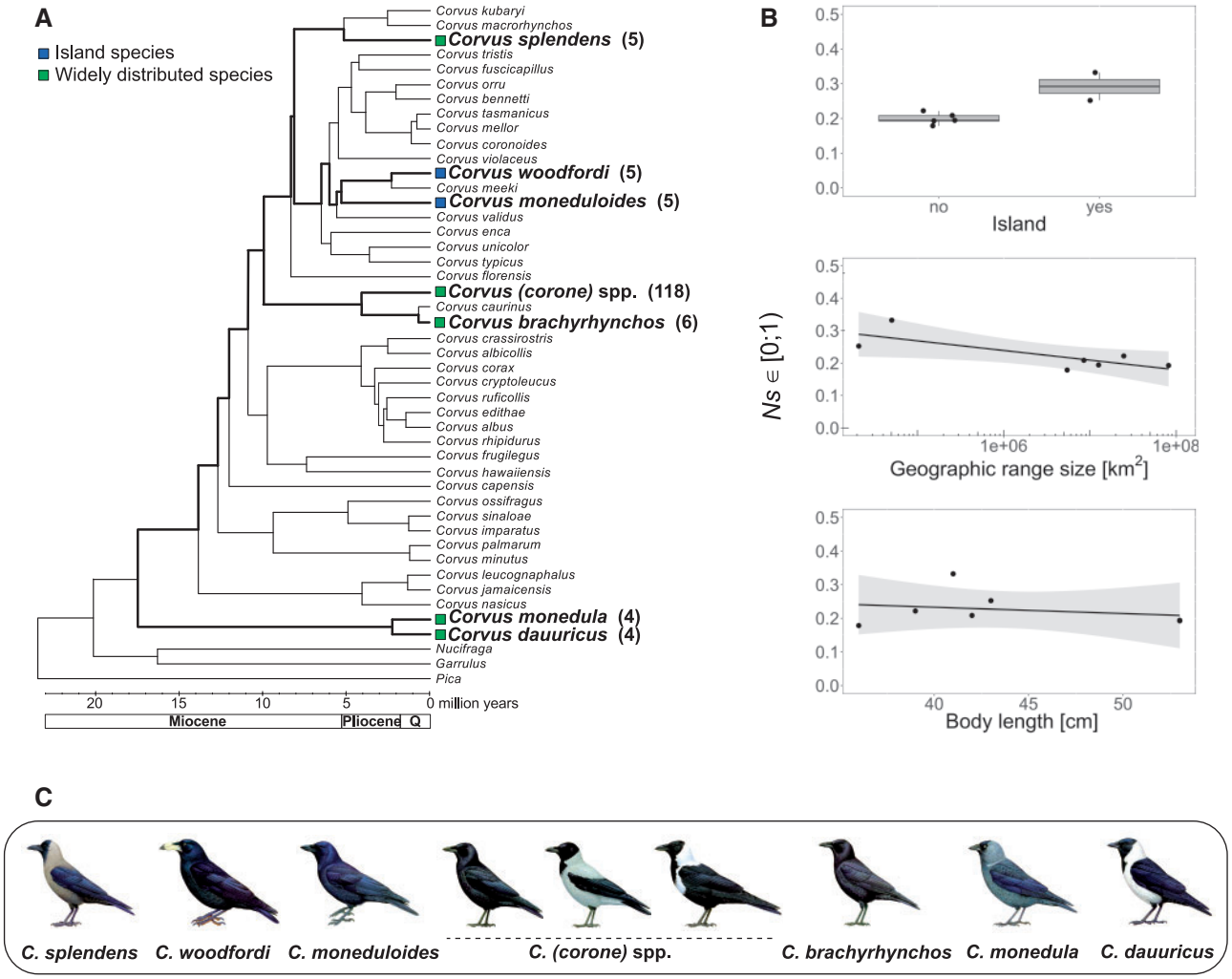

Fig. 1. Study system and mutation load. (*A*) Phylogeny of the genus *Corvus*, redrawn after Jønsson et al. (2012). Species included in this study are highlighted in bold, and numbers in brackets indicate the number of whole-genome resequenced individuals. The geographic origin is indicated by color (blue, island; green, continent). For the purpose of this article, we treat the up to five taxonomic groups that have been recognized within the *Corvus (corone)* species complex (Parkin et al. 2003) as a single species sharing recent ancestry with a substantial amount of cosegregating genetic variation (Vijay et al. 2016). (*B*) Mutation load of a species, estimated as the proportion of mildly deleterious mutations $Ns \in [0;1)$, differs by geographic origin (island vs. continent), and varies to a lesser extent with distributional range; the effect of body size is not statistically significant. (*C*) Species illustrations courtesy of *Handbook of the Birds of the World Alive.*

Table 1. Summary of All Statistical Multiple Linear Regression Models (numbered 0–7) Exploring Relationships between Mutation Load $Ns \in [0;1)$, As Estimated by the 2-Epoch Model of DFEalpha (Keightley and Eyre-Walker 2007), Geographic Range Size, Body Length, and the Mainland-Island Contrast.

Model	Type I Error				*k*	AIC_c	ΔAIC_c	wAIC	R^2_{adj}
	Intercept	Geographic Range Size	Body Length	Origin [mainland/island]					
3	0.000			0.013	3	−17.936	0	0.663	0.693
0	0.000				2	−15.404	2.531	0.187	0.000
1	0.002	0.041			3	−14.835	3.1	0.141	0.521
2	0.100		0.610		3	−8.807	9.129	0.007	−0.133
5	0.931	0.483		0.150	4	−4.908	13.028	0.001	0.665
4	0.069		0.782	0.029	4	−4.084	13.852	0.001	0.624
6	0.026	0.079	0.940		4	−0.847	17.089	0	0.402
7	0.935	0.357	0.461	0.153	5	35.6	53.536	0	0.640
Σ wAIC		0.142	0.008	0.665					

Note.—Type I errors are given for each model and response variable. The best supported model is highlighted in gray.
k, number of parameters; AIC_c, Akaike's information criterion adjusted for small sample sizes; ΔAIC_c, difference in AIC_c relative to the best model; $wAIC_c$, Akaike weights for small sample sizes; $\Sigma wAIC_c$, summed Akaike weights; R^2_{adj}, squared correlation coefficient, adjusted for the number of parameters.

consideration have largely reached evolutionary independence (supplementary table S5, Supplementary Material online). Moreover, reanalyses of the best-supported models (table 1) with explicit inclusion of phylogenetic distance showed no evidence for a phylogenetic contribution to the inference. Models with parameter estimates of zero for the covariance structure of the error received the highest likelihood.

Discussion

In this study, we took a comparative genomic approach to address the relationship between mutation load and population size proxies. Using genome-wide data from 118 individuals of the *Corvus (corone)* spp. species complex, we established that a moderate sample of less than ten chromosomal copies contains reliable information characterizing the species' DFE. Model-based correction of factors other than selection perturbing the "neutral SFS" stabilized the DFE across populations and demographic histories. Population stratification in the *Corvus (corone)* spp. species complex is moderate, and comparable to other species (F_{ST} range: autosomes, 0.02–0.25; sex chromosome, 0.04–0.51) (Poelstra et al. 2014). Populations within this species complex coalesce more recently with each other than with the American sister species *C. brachyrhynchos* (Haring et al. 2012; Vijay et al. 2016). It is thus reasonable to assume that representative population samples sharing recent common ancestry approximate the selective processes acting during much of the history of the species as a whole.

Under this premise, we quantified mutation load as the proportion of mildly deleterious sites ($Ns \in [0;1)$), and related it to life-history traits for all seven species in our data set. The use of genome-scale data for multiple individuals of several species, though moderate in number, allowed for statistical treatment in a comparative framework. Moreover, in contrast to studies leveraging information on the degree of purifying selection in natural populations from broad taxonomic sampling (Chen et al. 2017), we focused our analyses on a single avian genus. This allowed us to explore the effects of N_e as approximated by variation in body size and geographic range size, while keeping variation in other relevant life-history traits—such as the genetic system or propagule size—to a minimum (Owens and Bennett 1995). It also circumvents phylogenetic inertia that may otherwise hamper the interpretation of differences in DFE across large evolutionary distances. With due caution concerning the relatively small number of species, we found a statically supported relationship between the mutation load of a species and its distributional range. This finding is in accordance with the observation that species with large population sizes appear to purge deleterious mutations more efficiently (Glémin 2003). Furthermore, our study provides evidence that species living on islands accumulate mildly deleterious mutations more readily than more widely distributed species (Johnson and Seger 2001; Woolfit and Bromham 2005; Gardiner et al. 2008). Similar to what has been reported elsewhere for the degree of genetic polymorphism (James et al. 2016), the difference between island and continental species with respect to the inferred mutation load was relatively small. Moreover, island species show a higher degree of variation (supplementary table S1, Supplementary Material online). Although this observation clearly needs to be substantiated for a larger sample of species, it may suggest a disproportionate effect of specific colonization history (bottleneck, degree of postcolonization gene flow) that is difficult to capture with simple proxies for N_e.

There have been recent claims (Kern and Hahn 2018) and counter-claims (Jensen et al. 2019) about the extent to which comparative genomics studies are rejecting the explanatory power of the neutral theory. Our findings are consistent with the central prediction from nearly neutral theory of negative scaling between the proportion of deleterious mutations and effective population size (Ohta 1992). Our results further imply that populations of mainland species with large distributional ranges harbor fewer deleterious mutations than their island congeners making the latter potentially more vulnerable to mutational meltdown (Frankham 1998). To corroborate the generality of this observation studies exploring the distribution of fitness effects in a large number of closely related species across a variety of taxa are encouraged.

Materials and Methods

A detailed account of the methods can be found in the supplementary text, Supplementary Material online.

Data Collection and Processing

We collated whole-genome resequencing data for a total of 147 individuals from population samples of seven *Corvus* species (see fig. 1 and supplementary table S6, Supplementary Material online). After quality assessment and adapter trimming, sequencing reads were aligned with bwa v0.7.13 (Li and Durbin 2009) to the closest available reference genome, either *C. moneduloides* (accession number: VRTO00000000) or *C. (corone) cornix* (Poelstra et al. 2014). Syntenic regions between the two reference assemblies were determined from lift-overs using SatsumaSynteny (Grabherr et al. 2010). Variant discovery and genotyping were performed using GATK v3.4.0 (Auwera et al. 2013). CpG-prone sites and sites with missing data were removed from *vcf* files.

Estimation of the DFE

We approximated the DFE by a gamma distribution using DFEalpha v.2.15 (Keightley and Eyre-Walker 2007) and DoFE v3.1 (Eyre-Walker et al. 2006). We further calculated the ratio of nucleotide diversity at nonsynonymous, 0-fold degenerate sites, and at synonymous, 4-fold degenerate sites (π_0/π_4), reflecting the strength of selection (Charlesworth and Eyre-Walker 2008).

Life-History Parameters, Demographic Inference, and Statistical Model Selection

For each species, we extracted data on geographical range size and mean body size (see supplementary table S6, Supplementary Material online, for data and references). Body length and body mass showed a high degree of collinearity ($r = 0.88$) and we report results for the former. Body

mass, however, yielded qualitatively the same results (supplementary table S4d, Supplementary Material online). Statistical analyses of relationships between the DFE and life-history parameters were based on model selection of multiple linear regression models using Akaike's information criterion (Burnham and Anderson 2002) and were performed in R version 3.5.1 (R Core Team 2018). To test for possible effects of genealogical nonindependence, we fitted a linear model allowing for correlation of errors as implemented in the package nlme v. 3.1-131.1 (Pinheiro et al. 2017). The correlation structure was specified by the phylogenetic distance matrix (based on both Jønsson et al. 2012, 2016) and modeled assuming Brownian motion (Freckleton et al. 2002). For all individuals with a minimum average sequencing coverage of 18×, we also estimated demographic history through time, using the program MSMC (Schiffels and Durbin 2014).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Carina Mugal for valuable discussions and input on the article. Funding was provided by the European Research Council (ERCStG-336536 FuncSpecGen to J.B.W.W.), the Swedish Research Council Vetenskapsrådet (621-2013-4510 to J.B.W.W.), the Knut and Alice Wallenberg Foundation (to J.B.W.W.), the Lawski foundation (to V.E.K. and J.B.W.W.), the German Research Foundation (KU 3402/1-1 to V.E.K.), the UK's Biotechnology and Biological Sciences Research Council (BB/G023913/2 to C.R.), and the New Zealand Marsden Fund (to G.R.H.). Analyses were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

Author Contributions

V.E.K. and J.B.W.W. conceived of the study and wrote the article. V.E.K. performed all bioinformatic analyses with help from N.D. and R.A.W.W. for data preprocessing and F.B.C. for estimates of shared polymorphism. J.W.P. performed sequentially Markovian coalescent analyses, J.B.W.W. conducted downstream statistical analyses. N.G., C.R., G.R.H., and M.G.R. contributed funding, samples, genomic data, technical expertise, and input on the draft article.

References

Akashi H, Osada N, Ohta T. 2012. Weak selection and protein evolution. *Genetics* 192(1):15–31.

Auwera GAV der, Carneiro MO, Hartl C, Poplin R, Angel G del, Levy–Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr Protoc Bioinformatics*. 43:11.10.1–11.10.33.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York: Springer Verlag.

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* 302494.2019.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet*. 10(3):195–205.

Charlesworth J, Eyre-Walker A. 2008. The McDonald–Kreitman test and slightly deleterious mutations. *Mol Biol Evol*. 25(6):1007–1015.

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.

Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol*. 13(4):e1002112.

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.

del Hoyo J, Elliott, A, Sargatal J, Christie DA, de Juana E, editors. 2017. Handbook of the birds of the world alive. Barcelona: Lynx Edicions. Available from: http://www.hbw.com.

Dunning JB. 1992. CRC handbook of avian body masses. Boca Raton, FL: CRC Press LLC.

Elena SF, Ekunwe L, Hajela N, Oden SA, Lenski RE. 1998. Distribution of fitness effects caused by random insertion mutations in *Escherichia coli*. *Genetica* 102:349.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet*. 8(8):610–618.

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.

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(6):1517–1527.

Frankham R. 1998. Inbreeding and extinction: island populations. *Conserv Biol*. 12(3):665–675.

Frankham R. 2015. Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Mol Ecol*. 24(11):2610–2618.

Freckleton RP, Harvey PH, Pagel M, Losos A. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat*. 160(6):712–726.

Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG. 2008. *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol Ecol*. 17(7):1648–1657.

Glémin S. 2003. How are deleterious mutations purged? Drift versus nonrandom mating. *Evolution* 57(12):2678–2687.

Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, Palma FD, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* 26(9):1145–1151.

Haring E, Däubl B, Pinsker W, Kryukov A, Gamauf A. 2012. Genetic divergences and intraspecific variation in corvids of the genus *Corvus* (Aves: Passeriformes: Corvidae) – a first survey based on museum specimens. *J Zool Syst Evol Res*. 50(3):230–246.

Huber CD, Kim BY, Marsden CD, Lohmueller KE. 2017. Determining the factors driving selective effects of new nonsynonymous mutations. *Proc Natl Acad Sci U S A*. 114(17):4465–4470.

James JE, Lanfear R, Eyre-Walker A. 2016. Molecular evolutionary consequences of island colonization. *Genome Biol Evol*. 8(6):1876–1888.

Jensen JD, Payseur BA, Stephan W, Aquadro CF, Lynch M, Charlesworth D, Charlesworth B. 2019. The importance of the neutral theory in 1968 and 50 years on: a response to Kern and Hahn 2018. *Evolution* 73(1):111–114.

Johnson KP, Seger J. 2001. Elevated rates of nonsynonymous substitution in island birds. *Mol Biol Evol*. 18(5):874–881.

Jønsson KA, Fabre P-H, Irestedt M. 2012. Brains, tools, innovation and biogeography in crows and ravens. *BMC Evol Biol*. 12:72.

Jønsson KA, Fabre P-H, Kennedy JD, Holt BG, Borregaard MK, Rahbek C, Fjeldså J. 2016. A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). *Mol Phylogenet Evol*. 94:87–94.

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.

Keightley PD, Eyre-Walker A. 2010. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philos Trans R Soc B*. 365(1544):1187–1193.

Keightley PD, Lynch M. 2003. Toward a realistic model of mutations affecting fitness. *Evolution* 57(3):683–685.

Kern AD, Hahn MW. 2018. The neutral theory in light of natural selection. *Mol Biol Evol*. 35(6):1366–1371.

Lande R. 1994. Risk of population extinction from fixation of new deleterious mutations. *Evolution* 48(5):1460–1469.

Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, Andolfatto P, Przeworski M. 2012. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol*. 10(9):e1001388.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.

Ohta T. 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst*. 23(1):263–286.

Owens IPF, Bennett PM. 1995. Ancient ecological diversification explains life-history variation among living birds. *Proc R Soc Lond B Biol Sci*. 261:227–232.

Parkin DT, Collinson M, Helbig AJ, Knox AG, Sangster G. 2003. The taxonomic status of Carrion and Hooded Crows. *Br Birds*. 96:274–290.

Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. 2017. nlme: linear and nonlinear mixed effects models. Available from: https://CRAN.R-project.org/package=nlme.

Poelstra JW, Vijay N, Bossu C, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr M, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344(6190):1410–1414.

R Core Team. 2018. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

Racimo F, Schraiber JG. 2014. Approximation to the distribution of fitness effects across functional categories in human segregating polymorphisms. *PLoS Genet*. 10(11):e1004697.

Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenuil A, Chiari Y, Dernat R, Duret L, Faivre N, et al. 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515(7526):261–263.

Schiffels S, Durbin R. 2014. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 46(8):919–925.

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.

Soulé ME, Mills LS. 1998. No need to isolate genetics. *Science* 282(5394):1658–1659.

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

Woolfit M, Bromham L. 2005. Population size and molecular evolution on islands. *Proc R Soc B*. 272(1578):2277–2282.