Unlinked background selection reduces neutral diversity more than linked background selection

Joseph Matheson[1], Joanna Masel[1]

[1]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA

Corresponding author: Joanna Masel

ABSTRACT

Background selection describes the reduction in neutral diversity caused by selection against deleterious alleles at other loci. It is typically assumed that the purging of deleterious alleles affects linked neutral variants, and indeed simulations typically only treat a genomic window. However, background selection at unlinked loci also depresses neutral diversity. We find that for a realistically high genome-wide deleterious mutation rate, the effects of unlinked background selection exceed those of linked background selection. This can be derived from previous approximate analytical approximations of the effects of both kinds of background selection, and we confirm it with simulations that treat multi-locus complexities in a human-like genome. Background selection reduces neutral genetic diversity by a factor that is independent of census population size. Outside of genic regions, the strength of background selection increases with the mean selection coefficient, contradicting the linked theory but in agreement with the unlinked theory. Neutral diversity within genic regions is fairly independent of the strength of selection.

Keywords: linkage disequilibrium, population genetics, nearly neutral theory, forward time simulation, effective population size, expected heterozygosity

INTRODUCTION

The neutral theory of molecular evolution postulates that i) most genetic diversity observed in natural populations is neutral with respect to an organism's fitness (Kimura 1968; King and Jukes 1969), and ii) dynamics are well-described by models of a single neutral locus in a population of a specified "effective" population size (Ewens 2004; Charlesworth 2009; Masel 2011; Kern and Hahn 2018). A crucial extension of this theory uses a reduction in the effective population size to incorporate the fact that deleterious mutations depress genetic diversity as they are removed, in a process known as background selection (Charlesworth *et al.* 1993). While the removal of a deleterious allele will depress genetic diversity at neutral loci across the entire genome, neutral loci linked to the deleterious allele will be particularly affected (Charlesworth 2012). Indeed, the effect of unlinked background selection is often considered small enough to be ignored (Charlesworth 2012). Here we revisit this assumption given realistically high genome-wide deleterious mutation rates, by using simulations to evaluate previously derived equations for linked vs. unlinked background selection.

In practice, observed genetic diversity at putatively neutral sites is used to estimate the coalescent effective population size N_e as the census size of an idealized population that produces the same neutral genetic diversity given demography (Charlesworth 2009) or background selection (Charlesworth *et al.* 1993). In an idealized population obeying Wright-Fisher or Moran dynamics, genetic diversity depends only on the product of the neutral mutation rate at that locus and the census size of the population (Kimura 1969). Neutral theory considers mutations that either are strictly neutral or so deleterious that

they can be ignored. Nearly neutral theory retains the binary distinction between rapidly purged versus neutral mutations, but allows the ratio of mutations of these two types to vary among species. Both models assume many independent single loci.

The problem with this binary distinction and independent loci approach is that slightly deleterious mutations are purged only slowly from populations. During this removal process, they depress genetic variation in the genome (Charlesworth *et al.* 1993). This depression in genetic variation caused by background selection is typically modeled as a decrease in the effective population size for a neutral locus linked to deleterious variants (Hudson and Kaplan 1995; Lohmueller *et al.* 2011; Comeron 2014). In a population with no recombination and sufficiently large effect deleterious mutations ($N_e s \gg 1$), the coalescent effective population size would decrease from N_e to $f_0 N_e$, where f_0 is the equilibrium frequency of individuals with no deleterious mutations, because any neutral variants linked to deleterious variants would be doomed (Charlesworth *et al.* 1993).

Recombination can decouple neutral variants from deleterious variants, reducing the degree to which background selections depresses neutral variation (Cutter and Payseur 2013). For a single neutral locus linked to a single locus where deleterious mutations with fixed heterozygous effect size sh occur at rate u per diploid individual per generation, and with recombination between the loci occurring at rate r, heterozygosity at the neutral locus is reduced by a factor $F \approx 1 - \frac{ush}{2(sh+r)^2}$ (Hudson and Kaplan 1994). This result can be straightforwardly extended to any number of deleterious sites linked to the focal neutral site by assuming that mutation and recombination rates are uniform across a genomic window, and that there is linkage equilibrium among deleterious variants e.g. because multiple significantly linked deleterious mutations are not present at the same time (Hudson and Kaplan 1995; Nordborg *et al.* 1996). In this case, in a genomic window we have

$$\frac{N_e}{N} = e^{-U_w/2sh+R_w} \quad (1)$$

where U_w is the total diploid deleterious mutation rate across the entire window, and R_w is the map length between the ends of the window (Hudson and Kaplan 1995). With the same assumptions, similar results can be obtained in the case where both deleterious and beneficial mutations occur (Kim and Stephan 2000).

The results above do not apply across different chromosomes or given free recombination between the ends of the window (Hudson and Kaplan 1995). However, deleterious variants do still depress neutral diversity even at unlinked sites: at a bare minimum, a deleterious mutation will eliminate any unique neutral genetic variants in a single individual who dies due to that deleterious mutation. The ratio of N_e to N with free recombination has been derived as

$$\frac{N_e}{N} = e^{-8Ush} \qquad (2)$$

where U is the total deleterious mutation rate at unlinked loci and sh is the fitness effect of a deleterious allele when in the heterozygous state (Charlesworth 2012).

It has previously been presumed that the reduction in diversity due to unlinked deleterious loci will be much smaller than that from linked loci (Charlesworth 2012). That is, for any given neutral locus, even though U across all unlinked deleterious loci is much larger than U_w across the window of linked

deleterious loci, $1/(2sh + R_w)$ will be so much larger than $8sh$ that the latter comparison will overwhelm the former comparison, and linked background selection will dominate. Indeed, models of background selection often simply omit the effects of unlinked background selection (e.g. Elyashiv *et al.* 2016; Ewing and Jensen 2016; Torres *et al.* 2018; Mcgee *et al.* 2022).

But omitting the effects of unlinked background selection may be a concern given the sheer quantity of deleterious mutations entering populations, and the greater dependence of Equation 2 on this rate. For example, the average human is estimated to begin life with an average of two new deleterious mutations not present in either parent (Lesecque *et al.* 2012), and high deleterious mutation rate estimates are not unique to humans (Haag-Liautard *et al.* 2007). Some argue that deleterious mutation rates are even higher, closer to 10 new deleterious mutations per person in humans (Kondrashov 2017).

Here we compare these simple analytical models of linked and unlinked background selection to more complex computational simulations. We perform a multi-locus simulation using the fwdpy11 package (Thornton 2014, 2019), which efficiently handles large numbers of non-neutral mutations in relatively large census size populations (Haller and Messer 2017). The addition of tree-sequence recording (Kelleher *et al.* 2018) to fwdpy11 allows the calculation of a coalescent effective population size without needing to explicitly simulate and track neutral mutations. This approach allows us to compare the 'observed' coalescent effective population size from our high-deleterious-mutation-rate simulations to the analytic expectations for linked (Hudson and Kaplan 1995) and unlinked (Charlesworth 2012) background selection. We then determine whether computational results can be approximated with linked background selection alone, and if not, the amount of disagreement attributable to unlinked background selection.

METHODS

Joint model

To calculate the expected combined effects of background selection at linked and unlinked sites, we simply multiply together the respective reductions in neutral diversity from Equations 1 and 2. For a given window with $R_w \leq 0.5$, and considering constant sh for convenience, this joint model is given by

$$\frac{N_e}{N} = e^{-8(U-U_w)sh} e^{-U_w/2sh+R_w} \quad (3)$$

where U is the total genome-wide deleterious mutation rate. We break the genome into 20,000 windows (this number is chosen to roughly match the number of genes in humans).

Multi-locus simulations

All simulations were written in Python using fwdpy11 (Thornton 2014, 2019). We simulated populations of N diploid individuals undergoing selection against deleterious mutations using a standard Wright-Fisher infinite-sites model for $10N$ generations. Each individual's genome was made up of 23 chromosomes of equal length, with recombination occurring via exactly two cross-overs per chromosome, matching data for humans (Pardo-Manuel De Villena and Sapienza 2001). We allow recombination to occur anywhere rather than explicitly simulating a centrosome or recombination hotspots.

Deleterious mutations occur with genome-wide rate U. In the "no genes" condition, they are located uniformly at random along the chromosomes, while in the "genes" condition they occur only in "genes". We simulate 1,000 genes, accounting for 10 percent of the genome, interspersed at regular intervals throughout the genome. Assumptions about "genes" were not chosen to be representative of any particular species, but simply to capture the qualitative consequences of clustering the sites that are subject to deleterious mutations and hence background selection.

A recent study of a large sample of modern European humans estimated a gamma distribution of fitness effects of new mutations with mean $sN_e = -224.33$ and $N_e = 23{,}646$, implying a mean sh≈ -0.01 (Kim *et al.* 2017). In our main results, we simplify to use a constant sh = -0.01 to avoid complications from deleterious mutations with sN_e so near to 1 that they are effectively neutral. We also explore higher and lower values of s, and a gamma distribution with the same mean and shape parameter $\alpha = 0.169$ (Kim *et al.* 2017). All mutations have $h = 0.5$.

While our forward-time simulations track only deleterious mutations, tree-sequence recording (Kelleher *et al.* 2018) during the simulation allows neutral mutations to be projected backwards onto the genealogical histories of different genomic regions, enabling us to later compute neutral genetic diversity and hence effective population size. In all simulations, neutral mutations occur uniformly at random on the entire genome at an arbitrary rate 10^{-4} per genomic 'unit', for a total rate of 0.23 per genome. This low value provides sufficient resolution of N_e at low computational cost. We use msprime (Kelleher *et al.* 2016) to calculate neutral diversity θ on the resulting tree sequence using an infinite-alleles model, and then calculate the effective population size for a simulation using $\theta = \frac{4N_e\mu}{1+4N_e\mu}$ and solving for N_e. In simulations with genes, we used a slightly different, infinite sites model, calculating the effective population size using $\theta = 4N_e\mu$ and solving for N_e.

We simulated census population sizes N ranging from 2,000 to 10,000. This is compatible with the range of inferred estimates for human effective population sizes (Takahata 1993; Tenesa *et al.* 2007; McEvoy *et al.* 2011). We calculate the 'observed' value of $\frac{N_e}{N}$ from the neutral diversity, to compare with analytical expectations.

RESULTS

Unlinked background selection reduces neutral diversity more than linked background selection does (Figure 1A) when the genome-wide mutation rate U is high, specifically when $U > 1$, as is estimated to be the case for humans (Lesecque *et al.* 2012). This main result is independent of N and s. A five-fold change in census population size N has no significant effect on the $\frac{N_e}{N}$ ratio (Figure 1A).

Larger selection coefficients result in stronger background selection for high U but not low U (Figure 1B). This is consistent with a dominant role for unlinked background selection (Eq. 2); linked background selection predicts the opposite effect (Eq. 1). Our approximate analytical joint model exaggerates the impact of selection strength; for our focal value of $sh = -0.01$, our joint model slightly underestimates background selection (Figure 1A and C), distinctly underestimates it for smaller selection coefficients (Figure 1B green), and slightly overestimates it for large selection coefficients (Figure 1B blue). A full distribution of selective effect sizes behaves similarly to a single sh value with the same mean (Figure 1C).

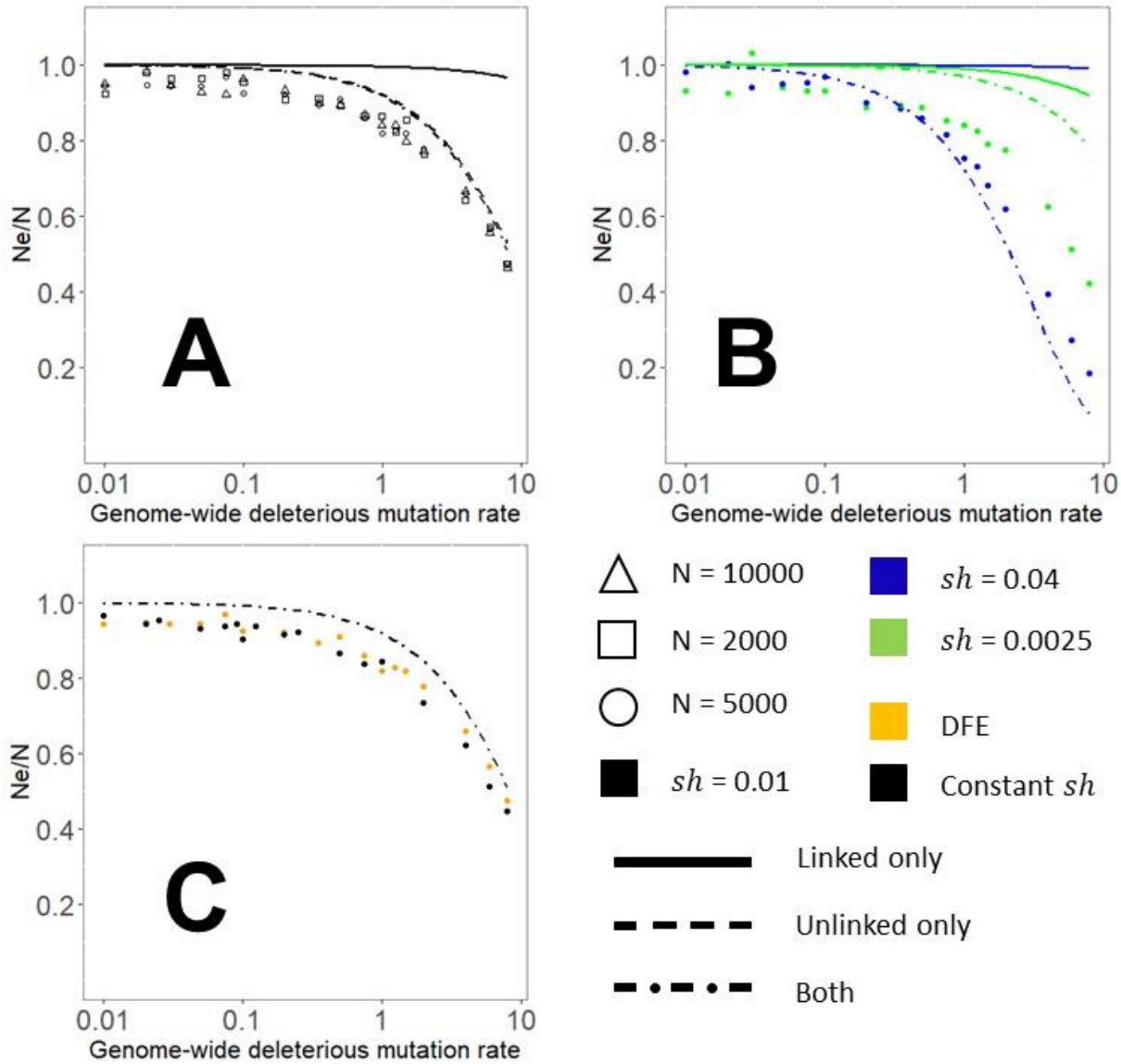

Figure 1. At high deleterious mutation rates, unlinked background selection reduces neutral diversity to a greater extent than linked background selection does. Where not shown, $sh = -0.01$ and $N = 5000$. Each point represents a single simulation — we chose to allocate computation to a denser grid of parameter values rather than to replicates of the same parameter values. (A) Unlinked background selection alone (dashed line) is a closer match than linked background selection alone (solid line) to the joint model (dot-dashed line) and to simulations (points). Census population size has no effect on the $\frac{N_e}{N}$ ratio across a five-fold change in simulated census population size. (B) Background selection at high mutation rates is stronger with larger selection coefficients, although our joint model exaggerates this effect. The model with linked background selection alone predicts the opposite dependence on sh, and strongly underestimates background selection. (C) The relationship between deleterious mutation rate and N_e is similar whether the effect size of new deleterious mutations is constant versus drawn from a distribution of effect sizes with the same mean value of -0.01.

The simulations above assume deleterious mutations occur uniformly at random across the genome. A more realistic scenario would be for deleterious mutations to be clustered within a functional subset of the genome. We modify our simulations to model genomes where only 10% of the genome is made up of 'genes' subject to deleterious mutations. Concentrating deleterious mutations into more tightly linked 'genes' results in a smaller reduction in neutral diversity than found in simulations where deleterious mutations occur across the genome (Figure 2, red vs. blue circles). This does not depend on population size (vertical comparison on Figure 2A, B, D).

Simulations quantify the degree to which neutral diversity is more depressed in genes than in intergenic regions (Figure 2, red squares vs. red triangles in panels C, D, and E). At low mutation rates, there is little depression in neutral diversity, with no appreciable difference between linked vs. unlinked sites. There is also little difference with strong selection (Figure 2, panel E); this is expected as we approach the limit at which each deleterious mutation immediately dooms the genome it appears on (Charlesworth *et al.* 1993). While larger selection coefficients increase overall background selection (discussed above for Fig. 1 and seen for circles in Figure 2 C-E), background selection in genic regions is relatively independent of selection coefficient (red triangles in Figure 2 C-E) and corresponds well to the joint model for $sh = -0.01$ (Figure 2D).

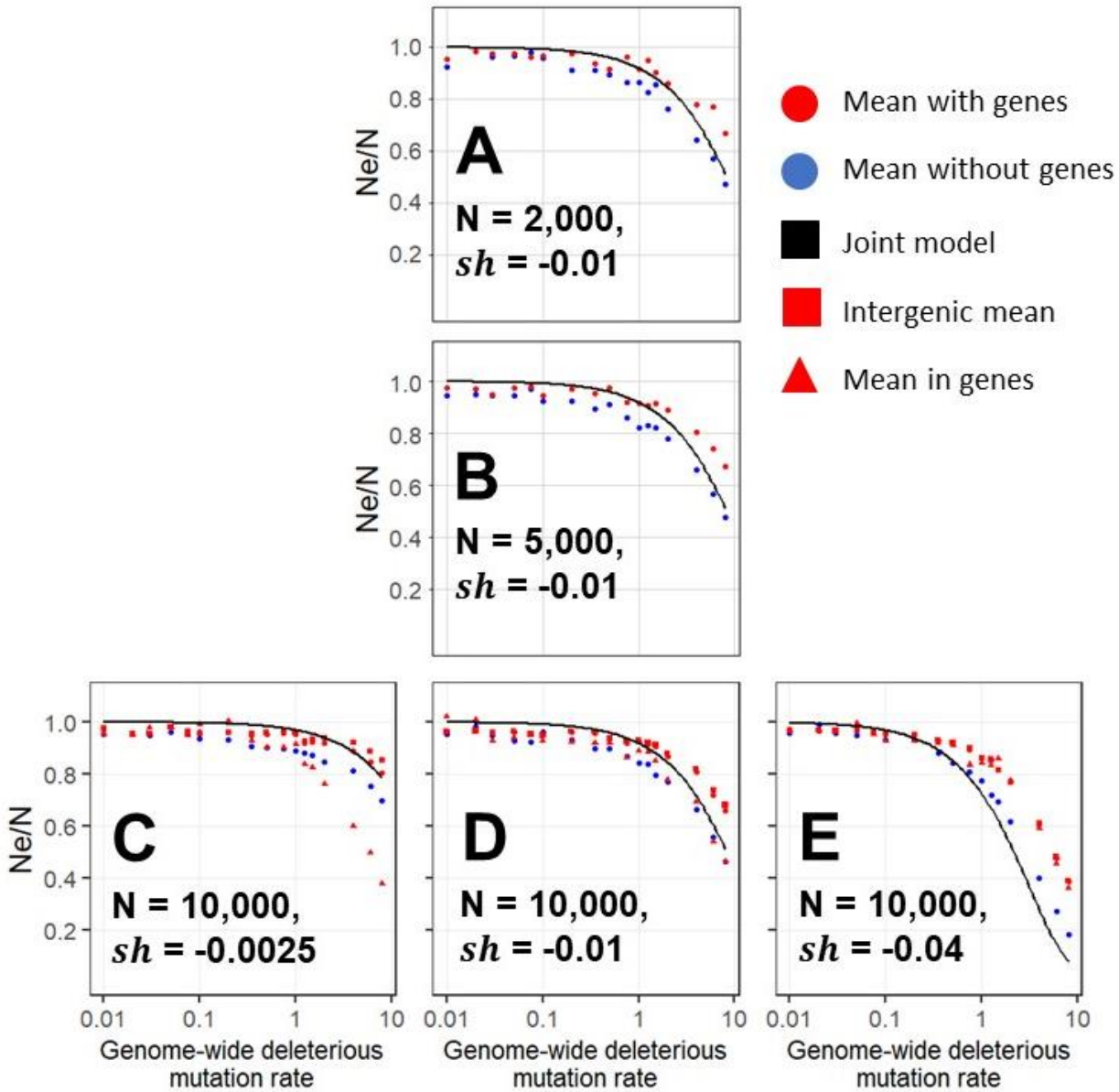

Figure 2. Concentrating deleterious mutations into 'genes' slightly weakens overall background selection, and for genic sites, removes the dependence on selection coefficients. As in Figure 1, background selection becomes significant only at high mutation rates, and census population size does not affect results (A vs B vs D). For large selection coefficients, background selection is similar at genic vs intergenic sites (E, red triangles vs squares). For small selection coefficients, the gap is considerable (C, red circles vs squares). The dependence on the strength of selection (circles in C vs D vs E, blue matches Figure 1) does not apply to genic sites (red triangles in C vs D vs E). In all panels, the solid black line is the theoretical expectation incorporating both linked and unlinked background selection, blue dots are simulations where deleterious mutations occur uniformly at random on the genome, and red dots are simulations where deleterious mutations are clustered into 'genes' (see Methods). Within the genes condition, neutral sites within genes are indicated as red triangles, and outside them are indicated as red squares.

Strong background selection might be well captured by traditional one locus models with lower N_e, if deleterious sites evolve independently. However, at high deleterious mutation rates, linkage disequilibrium occurs (Barton 1998; Barton and Otto 2005), breaking this assumption. As another test for independence, we asked whether the variance in the number of deleterious alleles is lower than that expected from the mean under a Poisson distribution (Figure 3). Independent evolution of sites breaks down for $U > 1$. Underdispersion among deleterious alleles has also been found empirically for humans and fruit flies (Sohail *et al.* 2017)(both with $U > 1$ (Haag-Liautard *et al.* 2007; Lesecque *et al.* 2012)). The index of dispersion was 0.94 for missense variants in the human "crucial genome" in which these variants are most likely to be deleterious (Sohail *et al.* 2017). Note that Sohail et al. (2017) interpreted underdispersion as evidence for synergistic epistasis, but we find underdispersion even in a non-epistatic model.

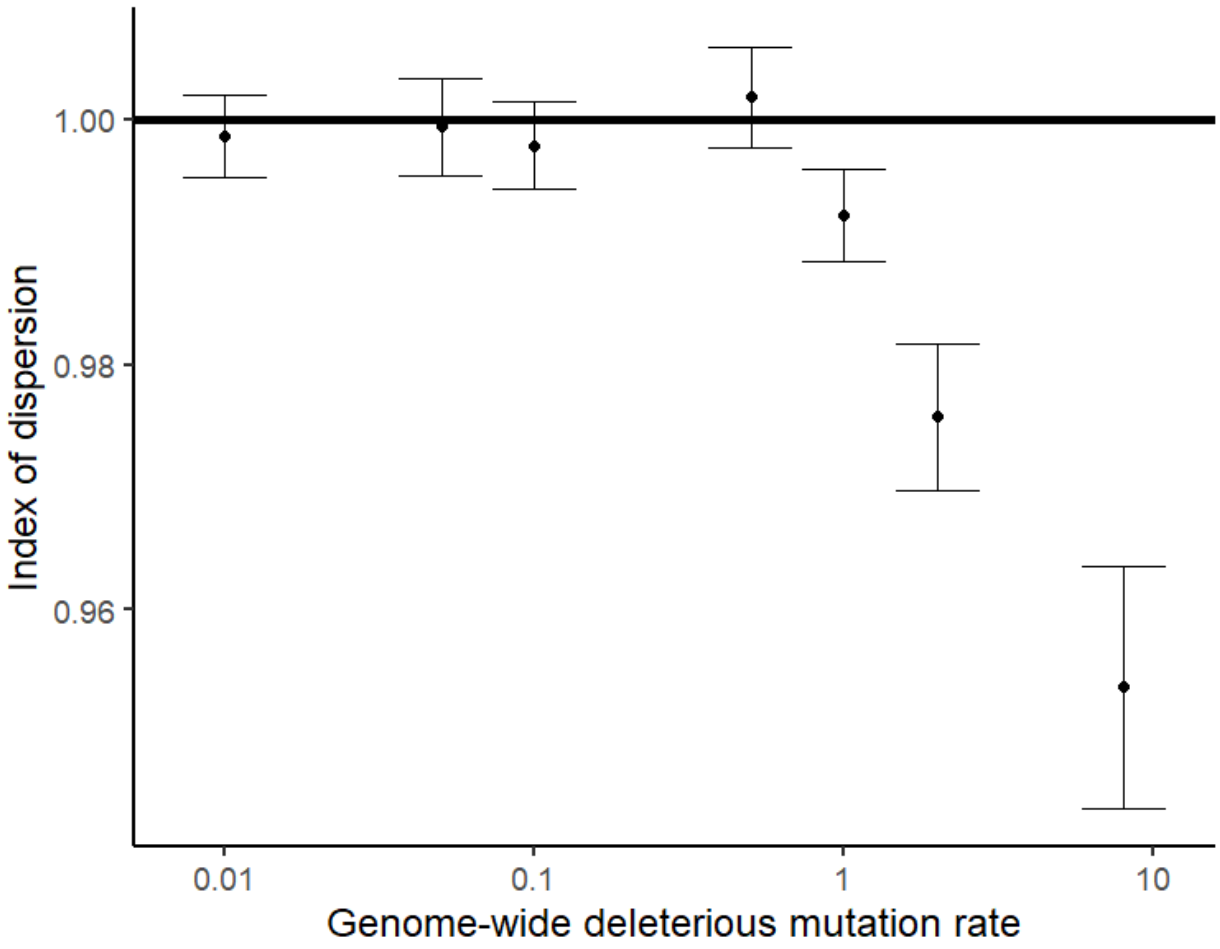

Figure 3. Deleterious mutation rates above 1 create non-independence among segregating sites. The index of dispersion is calculated as the variance in the number of deleterious alleles per haplotype divided by the mean number of deleterious alleles per haplotype. A haplotype is defined as all the chromosomes from one parent, i.e. each individual contains two haplotypes. A Poisson distribution has an index of dispersion of 1, shown as a horizontal line. Points are the mean index of dispersion from eleven replicate simulations with $N = 10{,}000$, $sh = -0.01$, and bars show the standard error in the estimate of the mean.

DISCUSSION

High deleterious mutation rates are well-established empirically in humans and a variety of other species. When deleterious mutation rates are realistically high, unlinked background selection reduces neutral diversity more than linked background selection does. Background selection is negligible at low genome-wide deleterious mutation rates, at which point it does not matter which model one uses to capture it. These findings do not depend on the census population size. Background selection is stronger with larger selection coefficients when deleterious sites are distributed uniformly at random, but when deleterious mutations are clustered into genes, this dependence disappears at neutral sites within genic regions. Existing analytical models exaggerate the dependence on selection coefficients, highlighting our incomplete understanding of unlinked background selection. The original view was that high-*s* mutations at unlinked sites that are in linkage disequilibrium with the focal site would exclude an effectively random set of individuals in each generation from the effective population size. This heuristic does not easily explain our findings.

Our multilocus simulations neglect some population features known to affect neutral diversity (e.g. adaptive evolution (Maynard Smith and Haigh 1974) and temporal changes in population size (Torres *et al.* 2020)), and simplify others (e.g. variation in dominance coefficients among deleterious variants (Gilbert *et al.* 2020) and heterogeneity in recombination rates (Kulathinal *et al.* 2008)). The purpose of our simulations is to isolate the effects of background selection with high mutation rates, rather than to accurately reflect the genetics of specific populations. Incorporating additional complications into the model might change the quantitative strength of background selection. However, we do not expect adding new complications to change the broader conclusion that unlinked background selection cannot be safely ignored.

It is already known that the effects of linked background selection are critical for explaining differences in neutral genetic diversity among genomic regions, and that failure to do so is a problem for demographic inference (Ewing and Jensen 2016; Pouyet *et al.* 2018; Johri *et al.* 2021). Our results bring up the possibility that unlinked background selection could also be important. At low mutation rates ($U \ll 1$, see Figure 3), the effects of unlinked background selection are trivial in magnitude. But at realistically high mutation rates ($U > 1$), unlinked background selection will result in non-independent evolution among sites, which cannot be accounted for with a fudge factor in the effective population size of a one-locus model. In a one-locus model, random changes to allele frequencies are modelled as white noise, i.e. there is no autocorrelation over time. When sites are not independent, being on a bad/good genetic background in one generation will predict being on a bad/good genetic background in the next generation, producing colored noise [cite Masel 2011]. We therefore cannot assume that a 1-locus model with lower Ne can accurately capture background selection at unlinked loci.

To save computational cost, previous simulations of background selection treated only a section of a chromosome. The local deleterious mutation is set to a value corresponding to high genome-wide *U*, but mutations outside the local window are neglected. We have shown that neglecting them is problematic. This could matter both in the context of demographic inference, and when the goal is to use neutral diversity to distinguish between the effects of linked background selection vs. selective sweeps (McVicker *et al.* 2009; Elyashiv *et al.* 2016; Murphy *et al.* 2021). Neglecting unlinked background selection is also a concern for papers investigating only negative selection; e.g. Torres et al. (2018) and

Beissinger et al (2016) look for differences in background selection between human or between maize populations, but their measure of background selection assumes a reference class of neutral sites that are unaffected, to be compared to sites subject to strong background selection.

Another common practice in evolutionary simulations is to rescale parameters (e.g. N, s, U, r) in a manner that keeps products of interest (e.g. Ns, Nr) constant (Hill and Robertson 1966; Comeron and Kreitman 2002; Hoggart *et al.* 2007; Kaiser and Charlesworth 2009; Uricchio and Hernandez 2014; Campos and Charlesworth 2019). We found that the strength of background selection is independent of N, but is stronger for large s (at least outside of genic regions). Reducing N while increasing s will therefore exaggerate the effect of background selection, i.e. the reduction in N_e will be greater than that expected as proportional to the reduction in simulated N. The net result of rescaling as it is normally conducted would thus be a reduction in sN_e in populations with high U.

Accounting for background selection is often conceived as a necessary step prior to inferring demography and/or adaptation. But unlinked background selection in species with high mutation rates globally suppresses neutral diversity to such a degree that it is worthy of scientific attention in its own right. To the extent that currently neutral genetic diversity might become relevant in new environments, background selection imposes a genome-wide penalty caused by the constant deluge of deleterious mutations. Massive background selection is thus a sister question to that of overwhelming deleterious mutation load, which has not yet been fully resolved (Kondrashov 1995, 2017; Agrawal and Whitlock 2012; Goyal *et al.* 2012; Simons *et al.* 2014).

DATA AVAILABILITY

Simulation code written in python, and graphs produced with R. Scripts available on github at www.github.com/MaselLab/BackgroundSelection.

ACKNOWLEDGEMENTS

We thank Bruce Walsh for pointing us to Hudson & Kaplan's analytical results on linked background selection that sparked this work; Denis Roze for a wealth of advice (including pointing us to the equation for unlinked background selection), detailed comments on the manuscript, and catching many errors on our part; two anonymous reviewers of an earlier version of this work who also pointed us to unlinked background selection; Kevin Thornton for pointing out a possible connection with implausibly high empirical estimates of the deleterious mutation rate, and for assistance with fwdpy simulations; and Ryan Gutenkunst, Parul Johri, and David Enard for helpful discussions.

FUNDING

This work was supported by the John Templeton Foundation [62028].

CONFLICT OF INTEREST

The authors declare no conflict of interest.

BIBLIOGRAPHY

Agrawal A. F., and M. C. Whitlock, 2012 Mutation Load: The Fitness of Individuals in Populations

Where Deleterious Alleles Are Abundant. Annu. Rev. Ecol. Evol. Syst 43: 115–35. https://doi.org/10.1146/annurev-ecolsys-110411-160257

Barton N. H., 1998 The effect of hitch-hiking on neutral genealogies. Genet. Res. 72: 123–133. https://doi.org/10.1017/S0016672398003462

Barton N. H., and S. P. Otto, 2005 Evolution of recombination due to random drift. Genetics 169: 2353–2370. https://doi.org/10.1534/genetics.104.032821

Beissinger T. M., L. Wang, K. Crosby, A. Durvasula, M. B. Hufford, *et al.*, 2016 Recent demography drives changes in linked selection across the maize genome. Nat. Plants 2: 1–7. https://doi.org/10.1038/NPLANTS.2016.84

Campos J. L., and B. Charlesworth, 2019 The effects on neutral variability of recurrent selective sweeps and background selection. Genetics 212: 287–303. https://doi.org/10.1534/genetics.119.301951

Charlesworth B., M. T. Morgan, and D. Charlesworth, 1993 The Effect of Deleterious Mutations on Neutral Molecular Variation. Genetics 134: 1289–1303.

Charlesworth B., 2009 Effective population size and patterns of molecular evolution and variation. Nat. Rev. Genet. 10: 195–205. https://doi.org/10.1038/nrg2526

Charlesworth B., 2012 The Effects of Deleterious Mutations on Evolution at Linked Sites. Genetics 190: 5–22. https://doi.org/10.1534/genetics.111.134288

Comeron J. M., and M. Kreitman, 2002 Population, evolutionary and genomic consequences of interference selection. Genetics 161: 389–410. https://doi.org/10.1093/genetics/161.1.389

Comeron J. M., 2014 Background Selection as Baseline for Nucleotide Variation across the Drosophila Genome. PLoS Genet. 10: e1004434. https://doi.org/10.1371/journal.pgen.1004434

Cutter A. D., and B. A. Payseur, 2013 Genomic signatures of selection at linked sites: unifying the disparity among species. Nat. Rev. Genet. 14: 262–274. https://doi.org/10.1038/nrg3425.Genomic

Elyashiv E., S. Sattath, T. T. Hu, A. Strutsovsky, G. McVicker, *et al.*, 2016 A Genomic Map of the Effects of Linked Selection in Drosophila. PLoS Genet. 12: e1006130. https://doi.org/10.1371/journal.pgen.1006130

Ewens W., 2004 *Mathematical Population Genetics*. Springer US, New York.

Ewing G. B., and J. D. Jensen, 2016 The consequences of not accounting for background selection in demographic inference. Mol. Ecol. 25: 135–141. https://doi.org/10.1111/mec.13390

Gilbert K. J., F. Pouyet, L. Excoffier, and S. Peischl, 2020 Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination. Curr. Biol. 30: 101-107.e3. https://doi.org/10.1016/j.cub.2019.11.063

Goyal S., D. J. Balick, E. R. Jerison, R. A. Neher, B. I. Shraiman, *et al.*, 2012 Dynamic Mutation–Selection Balance as an Evolutionary Attractor. Genetics 191: 1309–1319. https://doi.org/10.1534/genetics.112.141291

Haag-Liautard C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, *et al.*, 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in Drosophila. Nature 445: 82–85. https://doi.org/10.1038/nature05388

Haller B. C., and P. W. Messer, 2017 SLiM 2: Flexible, interactive forward genetic simulations. Mol. Biol. Evol. 34: 230–240. https://doi.org/10.1093/molbev/msw211

Hill W., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294. https://doi.org/10.1017/S001667230800949X

Hoggart C. J., M. Chadeau-Hyam, T. G. Clark, R. Lampariello, J. C. Whittaker, *et al.*, 2007 Sequence-level population simulations over large genomic regions. Genetics 177: 1725–1731. https://doi.org/10.1534/genetics.106.069088

Hudson R. R., and N. L. Kaplan, 1994 Gene Trees with Background Selection, pp. 140–153 in *Non-neutral Evolution: Theories and Molecular Data*, edited by Golding B. Springer US, Boston, MA.

Hudson R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics 141: 1605–1617.

Johri P., K. Riall, H. Becher, L. Excoffier, B. Charlesworth, *et al.*, 2021 The Impact of Purifying and Background Selection on the Inference of Population History: Problems and Prospects. Mol. Biol. Evol. 38: 2986–3003. https://doi.org/10.1093/molbev/msab050

Kaiser V. B., and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. Trends Genet. 25: 9–12.

Kelleher J., A. M. Etheridge, and G. McVean, 2016 Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes. PLoS Comput. Biol. 12: e1004842. https://doi.org/10.1371/journal.pcbi.1004842

Kelleher J., K. R. Thornton, J. Ashander, and P. L. Ralph, 2018 Efficient pedigree recording for fast population genetics simulation. PLoS Comput. Biol. 14: e1006581.

Kern A. D., and M. W. Hahn, 2018 The Neutral Theory in Light of Natural Selection. Mol. Biol. Evol. 35: 1366–1371. https://doi.org/10.1093/molbev/msy092

Kim Y., and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neural variation. Genetics 155: 1415–1427. https://doi.org/10.1093/genetics/155.3.1415

Kim B., C. Huber, and K. Lohmueller, 2017 Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples. Genetics 206: 345–361. https://doi.org/10.1534/genetics.116.197145/-/DC1.1

Kimura M., 1968 Evolutionary Rate at the Molecular Level. Nature 217: 624–626.

Kimura M., 1969 The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61: 893–903.

King J. L., and T. H. Jukes, 1969 Non-Darwinian Evolution. Science (80-.). 164: 788–798.

Kondrashov A. S., 1995 Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? J. Theor. Biol. 175: 583–594. https://doi.org/10.1006/jtbi.1995.0167

Kondrashov A. S., 2017 *Crumbling Genome: The Impact of Deleterious Mutations on Humans*. John Wiley & Sons, Inc, Hoboken, NJ.

Kulathinal R. J., S. M. Bennett, C. L. Fitzpatrick, and M. A. F. Noor, 2008 Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. Proc. Natl. Acad. Sci. U. S. A. 105: 10051–10056. https://doi.org/10.1073/pnas.0801848105

Lesecque Y., P. D. Keightley, and A. Eyre-Walker, 2012 A resolution of the mutation load paradox in humans. Genetics 191: 1321–1330. https://doi.org/10.1534/genetics.112.140343

Lohmueller K. E., A. Albrechtsen, Y. Li, S. Y. Kim, T. Korneliussen, *et al.*, 2011 Natural Selection Affects Multiple Aspects of Genetic Variation at Putatively Neutral Sites across the Human Genome. PLoS Genet. 7: e1002326. https://doi.org/10.1371/journal.pgen.1002326

Masel J., 2011 Genetic drift. Curr. Biol. 21: R837–R838. https://doi.org/10.1016/j.cub.2011.08.007

Maynard Smith J., and J. Haigh, 1974 The hitch-hiking effect of a favourable gene. Genet. Res. (Camb). 23: 23–35. https://doi.org/10.1017/S0016672308009579

McEvoy B. P., J. E. Powell, M. E. Goddard, and P. M. Visscher, 2011 Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res. 21: 821–829. https://doi.org/10.1101/gr.119636.110

McGee R. S., O. Kosterlitz, A. Kaznatcheev, B. Kerr, and C. T. Bergstrom, 2022 The cost of information acquisition by natural selection. biorxiv 2022.07.02. https://doi.org/10.1101/2022.07.02.498577

McVicker G., D. Gordon, C. Davis, and P. Green, 2009 Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5: e1000471. https://doi.org/10.1371/journal.pgen.1000471

Murphy D., E. Elyashiv, G. Amster, and G. Sella, 2021 Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements. bioRxiv 2021.07.02.450762. https://doi.org/10.1101/2021.07.02.450762

Nordborg M., B. Charlesworth, and D. Charlesworth, 1996 The effect of recombination on background selection. Genet. Res. (Camb). 67: 159–174.

Pardo-Manuel De Villena F., and C. Sapienza, 2001 Recombination is proportional to the number of chromosome arms in mammals. Mamm. Genome 12: 318–322. https://doi.org/10.1007/s003350020005

Pouyet F., S. Aeschbacher, A. Thiéry, and L. Excoffier, 2018 Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. Elife 7: 1–21. https://doi.org/10.7554/eLife.36317

Simons Y. B., M. C. Turchin, J. K. Pritchard, and G. Sella, 2014 The deleterious mutation load is insensitive to recent population history. Nat. Genet. 46: 220–224. https://doi.org/10.1038/ng.2896

Sohail M., O. A. Vakhrusheva, J. H. Sul, S. L. Pulit, L. C. Francioli, *et al.*, 2017 Negative selection in humans and fruit flies involves synergistic epistasis. Science (80-.). 356: 539–542. https://doi.org/10.1126/science.aah5238

Takahata N., 1993 Allelic genealogy and human evolution. Mol. Biol. Evol. 10: 2–22. https://doi.org/10.1093/oxfordjournals.molbev.a039995

Tenesa A., P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, *et al.*, 2007 Recent human effective population size estimated from linkage disequilibrium. Cold Spring Harb. Lab. Press Hum. 2: 520–526. https://doi.org/10.1101/gr.6023607.8

Thornton K. R., 2014 A c++ template library for efficient forward-time population genetic simulation of large populations. Genetics 198: 157–166. https://doi.org/10.1534/genetics.114.165019

Thornton K. R., 2019 Polygenic adaptation to an environmental shift: Temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. Genetics 213: 1513–1530. https://doi.org/10.1101/505750

Torres R., Z. A. Szpiech, and R. D. Hernandez, 2018 Human demographic history has amplified the

effects of background selection across the genome. PLoS Genet. 14: 1–27. https://doi.org/10.1371/journal.pgen.1007387

Torres R., M. G. Stetter, R. D. Hernandez, and J. Ross-Ibarra, 2020 The Temporal Dynamics of Background Selection in Nonequilibrium Populations. Genetics 214: 1019–1030.

Uricchio L. H., and R. D. Hernandez, 2014 Robust forward simulations of recurrent hitchhiking. Genetics 197: 221–236. https://doi.org/10.1534/genetics.113.156935