Title
Understanding Deleterious Variation in Complex Populations

Author
Kim, Bernard Youngsoo

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Understanding Deleterious Variation in Complex Populations

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Biology

by

Bernard Youngsoo Kim

2018

ABSTRACT OF THE DISSERTATION

Understanding Deleterious Variation

in Complex Populations

by

Bernard Youngsoo Kim

Doctor of Philosophy in Biology

University of California, Los Angeles, 2018

Professor Kirk Edward Lohmueller, Chair

Complex population demography can have subtle yet significant impacts on the genetic variation of populations. Furthermore, complex demography can subtly affect natural selection and therefore shapes the distribution of deleterious genetic variation. In my dissertation, I utilize a variety of computational tools to model the impact of deleterious variation in complex populations. In the first chapter, I investigated why among all human populations East Asians have the most Neanderthal ancestry. I found that multiple interbreeding events between Neanderthals and East Asians are required to explain the data, revising current models of human history. In the second chapter, I developed new computational tools for estimating the distribution of fitness effects using large datasets of genetic variants and estimated the amount of selection on amino acid changing mutations in humans. Here I found fewer strongly deleterious mutations compared to previous smaller studies, suggesting that neutral forces may play a greater role in human evolution than previously appreciated. In Chapter 3, I investigated the dynamics of deleterious genetic variation in hybrid populations using simulations and found

that differences in standing deleterious variation between parent populations can significantly impact the evolution of hybrids. It is therefore essential that null models of hybrid evolution consider the effects of deleterious variation before invoking processes such as hybrid incompatibility or adaptive introgression to explain unusual patterns of genetic variation.

The dissertation of Bernard Youngsoo Kim is approved.

Howard Bradley Shaffer

John P. Novembre

Kirk Edward Lohmueller, Committee Chair

University of California, Los Angeles

Dedicated,

with much love and gratitude, to

my parents and grandfather

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the generosity of all the people who have supported me along the way.

First, I would like to thank my advisor, Kirk Lohmueller. It has been quite the experience watching the lab grow from just a few of us to what is it today. So much has changed – for the better – in the last five years. At the beginning of it all, you were willing to take a risk on me as a student, when I was full of enthusiasm but not much else, and I will always be grateful for the opportunity you gave me. Through grad school, you have been an unending spring of kindness and support. Your constant positivity helped me get through both good and bad times. I have learned so much from you, about being an excellent scientist but also about being an excellent mentor, and I don't think I could've joined a better lab.

I would also like to thank the other faculty and staff who have contributed to my education. John Novembre, thank you for inspiring me when I was a clueless first year who barely knew anything about genetics. Your work is a continuous source of inspiration. Bob Wayne and Brad Shaffer, thank you for taking the time to provide me with valuable feedback along the way. Victoria Sork, thank you for your mentorship and your enthusiasm for working with me, even with your busy schedule. I am so glad I got to learn a little about how unique the California white oaks are. Mike Alfaro, thank you for your willingness to foster my interest in branching out beyond population genetics during my last months at UCLA. Jocelyn Yamadera and Tessa Villasenor, I couldn't imagine going through grad school without your help on all the little things along the way.

Grad school would not have been the same without people to turn around and talk to, whether about science or to groan about Kirk's terrible jokes. For that, I am grateful for my lab mates. Christian, thank you for your willingness to answer my statistics questions and for being a sounding board for ideas throughout all of grad school. Without you, my grad school experience would have been severely diminished. Clare, fellow founding member of the lab,

thank you for the innumerable things you helped me with, but especially your help with all the Hoffman issues. Diego and Ying, best of luck with your new labs and I hope I will see you at conferences in the future. Tanya, you are a machine and your work ethic inspires me to do better. Arun, Jazlyn, and Jesse, let's keep the lab Slack alive with papers and memes. Best of luck on your own journeys. Chris, thanks for putting up with my workaholism this last year and hopefully I'll see you on the trail/water soon. Annabel, I'm glad you ended up joining our lab and thank you for all the wonderful hangouts and conversations over the years. Last but not least, Max, you taught me how to be a better mentor and I thoroughly enjoyed working with you.

I consider myself incredibly fortunate to have never felt alone at UCLA. For that I thank my dearest friends. Jonathan Chang, to whom I bear an uncanny resemblance, I'll never forget that you and Zac were the only ones willing to get a beer after R boot camp. Your enthusiasm for hanging out hasn't changed since then. Grace John, I've never had a dull moment hanging out with you. Thank you for making grad school a brighter place. To David Surovec, thanks for being my lifting and backpacking buddy through all of grad school. To Jacqueline Robinson, my lab mate, roommate, best friend, camping buddy, fishing buddy, and Shakespeare companion, thank you for your love and for suffering through the long nights with me. To the rest of my friends on and off campus, thank you for your support, hang outs, and all the beers and Scotches we've shared through the years.

Finally, I would like to thank my family for their support through this journey, even at the beginning when I was only a source of disappointment. To my mom, dad and May, it is only with your help that I was able to overcome the challenges of my life and discover my passion for research. Thank you for your love and support through it all. To my grandfather, who always stood up for me even when everyone else had given up, I hope you're proud of what I've accomplished. You always knew that this was what I would end up doing.

Co-authored work: **Chapter 1** is a reprint of a published paper- Kim, BY and KE Lohmueller. Selection and reduced population size cannot explain higher amounts of

Neandertal ancestry in East Asian than in European human populations. The paper is used here with the permission of all co-authors and Cell Press. KEL provided guidance through all stages of the project and wrote part of the manuscript. **Chapter 2** is a reprint of a published paper- Kim, BY, Huber, CD, and KE Lohmueller. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. The paper is used here with the permission of all co-authors and the Genetics Society of America. CDH contributed some analyses and wrote part of the manuscript. KEL wrote part of the manuscript and provided guidance through all stages of the project. **Chapter 3** is a submitted paper- Kim, BY, Huber, CD, and KE Lohmueller. Deleterious variation shapes the genomic landscape of introgression. The paper is used here with the permission of all co-authors. CDH provided consultation innumerable times. KEL wrote part of the manuscript and provided guidance through all stages of the project.

Funding: This work was supported by a UCLA Dissertation Year Fellowship, a Searle Scholars Fellowship (to K. Lohmueller) and the National Institutes of Health (R35 GM-119856 to K. Lohmueller).

VITA

BERNARD YOUNGSOO KIM

EDUCATION

MS, Biology, UCLA, Los Angeles, CA, 2013

BS, Biology, UCLA, Los Angeles, CA, 2010

FELLOWSHIPS AND AWARDS

Google Cloud Platform Research Grant, 2018 ($19,800)

UCLA Dissertation Year Fellowship, 2017 ($20,000)

UCLA EEB Departmental Fellowship, 2013, 2014, 2017 ($26,000 total)

UCLA EEB Research and Travel Grant, 2016, 2017, 2018 ($4000 total)

PUBLICATIONS

Published and accepted work

Robinson, J.A., Brown, C., **Kim, B.Y.**, Lohmueller, K.E., and R. Wayne. (Accepted) Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Current Biology*.

Kim, B.Y., Wei, X., Fitz-Gibbon, S., Lohmueller, K.E., Ortego, J., Gugger, P.F., and V.L. Sork. (Accepted) RADseq data reveal pervasive, but not promiscuous, ancient introgression within Californian white oak species (*Quercus*: Fagaceae). *Molecular Ecology*.

Huber, C.D., **Kim, B.Y.**, Marsden, C.D., and K.E. Lohmueller. (2017) Determining the factors driving selective effects of new nonsynonymous mutations. *PNAS* 114(17):4465-4470

Kim, B.Y., Huber, C.D. and K.E. Lohmueller. (2017) Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics* 206(1):345-361

Robinson, J.A., Ortega-Del Vecchyo, D., Fan, Z., **Kim, B.Y.**, Marsden, C.D., Lohmueller, K.E., and R. Wayne. (2016) Genomic flatlining in the endangered island fox. *Current Biology* 26(9):1183-1189

Kim, B.Y. and K.E. Lohmueller (2015) Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. *American Journal of Human Genetics* 96(3):454-461.

Burreson, E.M., Passarelli, J.K., and **B.Y. Kim**. (2012) A new species of marine leech (Hirudinida: Piscicolidae) from giant kelpfish, *Heterostichus rostratus* Girard, in southern California. Journal of Parasitology 98(4):788-790.

Burreson, E.M., **Kim, B.Y.**, and J.K. Passarelli. (2012) A new species of *Mysidobdella* (Hirudinida: Piscicolidae) from mysids along the California coast. Journal of Parasitology 98(2):341-343.

Papers in review or revision

Kim, B.Y., Huber, C.D. and K.E. Lohmueller. (In review) Deleterious variation shapes the genomic landscape of introgression.

CHAPTER 1:

SELECTION AND REDUCED POPULATION SIZE CANNOT EXPLAIN HIGHER AMOUNTS OF NEANDERTAL ANCESTRY IN EAST ASIAN THAN IN EUROPEAN POPULATIONS

REPORT

Selection and Reduced Population Size Cannot Explain Higher Amounts of Neandertal Ancestry in East Asian than in European Human Populations

Bernard Y. Kim[1] and Kirk E. Lohmueller[1,2,*]

It has been hypothesized that the greater proportion of Neandertal ancestry in East Asians than in Europeans is due to the fact that purifying selection is less effective at removing weakly deleterious Neandertal alleles from East Asian populations. Using simulations of a broad range of models of selection and demography, we have shown that this hypothesis cannot account for the higher proportion of Neandertal ancestry in East Asians than in Europeans. Instead, more complex demographic scenarios, most likely involving multiple pulses of Neandertal admixture, are required to explain the data.

Initial genomic studies found Neandertal ancestry in non-African populations, suggesting that some ancestral admixture occurred between Neandertals and the ancestors of modern Eurasian populations.[1,2] One proposed explanation for this observation is that there was one pulse of Neandertal admixture in the Levant before humans migrated further into Europe and Asia.[2–4] However, more recent genomic studies[2–5] show that there are higher levels of Neandertal ancestry in East Asian populations than in Europeans. Initially, such a finding would appear to contradict the one-pulse admixture model. Additional pulses of Neandertal admixture into East Asian populations would be required to explain the increased Neandertal ancestry in East Asian populations.[5–8]

Recently, Sankararaman et al.[9] proposed a provocative hypothesis that could potentially rescue the one-pulse admixture model. They hypothesized that Neandertal alleles were weakly deleterious in humans. Because current evidence suggests that East Asian populations experienced stronger historical bottlenecks and had smaller effective population sizes,[10–14] the ability of purifying selection to remove weakly deleterious alleles from the population might have been less effective in East Asians than in Europeans.[15] The reason for this is that in the smaller East Asian population, weakly deleterious alleles might have drifted to higher frequencies. In the larger European population, however, the effect of drift would be smaller. Thus, there could have been a single pulse of Neandertal admixture in the ancestral Eurasian population, but because Europeans were better able than East Asians to remove weakly deleterious Neandertal alleles, Neandertal ancestry appears to have increased in East Asians.

Here, we used forward-in-time Wright-Fisher simulations to explicitly test this hypothesis (Figure S1). To do this, we wrote our own custom Python simulations, called "Forward_Neanderthal" (see Web Resources). We simulated 1,000,000 ancestry-informative sites as independent loci, all of which received a single pulse of Neandertal admixture at t_{admix} generations ago. However, each of these ancestry-informative sites could biologically correspond to a larger segment of Neandertal ancestry.

We assumed that a single admixture event between humans and Neandertals occurred $t_{admix} = 1{,}900$ generations ago. This time corresponds to 47,500 years ago if we assume 25 years/generation. We chose this time to reflect a plausible time at which admixture could have occurred between Neandertals and humans.[9,16] At each locus, at the start of the simulation (at time t_{admix}), we assumed that a proportion (f) of the chromosomes contained Neandertal ancestry. In practice, each of the 1,000,000 loci began the simulation with the Neandertal ancestry at frequency f. We examined $f = \{0.02, 0.04\}$, corresponding to plausible amounts of Neandertal admixture in human populations.[2,5,9] Although we note that there might have been a distribution of values of initial Neandertal ancestry across the genome, this variability should not affect our results unless the initial starting frequency of Neandertal ancestry were to differ between European and East Asian populations. Given that our models assume a single pulse of Neandertal admixture in the ancestral Eurasian population, which itself is randomly mating, there is little reason to conclude that f should vary between the populations under the models we are testing.

We then allowed the populations to evolve to the present day under demographic models with parameters estimated from data (see below). We did this by adjusting the frequencies of the alleles deterministically according to the standard selection equations (see below) and by binomial sampling to model genetic drift. The total number of chromosomes drawn to form the next generation varied over time to reflect the changes in population size over time. At the end of the simulation, we examined the remaining amount of Neandertal ancestry in each population. In order to have a fair basis of comparison to Sankararaman

[1]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90095, USA; [2]Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90095, USA
*Correspondence: klohmueller@ucla.edu
http://dx.doi.org/10.1016/j.ajhg.2014.12.029.

et al.,[9] for each site we sampled the same number of chromosomes from our simulated populations as in the 1000 Genomes Project[17] CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection; 174 chromosomes) and CHB (Han Chinese in Beijing, China; 190 chromosomes) populations. Under the assumption of independence and exchangeability among sites, a single haplotype can be modeled as a series of Bernoulli draws with success $p_1 \ldots p_k$ over the k sites in the genome (p_i is the frequency in the sample of the Neandertal allele at the i^{th} site). Therefore, the expected Neandertal ancestry per haplotype is equivalent to the mean frequency of Neandertal alleles ($p_1 \ldots p_k$) in the sample. In other words, $E[\text{Neandertal}] = (1/k)\sum_{i=1}^{k} p_i$. Thus, we computed the average Neandertal ancestry per genome (p_{all}) by averaging the per-site frequencies of Neandertal alleles in the sample over all 1,000,000 sites. Our approach is also analogous to that used in Sankararaman et al.,[9] except that we assume that Neandertal ancestry is known rather than inferred (see below for further discussion) and that all sites are independent. We calculated the ratio of Neandertal ancestry in the East Asian population to that in the European population (R) by dividing the average ancestry in the East Asian population by the average in the European population ($R = p_{\text{all\_ASN}} / p_{\text{all\_EUR}}$). We also recorded the proportion of sites still polymorphic for Neandertal ancestry in the sample (p_{var}), as well as the frequency of Neandertal alleles only at those sites where the Neandertal alleles were still segregating (p_{seg}). We assessed simulation variance by replicating the entire simulation process for a given model 20 times. 95% confidence intervals (CIs) were calculated as $\text{CI} = \bar{p}_{\text{all}} \pm 1.96\sigma$, where $\bar{p}_{\text{all}}$ and σ denote the mean and SD, respectively, of Neandertal ancestry per individual over the 20 simulation replicates.

Because the effects that Neandertal alleles have on human fitness are unclear, we allowed Neandertal alleles to have a range of effects from neutral to strongly deleterious. We defined the relative fitness of individuals heterozygous for Neandertal and human ancestry as $1 + hs$ and the fitness of individuals homozygous for Neandertal ancestry as $1 + s$ (s is the selection coefficient, and h is the dominance coefficient). First, we used scalar values of $s = \{0, -10^{-5}, -10^{-4}, -10^{-3}, -10^{-2}\}$. Additionally, we assumed that the selection coefficients (s) of the Neandertal alleles were drawn from a negative gamma distribution with parameters inferred from nonsynonymous SNPs by Boyko et al.[18] In particular, for the population-scaled selection coefficient, we used a gamma distribution that had a shape parameter ($\alpha = 0.184$) and a scale parameter ($\beta = 8{,}200$). Because this gamma distribution describes the distribution of $2Ns$, we divided the value of $2Ns$ by $2 \times 25{,}636$ (the value of N used in Boyko et al.) to obtain the distribution of the selection coefficient, s. The parameters of this gamma distribution were estimated for new nonsynonymous mutations and might not necessarily reflect the distribution of fitness effects for Neandertal alleles in humans. However, given the extremely limited information regarding the distribution of fitness effects of Neandertal alleles in humans, this gamma distribution is a reasonable first approximation because it includes a mixture of nearly neutral, weakly deleterious, and strongly deleterious alleles.

We investigated multiple models of dominance (h). We considered the standard models of codominance ($h = 0.5$) and recessive effects ($h = 0$). We also examined models of underdominance ($h = 2$) and overdominance, where individuals who are heterozygous for Neandertal ancestry have the lowest and highest fitnesses, respectively. Some special care was needed when we used the gamma distribution of selective effects. The value of s from the Boyko et al.[18] gamma distribution refers to the fitness effect of the heterozygous genotype, and $2s$ refers to the fitness of the homozygous genotype. In our simulations, s refers to the fitness effect of the homozygous genotype. Consequently, for simulations where $h = 0.5$ and $h = 0$, we multiplied the value of s obtained from the gamma distribution by 2.

We examined several different demographic models that have been fit to the East Asian and European populations (Tables S1 and S2). We first used the bottleneck models fit by Keinan et al.[11] (Table S1). The Keinan et al. bottleneck model assumes an ancestral human population size of N that then experienced two different bottlenecks, one of which was at approximately the same time in the European and East Asian populations (about 4,000 generations ago). However, this first bottleneck was older than the pulse of Neandertal admixture ($t_{\text{admix}} = 1{,}900$ generations ago). Because this earlier bottleneck was completed prior to the start of the simulations, we did not include it in the model. Rather, we assumed that the population remained at a constant size (N) until t_B generations ago, when a bottleneck occurred. The duration of the bottleneck is described by t_{Bleni}, and the population size during the bottleneck is N_B individuals. After the bottleneck, the population recovered to N individuals and remained that size until the simulation finished. Note that the Keinan et al.[11] model considers the European and East Asian populations separately from each other. As such, we also simulated the two populations separately (Figure S1).

The degree to which the different models matched the observed proportion of Neandertal ancestry in either population was quite variable (Figure 1A). In models where the observed present-day Neandertal ancestry was approximately compatible with the amounts observed in empirical data (between 0.5% and 5%), the ratio of Neandertal ancestry in East Asians to Neandertal ancestry in Europeans (R) was close to 1 (Figure 1B). It never matched the R values estimated from empirical data[9] ($R = 1.14$–1.31). This same result held regardless of the dominance coefficient, strength of selection, or initial proportion of Neandertal ancestry in the ancestral population ($f = 4\%$; Figures S2 and S3).

In order to investigate the sensitivity of our results to the precise demographic model assumed, we performed

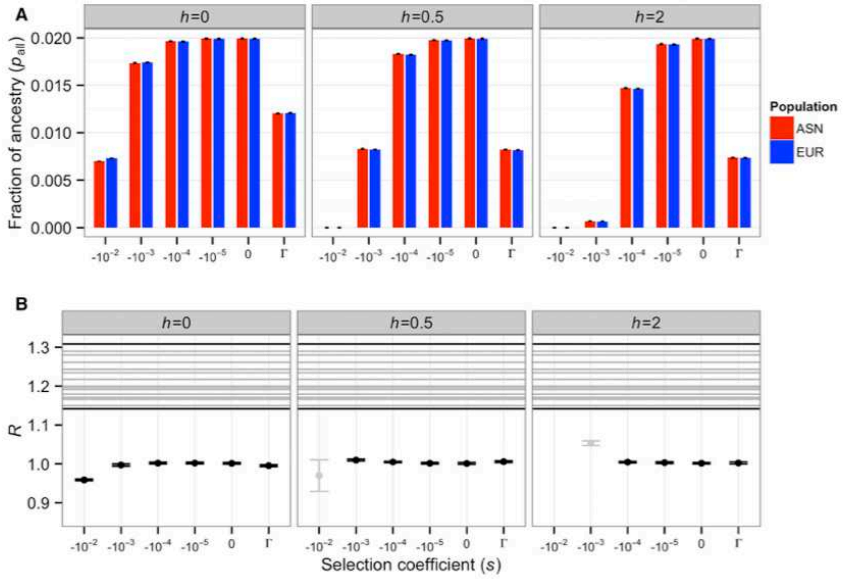

Figure 1. Predicted Neandertal Ancestry in East Asian and European Populations under the Keinan et al. Demographic Model when $f = 2\%$

Each column depicts results for a different dominance coefficient (h). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations.

(A) The fraction of Neandertal ancestry in East Asian (ASN) and European (EUR) populations.

(B) Ratio of Neandertal ancestry in East Asians to Neandertal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neandertal ancestry observed in empirical comparisons of an East Asian and a European population.[9] Models where the final proportion of Neandertal ancestry is concordant with the empirical data (between 0.5% and 5% in A) are colored black. Otherwise, they are colored gray. Note that across these models, the maximum value of R is only slightly higher than 1.0. However, the lowest observed value of R in the empirical data[9] (in a comparison of IBS [Iberian population in Spain] and CHS [Southern Han Chinese]) is 1.14. Thus, demography differences combined with purifying selection cannot generate an excess amount of Neandertal ancestry in East Asians relative to Europeans as large as that seen in the empirical data.

additional simulations where we varied some of the bottleneck parameters. First, we investigated whether changing the duration of the bottleneck (t_{Blen}) would affect our results. In the initial model, we assumed that $t_{Blen} = 100$ generations. We conducted additional simulations with $t_{Blen} = 50$ generations and $t_{Blen} = 200$ generations. Importantly, in both cases, we kept the overall severity of the bottleneck ($F = t_{Blen} / 2N_B$) the same as in the original Keinan et al. study.[11] In order to do this, we changed the number of individuals in the bottleneck (Table S1). We found that the length of the bottleneck had little impact on our results (Figures S4 and S5). For the models where the observed present-day Neandertal ancestry was approximately similar to the amount observed in empirical data (between 0.5% and 5%), the ratio of Neandertal ancestry in East Asians to Neandertal ancestry in Europeans (R) remained close to 1 and did not match the R values estimated from empirical data,[9] regardless of the dominance coefficient or strength of selection (Figures S4 and S5).

Second, we wanted to determine whether our results would be qualitatively different if the bottleneck in East Asia was actually more severe than estimated by Keinan et al.[11] We investigated models with bottlenecks 2- and 5-fold more severe than that estimated by Keinan et al.[11] Here, we kept the length of the bottleneck fixed at 100 generations (Table S1). Importantly, we did not change the severity of the bottleneck in the European population; we kept it at the original severity as estimated by Keinan et al. When Neandertal sites were weakly deleterious, nearly neutral, or recessive, increasing the severity of the bottleneck had little effect on our results (Figures S6 and S7). These models predicted R values close to 1, which is too low to be compatible with the observed ratio of East Asian to European Neandertal ancestry.[9] When $h \geq 0.5$ and $s \leq -0.001$, some R values were in the range of, or even greater than, those seen in the empirical data (Figures S6 and S7). However, the predicted proportion of Neandertal ancestry in modern humans was too low in these models (<0.5%; Figures S6A and S7A) to be compatible with the observed data (>1%).[9] Thus, although the more severe bottleneck might allow for some strongly selected Neandertal sites to drift to higher frequency in East Asians

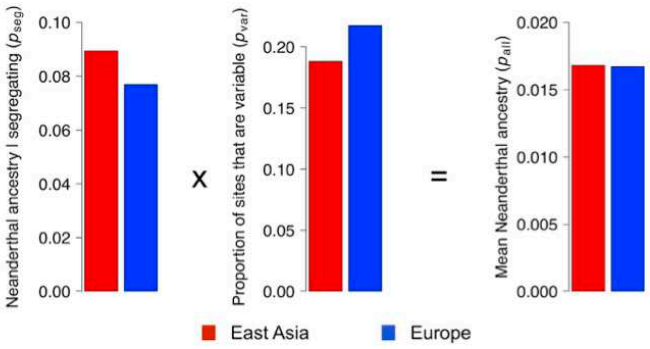

Figure 2. The Smaller Effective Population Size in East Asians Than in Europeans Has Two Competing Effects on Patterns of Neandertal Ancestry

(Left) The average Neandertal allele frequency at the end of the simulation given that the site segregates for the Neandertal and human allele (p_{seg}). Note that here, the average allele frequency in East Asia is higher than that seen in Europe as a result of the greater effects of genetic drift in East Asia than in Europe.

(Center) The percentage of sites (out of a total of 1,000,000 sites) where a Neandertal allele and a human allele are both still segregating at the end of the simulation (p_{var}). Note that fewer sites are segregating in the East Asian population because more were lost by genetic drift in this population.

(Right) The mean Neandertal ancestry per individual (p_{all}) is the product of both the mean frequency of alleles given that they are segregating and the percentage of sites that are segregating. Note that these two effects cancel each other out. These results suggest that East Asian and European individuals will have similar amounts of Neandertal ancestry under this model of demography and selection.

than in Europeans, such a model does not fit all aspects of the data. In summary, even if the East Asian bottleneck was 2- to 5-fold more severe than estimated, if we assume that the severity of the bottleneck in Europe was accurately estimated, purifying selection combined with the greater effect of genetic drift in the East Asian population cannot explain the higher proportion of Neandertal ancestry in East Asians than in Europeans.

Our findings suggest that reduced efficacy of purifying selection, due to greater genetic drift, in East Asians relative to Europeans cannot explain the observed increase in the proportion of Neandertal ancestry in East Asians. The reason for this is that greater drift in East Asians had two competing effects on Neandertal ancestry (Figure 2). For sites where both the Neandertal and human alleles were segregating at the end of the simulation, the Neandertal alleles tended to be at higher frequency in East Asians than in Europeans (Figure 2; Table S3). However, greater drift in East Asians also means that Neandertal alleles are lost from the population at a faster rate. Our simulations predicted that East Asian populations should have fewer sites with segregating Neandertal alleles than European populations (Figure 2; Table S3). These two competing effects of drift canceled each other out, yielding R values close to 1. For neutral alleles, this cancellation followed exactly from the mathematical formulation of the Wright-Fisher model. The expected value of the frequency of an allele at initial frequency f does not change after a generation of genetic drift, regardless of the population size.[19,20]

To examine the mechanism of allele-frequency change with selection, we conducted additional simulations in which the population was set to the size of the bottlenecks estimated in Keinan et al.[11] We ran these simulations for 100 generations and recorded the average frequency of the Neandertal alleles at the end of the simulation (which would correspond to the end of the population bottlenecks in the full demographic model). For the bottlenecks estimated by Keinan et al.,[11] the average Neandertal allele frequencies were essentially the same in both populations (Figure 3A; Figure S8). The nearly neutral theory predicts that mutations where $-1 < Ns < 0$ (according to our scaling of the relative fitnesses) are nearly neutral and are primarily affected by drift rather than selection.[15,21,22] Thus, Neandertal alleles where $s > -0.0018$ are predicted to be nearly neutral and primarily affected by drift in both populations, suggesting that the analytical predictions for neutral alleles approximately hold here as well. More strongly deleterious alleles also showed similar frequencies between the two populations, indicating that the subtle difference in the population size during the East Asian and European bottlenecks is too small to show a change in the effect of selection between the two populations in such a short time period. Because the bottleneck was estimated to be only slightly more severe in East Asia, the threshold at which alleles were nearly neutral was fairly similar between the populations (bottom panel of Figure 3A).

To examine whether the pattern seen in Figure 3A would hold with a stronger bottleneck in East Asia, we made the East Asian population size 5-fold smaller than that estimated by Keinan et al.[11] while keeping the European population size the same as originally estimated (Table S1). Again, nearly neutral alleles ($s > -0.0018$) were primarily affected by drift. As such, the Neandertal frequencies in East Asian and Europeans were predicted to be the same for the reasons discussed above (Figure 3B; Figure S8). Only when the selection coefficients for Neandertal alleles became more deleterious did we see a difference in allele frequency. When $s < -0.0018$, we saw that East Asians had a slightly higher frequency of Neandertal alleles than did Europeans (Figure 3B). Here, Neandertal alleles were predicted to be nearly neutral in East Asians but more affected by selection in Europeans (bottom panel of Figure 3B). This is the effect that Sankararaman et al.[9]

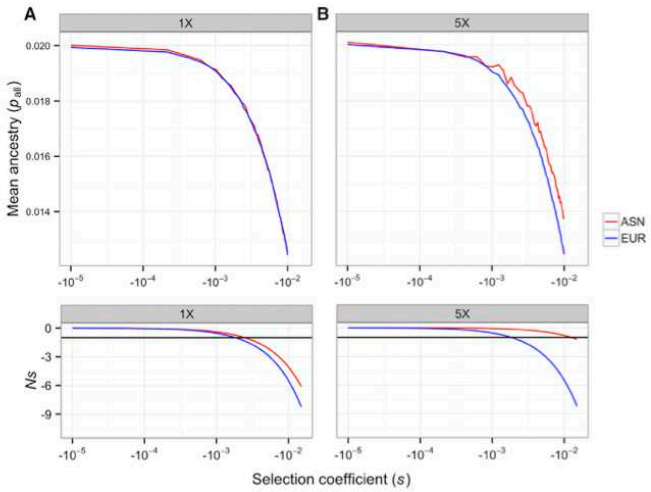

Figure 3. Predicted Mean Neandertal Allele Frequency at the End of the Population Bottlenecks in East Asian and European Populations for the Additive Case
(A) Population sizes were set to those inferred in Keinan et al.[11]
(B) The ASN population size was assumed to be 5-fold smaller than that estimated in Keinan et al.[11] In all cases, constant-sized populations were simulated for 100 generations. The bottom plots show how Ns changes as a function of s.
In (A), both populations have a similar value of Ns across the range of s. Alleles with $s > -0.0018$ are nearly neutral ($Ns > -1$) in both populations. In (B), when $s < -0.0018$, alleles in the ASN population remain nearly neutral, whereas those in the EUR population are more strongly selected. Here, $f = 2\%$.

hypothesized could explain the higher Neandertal ancestry in East Asians. But, our simulations suggest that this effect is unlikely to occur in practice because it requires a stronger bottleneck than that estimated for East Asia and a selection too strong to be compatible with observed amounts of Neandertal ancestry (see below).

Next, we wanted to assess whether other demographic features not included in the Keinan et al.[11] bottleneck model would influence our conclusions. Specifically, the Keinan et al.[11] model does not consider shared ancestry between the East Asian and European populations, migration between populations, or recent population growth. Thus, we performed additional simulations under a different human demographic model fit to the site-frequency spectrum of East Asian, European, and African populations.[23] This model jointly considers both the European and East Asian populations with migration between them and includes recent exponential population growth in both populations. This model also includes an unsampled African population that exchanges migrants with the European and East Asian populations. We included the African population because we wanted to investigate whether a higher migration rate between Africa and Europe than between Africa and East Asia could increase the values of R. Because the African population does not start with any Neandertal ancestry, migrants from Africa would be unlikely to carry Neandertal ancestry and would thus decrease the overall proportion of Neandertal ancestry in the population into which they migrate.

As before, we assumed that the Neandertal admixture occurred at time $t_{admix} = 1{,}900$ generations ago. In the Gravel et al.[23] model, this time occurred during the Eurasian population bottleneck, after the ancestral African population split from the ancestral Eurasian population. Thus, we started our simulation by introducing Neandertal ancestry at $f = \{0.02, 0.04\}$ into the ancestral Eurasian population, which had size N_b. After t_1 generations, this population split into European and East Asian populations with initial population sizes N_{EUR0} and N_{ASN0}, respectively, and growth rates r_{EUR} and r_{ASN}, respectively. The probabilities of migration, m, were assumed to be symmetric and were set to the previously estimated values.[23] Migration was assumed to be conservative, meaning that it does not change the populations sizes.[24] The frequency of the Neandertal allele in the European population after migration (f'_{EUR}) was $f'_{EUR} = f_{EUR}(1 - m_{EUR\_ASN} - m_{EUR\_AFR}) + f_{ASN}(m_{EUR\_ASN}) + f_{AFR}(m_{EUR\_AFR})$, where f_{EUR} is the frequency in the European population before migration. These populations continued to grow exponentially for t_2 generations, at which time the simulation was concluded. Table S2 shows the parameter values used for these simulations.

This more complex demographic model[23] showed results similar to those from the Keinan et al.[11] model. The ratio of Neandertal ancestry in East Asians to Neandertal ancestry in Europeans (R) remained close to 1 (Figure 4; Figure S9) for the models where the observed present-day Neandertal ancestry was approximately similar to that observed in empirical data (between 0.5% and 5%). Again, the observed R values estimated from the empirical data fell outside the range predicted by our models. Importantly, our implementation of the multi-population demographic model[23] included a higher migration rate between Africa and Europe than between Africa and East Asia. Thus, the fact that this model did not yield R values consistent with the observed data (Figure 4B) suggests that the previously estimated[23] rates of differential migration between African and non-African populations are insufficient to dilute the Neandertal ancestry in Europeans in relation to the Neandertal ancestry East Asians.

Our analyses are predicated on the assumption that the amount of Neandertal ancestry in present-day East Asia is

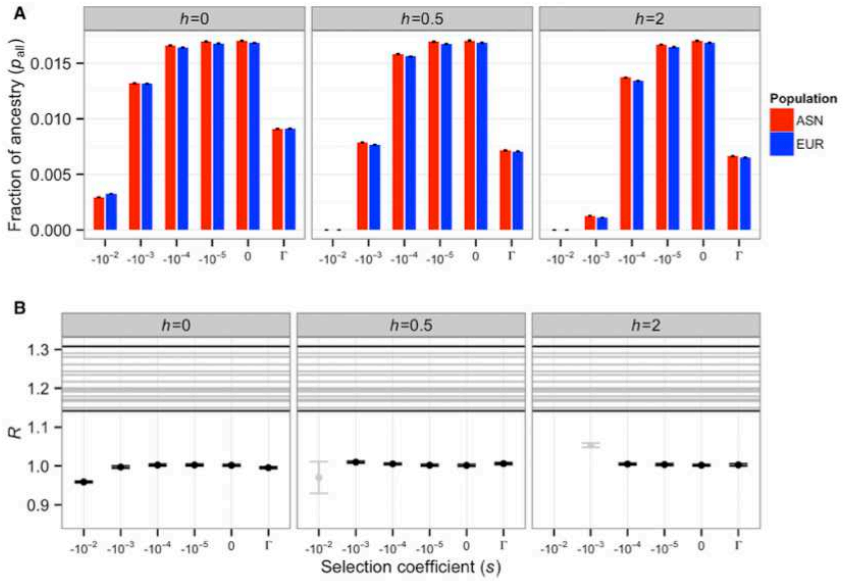

Figure 4. Predicted Neandertal Ancestry in East Asian and European Populations under the Gravel et al. Complex Demographic Model when $f = 2\%$

Each column depicts results for a different dominance coefficient (h). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations.

(A) The fraction of Neandertal ancestry in East Asian (ASN) and European (EUR) populations.

(B) Ratio of Neandertal ancestry in East Asians to Neandertal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neandertal ancestry observed in empirical comparisons of an East Asian and a European population.[9] Models where the final proportion of Neandertal ancestry is concordant with the empirical data (between 0.5% and 5% in A) are colored black. Otherwise, they are colored gray. Note that across these models, the maximum value of R is only slightly higher than 1.0. However, the lowest observed value of R in the empirical data[9] (in a comparison of IBS and CHS) is 1.14. Thus, demography differences combined with purifying selection cannot generate an excess amount of Neandertal ancestry in East Asians relative to Europeans as large as that seen in the empirical data.

truly higher than that in Europe. Our study did not assess whether there is differential performance of the statistical approaches to identifying Neandertal ancestry across different human populations. Multiple statistical approaches—including D statistics,[5,7] a conditional-random-field approach based on multiple summary statistics,[9] and methods based on linkage disequilibrium[5,6,25]—all suggest that East Asians have 15%–30% more Neandertal ancestry than European populations. These statistical methods measure different features of the data and have distinct underlying assumptions. Thus, the fact that they provide concordant results suggests that differential power is unlikely to explain the higher amount of Neandertal ancestry in East Asia. However, to better address whether the increased Neandertal ancestry in East Asia as inferred by the D statistic could be an artifact of complex demography, we conducted neutral coalescent simulations[26] under the Gravel et al.[23] demographic model, in which we included zero, one, or two pulses of Neandertal admixture[6] (Tables S4 and S5). Importantly, unlike our previous results that assumed that Neandertal ancestry could be unambiguously identified, the D statistics were applied to simulated genetic-variation data as done in practice. We found that higher migration rates between Europe and Africa than between East Asia and Africa in a model with one pulse of Neandertal admixture are not sufficient to generate the observed increase in Neandertal ancestry in East Asian populations (Table S4).

However, there are two possible ways a simple demographic model with one pulse of Neandertal admixture could still explain the patterns seen in the data. First, Neandertal alleles could have differential fitness effects in European and East Asian populations (i.e., s is different between Europeans and East Asians). Second, if all Neandertal sites are co-dominant or under-dominant and tend to be moderately to strongly deleterious ($s \leq -0.001$), R becomes larger (Figures 1 and 4; Figures S2 and S9 and Table S3). Yet, as discussed previously, R only matched the empirical data when the bottleneck in East Asia was 2- to 5-fold more severe than estimated (Figures S6 and S7). However, for such a model to be compatible with the amount of Neandertal ancestry observed in human populations,[9] the initial admixture proportion (f) would have to be substantially greater than 10% (Figure S10). Without

additional support, both of these models seem biologically less plausible than alternative demographic models.

In sum, our simulations suggest that across a wide range of biologically realistic models, a single pulse of Neandertal admixture, combined with the reduced efficacy of purifying selection against weakly deleterious alleles in East Asians, cannot explain the *R* values observed in empirical data. Instead, more complex demographic scenarios, possibly including an additional pulse or wave of Neandertal admixture into East Asian populations, must be invoked. Such two-pulse models have been shown to fit the observed data[5,6,8] better than the single-pulse-with-migration model,[6] even when only the genomic regions most likely to be neutrally evolving are considered.[25] In our simulations, across a range of different values for the strength of selection acting on Neandertal ancestry, a two-pulse model with realistic admixture proportions[6,25] could generate the *R* values observed in the actual data (Figures S11 and S12 and Table S3), suggesting that such a model is one viable explanation for differential patterns of Neandertal ancestry between East Asian and European populations.

Supplemental Data

Supplemental Data include 12 figures and 5 tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2014.12.029.

Acknowledgments

We thank Emilia Huerta-Sanchez, Bogdan Pasaniuc, Joshua M. Akey, Sriram Sankararaman, and members of the K.E.L. and Pasaniuc labs for helpful discussions and/or comments on the manuscript. K.E.L. is supported by a Searle Scholars Fellowship.

Received: September 18, 2014
Accepted: December 31, 2014
Published: February 12, 2015

Web Resources

The URL for data presented herein is as follows:

Forward_Neanderthal, https://github.com/LohmuellerLab/Forward_Neanderthal

References

1. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710–722.
2. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.
3. Wang, S., Lachance, J., Tishkoff, S.A., Hey, J., and Xing, J. (2013). Apparent variation in Neanderthal admixture among African populations is consistent with gene flow from Non-African populations. Genome Biol. Evol. *5*, 2075–2081.
4. Yang, M.A., Malaspinas, A.-S., Durand, E.Y., and Slatkin, M. (2012). Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. Mol. Biol. Evol. *29*, 2987–2995.
5. Wall, J.D., Yang, M.A., Jay, F., Kim, S.K., Durand, E.Y., Stevison, L.S., Gignoux, C., Woerner, A., Hammer, M.F., and Slatkin, M. (2013). Higher levels of neanderthal ancestry in East Asians than in Europeans. Genetics *194*, 199–209.
6. Vernot, B., and Akey, J.M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. Science *343*, 1017–1021.
7. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prüfer, K., de Filippo, C., et al. (2012). A high-coverage genome sequence from an archaic Denisovan individual. Science *338*, 222–226.
8. Currat, M., and Excoffier, L. (2011). Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. Proc. Natl. Acad. Sci. USA *108*, 15129–15134.
9. Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., and Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. Nature *507*, 354–357.
10. Voight, B.F., Adams, A.M., Frisse, L.A., Qian, Y., Hudson, R.R., and Di Rienzo, A. (2005). Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc. Natl. Acad. Sci. USA *102*, 18508–18513.
11. Keinan, A., Mullikin, J.C., Patterson, N., and Reich, D. (2007). Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nat. Genet. *39*, 1251–1255.
12. Conrad, D.F., Jakobsson, M., Coop, G., Wen, X., Wall, J.D., Rosenberg, N.A., and Pritchard, J.K. (2006). A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat. Genet. *38*, 1251–1260.
13. Jakobsson, M., Scholz, S.W., Scheet, P., Gibbs, J.R., VanLiere, J.M., Fung, H.C., Szpiech, Z.A., Degnan, J.H., Wang, K., Guerreiro, R., et al. (2008). Genotype, haplotype and copy-number variation in worldwide human populations. Nature *451*, 998–1003.
14. Schaffner, S.F., Foo, C., Gabriel, S., Reich, D., Daly, M.J., and Altshuler, D. (2005). Calibrating a coalescent simulation of human genome sequence variation. Genome Res. *15*, 1576–1583.
15. Ohta, T. (1992). The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. *23*, 263–286.
16. Sankararaman, S., Patterson, N., Li, H., Pääbo, S., and Reich, D. (2012). The date of interbreeding between Neandertals and modern humans. PLoS Genet. *8*, e1002947.
17. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
18. Boyko, A.R., Williamson, S.H., Indap, A.R., Degenhardt, J.D., Hernandez, R.D., Lohmueller, K.E., Adams, M.D., Schmidt, S., Sninsky, J.J., Sunyaev, S.R., et al. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. *4*, e1000083.
19. Kimura, M. (1964). Diffusion models in population genetics. J. Appl. Probab. *1*, 177–232.

20. Charlesworth, B., and Charlesworth, D. (2010). Elements of Evolutionary Genetics (Greenwood Village, CO: Roberts and Company).
21. Akashi, H., Osada, N., and Ohta, T. (2012). Weak selection and protein evolution. Genetics *192*, 15–31.
22. Ohta, T. (1972). Population size and rate of evolution. J. Mol. Evol. *1*, 305–314.
23. Gravel, S., Henn, B.M., Gutenkunst, R.N., Indap, A.R., Marth, G.T., Clark, A.G., Yu, F., Gibbs, R.A., and Bustamante, C.D.; 1000 Genomes Project (2011). Demographic history and rare allele sharing among human populations. Proc. Natl. Acad. Sci. USA *108*, 11983–11988.
24. Nagylaki, T. (1980). The strong-migration limit in geographically structured populations. J. Math. Biol. *9*, 101–114.
25. Vernot, B., and Akey, J.M. (2015). Complex history of admixture between modern humans and Neandertals. Am. J. Hum. Genet. *96*, 448–453.
26. Hudson, R.R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics *18*, 337–338.

Supplemental Results

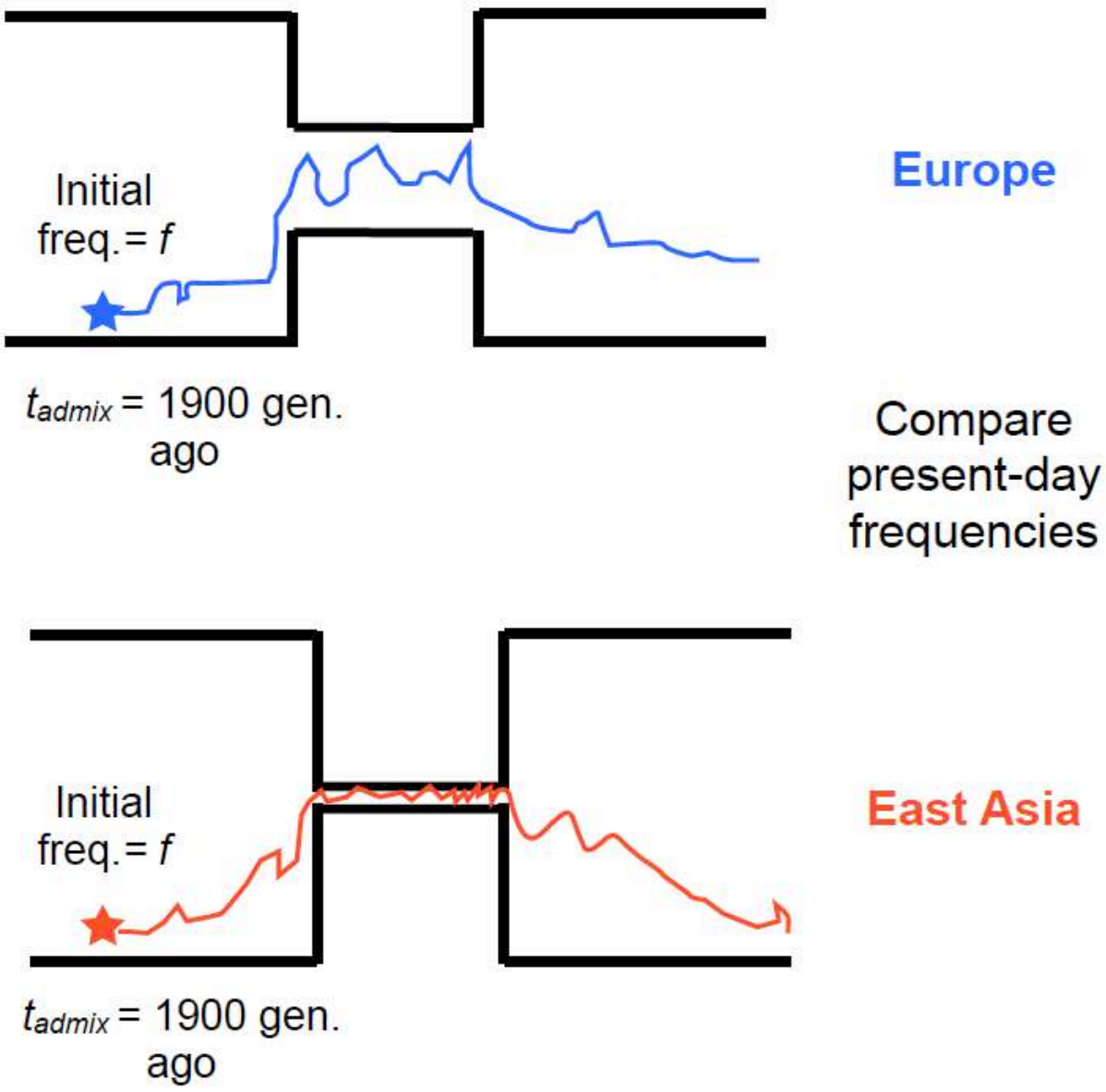

Figure S1: Simulation scenario. At time t_{admix}=1900 generations ago, a Neanderthal allele starts at frequency *f* and changes frequency each generation via selection and drift. Note the difference in bottleneck severity between the European and East Asian populations. See Tables S1 and S2 for a description of the parameters used.

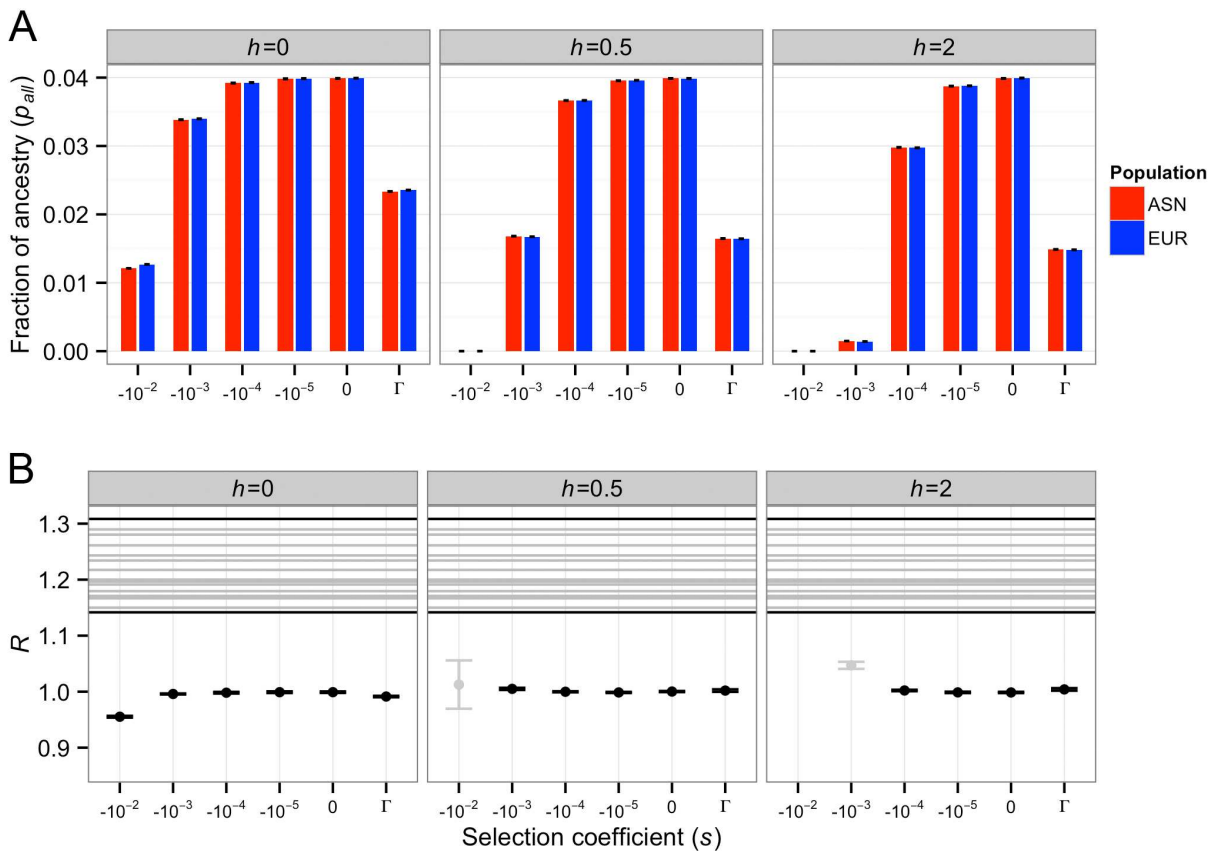

Figure S2: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model when *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

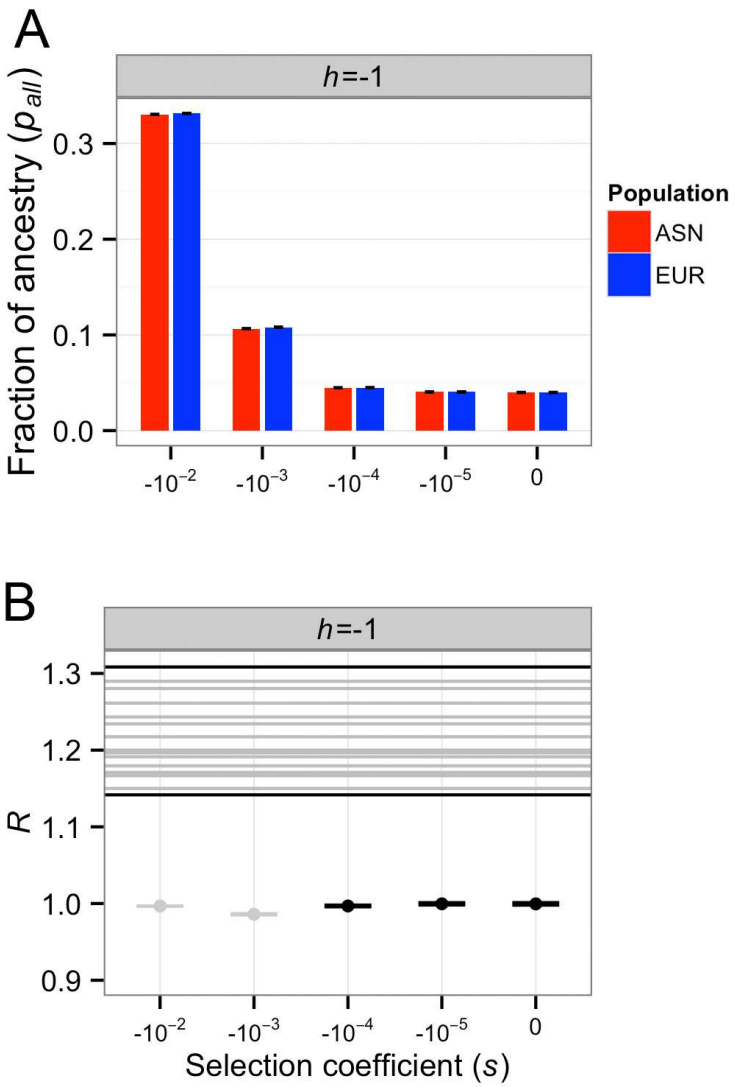

Figure S3: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model when *f*=4% with overdominance. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

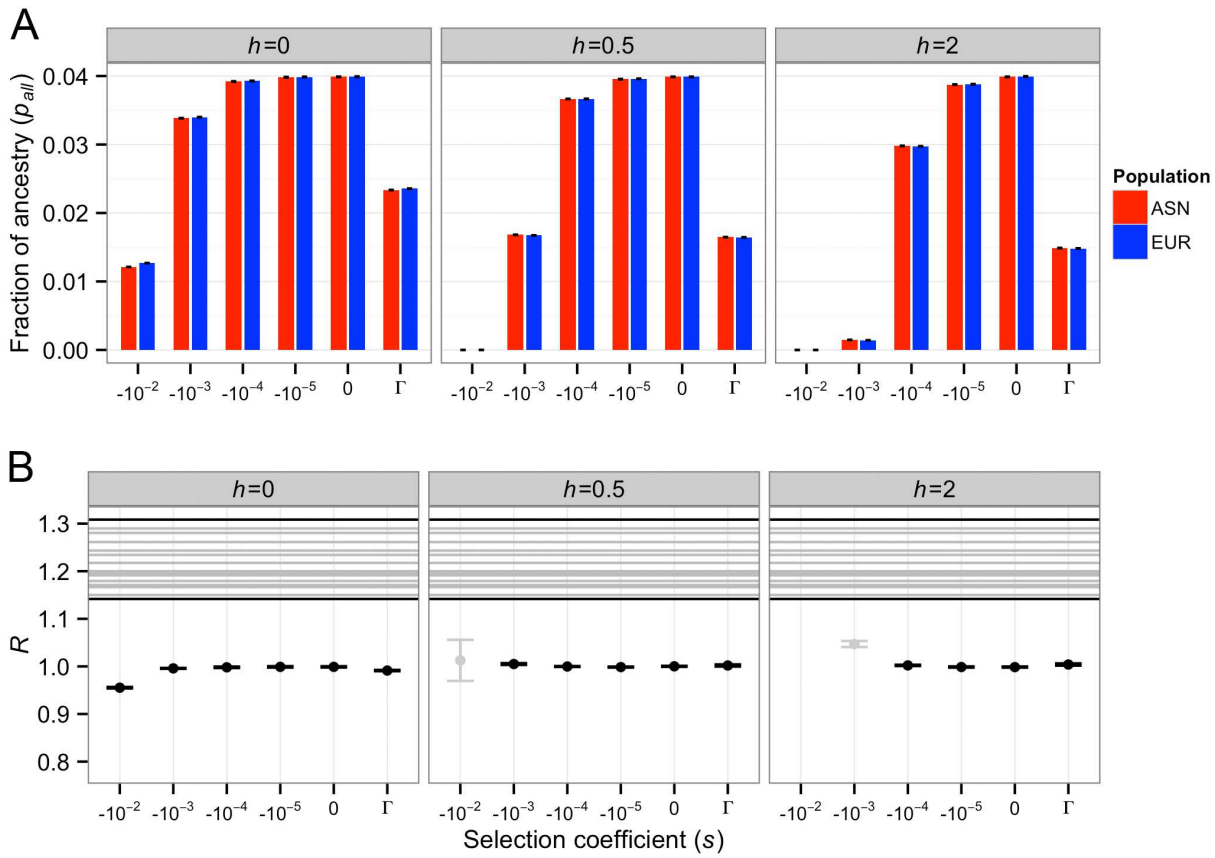

Figure S4: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model with a shorter bottleneck (t_{Blen}=50 generations*)*. The overall severity of the bottleneck (*F*) was equal to that estimated in Keinan et al. Here *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

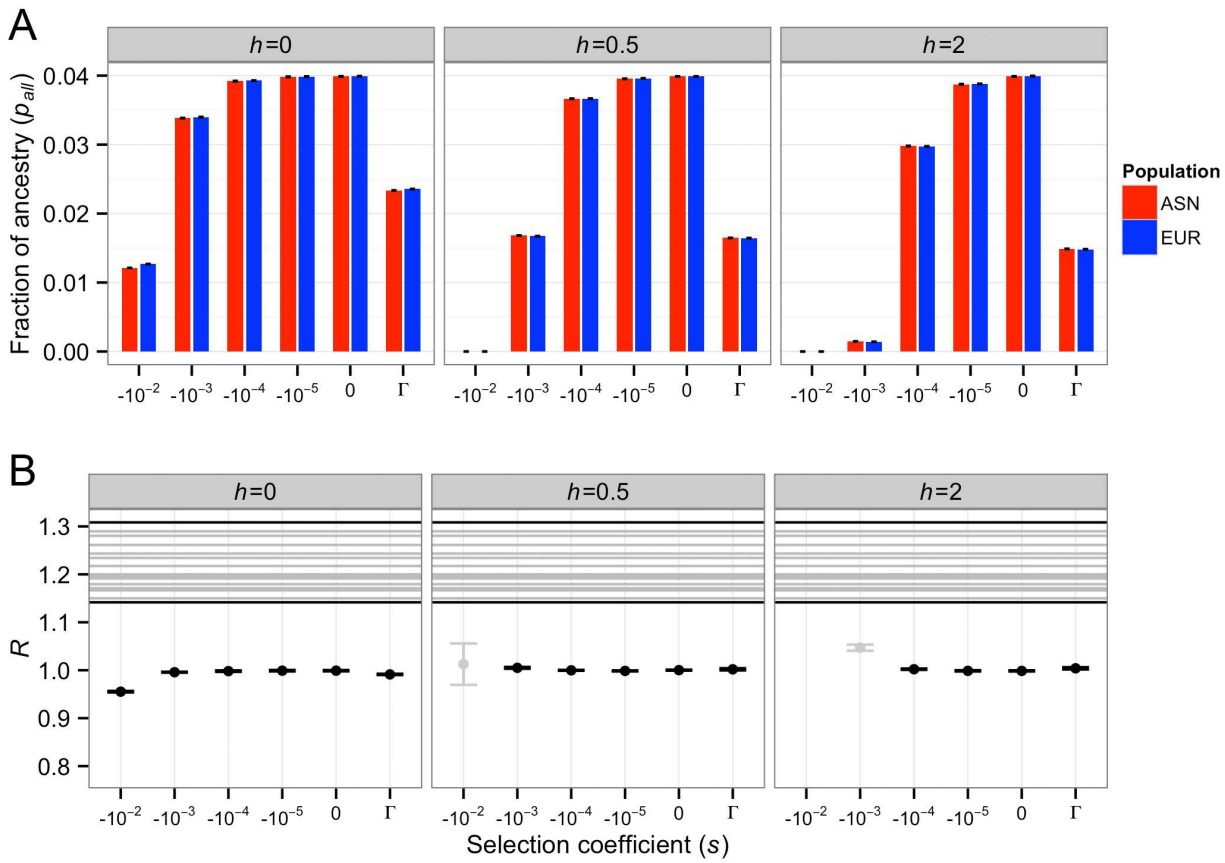

Figure S5: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model with a longer bottleneck (t_{Blen}=200 generations). The overall severity of the bottleneck (F) was equal to that estimated in Keinan et al. Here f=4%. Each column depicts results for a different dominance coefficient (h). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (R). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

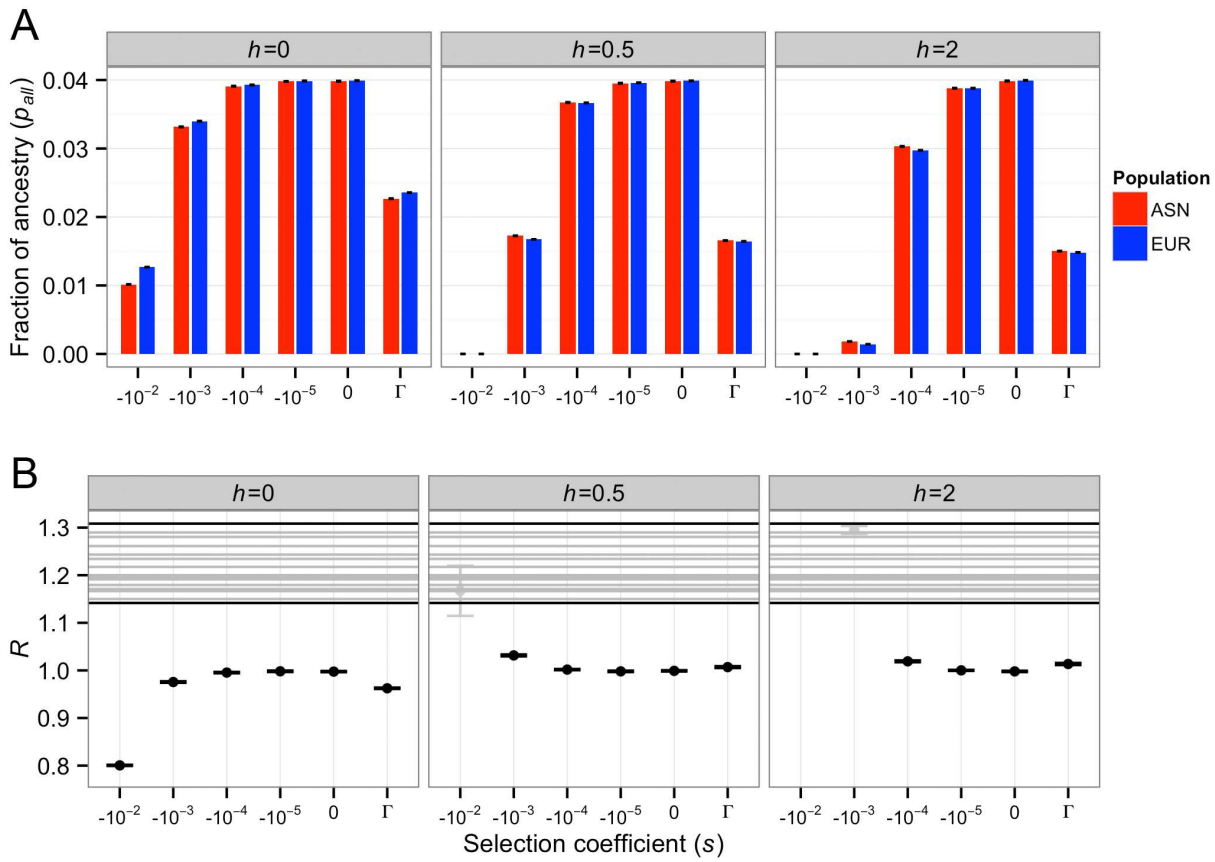

Figure S6: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model where the bottleneck in ASN was 2-times more severe than that estimated by Keinan et al. The severity of the EUR bottleneck was as estimated by Keinan et al. Here *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

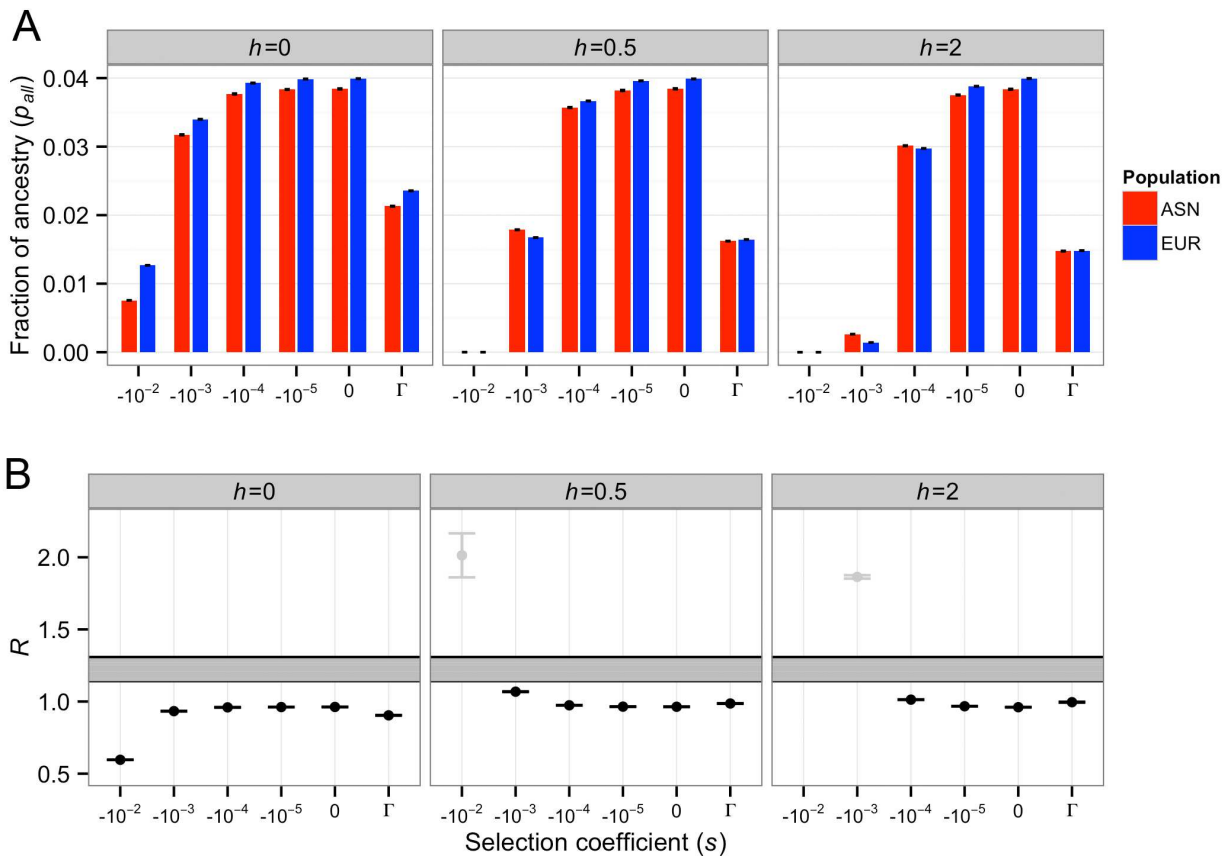

Figure S7: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model where the bottleneck in ASN was 5-times more severe than that estimated by Keinan et al. The severity of the EUR bottleneck was as estimated by Keinan et al. Here *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

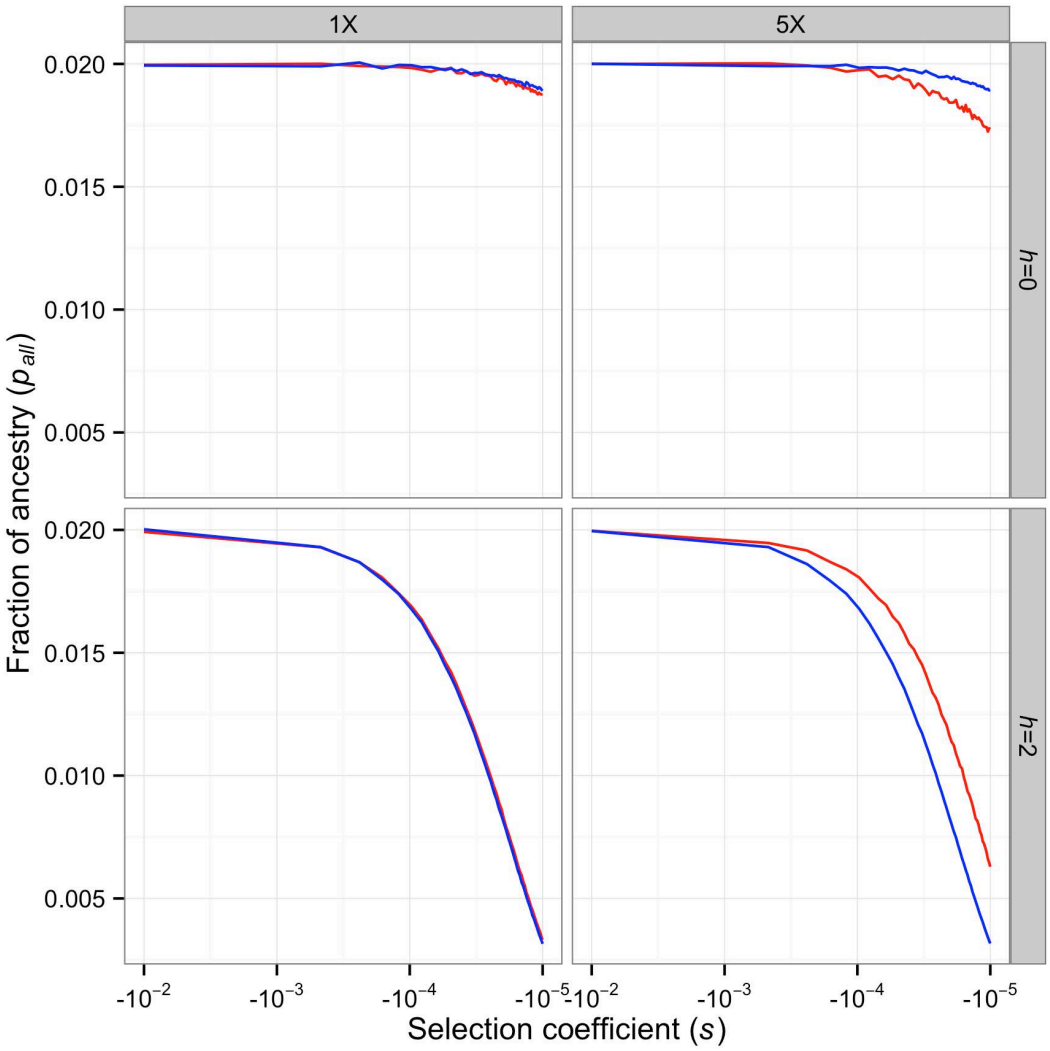

Figure S8: Predicted mean Neanderthal allele frequency at the end of the population bottlenecks in East Asia (ASN) and Europe (EUR) for the recessive and underdominant cases (h=0 and 2, respectively). (Left) Population sizes were set to those inferred in Keinan et al.[11] (Right) Population size in ASN was assumed to be 5-fold smaller than that estimated in Keinan et al.[11] In all cases, constant sized populations were simulated for 100 generations. Here f=2%.

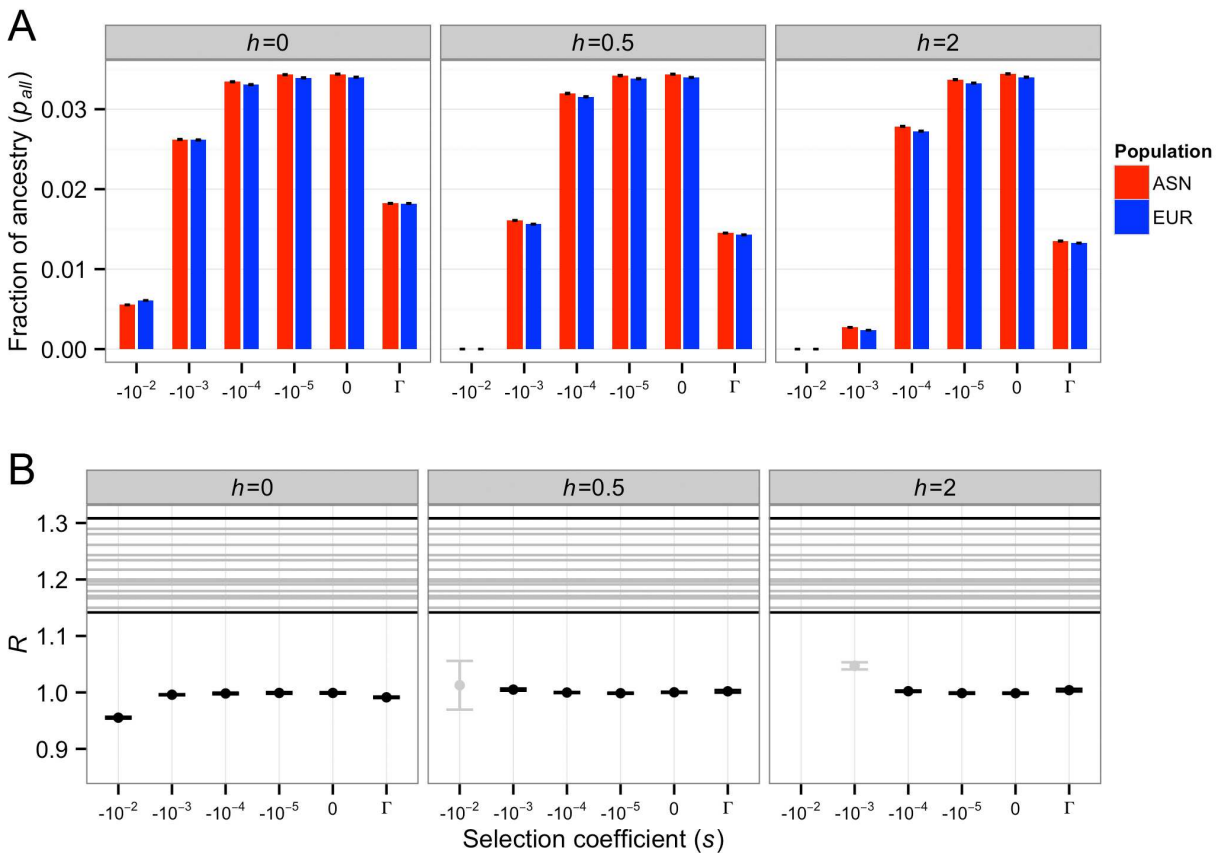

Figure S9: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Gravel et al.[22] complex demographic model when *f*=4%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

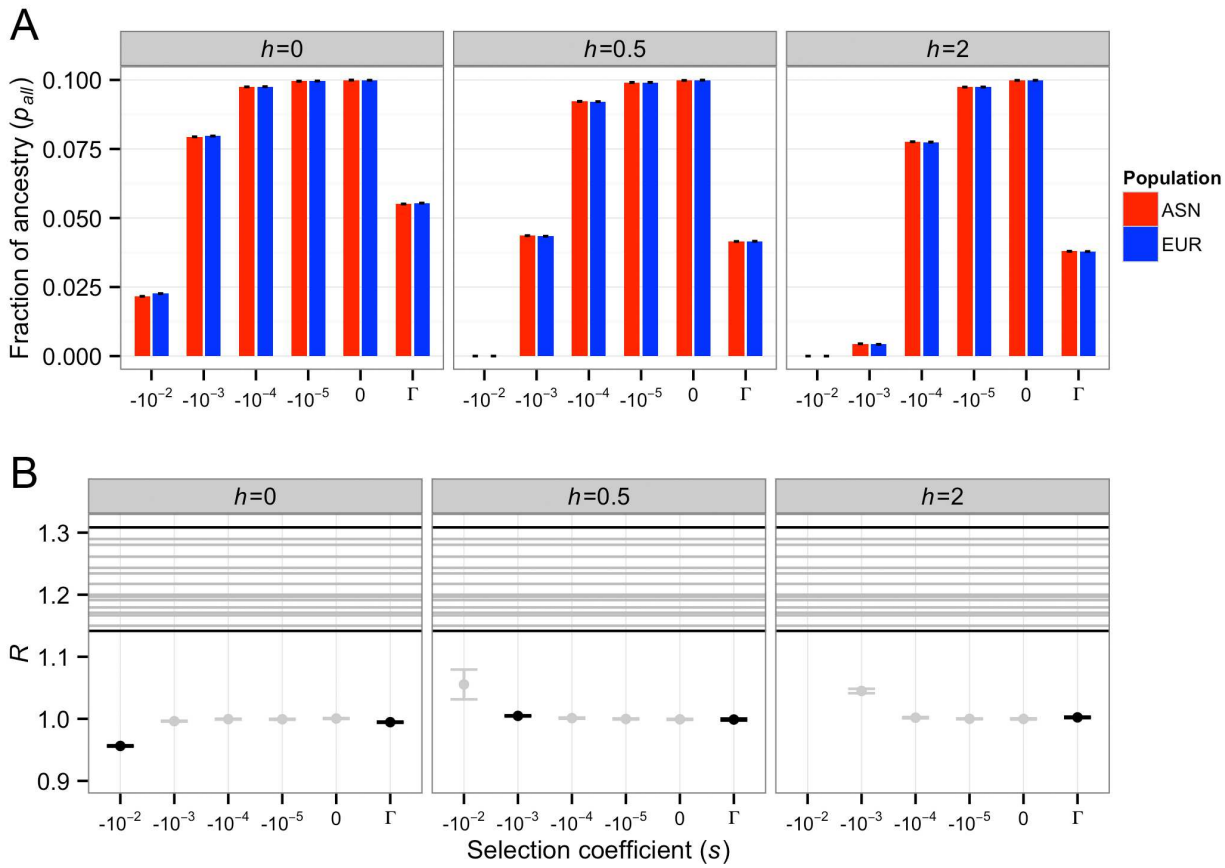

Figure S10: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Keinan et al.[11] demographic model when *f*=10%. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Models where the final proportion of Neanderthal ancestry is concordant with the empirical data (between 0.5-5% in (A)) are colored in black. Otherwise, they are colored gray.

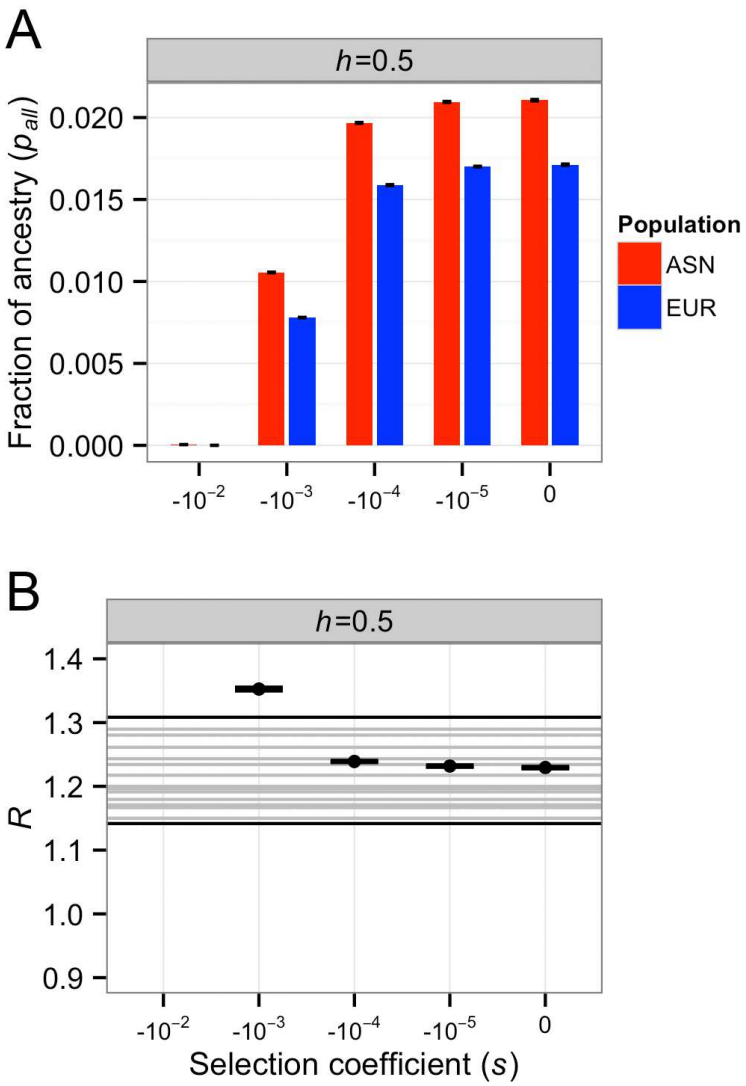

Figure S11: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Gravel et al.[22] complex demographic model when *f*=2% with a second pulse of Neanderthal admixture into East Asia. Specifically, 920 generations ago, the amount of Neanderthal ancestry at each site in East Asia was increased by 15% of the initial value of *f* (i.e. here 0.003 was added to the frequency of Neanderthal alleles in the East Asian population). Vernot and Akey[6,24] have estimated that that the second pulse of Neanderthal admixture into East Asia was about 15% of the initial admixture proportion, concordant with our present simulation. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Note that a broad range of selection coefficients provide values of *R* compatible with the observed ratio. The model where *s*=-0.01 predicts *R*=16. This point was omitted for plotting purposes.

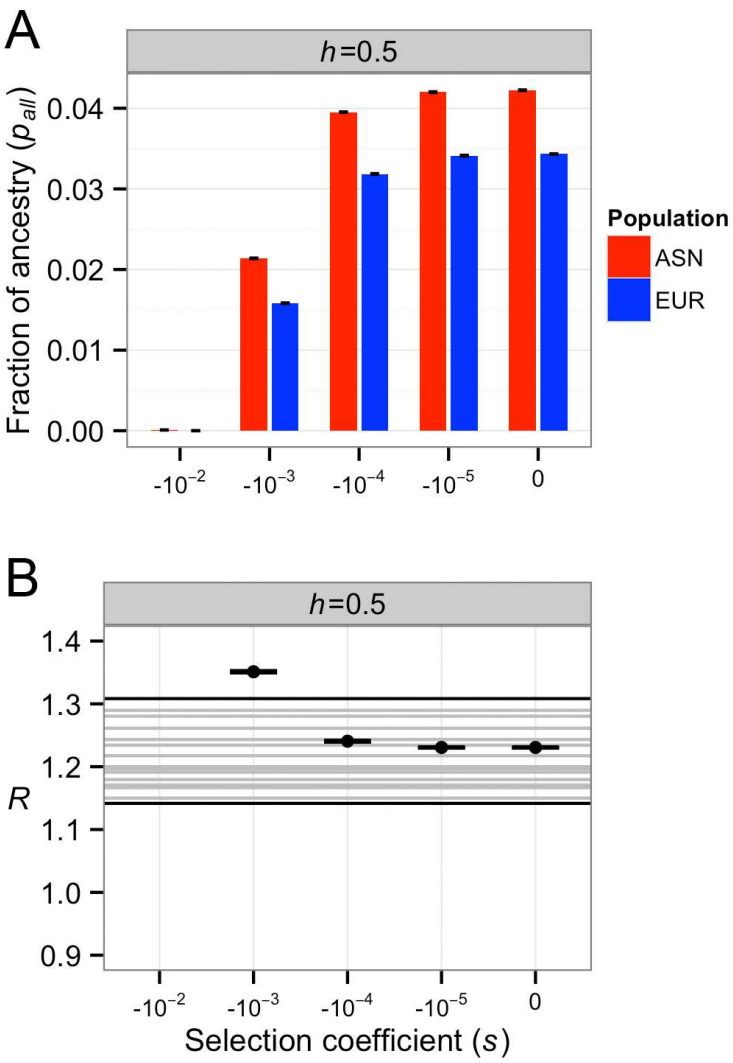

Figure S12: Predicted Neanderthal ancestry in East Asian (ASN) and European (EUR) populations under the Gravel et al.[22] complex demographic model when *f*=4% with a second pulse of Neanderthal admixture into East Asia. Specifically, 920 generations ago, the amount of Neanderthal ancestry at each site in East Asia was increased by 15% of the initial value of *f* (i.e. here 0.006 was added to the frequency of Neanderthal alleles in the East Asian population). Vernot and Akey[6,24] have estimated that that the second pulse of Neanderthal admixture into East Asia was about 15% of the initial admixture proportion, concordant with our present simulation. Each column depicts results for a different dominance coefficient (*h*). Γ denotes a gamma distribution of fitness effects. Error bars denote approximate 95% confidence intervals on our simulations. (A) The fraction of Neanderthal ancestry. (B) Ratio of Neanderthal ancestry in East Asians to Neanderthal ancestry in Europeans (*R*). Horizontal lines indicate the ratios of mean Neanderthal ancestry observed in empirical comparisons of an East Asian and a European population[7]. Note that a broad range of selection coefficients provide values of *R* compatible with the observed ratio. The model where *s*=-0.01 predicts *R*=16. This point was omitted for plotting purposes.

Table S1. Parameters used for the Keinan et al. bottleneck model

Population	*N*	*F*	t_B	t_{Blen}	N_B
Parameters inferred in Keinan et al.					
ASN	10063	0.123	720	100	407
EUR	10085	0.091	640	100	549
Shorter bottleneck					
ASN	10063	0.123	720	50	204
EUR	10085	0.091	640	50	275
Longer bottleneck					
ASN	10063	0.123	720	200	814
EUR	10085	0.091	640	200	1098
2-fold more severe bottleneck					
ASN	10063	0.246	720	100	275
EUR	10085	0.091	640	100	549
5-fold more severe bottleneck					
ASN	10063	0.615	720	100	110
EUR	10085	0.091	640	100	549

Table S2. Parameters used for the Gravel et al. model

Parameter	Value
t_1	980
t_2	920
N_{AFR}	14474
N_b	1861
N_{ASN0}	550
N_{EUR0}	1032
r_{ASN}	0.0048
r_{EUR}	0.0038
m_{ASN_AFR}	0.78e-5
m_{EUR_AFR}	2.5e-5
m_{EUR_ASN}	3.11e-5

Table S3: Summary of simulation results

model	*h*	seg_freq_ASN	seg_pct_ASN	seg_freq_EUR	seg_pct_EUR	anc_ASN	anc_EUR	ASN/EUR
Keinan bottleneck	0	0.11	0.22	0.09	0.25	0.02	0.02	0.99
Keinan bottleneck	0.5	0.12	0.14	0.10	0.16	0.02	0.02	1.00
Keinan bottleneck	2	0.12	0.13	0.10	0.14	0.01	0.01	1.00
Gravel	0	0.16	0.11	0.14	0.13	0.02	0.02	1.00
Gravel	0.5	0.18	0.08	0.16	0.09	0.01	0.01	1.02
Gravel	2	0.19	0.07	0.16	0.08	0.01	0.01	1.02

Description of the column headings: seg_freq_ASN, mean allele frequency in East Asia, conditional on the Neanderthal allele segregating at the end of the simulation; seg_pct_ASN, % of 1 million segments where the Neanderthal allele is segregating at the end of the simulation in East Asia; seg_freq_EUR, mean allele frequency in Europe, conditional on the Neanderthal allele segregating at the end of the simulation; seg_pct_EUR, % of 1 million segments where the Neanderthal allele is segregating at the end of the simulation in Europe; anc_ASN, mean Neanderthal ancestry in East Asia; anc_EUR, mean Neanderthal ancestry in Europe; ASN/EUR, Ratio of anc_ASN to anc_EUR

Table S4: Expected *D* statistics under realistic models of human history assuming 0,1, or 2 pulses of Neanderthal admixture

Model	P1	P2	Num P1	Num P2	D	SE
No admixture	ASN	EUR	84987	85221	-0.0014	0.0025
	ASN	AFR	90844	90332	-0.0028	0.0023
	EUR	AFR	90063	89785	-0.0015	0.0024
One pulse	ASN	EUR	102233	102151	0.0004	0.0022
	ASN	AFR	91001	109182	0.0908	0.0023
	EUR	AFR	90798	108897	0.0906	0.0022
Two pulse	ASN	EUR	106000	101487	0.0218	0.0021
	ASN	AFR	90837	113021	0.1088	0.0022
	EUR	AFR	90881	108552	0.0886	0.0022

D statistics were computed from data simulated using ms[25] under the demographic model estimated for human populations in Gravel et al.[22] with our own modifications and those suggested by Vernot and Akey[6,24]. Specifically, recent population growth, as used in Vernot and Akey was included in the model. We simulated the three human populations and a Neanderthal population that split from the human population 400,000 years ago. The Neanderthal population had a constant size of 1500 individuals. The one pulse model includes a 500-year period of migration between the ancestral non-African population and Neanderthals. The two-pulse model includes the same migration as in the one pulse model, except it includes an additional 500 years of migration between the Neanderthal and East Asian populations. Note, we decreased the human-Neanderthal migration rates by 2 relative to the values given in Vernot and Akey[6] to give *D* statistics more comparable to those observed in actual data. The precise ms commands for these models are given in Table S5.

The *D* test was computed as: *D*=(Num_P1-Num_P2)/(Num_P1+Num_P2). Our simulations assume that the derived allele can be accurately inferred. As such, we simulated the three human populations (EUR, AFR, ASN) and a Neanderthal population.

Standard errors were computed using a nonparametric bootstrap of the values shown in the table. This is appropriate as each site was simulated independently of the others.

The *D* statistics for all the simulations without any Neanderthal admixture are within 2 standard errors of 0. Further, the *D* statistic computed using ASN and EUR under the one pulse model also is within 2 standard errors of 0. This suggests that a model with one pulse of Neanderthal admixture cannot explain the higher Neanderthal ancestry in East Asia, even with a higher migration rate between African and Europe than between Africa and East Asia. The two-pulse model, however, predicts *D* statistics significantly >0.

Table S5: ms commands for neutral coalescent simulations in Table S4

Model	Command
No admixture	ms 4 1 -s 1 -I 4 1 1 1 1 0 -n 4 2.051984e-01 -n 1 58.002735978 -n 2 70.041039672 -n 3 187.55 -eg 0 1 482.46 -eg 0 2 570.18 -eg 0 3 720.23 -em 0 2 4 0 -em 0 2 4 0 -em 0 3 4 0 -em 0 3 4 0 -em 0 1 2 0.7310 -em 0 2 1 0.7310 -em 0 1 3 0.228072 -em 0 3 1 0.228072 -em 0 2 3 0.909364 -em 0 3 2 0.909364 -eg 0.006997264 1 0 -eg 0.006997264 2 2.089166e+01 -eg 0.006997264 3 3.006376e+01 -en 0.006997264 1 1.98002736 -en 0.031463748 2 7.774282e-01 -en 0.031463748 3 5.820793e-01 -ej 5.453352e-02 3 2 -en 5.453352e-02 2 7.774282e-01 -em 5.453352e-02 1 2 4.386 -em 5.453352e-02 2 1 4.386 -ej 8.207934e-02 2 1 -en 8.207934e-02 1 1.98002736 -en 0.20246238 1 1 -ej 9.575923e-01 4 1
One pulse	ms 4 1 -s 1 -I 4 1 1 1 1 0 -n 4 2.051984e-01 -n 1 58.002735978 -n 2 70.041039672 -n 3 187.55 -eg 0 1 482.46 -eg 0 2 570.18 -eg 0 3 720.23 -em 6.635294e-02 2 4 0 -em 0 3 4 0 -em 0 3 4 0 -em 0 1 2 0.7310 -em 0 2 1 0.7310 -em 0 1 3 0.228072 -em 0 3 1 0.228072 -em 0 2 3 0.909364 -em 0 3 2 0.909364 -eg 0.006997264 1 0 -eg 0.006997264 2 2.089166e+01 -eg 0.006997264 3 3.006376e+01 -en 0.006997264 1 1.98002736 -en 0.031463748 2 7.774282e-01 -en 0.031463748 3 5.820793e-01 -ej 5.453352e-02 3 2 -en 5.453352e-02 2 7.774282e-01 -em 5.453352e-02 1 2 4.386 -em 5.453352e-02 2 1 4.386 -ej 8.207934e-02 2 1 -en 8.207934e-02 1 1.98002736 -en 0.20246238 1 1 -em 6.566895e-02 2 4 4.386000e+01 -ej 9.575923e-01 4 1
Two pulse	ms 4 1 -s 1 -I 4 1 1 1 1 0 -n 4 2.051984e-01 -n 1 58.002735978 -n 2 70.041039672 -n 3 187.55 -eg 0 1 482.46 -eg 0 2 570.18 -eg 0 3 720.23 -em 0 1 2 0.7310 -em 0 2 1 0.7310 -em 0 1 3 0.228072 -em 0 3 1 0.228072 -em 0 2 3 0.909364 -em 0 3 2 0.909364 -eg 0.006997264 1 0 -eg 0.006997264 2 2.089166e+01 -eg 0.006997264 3 3.006376e+01 -en 0.006997264 1 1.98002736 -en 0.031463748 2 7.774282e-01 -en 0.031463748 3 5.820793e-01 -ej 5.453352e-02 3 2 -en 5.453352e-02 2 7.774282e-01 -em 5.453352e-02 1 2 4.386 -em 5.453352e-02 2 1 4.386 -ej 8.207934e-02 2 1 -en 8.207934e-02 1 1.98002736 -en 0.20246238 1 1 -em 6.566895e-02 2 4 4.386000e+01 -em 6.635294e-02 2 4 0 -em 5.316553e-02 3 4 8.832178e+00 -em 5.384952e-02 3 4 0 -ej 9.575923e-01 4 1

See Table S4 for a description of the demographic model

CHAPTER 2:

INFERENCE OF THE DISTRIBUTION OF SELECTION COEFFICENTS FOR NEW NONSYNONYMOUS MUTATIONS USING LARGE SAMPLES

Inference of the Distribution of Selection Coefficients for New Nonsynonymous Mutations Using Large Samples

Bernard Y. Kim,* Christian D. Huber,* and Kirk E. Lohmueller*,†,‡,1
*Department of Ecology and Evolutionary Biology, †Interdepartmental Program in Bioinformatics, and ‡Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California 90095

ABSTRACT The distribution of fitness effects (DFE) has considerable importance in population genetics. To date, estimates of the DFE come from studies using a small number of individuals. Thus, estimates of the proportion of moderately to strongly deleterious new mutations may be unreliable because such variants are unlikely to be segregating in the data. Additionally, the true functional form of the DFE is unknown, and estimates of the DFE differ significantly between studies. Here we present a flexible and computationally tractable method, called Fit$\partial a \partial i$, to estimate the DFE of new mutations using the site frequency spectrum from a large number of individuals. We apply our approach to the frequency spectrum of 1300 Europeans from the Exome Sequencing Project ESP6400 data set, 1298 Danes from the LuCamp data set, and 432 Europeans from the 1000 Genomes Project to estimate the DFE of deleterious nonsynonymous mutations. We infer significantly fewer (0.38–0.84 fold) strongly deleterious mutations with selection coefficient $|s| > 0.01$ and more (1.24–1.43 fold) weakly deleterious mutations with selection coefficient $|s| < 0.001$ compared to previous estimates. Furthermore, a DFE that is a mixture distribution of a point mass at neutrality plus a gamma distribution fits better than a gamma distribution in two of the three data sets. Our results suggest that nearly neutral forces play a larger role in human evolution than previously thought.

KEYWORDS deleterious mutations; diffusion theory; population genetics; site frequency spectrum

A fundamental concept in population genetics is the distribution of fitness effects (DFE) of new mutations. The DFE refers to the proportion of new mutations that have particular effects on fitness. The DFE is a crucial quantity in evolutionary genetics because it determines how selection affects genetic variation (Eyre-Walker and Keightley 2007), the conditions under which recombination could evolve (Keightley and Otto 2006), and the spectrum of mutations potentially involved in genetic diseases (Eyre-Walker 2010). Further, an accurate DFE is required for robust inference of the amount of adaptive evolution across taxa (Boyko *et al.* 2008; Gossmann *et al.* 2012; Castellano *et al.* 2016; Galtier 2016); a topic of widespread interest. Because this distribution is so important, considerable effort has been put forth toward estimating it in several species.

In organisms that allow experimental manipulation, the DFE can be directly estimated. Here, the DFE is either derived from direct measurements of fitness from a collection of single-step mutants, or indirectly inferred from observed changes in population fitness in mutation accumulation (MA) experiments (Eyre-Walker and Keightley 2007; Bataillon and Bailey 2014). The first approach, in combination with high-throughput methods, has been successfully applied to examine the full spectrum of (even weak) selection coefficients of mutations in small mutational target regions in a number of viral, bacterial, and yeast systems (Fowler *et al.* 2010; Hietpas *et al.* 2011; Boucher *et al.* 2014). They frequently report a gamma or unimodal, similarly shaped distribution of fitness effects (Bataillon and Bailey 2014), or a bimodal distribution with a second cluster of highly deleterious mutations (Acevedo *et al.* 2014; Bank *et al.* 2014;

doi: https://doi.org/10.1534/genetics.116.197145
Manuscript received October 23, 2016; accepted for publication February 14, 2017; published Early Online February 27, 2017.
Available freely online through the author-supported open access option.
Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.197145/-/DC1.
†Corresponding author: Department of Ecology and Evolutionary Biology, University of California, Los Angeles, 621 Charles E. Young Dr. South, Los Angeles, CA 90095-1606. E-mail: klohmueller@ucla.edu

Boucher *et al.* 2014). The second approach infers the DFE from fitness trajectories of a collection of populations over time in MA experiments, without directly identifying the mutations involved. Assuming that the true DFE is gamma distributed, they estimate the parameters of a gamma DFE that best fit to the observed changes in the mean and variance of fitness among replicate populations (Halligan and Keightley 2009). These studies point to a shape of the DFE that is less leptokurtic than an exponential distribution, with the mode different from zero. This could indicate that the true underlying DFE is more complex than the gamma distribution (Halligan and Keightley 2009), or reflect a bias of MA-based methods toward mutations with large fitness effects (Eyre-Walker and Keightley 2007). In summary, experimental estimates of the DFE suggest that a substantial proportion of new mutations are strongly deleterious. However, due to the inherent limitations of these methods, inference of the exact functional form of the genome-wide DFE is challenging.

A second category of methods to infer the DFE involves examining patterns of neutral and putatively deleterious genetic variation in natural populations, and finding the model of demographic history and purifying selection that can match the observed level of variation. This framework has been applied to many species including humans (Eyre-Walker *et al.* 2006; Keightley and Eyre-Walker 2007; Boyko *et al.* 2008; Li *et al.* 2010), *Drosophila* (Keightley and Eyre-Walker 2007; Kousathanas and Keightley 2013), yeast (Koufopanou *et al.* 2015), orangutans (Ma *et al.* 2013), gorillas (McManus *et al.* 2015), and mice (Halligan *et al.* 2013). Many of these studies suggest that the DFE has a strongly leptokurtic distribution, conflicting with the MA-based estimates. Consistent with the bimodal DFE found by many site-directed mutagenesis studies (Bataillon and Bailey 2014; Boucher *et al.* 2014), they find a large proportion of nearly neutral mutations, as well as many strongly deleterious mutations. For example, previous studies in humans (Eyre-Walker *et al.* 2006; Boyko *et al.* 2008) have estimated the parameters of a gamma distribution for the DFE of new nonsynonymous mutations. These studies have found ~57–61% of new nonsynonymous mutations to be moderately to strongly deleterious ($|s| \geq 10^{-3}$), about 15–16% to be weakly deleterious ($10^{-4} \leq |s| < 10^{-3}$), and the remainder (24–28%) to be nearly neutral (Figure 1).

The estimates of the DFE from genetic variation data for humans by Eyre-Walker *et al.* (2006) and Boyko *et al.* (2008) have been widely used in human population genetic studies. For example, these DFEs were used to estimate differences in the genetic load across human populations (Henn *et al.* 2016), to model the ancient introgression of Neanderthal alleles into humans (Harris and Nielsen 2016), as a model for the frequency spectrum of deleterious polymorphisms in simulating data for disease studies (Uricchio *et al.* 2016), to evaluate the contribution of background selection to diversity on the *Y* chromosome in humans (Wilson Sayres *et al.* 2014), and to estimate the strength of selection acting on disease genes (Moon and Akey 2016). While the Boyko *et al.* (2008) study has had considerable impact in the field, it is important to appreciate that the estimates of the DFE were made using a sample of a small number of individuals. As such, most of the variation segregating in those samples is likely to be neutral or nearly neutral. Inferences about the proportion of moderately and strongly deleterious mutations largely come from assuming the gamma distribution approximates the DFE of new mutations well, and then tabulating the proportion in those categories predicted by the gamma distribution. In other words, the second mode of strongly deleterious and lethal mutations observed by experimental studies is unlikely to be directly observed in polymorphism data sets, and these proportions are extrapolated from the long tail of the DFE.

This extrapolation of the proportion of strongly deleterious mutations may not be accurate. A more recent study using exome sequencing data from 200 Danish individuals (Li *et al.* 2010) estimated a DFE that differs considerably from that inferred in Boyko *et al.* (2008) or from the experimental estimates in lower organisms. Specifically, Li *et al.* (2010) found a mixture distribution consisting of a neutral point mass and gamma distribution fit best to their data (Figure 1). Additionally, they estimated that only 1% of new mutations have $|s| > 10^{-4}$ (compared to 57% in Boyko *et al.* 2008), and 78% of new mutations fall in the $10^{-4} \leq |s| < 10^{-3}$ range (compared to 15% in Boyko *et al.* 2008). Li *et al.* (2010) attributed this difference in the DFEs to their study considering a larger sample of individuals. As such, they surmised that they sampled more weakly deleterious variants, allowing more accurate inferences. However, this explanation has not been tested using simulations or larger data sets. Thus, the proportions of moderately *vs.* strongly deleterious nonsynonymous mutations in humans, as well as the functional form of the DFE, remain elusive.

Due to large-scale genome and exome sequencing projects, samples of hundreds to thousands of individuals are available (Tennessen *et al.* 2012; Fu *et al.* 2013; Lohmueller *et al.* 2013; The 1000 Genomes Project Consortium 2015). These large data sets should yield more reliable inferences of the DFE because moderately deleterious polymorphisms should be segregating, albeit at low frequency, in these samples (Supplemental Material, Figure S1 in File S1). As such, it should be possible to determine the functional form of the DFE and directly estimate the proportion of moderately and strongly deleterious mutations.

However, a major roadblock to using these new data sets for inference of the DFE is a lack of suitable software for inference from large samples. Generally, methods to infer the DFE summarize the allele frequency information of two classes of sites, one assumed to be neutral and the other selected by the site frequency spectrum (SFS). Then, they find the DFE that, under the inferred model of demography fit to the SFS from neutral sites, fits the observed SFS from selected sites. The method of Keightley and Eyre-Walker (2007), DFE-alpha, models demography using a Wright–Fisher transition matrix. It can only consider demographic models with one or

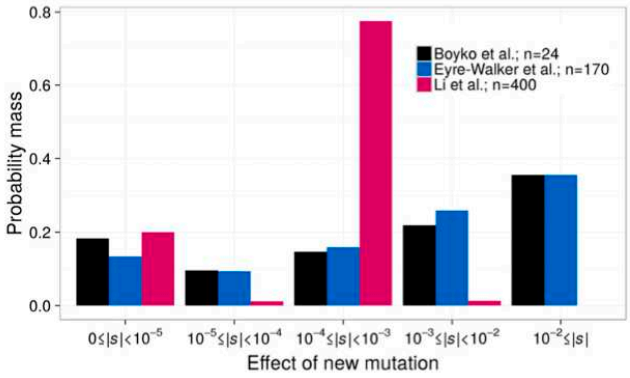

Figure 1 Previously inferred DFEs differ across studies. We rescaled the DFE in terms of the population size assumed or inferred in each study. A population size of 10,000 diploids is used to rescale the distribution of $2Ns$ to s for Eyre-Walker *et al.* (2006). For Boyko *et al.* (2008) and Li *et al.* (2010), we rescale the DFE from $2Ns$ to s using population sizes of 25,636 and 52,097 diploids, respectively (see *Materials and Methods*).

two size changes due to computational complexity, and it can be slow for the two-size-change model. This is particularly limiting in large samples of human genetic variation since a single-size-change demographic model is insufficient for capturing the excess of rare variation in human populations (Keightley and Eyre-Walker 2007; Kousathanas and Keightley 2013). Another class of methods to infer the DFE uses the Poisson random field (PRF) approach (Sawyer and Hartl 1992; Hartl *et al.* 1994; Williamson *et al.* 2005; Eyre-Walker *et al.* 2006; Boyko *et al.* 2008; Li *et al.* 2010). This approach has been implemented in the program PRFREQ (Boyko *et al.* 2008), but that implementation becomes numerically unstable when applied to samples larger than a few hundred individuals. The program ∂a∂i (Gutenkunst *et al.* 2009) uses a similar framework, but implementing a DFE is slow due to the way that the DFE is repeatedly integrated (Figure S2 in File S1). Thus, there is a need for a new software tool to infer the DFE from large samples.

In this study, we first extend the program ∂a∂i to analyze arbitrary DFEs in a computationally efficient manner. We implement these features in a module for ∂a∂i, which we call Fit∂a∂i. We then use this approach to estimate the DFE of deleterious, nonsynonymous mutations from multiple large human data sets. We consider several different functional forms for the DFE. We find that across the multiple data sets, a mixture distribution where a proportion of mutations are neutral and the remainder are gamma distributed fits best. Analysis of multiple data sets suggests there are fewer strongly deleterious mutations where $|s| > 10^{-2}$ (0.38–0.84 fold) than previously estimated in Boyko *et al.* (2008) (35%), regardless of the functional form of the DFE. Further, our results are not consistent with a model where 99% of new mutations have a selection coefficient weaker than 10^{-3}, as suggested by Li *et al.* (2010). Because we anticipate that our estimates of the DFE will be useful in subsequent simulation studies, we provide SFS_CODE (Hernandez 2008) and SLiM (Messer 2013) commands to simulate data where mutations have fitness effects from these DFEs.

Materials and Methods

Fit∂a∂i: software to infer the DFE

Here we present our new software, Fit∂a∂i, to infer distributions of selection coefficients of new mutations under the PRF model using the SFS. Fit∂a∂i is a module that extends the functionality of the Python package, ∂a∂i (Gutenkunst *et al.* 2009). Specifically, ∂a∂i uses diffusion theory to compute the expected SFS for a set of demographic parameters and selection coefficients. Fit∂a∂i offers a substantial computational improvement over the existing implementation of ∂a∂i for models involving more than a single selection coefficient. To do this, Fit∂a∂i computes SFSs for a range of selection coefficients and saves each SFS into an array. Then, subsequent integrations of the DFE are done using the array of precomputed SFSs. This process results in a substantial improvement in computational speed compared to the existing implementation of ∂a∂i, which recomputes the SFS for each selection coefficient in each step of the optimization process (Figure S2 in File S1). Fit∂a∂i also allows parallel computation of the SFS by using multiple cores or a cluster. Importantly, Fit∂a∂i leverages the modular nature of ∂a∂i to allow the user to define any arbitrary demographic model and DFE, including DFEs that are complex mixture distributions. Lastly, we incorporated an optimization routine that allows for constrained optimization of complex mixture distributions. Below we describe our inference procedure in greater detail, starting with inference of demography, followed by the details of Fit∂a∂i. We then discuss a simulation study to assess its performance, both under the assumptions of the PRF model as well as when some are violated, and then the data sets that we use to infer the DFE in humans.

Inference using the SFS

We inferred demography and selection from segregating sites in a maximum likelihood framework (Williamson *et al.* 2005, Boyko *et al.* 2008, Gutenkunst *et al.* 2009). Because both demography and selection affect patterns of deleterious mutations, our inference of the DFE begins (as done in Williamson *et al.* 2005 and Boyko *et al.* 2008) by first

estimating a demographic model for putatively neutral, synonymous sites. Then, conditional on the demographic parameter estimates, we infer the parameters for the DFE of nonsynonymous mutations.

To do this, we summarized synonymous and nonsynonymous sites with the SFS. The SFS can be described as a vector, $\boldsymbol{X} = [X_1, X_2, \ldots, X_{n-1}]$, in which each entry X_i describes the number of SNPs at frequency i in a data set of size n chromosomes. In the PRF framework, each entry in the SFS is assumed to be comprised of independent sites (Sawyer and Hartl 1992; Hartl *et al.* 1994).

Additionally, we folded the SFS to avoid difficulties with misidentification of the ancestral state (Hernandez *et al.* 2007). This form of the SFS counts the number of SNPs of minor allele frequencies (MAFs) 1 to $n/2$ without taking the ancestral state into account. The folded SFS has been shown to perform well at inferring the DFE of deleterious mutations (Keightley and Eyre-Walker 2007; Boyko *et al.* 2008; Tataru *et al.* 2016).

Inference of demography

A demographic model, the parameters of which are denoted as Θ_D, was fit to the SFS of synonymous sites with *∂a∂i* (Gutenkunst *et al.* 2009). Here, *∂a∂i* uses a diffusion approximation to compute the distribution of allele frequencies given some demographic model, $f(x; \Theta_D)$. Then, the expected number of SNPs at frequency i in a sample of size n chromosomes can be written as:

$$E[X_i] = \frac{\theta}{2}\int_0^1 x^i(1-x)^{n-i} f(x;\Theta_D)dx. \quad (1)$$

The multinomial likelihood, computed with the folded SFS, is maximized to estimate the demographic parameters:

$$\mathcal{L}(\boldsymbol{X}^*|\Theta_D) = \prod_i \left(\frac{E[X_i^*|\Theta_D]}{\sum_i E[X_i^*|\Theta_D]}\right)^{X_i^*}, \quad (2)$$

where X_i^* denotes the observed count of SNPs at frequency i in the folded SFS. The multinomial likelihood uses the proportions of SNPs at a particular frequency in the sample rather than the counts from the model. Therefore, it does not require an *a priori* assumption about the mutation rate or ancestral population size. The mutation rate of synonymous sites, denoted θ_S, was then computed as the scaling factor difference between the optimized SFS and the data.

When fitting models incorporating periods of rapid exponential growth with *∂a∂i*, we set the program parameter *dadi.Inference.set_timescale_factor*$=10^{-6}$, in contrast to the default setting of 10^{-3}. In *∂a∂i*, periods of exponential growth are approximated with a series of instantaneous size changes and, if the time steps are not small enough, parameters related to exponential growth will not be inferred correctly. This causes the demographic model to inaccurately predict the expected numbers of rare variants, biasing downstream inference of selection.

Inference of selection

To infer the DFE, we developed the Fit*∂a∂i* module, which uses *∂a∂i* and some of the methodological improvements of Ragsdale *et al.* (2016). First, we condition on the demographic model that was fit to synonymous sites using the procedure described above. Given that demography, *∂a∂i* is used to compute a distribution of allele frequencies $f(x; \Theta_D, \gamma)$, where Θ_D is a vector containing the demographic parameters inferred from synonymous sites and γ is a single population-scaled selection coefficient. Specifically, $\gamma = 2N_A s$, where N_A is the ancestral population size, but it is rescaled in each time period of the demographic model by the fold size change relative to the ancestral population. A DFE, denoted $g(\gamma)$, can be incorporated by generating $f(x; \Theta_D, \gamma)$ for a range of γ, then weighting the contribution of each of these frequency spectra by $g(\gamma)$ (Boyko *et al.* 2008):

$$E[X_i] = \frac{\theta}{2}\int_{-\infty}^{\infty}\int_0^1 x^i(1-x)^{n-i} f(x;\Theta_D,\gamma)g(\gamma|\Theta_{DFE})dxd\gamma. \quad (3)$$

In the standard implementation of *∂a∂i*, this process is time consuming because the SFS must be computed repeatedly during each step of optimization. In other words, $f(x; \Theta_D, \gamma)$ is computed each time a given value of γ is evaluated in a DFE. This process can be especially slow for large ranges of γ and for large sample sizes. Therefore, similar to Ragsdale *et al.* (2016), we initially computed the SFS for a range of selection coefficients, and then cached these results to avoid recomputing the frequency spectra (Figure S2 in File S1). In addition, we computed many frequency spectra in parallel to save time, added compatibility for user-defined DFEs, modified the optimization routines available in *∂a∂i* to work with cached spectra, and added the option to use constrained optimization for the inference of complex mixture distributions. These extensions are part of the Fit*∂a∂i* module.

To infer selection, we fixed the demographic parameters to the maximum-likelihood estimates (MLEs) inferred from synonymous sites, $\hat{\Theta}_D$. Then, we fit a DFE, the parameters of which are denoted as Θ_{DFE}, to the folded SFS of nonsynonymous sites by maximizing the Poisson likelihood:

$$\mathcal{L}(\boldsymbol{X}^*|\hat{\Theta}_D,\Theta_{DFE}) = \prod_i \exp\left(-E\left[X_i^*|\hat{\Theta}_D,\Theta_{DFE}\right]\right) \times \frac{E[X_i^*|\hat{\Theta}_D,\Theta_{DFE}]}{X_i^*!}. \quad (4)$$

Unlike the multinomial likelihood, the Poisson likelihood requires an *a priori* assumption about the mutation rate for nonsynonymous sites, θ_{NS}. To obtain this, we multiplied our estimate of θ_S by an assumed ratio of the nonsynonymous to synonymous mutation rate, θ_{NS}/θ_S, to obtain θ_{NS}. Specifically, we assumed the ratio to be $\theta_{NS}/\theta_S = 2.31$ (Huber *et al.* 2016), but also estimated the DFE using $\theta_{NS}/\theta_S = 2.5$ to provide a fair comparison to Boyko *et al.* (2008).

Each DFE is defined as an integrable function over a log-spaced range of 600 selection coefficients over intervals between $|s| = [10^{-8}, 0.5]$. We considered any portion of the DFE smaller than $|s| = 10^{-8}$ to be effectively neutral ($|s| = 0$), and any variants of $|s| > 0.5$ to have negligible probability of being found in polymorphism data (*i.e.*, not found in the data). Note that here we only consider the deleterious DFE but this function can easily be extended to incorporate positive selection (Huber *et al.* 2016).

Fit∂a∂i includes many of the standard DFEs (Boyko *et al.* 2008; Kousathanas and Keightley 2013), such as a gamma distribution and several mixture distributions. Specifically, we investigated mixture distributions where a proportion of mutations are neutral; with the rest following a gamma distribution as well as a mixture distribution where a fraction is neutral, a fraction is lethal, and the remainder follows an exponential distribution of fitness effects. Lastly, Fit∂a∂i includes arbitrary mixture distributions with a fixed number of fitness classes, or bins, where each bin can have its own range of selection coefficients (called the "discrete DFE"). Fit∂a∂i infers the proportion of new mutations in each fitness class. For mixture distributions incorporating a point mass at neutrality or lethality, we define the DFE so it can be treated as a single integrable function. We add the area of the point mass to a part of the distribution that is assumed to be neutral or lethal. For example, to add a point mass of neutral mutations to the "neutral+gamma DFE," we add the probability mass of neutral mutations, p_{neu}, to the $|s| = [0, 10^{-8}]$ portion of the distribution. Then, we integrate the gamma DFE between $|s| = [0, 10^{-8}]$ and sum it with p_{neu} to obtain the total mass of neutral mutations. Additionally, we used the SLSQP algorithm (Kraft 1988) as implemented in SciPy 0.17.1 to perform constrained optimization for mixture distributions incorporating more than two components. Throughout, we will use α and β to denote the shape and scale parameters of the gamma distribution, respectively, and λ to denote the rate parameter of the exponential distribution. The DFEs we report will be scaled by the ancestral population size. To estimate confidence intervals (C.I.'s) for our data, we Poisson resampled the nonsynonymous SFS and refit the DFE to the resampled data (Boyko *et al.* 2008). We note these C.I.'s are likely too narrow because they assume independence between all sites and do not account for the uncertainty in the demographic inference.

Simulations

To assess the performance of Fit∂a∂i, we performed forward-in-time simulations under different models of selection and demography. Simulations of independent sites were done using the program PReFerSIM (Ortega-Del Vecchyo *et al.* 2016), which simulates unlinked SNPs under the PRF model. We simulated synonymous sites separately with a population-scaled mutation rate of $\theta_S = 4000$ to approximately match the amount of synonymous genetic diversity in our data sets. We simulated nonsynonymous sites at a ratio of 2.5 nonsynonymous to 1 synonymous site, in other words using $L_{NS}/L_S = 2.5$. These simulations included sample sizes of $n = 24$ and $n = 2596$ chromosomes using a demographic model of constant size, a twofold instantaneous size change, and the demography inferred from the LuCamp data. We considered a variety of DFEs, which are described in more detail in specific instances below.

Because the PRF model makes several restrictive assumptions that are likely to not apply to real data sets, we performed an additional set of forward simulations violating these assumptions, and assessed the effect on inferences using Fit∂a∂i. Specifically, we investigated the effects that unmodeled linkage, background selection, and population structure might have on our inference of the DFE. To do this, we simulated 100–150 Mb regions using the recombination rate and arrangement of functional elements on human chromosome *1* (Huber *et al.* 2016) using the forward simulation program SLiM (Messer 2013). We assumed a gamma DFE for both nonsynonymous ($\alpha = 0.2$, $\beta = 200$) and conserved noncoding sites ($\alpha = 0.0415$, $\beta = 50$; see Torgerson *et al.* 2009). We assume that 400 generations ago, the ancestral population expanded eightfold and split into eight genetically isolated populations. This population size increase reflects the Neolithic expansion into Europe under the demic diffusion model (Chikhi *et al.* 2002; Gazave *et al.* 2014). We then sampled 100 chromosomes equally across the eight populations, combined them together, and analyzed them as though they were from a single population. The ancestral population was simulated for a burn-in period of 10*N* generations. To avoid prohibitively slow forward simulations, we simulated with an ancestral effective population size of 1000 and scaled mutation rate, recombination rate, selection coefficients, and demographic parameters accordingly (Aberer and Stamatakis 2013). We then fit a single population demographic model (which is the incorrect model) to the synonymous SNPs in each simulated data set using ∂a∂i. Then, conditional on these demographic parameters, we inferred the DFE using Fit∂a∂i. Our goal with these simulations is to mimic what researchers do in practice; where they do not know the true demographic model, but try to fit a simplified model to the data.

Data

We downloaded SNP data for 432 unrelated European (EUR) individuals from the 1000 Genomes Project phase 3 release (The 1000 Genomes Project Consortium 2015); 6503 individuals from the National Heart, Lung, and Blood Institute ESP6500SI-V2 European American (EUR) data set (Tennessen *et al.* 2012; Fu *et al.* 2013); and 2000 Danish individuals from the LuCamp project (Lohmueller *et al.* 2013). The 1000 Genomes phase 3 data were downloaded from the European Bioinformatics Institute file transfer protocol site http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/. Related individuals were removed by sampling only mothers and fathers from trios or extended families. Only SNPs from the phase 3, exome-targeted sequencing which passed the strict

mask filter were used. The total length of sites considered in the analysis, $L_S + L_{NS}$, was computed by taking this filtering into account. Variants were annotated using the 1000 Genomes Project-filtered annotations. The Exome Sequencing Project (ESP) ESP6400 data set was downloaded from the Exome Variant Server (http://evs.gs.washington.edu/EVS/). Only sites with 1800 or more European individuals sequenced, according to the site-by-site annotations, were used for the analysis. The LuCamp data were obtained from Lohmueller *et al.* (2013). For computational tractability, a hypergeometric distribution was used to project the LuCamp and ESP data sets down to sample sizes of 1298 and 1300 diploids (Gutenkunst *et al.* 2009), respectively, after filtering problematic individuals. All 432 unrelated European individuals from the 1000 Genomes Project were used. From these data, we assembled the folded SFS of synonymous and nonsynonymous sites, which were used for subsequent inference. To examine the effect of a smaller sample size on inference of demography and selection, we also projected the data down to a sample size of 24 chromosomes.

Estimating s from 2Ns

The DFEs inferred using the approach described above were for the population-scaled selection coefficient, scaled by twice the ancestral population size ($\gamma = 2N_A s$). Because we were interested in the distribution of s, we needed to estimate N_A. We computed N_A from the value of θ_S inferred from synonymous sites (Table S1 in File S1) using the equation $\theta_S = 4N_A\mu L_S$. Detailed information about these parameters used for our analysis can be found in Table S2 in File S1. However, this value of N_A depends on assumptions about the per-base-pair mutation rate and the ratio of possible nonsynonymous to synonymous sites, L_{NS}/L_S, since these quantities are computed from the total number of coding sites, $L_S + L_{NS}$. We assumed the mutation rate to be $\mu = 1.5 \times 10^{-8}$ to reflect estimates of the mutation rate in the human exome (Ségurel *et al.* 2014). For comparison to results from Boyko *et al.* (2008), we assumed the mutation rate to be $\mu = 1.8 \times 10^{-8}$. For the reanalysis of the Boyko *et al.* (2008) data set, we assumed the same ancestral population size, $N = 7778$ diploids, instead of computing it. To compute the total number of coding sites, $L_S + L_{NS}$, in each data set, we intersected the coding exons from the University of California Santa Cruz canonical transcript with the relevant filters for each data set. For the 1000 Genomes data, we intersected the phase 3 strict mask, the exome targets, and the hg19 coding exons. For the analysis of the ESP data set, we used the intersection of the hg19 coding exons and the site-by-site annotations to count the total number of coding sites for which $n \geq 2600$ alleles had been sequenced. For the LuCamp data, we obtained the value of $L_S + L_{NS}$ from Lohmueller *et al.* (2013).

Data availability

This research uses previously published data sets obtained as previously described. The Fit∂a∂i software is available at https://github.com/LohmuellerLab/fitdadi.

Results

Validation of Fit∂a∂i by comparison to previous analyses

We first examined the performance of Fit∂a∂i by fitting a demographic model and DFE to the African-American SFS from Boyko *et al.* (2008). Fit∂a∂i produces similar estimates of the shape and scale parameters of the gamma distribution compared to Boyko *et al.* (2008) (Boyko: $\alpha = 0.184$, $\beta = 2488$; Fit∂a∂i: $\alpha = 0.179$, $\beta = 3161$). Additionally, Fit∂a∂i produced similar estimates of the proportions of mutations in different bins of the DFE (Table S3 in File S1).

Performance on simulated data

We further investigated the performance of Fit∂a∂i by performing forward simulations under the PRF model (see *Materials and Methods*). We first considered the best-fit DFE of Boyko *et al.* (2008), rescaled to have an ancestral population size of $N = 10{,}085$ diploids ($\alpha = 0.184$, $\beta = 3226$). Fit∂a∂i is able to accurately infer the DFE from our simulated data sets (Table 1). Predictably, the variance of our estimates of the most deleterious portion of the DFE ($|s| > 10^{-2}$) is five- to sixfold greater for the small ($n = 24$) samples. However, for the samples of size $n = 2596$, the variance in the estimates of this portion of the DFE is significantly reduced and overall estimates of the proportions of the DFE where $|s| > 10^{-3}$ are accurate. Therefore, this sample size should allow us to accurately infer the most deleterious portions of the DFE.

Because it is not certain that the DFE is truly gamma distributed, we simulated data sets of 2596 chromosomes with other DFEs. Again, we scaled these DFEs to an ancestral population size of 10,085 diploids. We considered the mixture distribution of Li *et al.* (2010), which consists of 20% neutral and 80% gamma-distributed ($\alpha = 4$, $\beta = 2.148$) selection coefficients (the neutral+gamma DFE). We also considered a mixture distribution consisting of five bins, (the discrete DFE) with breaks at $|s| = [0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 1]$. Within each bin, the values of s were uniformly distributed. We examined three weighting schemes for this distribution. First, we computed the probability mass in each bin from the mixture distribution of Li *et al.* (2010). Then, we computed the probability mass in each bin from a gamma distribution with parameters $\alpha = 0.203$ and $\beta = 1082.1$, but where the mass in the $|s| = [10^{-2}, 1]$ bin was placed into the $|s| = [10^{-3}, 10^{-2})$ bin, and the opposite case where the mass in the $|s| = [10^{-3}, 10^{-2})$ bin was placed into the $|s| = [10^{-2}, 1]$ bin. This was done to evaluate whether we could distinguish between these discrete DFEs and to evaluate our ability to distinguish strongly deleterious mutations from moderately deleterious mutations in a large sample.

We find that if the true underlying DFE is distributed according to the Li *et al.* (2010) neutral+gamma distribution, the discrete DFE is able to estimate the true DFE, albeit with some limitations (Figure 2, A and B). For example, when the true DFE follows Li *et al.* (2010), our inference under the

Table 1 Performance of Fit∂a∂i on simulated data sets

Demography	n_{chr}	α (shape)	β (scale)	$0 \le \|s\| < 10^{-5}$	$10^{-5} \le \|s\| < 10^{-4}$	$10^{-4} \le \|s\| < 10^{-3}$	$10^{-3} \le \|s\| < 10^{-2}$	$10^{-2} \le \|s\|$
True[a]	—	0.184	3266[b]	0.182	0.096	0.146	0.219	0.357
Constant	2596	0.180 ± 0.010	3712.2 ± 980.2	0.186 ± 0.009	0.095 ± 0.002	0.144 ± 0.006	0.213 ± 0.013	0.363 ± 0.016
size	24	0.185 ± 0.028	3613.1 ± 4196.7	0.182 ± 0.017	0.097 ± 0.009	0.148 ± 0.023	0.221 ± 0.043	0.353 ± 0.060
Two-fold	2596	0.191 ± 0.007	2606.0 ± 410.7	0.178 ± 0.008	0.098 ± 0.002	0.152 ± 0.004	0.230 ± 0.009	0.341 ± 0.010
expansion	24	0.187 ± 0.023	3259.8 ± 2612.6	0.181 ± 0.016	0.097 ± 0.008	0.149 ± 0.019	0.223 ± 0.036	0.350 ± 0.050
LuCamp	2596	0.182 ± 0.008	3411.9 ± 558.5	0.184 ± 0.008	0.096 ± 0.001	0.145 ± 0.004	0.216 ± 0.008	0.358 ± 0.009
	24	0.186 ± 0.027	3435.4 ± 3249.9	0.182 ± 0.017	0.097 ± 0.009	0.148 ± 0.022	0.222 ± 0.042	0.351 ± 0.060

95% intervals were estimated as ± 1.96 SD of 100 replicates in each simulation set. chr, chromosome.
[a] Values show the true DFE used to simulate the data.
[b] Here the simulation was scaled to an ancestral population size of N = 10,085 diploids.

discrete DFE correctly estimates a negligible fraction of moderately or strongly deleterious new mutations ($|s| > 10^{-3}$), and correctly infers a mode of weakly deleterious mutations ($10^{-4} \le |s| < 10^{-3}$). However, estimates of the proportion of nearly neutral and neutral mutations ($|s| < 10^{-4}$) are less accurate (Figure 2A). When we simulate with the discretized distribution of Li *et al.* (2010), our estimates of the proportions of the discrete DFE are unbiased (Figure 2B). Additionally, we can distinguish between DFEs with varying proportions of moderately and strongly deleterious mutations (Figure 2, C and D). Although it is unlikely that the DFE of any natural population is discretized in this manner, these results show that the discrete DFE can help to approximate the general form of the underlying DFE, even if the true DFE is multimodal. This mimics what would be done in practice, where the precise functional form of the DFE is not known *a priori*. Therefore, fitting the discrete DFE should provide a general idea of the true DFE, especially if the true DFE is significantly multimodal. Notably, the discrete distribution can distinguish between strongly and moderately deleterious mutations at our sample size of 2596 chromosomes.

Simulations with linkage and population structure

The procedure of first inferring demography from the synonymous SFS and then selection from the nonsynonymous SFS provides unbiased estimates of selection, even in the presence of linkage (Boyko *et al.* 2008; Messer and Petrov 2013; Comeron 2014). In other words, this methodology controls for the effects of selection at linked sites. However, it is unclear what effect population structure might have on inference of the DFE. It is well known that such cryptic structure affects the SFS and may bias demographic inference (Ptak and Przeworski 2002; Gazave *et al.* 2014). Further, large human resequencing data sets may contain cryptic population structure (Novembre and Ramachandran 2011). For example, the 1000 Genomes European sample is comprised of five different subpopulations. To examine the performance of Fit∂a∂i when applied to data sets where the assumptions of the PRF model and a single, unstructured population are violated, we performed forward simulations including background selection and population structure (see *Materials and Methods*). We fit a single population, single-size-change demographic model to synonymous sites; and then, conditional on the size-change demographic model, a gamma DFE to nonsynonymous sites for each simulation replicate. Even when using the incorrect demographic model, we accurately infer selection from simulated data in the presence of linkage and population structure (Figure 3). Importantly, the single-size-change demographic model provides a reasonable approximation to the SFS when there are both population expansions and structure (Figure 3A). This in turn allows for the accurate estimation of both the shape and scale parameters of the gamma DFE (Figure 3B).

Therefore, simulations and a comparison to existing empirical data suggest that Fit∂a∂i can reliably infer the DFE in the presence of complex demographic scenarios. Below we present additional simulation scenarios to examine the performance of Fit∂a∂i with varying sample sizes and when the assumed demography and DFE are misspecified.

Demographic inference

We begin by fitting a demographic model to the synonymous SFS of each of the three data sets (LuCamp, ESP, and 1000 Genomes) using ∂a∂i. Briefly, this demographic model incorporates an out-of-Africa bottleneck, a recovery period, and recent exponential population growth (Figures S3 and S4 in File S1). Our estimates of demography as well as the inferred population sizes are presented in Tables S1 and S2 in File S1. Predictably, the parameter describing the magnitude of recent population expansion is harder to infer when using a sample size of 24 chromosomes than when using the larger sample sizes (n = 2596 chromosomes). Although the demographic model we infer is biased by linked selection, this step controls for these effects when inferring selection (Boyko *et al.* 2008; Kousathanas and Keightley 2013; Messer and Petrov 2013; Huber *et al.* 2016).

Inference of the DFE from large data sets

Here we estimate the DFE for new nonsynonymous mutations using large samples. First, like previous studies, we fit a gamma distribution to the DFE (Table S4 in File S1). We infer a strongly leptokurtic distribution where there are many neutral and nearly neutral mutations (*i.e.*, 34–37% of new mutations have $|s| < 10^{-4}$), as well as a class of strongly deleterious mutations (*i.e.*, 15–22% of new mutations have $|s| > 10^{-2}$). Interestingly, the estimates from the three

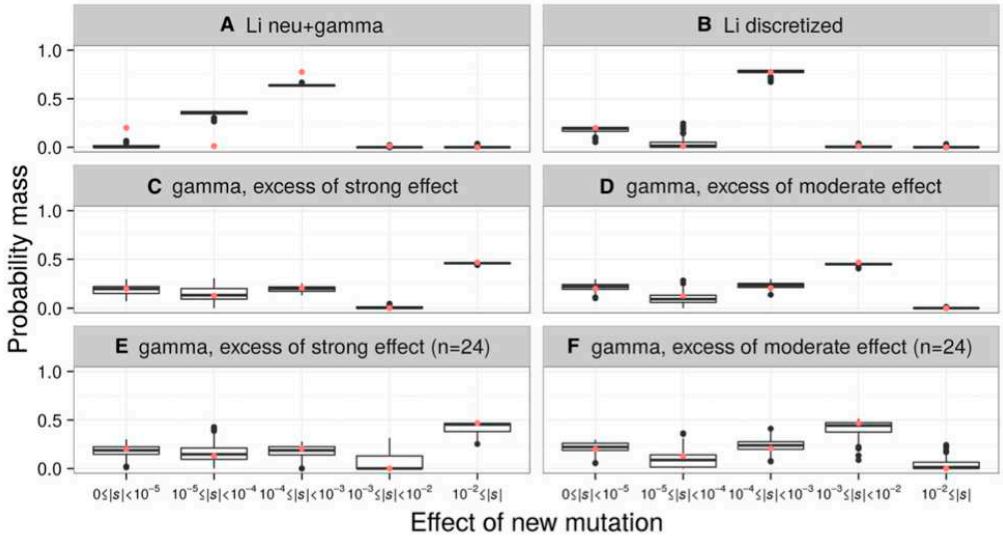

Figure 2 The discrete DFE can recover the approximate form of the DFE from simulated data. The distributions of the proportions of mutations with different selective effects, as inferred by the discrete DFE for 100 simulated data sets, are shown. Each simulation set assumed the demographic model fit to the LuCamp synonymous SFS. A red point depicts the true proportions of the simulated DFE. The true DFE for each set is: (A) the continuous neutral+gamma distribution of Li *et al.* (2010) (p_{neu} = 0.2, α = 4, β = 1.065 × 10^{-4}), (B) the discretized version of that distribution, (C–F) a gamma DFE (α = 0.215, β = 567.1), but where (C and E) the mass of the $10^{-3} \leq |s| < 10^{-2}$ bin was added to the $10^{-2} \leq |s|$ bin, and (D and F) where the mass of the $10^{-2} \leq |s|$ bin was added to the $10^{-3} \leq |s| < 10^{-2}$ bin. The data sets simulated for (C) and (D) had sample sizes of n = 2596 chromosomes, while the data sets for (E) and (F) had sample sizes of n = 24 chromosomes.

different data sets are generally concordant, though the 95% C.I.'s sometimes do not overlap. While this may suggest that the differences cannot be attributable to limited amounts of data in the SFS, we caution that these C.I.'s are likely too narrow because they do not account for the nonindependence of SNPs or the uncertainty of demographic estimates.

When compared directly to Boyko *et al.* (2008), the best-fit gamma DFEs inferred from all three data sets are generally shifted toward neutrality (Table 2 and Tables S4 and S5 in File S1), even when matching the mutation rates to those of Boyko *et al.* (2008) ($\mu = 1.8 \times 10^{-8}$ and $L_{NS}/L_S = 2.5$). We infer 19.2–22.9% of new mutations have a selection coefficient $|s| < 10^{-5}$, compared to the 18.3% observed by Boyko *et al.* (2008). This corresponds to a 1.05- to 1.25-fold increase. Additionally, we infer 24.5–29.8% of new mutations are strongly deleterious ($|s| > 10^{-2}$), which corresponds to a 0.69- to 0.84-fold decrease of the 35.5% inferred by Boyko *et al.* (2008). Taken together, when assuming a gamma distribution for the DFE, all three data sets suggest fewer strongly deleterious mutations than seen in Boyko *et al.* (2008).

Next, we explored the fit of complex DFEs to these large samples. Using the same combination of mutation rates as with the gamma, we fit the neutral+gamma mixture distribution; a mixture distribution of a point mass of neutral, a point mass of lethal, and exponentially distributed new

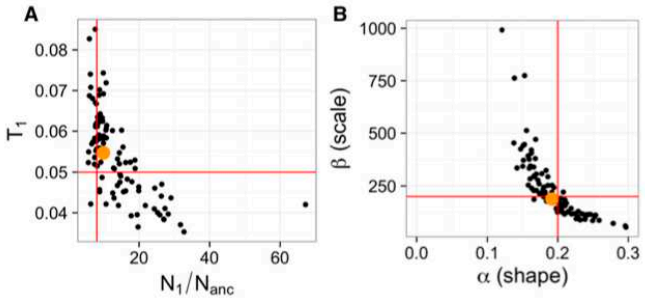

Figure 3 Inference of the DFE is robust to misspecification of the demographic model and background selection. Points show the MLEs of the (A) demographic parameters and (B) DFE parameters inferred from 100 simulated data sets with linkage and population structure. Red lines denote the true values and the yellow dots denote the median estimates across the 100 data sets. Estimates of time of expansion (T_1) and the ratio of current to ancestral population size (N_1/N_{ANC}) tend to be biased because demography is incorrectly modeled due to background selection, but estimates of the DFE are unbiased.

Table 2 MLEs of various DFEs

Data set	DFE	Parameter MLEs	Log-likelihood	AIC	|ΔAIC|[a]
LuCamp	Gamma	$\alpha = 0.215$, $\beta = 562.1$	−3334.7	6673.4	13
	Neu+gamma	$p_{neu} = 0.164$, $\alpha = 0.338$, $\beta = 367.7$	−3327.2	6660.4	0
	Neu+exp+let	$p_{neu} = 0.304$, $p_{exp} = 0.613$, $\lambda = 66.56$	−3337.8	6681.6	21.2
	Discrete[b]	$p_1 = 0.309$, $p_2 = 0.024$, $p_3 = 0.247$, $p_4 = 0.372$	−3334.1	6676.2	15.8
1kG EUR	Gamma	$\alpha = 0.186$, $\beta = 875.0$	−1450.5	2905.0	0
	Neu+gamma	$p_{neu} = 0.031$, $\alpha = 0.199$, $\beta = 820.6$	−1450.8	2907.6	2.6
	Neu+exp+let	$p_{neu} = 0.312$, $p_{exp} = 0.509$, $\lambda = 41.48$	−1472.0	2950.0	45
	Discrete[b]	$p_1 = 0.286$, $p_2 = 0.099$, $p_3 = 0.222$, $p_4 = 0.305$	−1453.4	2914.8	9.8
ESP EUR	Gamma	$\alpha = 0.169$, $\beta = 1327.4$	−3012.6	6029.2	2.6
	Neu+gamma	$p_{neu} = 0.092$, $\alpha = 0.207$, $\beta = 1082.3$	−3010.3	6026.6	0
	Neu+exp+let	$p_{neu} = 0.341$, $p_{exp} = 0.504$, $\lambda = 63.90$	−3071.6	6149.2	122.6
	Discrete[b]	$p_1 = 0.334$, $p_2 = 0.041$, $p_3 = 0.201$, $p_4 = 0.306$	−3029.5	6067.0	40.4

These results are reported assuming $L_{NS}/L_S = 2.31$ and $\mu = 1.5\times10^{-8}$. See Table S4 in File S1 for additional information. The shape and scale parameters of the gamma distribution are denoted with α and β, respectively, and the rate parameter of the exponential distribution is denoted with λ. Neu, neutral; exp, exponential; let, lethal; 1kG, 1000 Genomes.

[a] Change in AIC relative to the model with the lowest AIC.

[b] In terms of |s|, these parameters correspond to the ranges of |s| corresponding to: $[0, 10^{-5})$, $[10^{-5}, 10^{-4})$, $[10^{-4}, 10^{-3})$, and $[10^{-3}, 10^{-2})$, respectively. The mass in the $[10^{-2}, 1]$ range is the complement of the total mass of the four aforementioned categories.

mutations (the "neutral+exp+lethal" DFE); and the discrete DFE described previously. The MLEs, as well as the proportion of mutations with varying selection coefficients predicted by these distributions, are depicted in Table 2, Table 3, and Table S4 in File S1.

When we assume $\mu = 1.5 \times 10^{-8}$ and $L_{NS}/L_S = 2.31$, the neutral+gamma DFE fit best to the LuCamp and ESP data sets as reflected by the highest log-likelihood and Akaike information criterion (AIC) score (Table 3 and Table S4 in File S1). The gamma still fit best to the 1000 Genomes data set. Compared to the gamma DFEs inferred previously for two data sets, our best-fitting DFEs predict slightly fewer (0.92–0.98 fold) new mutations with $|s| > 10^{-2}$, and slightly more (1.06–1.18 fold) new mutations of $|s| < 10^{-5}$. When we matched the mutation rates of Boyko *et al.* (2008) with $\mu = 1.8 \times 10^{-8}$ and $L_{NS}/L_S = 2.5$, the discrete DFE fit best to the LuCamp data set (Table 3 and Table S4 in File S1). However, the gamma DFE continued to fit best to the 1000 Genomes and ESP data sets under these assumptions. The best fitting DFEs are depicted in Figure 4 and Figure S5 in File S1, and a comparison of the model to the SFS of the data are shown in Figure S6 in File S1. When using the larger mutation rate, we find the discrete DFE to fit best to the LuCamp data set, which predicts significantly fewer (0.54 fold) new mutations of $|s| < 10^{-2}$ than the gamma DFE fit using the same mutation rate.

One concern is that biases in SNP calling may affect these inferences. One way to test for this is by masking the singletons in the analysis, since singletons may be enriched for false SNPs due to sequencing errors (Boyko *et al.* 2008). We fit the gamma and neutral+gamma DFEs while masking the singleton category in the SFS and find little difference in the inferred DFEs (Table S6 in File S1). This finding suggests our inferences are robust to potential errors in SNP calling in these data sets.

The DFEs we have inferred thus far differ from that inferred in Boyko *et al.* (2008). In that study, 35.5% of new nonsynonymous mutations were inferred to be strongly deleterious in African-Americans, and 37.9% in Europeans. We infer fewer new strongly deleterious nonsynonymous mutations, even when matching the mutation rates used in Boyko *et al.* (2008) (Figure 4 and Figure S5 in File S1). Furthermore, the distribution of $2Ns$ also shows fewer strongly deleterious mutations (27.1–36.9% of mutations with $2Ns > 100$) than seen in Boyko *et al.* (2008) (40.4% of mutations with $2Ns > 100$; Figure S5 in File S1). Our results remain consistent across data sets and assumed forms of the DFE.

Additionally, our estimates of the DFE differ substantially from the estimates provided by Li *et al.* (2010). Specifically, Li *et al.* (2010) infer almost no strongly or moderately deleterious new nonsynonymous mutations. That is, 1% of new nonsynonymous mutations have selection coefficients of $10^{-3} < |s| < 10^{-2}$ and 0% have a selection coefficient $|s| > 10^{-2}$ (Figure 1). All of our estimates infer that at least ~30% of new nonsynonymous mutations have a selection coefficient $|s| > 10^{-3}$, even when the assumed mutation rate is small (Figure 4 and Table 3). Our simulations suggest if the true underlying DFE follows that suggested by Li *et al.* (2010), we should be able to estimate those proportions using both the neutral+gamma and the discrete DFE (Figure 2, A and B). The fact that our inferences did not show similar estimates to those inferred by Li *et al.* (2010) suggests that our data and analyses are not consistent with the distribution inferred by Li *et al.* (2010) (Table S5 in File S1). In the following sections, we explore several reasons why the different studies infer different DFEs.

Assessing the role of sample size and model misspecification using simulations

One possibility for the distinct estimates of the DFE is that the three studies used different sample sizes. Larger samples will have more moderately and strongly deleterious variants segregating than will smaller data sets (Figure S1 in File S1). To investigate the effect of sample size on our ability to infer the

Table 3 We infer more nearly neutral ($|s| < 10^{-5}$) and fewer strongly deleterious ($|s| \geq 10^{-2}$) new mutations than previous studies

Data set	μ, L_{NS}/L_S	Best fit DFE	$0 \leq \lvert s \rvert < 10^{-5}$	$10^{-5} \leq \lvert s \rvert < 10^{-4}$	$10^{-4} \leq \lvert s \rvert < 10^{-3}$	$10^{-3} \leq \lvert s \rvert < 10^{-2}$	$10^{-2} \leq \lvert s \rvert$
Boyko *et al.* (2008) (AA[a])	1.8×10^{-8}, 2.5[b]	Gamma	0.183	0.096	0.147	0.219	0.355
1000 Genomes	1.8×10^{-8}, 2.5[b]	Gamma	0.217 (0.212–0.223)	0.112 (0.111–0.113)	0.169 (0.165–0.172)	0.243 (0.235–0.249)	0.259 (0.252–0.266)
ESP		Gamma	0.229 (0.223–0.234)	0.105 (0.104–0.106)	0.152 (0.150–0.155)	0.216 (0.212–0.221)	0.298 (0.294–0.302)
LuCamp		Discrete	0.278 (0.221–0.303)	0.027 (0.001–0.110)	0.211 (0.167–0.234)	0.352 (0.330–0.373)	0.132 (0.124–0.142)
1000 Genomes	1.5×10^{-8}, 2.31	Gamma	0.237 (0.231–0.243)	0.127 (0.125–0.128)	0.192 (0.188–0.197)	0.266 (0.259–0.272)	0.178 (0.171–0.186)
ESP		Neu+gamma	0.263 (0.250–0.277)	0.104 (0.091–0.114)	0.167 (0.160–0.173)	0.249 (0.241–0.259)	0.217 (0.211–0.221)
LuCamp		Neu+gamma	0.242 (0.223–0.260)	0.091 (0.072–0.107)	0.194 (0.183–0.203)	0.332 (0.313–0.352)	0.141 (0.129–0.152)

Results in comparison with data from Boyko *et al.* (2008). C.I.'s were constructed by Poisson resampling the nonsynonymous SFS and fitting a DFE 200 times. See Table S5 in File S1 for additional information. Neu, neutral.

[a] African-American.

[b] The results presented with the assumptions $L_{NS}/L_S = 2.5$ and $\mu = 1.8 \times 10^{-8}$ match the mutation rate assumptions of Boyko *et al.* (2008).

DFE, we simulated 200 data sets, without linkage, of sample sizes n = 12, 24, 100, 250, and 500 chromosomes. Each data set was simulated using the demographic model and gamma DFE inferred from the LuCamp data set.

First, we simulated neutral synonymous sites and inferred the demographic parameters from each data set. This was done in two ways. First, we estimated the parameters from the full demographic model that was used to generate the data (herein the "full" model). Second, we inferred the parameters in a demographic model with only three instantaneous size changes (herein the "three-epoch" model). This is meant to mimic the situation in Boyko *et al.* (2008), where the true demography of the European population was likely complex, but simpler three-epoch demographic models could accurately fit the synonymous SFS. Next, as done in our inference and in previous studies, we estimated the parameters of a gamma distribution for the DFE of nonsynonymous mutations, conditioning separately upon the two demographic models.

When the full demographic model was fit to the simulated data sets, we found the variance of our estimates, both of demography and selection, decreased as sample size increased (Figure S7 in File S1). We were unable to correctly infer the magnitude of recent population growth with small sample sizes, consistent with previous work (Keinan and Clark 2012; Nelson *et al.* 2012; Tennessen *et al.* 2012; Fu *et al.* 2013). However, this did not affect the inference of selection as long as the demographic model fit reasonably well to synonymous sites (Figure S7 in File S1). At small sample sizes, the three-epoch model could fit the synonymous SFS well and thus estimates of selection were also unbiased. However, we found that for sample sizes >100 chromosomes, the three-epoch model increasingly became unable to account for the excess of rare variants caused by recent growth. The inability to account for the rare variants in the sample then biased the estimates of both the shape and scale parameters of the gamma distribution. However, this effect seems to be negligible at a sample size of 24 chromosomes (Figure S7 in File S1).

As long as the demographic model fits the observed SFS of synonymous sites, small sample sizes can estimate the parameters in a gamma distributed DFE, even when the demographic model is not the correct one. The accuracy of the estimates increases with sample size, especially for the scale parameter, and notably provides better estimates of the strongly deleterious portion of the DFE (Figure S7B in File S1 and Table 1). Thus, the results of Boyko *et al.* (2008) are unlikely to be affected by misspecification of demography due to small sample size.

Another possibility for the varying estimates of the DFE is that the DFE itself may be misspecified. Although parametric distributions are convenient for approximating the DFE, the true form of the DFE is unknown. Additionally, we have shown that the neutral+gamma DFE and the discrete DFE can fit large data sets better than the gamma DFE. To investigate an example of what would happen if the DFE is misspecified, we

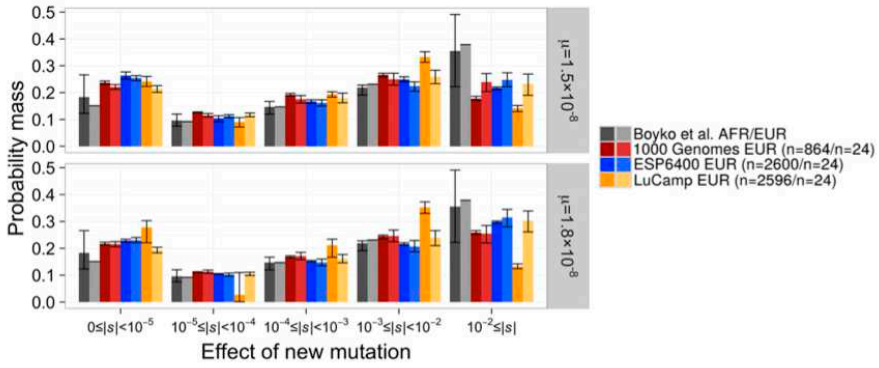

Figure 4 The distribution of selection coefficients of new mutations under our best-fit DFEs compared to Boyko *et al.* (2008). Results are presented for the best-fit DFE to each full data set and the best-fit DFE when the data were projected down to $n = 24$ chromosomes. C.I.'s were estimated by Poisson resampling the nonsynonymous SFS and fitting a DFE 200 times. C.I.'s for the DFE fit to the Boyko *et al.* (2008) European data set were unavailable. Note that our models predict more nearly neutral mutations ($0 \leq |s| < 10^{-5}$) and fewer strongly deleterious mutations ($10^{-2} \leq |s|$) than Boyko *et al.* (2008), across all mutation rates. Top panel denotes our favored mutation rate while the bottom panel denotes the mutation rate used by Boyko *et al.* (2008). See Figure S5 in File S1 for a comparison of the population-scaled selection coefficients ($2Ns$).

simulated 100 data sets without linkage for the best-fit neutral+gamma DFE inferred from the LuCamp data set, scaled to an ancestral population size of 10,085 diploids ($p_{neu} = 0.164$, $\alpha = 0.338$, $\beta = 358.8$). We also downsampled each data set to $n = 24$ chromosomes. Then, we fit a gamma and neutral+gamma DFE to each full and downsampled data set.

When the true DFE is neutral+gamma distributed, inference of the DFE from small samples overestimates the proportion of strongly deleterious mutations (Figure 5). When the DFE is correctly specified, we obtain unbiased estimates of the DFE even from small samples. However, at a sample size of $n = 24$ chromosomes, both the gamma and neutral+gamma distributions have a similar log-likelihood (Figure 5A). This was also observed in Boyko *et al.* (2008). Then, the extra parameter in the neutral+gamma distribution penalizes the true DFE when choosing the best-fit DFE by AIC score. This leads one to choose the gamma distribution as the best-fit DFE to the small sample, even when the true DFE follows a neutral+gamma distribution. Fitting the gamma distribution yields a DFE with more new mutations with $|s| > 10^{-2}$ and a decrease in the proportion of new mutations with $|s| < 10^{-5}$ compared to the true DFE (Figure 5B).

Assessing the role of sample size using real data

Next, we investigated the role that sample size has on inference of the DFE from real data. To do this, we projected our synonymous and nonsynonymous frequency spectra down to a sample size of $n = 24$ chromosomes to match the sample size of Boyko *et al.* (2008), then fit a demographic model and DFEs as previously described. Here we used the mutation rate assumptions $\mu = 1.5 \times 10^{-8}$ and $L_{NS}/L_S = 2.31$, but also matched the mutation rate of Boyko *et al.* (2008) ($\mu = 1.8 \times 10^{-8}$ and $L_{NS}/L_S = 2.5$). Then, we fit the gamma, neutral+gamma, and discrete DFEs—which were the best-fitting distributions to the full data—to the downsampled data sets.

As predicted by our simulations, there is generally little difference in the fit of the different DFEs to the downsampled data sets in terms of log-likelihood (Table S5 in File S1). The neutral+gamma and discrete DFEs often fit better than the gamma, but the difference in log-likelihood is small (0.1–0.6 log-likelihood units). Thus, the gamma DFE is selected as the best-fit DFE for all downsampled data sets by AIC. These results mimic what was observed in our simulations, although the pattern is not wholly consistent across data sets and mutation rates. When we assume $\mu = 1.8 \times 10^{-5}$ and $L_{NS}/L_S = 2.5$, the gamma DFE fits best to both the full and downsampled 1000 Genomes and ESP data sets (Figure 4 and Table S5 in File S1). There also appears to be little difference between the gamma DFE fit to the full and downsampled data. By contrast, the discrete DFE fits best to the LuCamp data under these mutation rates. Additionally, the neutral+gamma fit best to the full ESP and LuCamp data when we assume $\mu = 1.5 \times 10^{-5}$ and $L_{NS}/L_S = 2.31$. The gamma DFE fit to the downsampled versions of these data sets predicts more strongly deleterious ($|s| > 10^{-2}$) and more nearly neutral ($|s| < 10^{-5}$) new mutations (Figure 4 and Table S5 in File S1). The DFE fit to the 1000 Genomes data using the lower mutation rates does not follow this pattern. The gamma DFE fits best to both versions of the data set, yet the estimates from the small data set still predict more strongly deleterious new mutations. These results seem to corroborate the results from our simulations. That is, fitting a DFE using a small sample may result in misspecification of the DFE, which, in turn, may cause inaccuracies in the inferred proportions of the DFE. We believe this may explain some of the differences between the findings of Boyko *et al.* (2008) and the findings in this study.

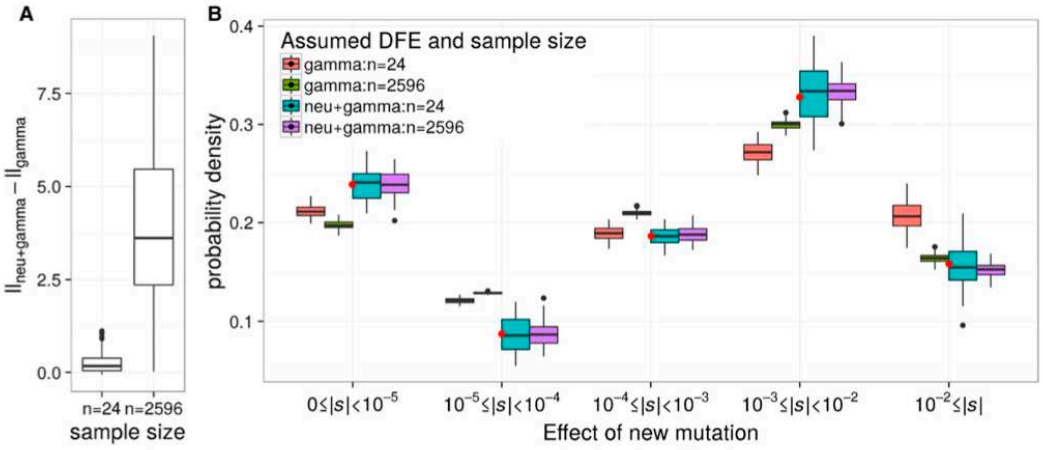

Figure 5 Small sample size and misspecification of the DFE can explain some of the differences between previous estimates and our estimates. Gamma and neutral+gamma DFEs were fit to 100 simulated data sets of sample sizes n = 24 and n = 2596 chromosomes, where the true DFE was neutral+gamma distributed (ρ_{neu} = 0.164, α = 0.338, β = 358.8). (A) The distributions of the difference in log-likelihood between the gamma and neutral+gamma distributions. When the sample size is large (n = 2596) the neutral+gamma distribution has a higher log-likelihood than the gamma distribution. However, the small samples (n = 24) are unable to distinguish between the gamma and neutral+gamma distributions. (B) The estimated proportions of new mutations having different selective effects when fitting the gamma and neutral+gamma distributions. Note that when n = 24, the gamma distribution overpredicts the proportion of strongly deleterious mutations ($|s| \geq 0.01$). Red dots denote the true proportion of mutations in each bin. The boxes cover the first and third quartiles, and the band represents the median. The whiskers cover the highest and lowest datum within 1.5 times the interquartile range from the first and third quartiles. Lastly, any data outside that region are plotted as outlier points.

Assessing the role of the likelihood function using simulations

Next, we investigated the performance of the multinomial *vs.* Poisson likelihoods at inferring the DFE. In this study, as well as in Boyko *et al.* (2008), we fit the DFE using the Poisson likelihood, which uses an *a priori* estimate of the population-scaled mutation rate, θ, to fit the curvature of the SFS as well as the total number of SNPs. Too few segregating variants would suggest the presence of strongly deleterious variants that are not segregating in the sample (Boyko *et al.* 2008).

The multinomial likelihood fits the curvature of the SFS while conditioning on the total number of SNPs. As such, the number of SNPs provides no additional information. The multinomial inference is similar to how the DFE was inferred by Li *et al.* (2010) in that they only used information from the curvature of the SFS. Note, however, Li *et al.* (2010) fit the population frequency spectrum using a least-squares approach while the multinomial likelihood fits the sample frequency spectrum. As such, the multinomial likelihood function does not strictly mirror the procedure of Li *et al.* (2010). Using the multinomial likelihood is convenient because it does not require any prior assumptions about the population scaled mutation rate, θ, yet may be underpowered when trying to identify the proportion of strongly deleterious mutations, unless such variants are segregating in the sample.

To compare the two likelihood methods at varying sample sizes, we fit the full model to simulated data sets of n = 12, 24, 50, 100, 150, 200, and 250 chromosomes using both the multinomial and Poisson likelihoods (Figure S8 in File S1). Again, we simulated 200 data sets at each sample size with the LuCamp demography and a gamma DFE with parameters α = 0.203 and β = 1082.1. In general, the accuracy of our shape parameter estimate improves as the sample size increases, and we find the multinomial and Poisson likelihoods can both be used to reliably estimate the shape parameter. While this trend holds true for the scale parameter using the Poisson likelihood, we find that we are unable to accurately infer the scale parameter using the multinomial likelihood, even for a sample of n = 250 chromosomes. For example, nearly 50% of all the parameter estimates lie close to the maximum bound and 25% lie close to the minimum bound allowed during optimization. We found that this result can be explained by the likelihood surface being exceptionally flat with respect to the scale parameter. In other words, we cannot estimate the strength of purifying selection using only the curvature of the SFS with these sample sizes. Therefore, because Li *et al.* (2010) fit only the curvature of the SFS and excluded rare variants (<2% MAF) in a sample of size of n = 400 chromosomes, the power to detect strongly deleterious variants may be quite low, resulting in different parameter estimates from those in Boyko *et al.* (2008) and our present work.

Discussion

We developed a computational method to infer the DFE of new mutations from large data sets, and then estimated the DFEs

for nonsynonymous mutations using the SFS of 432 Europeans from the 1000 Genomes Project, 1300 Europeans from the ESP, and 1298 Europeans from the LuCamp project. The new DFEs suggest there are fewer strongly deleterious mutations than previously estimated (Figure 4). Although we find a neutral+gamma mixture DFE fits best to the ESP and LuCamp data sets, a gamma DFE seems to be a better fit to the 1000 Genomes data (Table 3). Nevertheless, our best-fit DFEs predict 0.38 to 0.84 as many strongly deleterious new mutations compared to the current, widely used estimates of Boyko *et al.* (2008). Additionally, these findings are robust to assumptions about the mutation rate or the assumed functional form of the DFE. We show small sample size can lead to incorrect estimates of the DFE, specifically when the DFE is approximated with a parametric distribution that is not the true distribution (Figure 5). Therefore, our estimates provide an important update to previous studies of the DFE that used smaller sample sizes. Our current estimates of the DFE, particularly the estimates of the proportion of moderately and strongly deleterious mutations, should be more reliable and precise than previous estimates. To facilitate their utility for future researchers, we provide scripts for implementing these models on GitHub (see *Materials and Methods*).

Our results suggest misspecification of the DFE may explain some of the differences in the DFEs we infer from small and large data sets. This is particularly relevant because the true DFE is almost certainly not a parametric distribution. At small sample sizes, different forms of the DFE tended to similarly fit the data in terms of log-likelihood (Table S5 in File S1). Therefore, the DFE that had fewer parameters (*i.e.*, gamma) was selected as the best-fit DFE. Additionally, we infer more strongly deleterious ($|s| > 10^{-2}$) new mutations from the downsampled data sets. We showed through simulations that even if the true DFE is neutral+gamma distributed, a gamma DFE is selected as the best fit to a small sample. Furthermore, this leads to inaccuracy in recovering the true proportions of the DFE (Figure 5). While the neutral+gamma distribution is also unlikely to be the true DFE, our simulations reproduce the patterns observed when downsampling the real data. Therefore, large sample size not only helps to improve the precision of the estimated DFEs, but also helps to approximate the correct form of the DFE. We expect this question to be better resolved as additional and larger sequencing data sets continue to be generated in the future.

A gamma or similarly shaped distribution of deleterious mutations is well supported by experimental estimates of the DFE in laboratory organisms (Martin and Lenormand 2006; Bataillon and Bailey 2014), although some studies suggest more complex distributions (Halligan and Keightley 2009; Hietpas *et al.* 2011; Jacquier *et al.* 2013). A number of theoretical models also predict the functional form of the DFE (Huber *et al.* 2016). Most progress in this regard comes from phenotype fitness-landscape models such as Fisher's geometric model (FGM) (Martin and Lenormand 2006; Chevin *et al.* 2010; Lourenço *et al.* 2011; Tenaillon 2014; Huber *et al.* 2016) and biophysical models of protein stability (Cherry 1998; Goldstein 2013; Serohijos and Shakhnovich 2014). Under fairly general assumptions, the predicted DFE under these models for a perfectly adapted population is gamma distributed (Martin and Lenormand 2006; Martin 2014; Serohijos and Shakhnovich 2014), and a strongly leptokurtic shape would suggest that most mutations have low pleiotropy (Martin and Lenormand 2006; Lourenço *et al.* 2011). However, our finding of a neutral+gamma distribution suggests that the general FGM is inadequate, since it does not predict the neutral point mass. Alternatively, our support for a neutral point mass might not reflect truly neutral mutations, but rather slightly beneficial mutations that behave effectively neutral under the historically small human population size (Huber *et al.* 2016). Since these mutations would segregate at frequencies similar to neutral mutations, and since we do not explicitly model the effect of beneficial mutations on the SFS, our method would place these mutations at the neutral point mass. Such slightly beneficial mutations are predicted under FGM when deleterious mutations fix and move the population away from the optimal phenotype (Lourenço *et al.* 2011). Slightly deleterious mutations can fix in the population when the effect of drift is large, *i.e.*, the effective population size is small. Thus, our support for the neutral+gamma distribution might be consistent with a large effect of drift in the relatively small human population (Huber *et al.* 2016). Alternatively, a recent change in the selective environment could have moved the human population away from the phenotypic optimum at many genes, leading to a similar increase of the proportion of slightly beneficial mutations (Martin and Lenormand 2006; Chevin *et al.* 2010; Lourenço *et al.* 2011; Bank *et al.* 2014).

Additionally, our results show that estimates of the DFE are sensitive to the mutation rate. For any given data set, assuming a higher nonsynonymous mutation rate will result in the inference of stronger purifying selection due to the increased number of SNPs expected (but not observed) relative to the nonsynonymous mutation rate. There are two assumptions that factor into the calculation of the nonsynonymous mutation rate: First, it is unclear what the true mutation rate is. Whole genome, pedigree-based estimates suggest a mutation rate of about 10^{-8} per base pair per generation, exome-based estimates suggest rates of 1.5×10^{-8}, and phylogenetic estimates suggest a mutation rate in the range of $2.0–2.5 \times 10^{-8}$ (Ségurel *et al.* 2014). Second, we infer a mutation rate from synonymous sites, but use that mutation rate to make an *a priori* assumption about the nonsynonymous mutation rate. Many studies have the nonsynonymous mutation rate at 2.5 times the synonymous mutation rate, but we believe 2.31 to be a more accurate estimate, as this takes into account the CpG mutational bias and a 3:1 transition:transversion ratio (Huber *et al.* 2016). These two factors combined can result in large differences in the DFE. For example, the gamma DFE fit to the LuCamp data predicts 15% of mutations to be strongly deleterious ($|s| > 10^{-2}$) when assuming $\theta_{NS}/\theta_S = 2.31$ and $\mu = 1.5 \times 10^{-8}$, but 25% of new mutations to be strongly

deleterious when assuming $\theta_{NS}/\theta_S = 2.5$ and $\mu = 2.5\times10^{-8}$. Although our results remain qualitatively consistent across the range of mutation rates, uncertainty about the true rate leads to uncertainty in estimating the DFE.

Another important aspect of our results is the consistency of our estimates of the DFE between data sets. Our estimates of the DFE suggest a skew toward neutrality compared to previous studies, and we infer a consistent range of neutral ($|s| < 10^{-5}$, 24–26%), moderately deleterious ($10^{-3} < |s| < 10^{-2}$, 25–33%), and strongly deleterious ($|s| < 10^{-2}$, 14–22%) new mutations between the three data sets. The consistency of our results across data sets suggests our inferences are accurate and robust to sampling from different populations, sequencing, bioinformatic processing, and sample size. This suggests the DFEs we have inferred are reliable updates to the DFEs inferred by Eyre-Walker *et al.* (2006) and Boyko *et al.* (2008).

It is also worth noting that our methodology has key differences from that of Li *et al.* (2010). Li *et al.* (2010) estimated the DFE using the population frequency spectrum excluding rare variants (MAF < 2%), under a constant size demographic model using a least-squares method, and fit the curvature of the SFS while not considering the total number of SNPs in the sample. The extent to which these methodological differences as well as differences in sequencing or bioinformatic processing of the data between their study and our present study contribute to the different estimated DFEs remains unclear. However, we have shown that for small and moderately sized samples, fitting only the curvature of the SFS is insufficient for estimating the scale parameter of the DFE. In other words, for smaller samples, the number of SNPs in the data must be considered to estimate the proportions of moderately and strongly deleterious new mutations, since moderately to strongly deleterious mutations are unlikely to be found in the sample.

Fit∂a∂i infers the DFE for new mutations rather than segregating variants. Interestingly, even inference using the multinomial likelihood function, which only uses the frequencies of segregating variants, still infers the DFE for new mutations. We used simulations to compare the DFE of new mutations to that of segregating variants (Figure S9 and Table S7 in File S1). The DFE of segregating variants depends on the sample size and is shifted to be more neutral than that of new mutations. Approximating the DFE of segregating variants using a gamma distribution reveals a different shape parameter than that of new mutations (Table S7 in File S1); confirming that Fit∂a∂i, when applied to segregating variants, estimates the DFE of new mutations. While long-tailed distributions such as the gamma do not directly capture the mode of very strongly deleterious new mutations observed by many experimental studies (Eyre-Walker and Keightley 2007), the proportion could be extrapolated from the DFEs we inferred. For example, for a gamma-distributed DFE, the proportion of new mutations that is lethal would be computed as the mass of the distribution greater than $|s| > 1$.

Our results argue that, provided the demographic model fit to the data can adequately match the SFS of neutral synonymous sites, inference of the DFE should be robust to cryptic, unmodeled, population structure. In other words, the skew in the frequency spectrum due to demography must be reproduced accurately, but inference of selection is relatively robust to the way the skew is modeled. This result is consistent with the work of Ma *et al.* (2013). Alternately, other studies rescale the entries of the nonsynonymous SFS independently to match the skew of the synonymous SFS from the standard neutral model (Eyre-Walker *et al.* 2006; Galtier 2016; Tataru *et al.* 2016). However, this method is not always accurate for demographies including recent, rapid expansions since the skew on neutral and selected sites may differ (Eyre-Walker *et al.* 2006). Further, fitting many independent scaling parameters to large samples can be problematic (Tataru *et al.* 2016). Thus, Fit∂a∂i offers an advantage over the rescaling methods in these contexts.

Although Fit∂a∂i was developed to work with large sequencing data sets, it still has several limitations. The inference framework we use becomes increasingly slower for larger samples and requires significant computational resources and time to initially generate the SFS for the range of selection coefficients. Additionally, the frequency spectrum becomes difficult to compute for larger selection coefficients ($2Ns > 10{,}000$). This is mainly because finer integration grids must be used to accurately estimate low frequency variants. Also, like the method of Boyko *et al.* (2008), our method assumes additive selective effects and should be interpreted as averaging of selection over all heterozygotes and genetic backgrounds. Nevertheless, we anticipate that our method will be useful for estimating the DFE across the tree of life as polymorphism data sets from different species continue to accumulate.

Our results suggest that there may be more weakly and moderately deleterious nonsynonymous mutations than previously appreciated. This has a number of important implications for medical genetic studies. These variants could possibly contribute to disease risk. However, these mutations could also confound statistical tests that compare observed levels of variation to those predicted by population genetic models. For example, using the DFE of Boyko *et al.* (2008) would predict fewer segregating deleterious variants because more new mutations were estimated to be strongly deleterious and would not segregate in the sample. However, if those mutations were instead only moderately deleterious, some could drift up in frequency and actually segregate in the sample. Further, a common approach to modeling how deleterious variants affect complex traits (Eyre-Walker 2010) assigns mutational effects on a trait as a function of their effects on fitness. This approach has been widely used to quantify the architecture of complex traits (Morris *et al.* 2012; Mancuso *et al.* 2016), to study the effects of demography on traits (Lohmueller 2014a; Simons *et al.* 2014), and to assess the power of rare variant association tests (Uricchio *et al.* 2016). The accuracy and realism

of these models depend on having accurate estimates of the DFE.

Additionally, the DFE determines the extent to which background selection affects patterns of neutral variation. Accurately characterizing the reduction in diversity (*i.e.*, effective population size N_e) should reduce bias when trying to learn the true demography of a population using sites linked to selected variants (Ewing and Jensen 2016; Schrider *et al.* 2016). We used a deterministic approximation (Nicolaisen and Desai 2013) of the models of Zeng and Charlesworth (2011) to contrast the effects of background selection predicted from the DFE of Boyko *et al.* (2008) and the DFEs we inferred in our study (Figure S10 in File S1). We computed the reduction in N_e (*i.e.*, increase in the rate of coalescence) as a function of time due to background selection for the two different mutation rates used in our inferences ($\mu = 1.5 \times 10^{-8}$, $L_{NS}/L_S = 2.31$; $\mu = 1.8 \times 10^{-8}$ and $L_{NS}/L_S = 2.5$) as well as the higher deleterious mutation rate of McVicker *et al.* (2009): $\mu = 7.4 \times 10^{-8}$, also assuming $L_{NS}/L_S = 2.5$. Importantly, all the DFEs predict that background selection will reduce diversity and skew the SFS toward an excess of rare variants compared to models of constant population size (Figure S10 in File S1). However, DFEs with fewer strongly deleterious mutations, like the best fit DFEs to the ESP EUR and LuCamp data sets, predict less of an overall reduction in neutral diversity compared to Boyko *et al.* (2008). Further, the change in coalescent rates over time varies across DFEs, suggesting that the degree to which background selection affects the curvature of the SFS does depend on the specific DFE.

More broadly, our results have important implications for understanding and quantifying deleterious variants across human populations (Lohmueller *et al.* 2008; Lohmueller 2014b; Simons *et al.* 2014; Do *et al.* 2015). Specifically, the fate of strongly deleterious mutations is relatively insensitive to population demography. The fate of weakly and moderately deleterious mutations, however, is linked more closely with effective population size (Henn *et al.* 2016). Human evolution in particular is influenced by nearly neutral processes due to relatively small effective population sizes. Then, a DFE containing fewer strongly deleterious new mutations suggests the nature of purifying selection in humans may be different from what is currently understood. For example, larger proportions of moderately and weakly deleterious mutations may suggest greater differences in the proportion of segregating deleterious mutations and genetic load between human populations (Henn *et al.* 2016). Accurate inferences of the DFE are critical in this regard as researchers begin to use explicit models of demography and selection to quantify differences in the amounts of deleterious variants across populations (Brandvain and Wright 2016; Gravel 2016).

Acknowledgments

We thank Emilia Huerta-Sanchez, Diego Ortega-Del Vecchyo, and Noah Rosenberg, as well as two reviewers, for constructive comments on the manuscript. This work was supported by a Searle Scholars Fellowship, an Alfred P. Sloan Research Fellowship in Computational and Molecular Biology, and National Institutes of Health grant R35 GM-119856 to K.E.L.

Literature Cited

Aberer, A. J., and A. Stamatakis, 2013 Rapid forward-in-time simulation at the chromosome and genome level. BMC Bioinformatics 14: 216.

Acevedo, A., L. Brodsky, and R. Andino, 2014 Mutational and fitness landscapes of an RNA virus revealed through population sequencing. Nature 505: 686–690.

Bank, C., R. T. Hietpas, A. Wong, D. N. Bolon, and J. D. Jensen, 2014 A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. Genetics 196: 841–852.

Bataillon, T., and S. F. Bailey, 2014 Effects of new mutations on fitness: insights from models and data. Ann. N. Y. Acad. Sci. 1320: 76–92.

Boucher, J. I., P. Cote, J. Flynn, L. Jiang, A. Laban *et al.*, 2014 Viewing protein fitness landscapes through a next-gen lens. Genetics 198: 461–471.

Boyko, A. R., S. H. Williamson, A. R. Indap, J. D. Degenhardt, R. D. Hernandez *et al.*, 2008 Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4: e1000083.

Brandvain, Y., and S. I. Wright, 2016 The limits of natural selection in a nonequilibrium world. Trends Genet. 32: 201–210.

Castellano, D., M. Coronado-Zamora, J. L. Campos, A. Barbadilla, and A. Eyre-Walker, 2016 Adaptive evolution is substantially impeded by Hill–Robertson interference in *Drosophila*. Mol. Biol. Evol. 33: 442–455.

Cherry, J. L., 1998 Should we expect substitution rate to depend on population size? Genetics 150: 911–919.

Chevin, L.-M., R. Lande, and G. M. Mace, 2010 Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. PLoS Biol. 8: e1000357.

Chikhi, L., R. A. Nichols, G. Barbujani, and M. A. Beaumont, 2002 Y genetic data support the Neolithic demic diffusion model. Proc. Natl. Acad. Sci. USA 99: 11008–11013.

Comeron, J. M., 2014 Background selection as baseline for nucleotide variation across the *Drosophila* genome. PLoS Genet. 10: e1004434.

Do, R., D. Balick, H. Li, I. Adzhubei, S. Sunyaev *et al.*, 2015 No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nat. Genet. 47: 126–131.

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

Eyre-Walker, A., 2010 Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. Proc. Natl. Acad. Sci. USA 107: 1752–1756.

Eyre-Walker, A., and P. D. Keightley, 2007 The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8: 610–618.

Eyre-Walker, A., M. Woolfit, and T. Phelps, 2006 The distribution of fitness effects of new deleterious amino acid mutations in humans. Genetics 173: 891–900.

Fowler, D. M., C. L. Araya, S. J. Fleishman, E. H. Kellogg, J. J. Stephany *et al.*, 2010 High-resolution mapping of protein sequence-function relationships. Nat. Methods 7: 741–746.

Fu, W., T. D. O'Connor, G. Jun, H. M. Kang, G. Abecasis *et al.*, 2013 Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. Nature 493: 216–220.

Galtier, N., 2016 Adaptive protein evolution in animals and the effective population size hypothesis. PLoS Genet. 12: e1005774.

Gazave, E., L. Ma, D. Chang, A. Coventry, F. Gao *et al.*, 2014 Neutral genomic regions refine models of recent rapid human population growth. Proc. Natl. Acad. Sci. USA 111: 757–762.

Goldstein, R. A., 2013 Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability. Genome Biol. Evol. 5: 1584–1593.

Gossmann, T. I., P. D. Keightley, and A. Eyre-Walker, 2012 The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. Genome Biol. Evol. 4: 658–667.

Gravel, S., 2016 When is selection effective? Genetics 203: 451–462.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante, 2009 Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5: e1000695.

Halligan, D. L., and P. D. Keightley, 2009 Spontaneous mutation accumulation studies in evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. 40: 151–172.

Halligan, D. L., A. Kousathanas, R. W. Ness, B. Harr, L. Eöry *et al.*, 2013 Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. PLoS Genet. 9: e1003995.

Harris, K., and R. Nielsen, 2016 The genetic cost of Neanderthal introgression. Genetics 203: 881–891.

Hartl, D. L., E. N. Moriyama, and S. A. Sawyer, 1994 Selection intensity for codon bias. Genetics 138: 227–234.

Henn, B. M., L. R. Botigué, S. Peischl, I. Dupanloup, M. Lipatov *et al.*, 2016 Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc. Natl. Acad. Sci. USA 113: E440–E449.

Hernandez, R. D., 2008 A flexible forward simulator for populations subject to selection and demography. Bioinformatics 24: 2786–2787.

Hernandez, R. D., S. H. Williamson, and C. D. Bustamante, 2007 Context dependence, ancestral misidentification, and spurious signatures of natural selection. Mol. Biol. Evol. 24: 1792–1800.

Hietpas, R. T., J. D. Jensen, and D. N. A. Bolon, 2011 Experimental illumination of a fitness landscape. Proc. Natl. Acad. Sci. USA 108: 7896–7901.

Huber, C. D., B. Y. Kim, C. D. Marsden, and K. E. Lohmueller, 2016 Determining the factors driving selective effects of new nonsynonymous mutations. bioRxiv Available at: http://www.biorxiv.org/content/early/2016/08/23/071209.

Jacquier, H., A. Birgy, H. L. Nagard, Y. Mechulam, E. Schmitt *et al.*, 2013 Capturing the mutational landscape of the beta-lactamase TEM-1. Proc. Natl. Acad. Sci. USA 110: 13067–13072.

Keightley, P. D., and A. Eyre-Walker, 2007 Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics 177: 2251–2261.

Keightley, P. D., and S. P. Otto, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. Nature 443: 89–92.

Keinan, A., and A. G. Clark, 2012 Recent explosive human population growth has resulted in an excess of rare genetic variants. Science 336: 740–743.

Koufopanou, V., S. Lomas, I. J. Tsai, and A. Burt, 2015 Estimating the fitness effects of new mutations in the wild yeast *Saccharomyces paradoxus*. Genome Biol. Evol. 7: 1887–1895.

Kousathanas, A., and P. D. Keightley, 2013 A comparison of models to infer the distribution of fitness effects of new mutations. Genetics 193: 1197–1208.

Kraft, D., 1988 *A Software Package for Sequential Quadratic Programming*. Wiss. Berichtswesen d, DFVLR, Cologne.

Li, Y., N. Vinckenbosch, G. Tian, E. Huerta-Sanchez, T. Jiang *et al.*, 2010 Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. Nat. Genet. 42: 969–972.

Lohmueller, K. E., 2014a The impact of population demography and selection on the genetic architecture of complex traits. PLoS Genet. 10: e1004379.

Lohmueller, K. E., 2014b The distribution of deleterious genetic variation in human populations. Curr. Opin. Genet. Dev. 29: 139–146.

Lohmueller, K. E., A. R. Indap, S. Schmidt, A. R. Boyko, R. D. Hernandez *et al.*, 2008 Proportionally more deleterious genetic variation in European than in African populations. Nature 451: 994–997.

Lohmueller, K. E., T. Sparsø, Q. Li, E. Andersson, T. Korneliussen *et al.*, 2013 Whole-exome sequencing of 2,000 Danish individuals and the role of rare coding variants in type 2 diabetes. Am. J. Hum. Genet. 93: 1072–1086 (erratum: Am. J. Hum. Genet. 94: 479).

Lourenço, J., N. Galtier, and S. Glémin, 2011 Complexity, pleiotropy, and the fitness effect of mutations. Evolution 65: 1559–1571.

Ma, X., J. L. Kelley, K. Eilertson, S. Musharoff, J. D. Degenhardt *et al.*, 2013 Population genomic analysis reveals a rich speciation and demographic history of orang-utans (*Pongo pygmaeus* and *Pongo abelii*). PLoS One 8: e77175.

Mancuso, N., N. Rohland, K. A. Rand, A. Tandon, A. Allen *et al.*, 2016 The contribution of rare variation to prostate cancer heritability. Nat. Genet. 48: 30–35.

Martin, G., 2014 Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. Genetics 197: 237–255.

Martin, G., and T. Lenormand, 2006 A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution 60: 893–907.

McManus, K. F., J. L. Kelley, S. Song, K. R. Veeramah, A. E. Woerner *et al.*, 2015 Inference of gorilla demographic and selective history from whole-genome sequence data. Mol. Biol. Evol. 32: 600–612.

McVicker, G., D. Gordon, C. Davis, and P. Green, 2009 Widespread genomic signatures of natural selection in hominid evolution. PLoS Genet. 5: e1000471.

Messer, P. W., 2013 SLiM: simulating evolution with selection and linkage. Genetics 194: 1037–1039.

Messer, P. W., and D. A. Petrov, 2013 Frequent adaptation and the McDonald–Kreitman test. Proc. Natl. Acad. Sci. USA 110: 8615–8620.

Moon, S., and J. M. Akey, 2016 A flexible method for estimating the fraction of fitness influencing mutations from large sequencing data sets. Genome Res. 26: 834–843.

Morris, A. P., B. F. Voight, T. M. Teslovich, T. Ferreira, A. V. Segrè *et al.*, 2012 Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat. Genet. 44: 981–990.

Nelson, M. R., D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean *et al.*, 2012 An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 337: 100–104.

Nicolaisen, L. E., and M. M. Desai, 2013 Distortions in genealogies due to purifying selection and recombination. Genetics 195: 221–230.

Novembre, J., and S. Ramachandran, 2011 Perspectives on human population structure at the cusp of the sequencing era. Annu. Rev. Genomics Hum. Genet. 12: 245–274.

Ortega-Del Vecchyo, D., C. D. Marsden, and K. E. Lohmueller, 2016 PReFerSim: fast simulation of demography and selection under the Poisson random field model. Bioinformatics 32: 3516–3518.
Ptak, S. E., and M. Przeworski, 2002 Evidence for population growth in humans is confounded by fine-scale population structure. Trends Genet. 18: 559–563.
Ragsdale, A. P., A. J. Coffman, P. Hsieh, T. J. Struck, and R. N. Gutenkunst, 2016 Triallelic population genomics for inferring correlated fitness effects of same site nonsynonymous mutations. Genetics 203: 513–523.
Sawyer, S. A., and D. L. Hartl, 1992 Population genetics of polymorphism and divergence. Genetics 132: 1161–1176.
Schrider, D. R., A. G. Shanku, and A. D. Kern, 2016 Effects of linked selective sweeps on demographic inference and model selection. Genetics 204: 1207–1223.
Ségurel, L., M. J. Wyman, and M. Przeworski, 2014 Determinants of mutation rate variation in the human germline. Annu. Rev. Genomics Hum. Genet. 15: 47–70.
Serohijos, A. W. R., and E. I. Shakhnovich, 2014 Contribution of selection for protein folding stability in shaping the patterns of polymorphisms in coding regions. Mol. Biol. Evol. 31: 165–176.
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.
Tataru, P., M. Mollion, S. Glemin, and T. Bataillon, 2016 Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. bioRxiv Available at: http://biorxiv.org/content/early/2016/07/05/062216.
Tenaillon, O., 2014 The utility of Fisher's geometric model in evolutionary genetics. Annu. Rev. Ecol. Evol. Syst. 45: 179–201.
Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny *et al.*, 2012 Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337: 64–69.
The 1000 Genomes Project Consortium, 2015 A global reference for human genetic variation. Nature 526: 68–74.
Torgerson, D. G., A. R. Boyko, R. D. Hernandez, A. Indap, X. Hu *et al.*, 2009 Evolutionary processes acting on candidate cis-regulatory regions in humans inferred from patterns of polymorphism and divergence. PLoS Genet. 5: e1000592.
Uricchio, L. H., N. A. Zaitlen, C. J. Ye, J. S. Witte, and R. D. Hernandez, 2016 Selection and explosive growth alter genetic architecture and hamper the detection of causal rare variants. Genome Res. 26: 863–873.
Williamson, S. H., R. Hernandez, A. Fledel-Alon, L. Zhu, R. Nielsen *et al.*, 2005 Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc. Natl. Acad. Sci. USA 102: 7882–7887.
Wilson Sayres, M. A., K. E. Lohmueller, and R. Nielsen, 2014 Natural selection reduced diversity on human Y chromosomes. PLoS Genet. 10: e1004064.
Zeng, K., and B. Charlesworth, 2011 The joint effects of background selection and genetic recombination on local gene genealogies. Genetics 189: 251–266.

Communicating editor: N. A. Rosenberg

Supplemental Results

Table S1. Demographic model parameter estimates from synonymous sites.

	n_{chr}	θ_S	N_1	T_1[a]	N_2	T_2	N_C	T_C
LuCamp	2596	4261.2	0.08984	0.01	1.0512	0.07304	31.270	0.01158
1kG	864	5984.9	0.08469	0.01	1.1007	0.07043	53.283	0.02009
ESP	2600	6415.1	0.11949	0.01	1.3111	0.05254	98.65	0.01502
LuCamp	24	4438.9	0.07554	0.01	0.8703	0.10652	53.09[b]	0.01311
1kG	24	6122.4	0.07447	0.01	0.9606	0.08805	93.23[b]	0.01991
ESP	24	6735.6	0.08887	0.01	0.9787	0.08434	63.79[b]	0.01758

Note: See Figure S3 for a pictorial representation of the model.
Parameter descriptions:
N_i: Population size relative to the ancestral population
T_i: Time, in units of $2N_{ANC}$ generations (N_{ANC} is the ancestral population size)
θ_S: The population mutation rate of synonymous sites.
[a]T_1 is fixed in this model.
[b]These estimates are unreliable and can range from 50-200 with no appreciable change in the fit of the model.

Table S2. Parameters used to scale the DFE in terms of *s*.

	n_{chr}	θ_S	L_{NS}/L_S	θ_{NS}	μ	L_S+L_{NS}	L_S	L_{NS}	N_{ANC}
LuCamp	2596	4261.2	2.5	10652.9	1.8*e*-8	20,043,582	5,726,738	14,316,844	10334
1kG	864	5984.9		14962.3	1.8*e*-8	26,673,114	7,620,890	19,052,224	10907
ESP	2600	6415.1		16037.7	1.8*e*-8	31,427,992	8,979,426	22,448,566	9922
LuCamp	2596	4261.2	2.31	9843.3	1.5*e*-8	20,043,582	6,055,463	13,988,119	11728
1kG	864	5984.9		13825.1	1.5*e*-8	26,673,114	8,058,343	18,614,771	12378
ESP	2600	6415.1		14818.8	1.5*e*-8	31,427,992	9,494,862	21,933,130	11261
LuCamp	24	4438.9	2.5	11097.3	1.8*e*-8	20,043,582	5,726,738	14,316,844	10766
1kG	24	6122.4		15306.0	1.8*e*-8	26,673,114	7,620,890	19,052,224	11158
ESP	24	6735.6		16839.0	1.8*e*-8	31,427,992	8,979,426	22,448,566	10418
LuCamp	24	4438.9	2.31	10253.9	1.5*e*-8	20,043,582	6,055,463	13,988,119	12217
1kG	24	6122.4		14142.7	1.5*e*-8	26,673,114	8,058,343	18,614,771	12663
ESP	24	6735.6		15559.2	1.5*e*-8	31,427,992	9,494,862	21,933,130	11823

Parameter descriptions:
θ_S: The population scaled synonymous mutation rate
θ_{NS}: The population scaled nonsynonymous mutation rate
L_S: The number of synonymous sites
L_{NS}: The number of nonsynonymous sites
N_{ANC}: The ancestral population sizes computed from $\theta_S=4N_{ANC}\mu L_S$

Table S3. Inference using Fit∂a∂i on the African-American dataset from Boyko et al. (2008).

	∂a∂i[a]	Boyko et al.
N_{curr}/N_{anc}	3.352	3.296
T_{exp}	7067	6809
	fit∂a∂i	**Boyko et al.**
α **(shape)**	0.179	0.184
β **(scale)**	3161	2488
0 ≤ |s| < 1e-4	27.7%	27.9%
1e-4 ≤ |s| < 1e-4	14.1%	14.7%
1e-2 ≤ |s| < 1e-3	20.9%	21.9%
1e-2 ≤ |s|	37.3%	35.5%

Notes: Population sizes are reported relative to the ancestral population size, and times are reported in units of generations. The scale parameter of the gamma distribution is scaled in terms of the ancestral population size. The proportion of new mutations in each range of selective effects was computed from the gamma distribution.

[a]Demographic inference was done using the standard framework of ∂a∂i.

Table S4. Maximum likelihood estimates for the various DFEs under different mutation rate assumptions.

Dataset	n_{chr}	Distribution	Parameters[a]			0 ≤ |s| < 1e-5	1e-5 ≤ s < 1e-4	1e-4 ≤ |s| < 1e-3	1e-3 ≤ |s| < 1e-2	1e-2 ≤ |s|	Log-likelihood	|ΔAIC|[b]
μ=1.5e-8, L_{NS}/L_S=2.31												
1kG EUR	864	gamma	--	0.186	875	0.237	0.127	0.192	0.266	0.178	-1450.5	0
		neu+gamma	0.031	0.199	820.6	0.242	0.122	0.191	0.270	0.176	-1450.8	2.6
		neu+exp+let	0.312	0.509	41.48	0.216	0.026	0.199	0.279	0.279	-1472.0	45
		discrete	--	--	--	**0.286**	**0.099**	**0.222**	**0.305**	0.089	-1453.4	9.8
ESP EUR	2600	gamma	--	0.169	1327.4	0.248	0.118	0.173	0.240	0.220	-3012.6	2.6
		neu+gamma	0.092	0.207	1082.3	0.263	0.104	0.167	0.249	0.217	-3010.3	0
		neu+exp+let	0.341	0.504	63.90	0.343	0.016	0.132	0.340	0.169	-3071.6	122.6
		discrete	--	--	--	**0.334**	**0.041**	**0.201**	**0.306**	0.118	-3029.5	40.4
LuCamp	2596	gamma	--	0.215	562.1	0.205	0.131	0.212	0.298	0.154	-3334.7	13.8
		neu+gamma	0.164	0.338	367.7	0.241	0.091	0.194	0.332	0.141	-3326.8	0
		neu+exp+let	0.304	0.613	66.56	0.306	0.019	0.161	0.413	0.101	-3337.8	22
		discrete	--	--	--	**0.309**	**0.024**	**0.247**	**0.372**	0.049	-3334.2	16.8
μ=1.8e-8, L_{NS}/L_S=2.5												
1kG EUR	864	gamma	--	0.181	1585.1	0.217	0.112	0.169	0.243	0.259	-1452.1	0
		neu+gamma	0.014	0.184	1554.8	0.220	0.111	0.168	0.244	0.258	-1452.5	2.8
		neu+exp+let	0.288	0.470	41.53	0.290	0.022	0.168	0.275	0.244	-1472.0	41.8
		discrete	--	--	--	**0.259**	**0.088**	**0.201**	**0.279**	0.174	-1454.2	8.2

ESP EUR	2600	gamma	--	**0.164**	**2527.8**	0.229	0.105	0.152	0.216	0.298	-3008.6	0
		neu+gamma	**0.032**	**0.176**	**2326.6**	0.234	0.101	0.150	0.218	0.297	-3008.6	2
		neu+exp+let	**0.315**	**0.466**	**63.88**	0.316	0.013	0.110	0.321	0.240	-3071.6	128
		discrete	--	--	--	**0.307**	**0.032**	**0.178**	**0.284**	0.199	-3032.5	51.8
LuCamp	2596	gamma	--	**0.203**	**1082.1**	0.192	0.114	0.181	0.268	0.245	-3335.9	2.6
		neu+gamma	**0.027**	**0.216**	**1017.5**	0.197	0.109	0.178	0.271	0.244	-3335.9	4.6
		neu+exp+let	**0.281**	**0.566**	**66.51**	0.283	0.016	0.134	0.390	0.178	-3337.8	8.44
		discrete	--	--	--	**0.278**	**0.027**	**0.211**	**0.352**	0.132	-3332.6	0

[a]The bolded parameters were directly inferred and are denoted in the following order. Gamma: α (shape), β (scale). Neutral+gamma: p_{neu}, α (shape), β (scale). The scale parameters are in terms of $2N_{ANC}s$. Neutral+exp+lethal: p_{neu}, p_{exp}, λ (rate). For the discrete DFE, the proportion of each bin is directly inferred. Note, there are only four free bins because the mass in all of the bins together is constrained to sum to 1.

[b]Change in AIC relative to the model with the lowest AIC.

Table S5. A comparison of our three best-fitting DFEs to previously inferred DFEs.

n_{chr}	Dataset	DFE	p_{neu}	α (shape)	β (scale)	0 ≤ |s| < 1e-5	1e-5 ≤ |s| < 1e-4	1e-4 ≤ |s| < 1e-3	1e-3 ≤ |s| < 1e-2	1e-2 ≤ |s|	Log-likelihood
24	Boyko et al. (AFR)	gamma	--	0.184 (0.158 - 0.206)	2488 (1062 - 6159)	0.183 (0.123 - 0.266)	0.096 (0.075 -0.12)	0.147 (0.120 - 0.167)	0.219 (0.191 - 0.228)	0.355 (0.222-0.491)	--
		neu+gamma	0.148 (0.000 – 0.235)	0.344 (0.178 – 0.790)	1900 (280-12300)	0.189 (0.235 - 0.300)	0.050 (0.001 - 0.151)	0.110 (0.004 - 0.222)	0.232 (0.027 - 0.257)	0.419 (0.070-0.733)	--
	Boyko et al. (EUR)	gamma	--	0.206	2263	0.151	0.092	0.147	0.231	0.379	--
		neu+gamma	0.239	1.09	800	0.240	0.006	0.068	0.455	0.232	--
180*	Eyre-Walker et al.[a]	gamma	--	0.23 (0.19-0.27)	1845 (833-4032)	0.134 (0.076 - 0.223)	0.094 (0.066 - 0.122)	0.159 (0.122 - 0.188)	0.259 (0.223 - 0.266)	0.354 (0.202-0.513)	--
400	Li et al.[b]	neu+gamma	0.2	4	11.1*	0.200	0.012	0.775	0.013	0.000	--
				μ=1.8e-8, L_{NS}/L_S=2.5							
864	1000 Genomes (EUR)	gamma	--	0.181 (0.175 - 0.185)	1585.1 (1451.9 - 1796.6)	0.217 (0.212 - 0.223)	0.112 (0.111 - 0.113)	0.169 (0.165 - 0.172)	0.243 (0.235 - 0.249)	0.259 (0.252-0.266)	-1452.1
		neu+gamma	0.010 (0.001 - 0.036)	0.184 (0.178 - 0.200)	1554.8 (1340.5 - 1724.4)	0.220 (0.212 - 0.228)	0.111 (0.105 - 0.112)	0.168 (0.164 - 0.172)	0.244 (0.237 - 0.252)	0.258 (0.251-0.265)	-1452.5
		discrete	--	--	--	0.259 (0.203 - 0.311)	0.088 (0.005 - 0.165)	0.201 (0.163 - 0.243)	0.279 (0.251 - 0.305)	0.174 (0.160-0.188)	-1454.2

2600	ESP (EUR)	gamma	--	0.164 (0.160 - 0.168)	2527.8 (2284.9 - 2742.1)	0.229 (0.223 - 0.234)	0.105 (0.104 - 0.106)	0.152 (0.150 - 0.155)	0.216 (0.212 - 0.221)	0.298 (0.294-0.302)	-3008.6
		neu+gamma	0.032 (0.012 - 0.089)	0.176 (0.168 - 0.203)	2326.6 (1936.4 - 2591.9)	0.234 (0.228 - 0.244)	0.101 (0.091 - 0.104)	0.150 (0.145 - 0.153)	0.218 (0.212 - 0.225)	0.297 (0.294-0.301)	-3008.6
		discrete	--	--	--	0.307 (0.252 - 0.331)	0.032 (0.001 - 0.114)	0.178 (0.141 - 0.200)	0.284 (0.269 - 0.300)	0.199 (0.193-0.207)	-3032.5
2596	LuCamp	gamma	--	0.203 (0.197 - 0.207)	1082.1 (999.3-1184.4)	0.192 (0.188 - 0.198)	0.114 (0.113 - 0.115)	0.181 (0.177 - 0.184)	0.268 (0.261 - 0.273)	0.245 (0.238-0.251)	-3335.9
		neu+gamma	0.027 (0.005 - 0.104)	0.216 (0.201 - 0.262)	1017.5 (813.8-1150.4)	0.197 (0.190 - 0.215)	0.109 (0.093 - 0.113)	0.178 (0.169 - 0.182)	0.271 (0.263 - 0.283)	0.244 (0.237-0.250)	-3335.9
		discrete	--	--	--	0.278 (0.221 - 0.303)	0.027 (0.001 - 0.110)	0.211 (0.167 - 0.234)	0.352 (0.330 - 0.373)	0.132 (0.124-0.142)	-3332.6
24	000 Genomes (EUR)	gamma	--	0.183 (0.170 - 0.197)	1534.8 (1018.8 - 2311.4)	0.216 (0.206 - 0.225)	0.113 (0.107 - 0.119)	0.171 (0.158 - 0.185)	0.246 (0.225 - 0.268)	0.255 (0.220-0.285)	-53.8
		neu+gamma	0.051 (0.017 - 0.179)	0.209 (0.186 - 0.339)	1256.8 (451.4-1834.4)	0.221 (0.212 - 0.244)	0.105 (0.074 - 0.113)	0.169 (0.155 - 0.180)	0.256 (0.233 - 0.323)	0.248 (0.189-0.279)	-54.1
		discrete	--	--	--	0.155 (0.092 - 0.279)	0.262 (0.056 - 0.366)	0.053 (0.001 - 0.249)	0.495 (0.002 - 0.556)	0.034 (0.001-0.384)	-53.2

	ESP (EUR)	gamma	--	0.159 (0.147 - 0.172)	3349.2 (2020.4 - 5585.4)	0.230 (0.221 - 0.240)	0.102 (0.096 - 0.108)	0.147 (0.135 - 0.160)	0.207 (0.187 - 0.229)	0.315 (0.280-0.345)	-54.1
		neu+gamma	0.084 (0.027 - 0.221)	0.201 (0.161 - 0.388)	2112.3 (480.8-4144.3)	0.240 (0.232 - 0.260)	0.092 (0.058 - 0.102)	0.146 (0.132 - 0.154)	0.223 (0.192 - 0.297)	0.300 (0.224-0.327)	-54.2
		discrete	--	--	--	0.216 (0.130 - 0.323)	0.172 (0.001 - 0.304)	0.086 (0.001 - 0.251)	0.373 (0.001 - 0.577)	0.153 (0.001-0.428)	-54.0
	LuCamp	gamma	--	0.189 (0.172 - 0.205)	2050.3 (1256.3 - 3583.5)	0.193 (0.183 - 0.204)	0.105 (0.099 - 0.111)	0.161 (0.146 - 0.177)	0.239 (0.210 - 0.266)	0.302 (0.261-0.339)	-51.1
		neu+gamma	0.113 (0.024 - 0.205)	0.278 (0.195 - 0.509)	951.5 (267.7-2078.6)	0.208 (0.195 - 0.233)	0.085 (0.044 - 0.105)	0.161 (0.145 - 0.172)	0.280 (0.233 - 0.383)	0.266 (0.180-0.310)	-50.9
		discrete	--	--	--	0.262 (0.119 - 0.301)	0.049 (0.001 - 0.268)	0.209 (0.003 - 0.290)	0.257 (0.001 - 0.600)	0.222 (0.001-0.425)	-50.9
μ=1.5e-8, L_{NS}/L_S =2.31											
864	1000 Genomes (EUR)	gamma	--	0.186 (0.181 - 0.191)	875.0 (787.9-978.0)	0.237 (0.231 - 0.243)	0.127 (0.125 - 0.128)	0.192 (0.188 - 0.197)	0.266 (0.259 - 0.272)	0.178 (0.171-0.186)	-1450.5
		neu+gamma	0.031 (0.004 - 0.104)	0.199 (0.186 - 0.238)	820.6 (670.2-908.3)	0.242 (0.233 - 0.256)	0.122 (0.109 - 0.127)	0.191 (0.185 - 0.195)	0.270 (0.263 - 0.282)	0.176 (0.167-0.183)	-1450.8

		discrete	--	--	--	0.286 (0.240 - 0.333)	0.099 (0.012 - 0.169)	0.222 (0.182 - 0.264)	0.305 (0.271 - 0.334)	0.089 (0.071-0.107)	-1453.4
2600	ESP (EUR)	gamma	--	0.169 (0.165 - 0.174)	1327.4 (1211.9 - 1439.2)	0.248 (0.242 - 0.254)	0.118 (0.117 - 0.119)	0.173 (0.170 - 0.177)	0.240 (0.235 - 0.246)	0.220 (0.214-0.224)	-3012.6
		neu+gamma	0.092 (0.026 - 0.158)	0.207 (0.179 - 0.251)	1082.3 (880.3-1261.5)	0.263 (0.250 - 0.277)	0.104 (0.091 - 0.114)	0.167 (0.160 - 0.173)	0.249 (0.241 - 0.259)	0.217 (0.211-0.221)	-3010.3
		discrete	--	--	--	0.334 (0.284 - 0.363)	0.041 (0.001 - 0.115)	0.201 (0.165 - 0.228)	0.306 (0.285 - 0.324)	0.118 (0.110-0.126)	-3029.5
2596	LuCamp	gamma	--	0.215 (0.207 - 0.220)	562.1 (525.9-620.8)	0.205 (0.199 - 0.213)	0.131 (0.130 - 0.132)	0.212 (0.207 - 0.215)	0.298 (0.290 - 0.303)	0.154 (0.147-0.161)	-3334.7
		neu+gamma	0.164 (0.110 - 0.215)	0.338 (0.281 - 0.415)	367.7 (295.4-455.7)	0.242 (0.223 - 0.260)	0.091 (0.072 - 0.107)	0.194 (0.183 - 0.203)	0.332 (0.313 - 0.352)	0.141 (0.129-0.152)	-3326.8
		discrete	--	--	--	0.309 (0.260 - 0.332)	0.024 (0.001 - 0.099)	0.247 (0.201 - 0.269)	0.372 (0.347 - 0.395)	0.049 (0.039-0.060)	-3334.2
24	1000 Genomes (EUR)	gamma	--	0.183 (0.170 - 0.197)	1534.8 (1018.8 - 2311.4)	0.221 (0.212 - 0.230)	0.115 (0.109 - 0.122)	0.175 (0.161 - 0.189)	0.250 (0.228 - 0.272)	0.239 (0.204-0.271)	-53.8
		neu+gamma	0.051 (0.017 - 0.179)	0.209 (0.186 - 0.339)	1256.8 (451.4-1834.4)	0.226 (0.217 - 0.247)	0.108 (0.077 - 0.116)	0.173 (0.160 - 0.187)	0.260 (0.237 - 0.329)	0.232 (0.170-0.264)	-54.1

	discrete	--	--	--	0.155 (0.092 - 0.279)	0.262 (0.056 - 0.366)	0.053 (0.001 - 0.249)	0.495 (0.002 - 0.556)	0.034 (0.001-0.384)	-53.2
ESP (EUR)	gamma	--	0.159 (0.149 - 0.171)	2001.6 (1390.4 - 2898.3)	0.254 (0.244 - 0.263)	0.113 (0.106 - 0.118)	0.162 (0.150 - 0.173)	0.224 (0.205 - 0.240)	0.247 (0.222-0.274)	-54.1
	neu+gamma	0.127 (0.031 - 0.249)	0.231 (0.167 - 0.424)	1084.0 (338.2-2202.3)	0.264 (0.250 - 0.287)	0.096 (0.062 - 0.112)	0.162 (0.147 - 0.176)	0.254 (0.216 - 0.332)	0.224 (0.148-0.269)	-54.3
	discrete	--	--	--	0.256 (0.170 - 0.355)	0.169 (0.001 - 0.304)	0.101 (0.001 - 0.294)	0.425 (0.001 - 0.515)	0.050 (0.001-0.355)	-54.0
LuCamp	gamma	--	0.189 (0.175 - 0.206)	1307.8 (806.7-2071.6)	0.214 (0.201 - 0.226)	0.117 (0.110 - 0.124)	0.179 (0.162 - 0.198)	0.258 (0.233 - 0.283)	0.232 (0.190-0.269)	-51.1
	neu+gamma	0.113 (0.024 - 0.205)	0.278 (0.195 - 0.509)	951.5 (267.7-2078.6)	0.211 (0.199 - 0.235)	0.089 (0.048 - 0.108)	0.166 (0.150 - 0.177)	0.286 (0.237 - 0.391)	0.248 (0.156-0.295)	-50.9
	discrete	--	--	--	0.262 (0.119 - 0.301)	0.049 (0.001 - 0.268)	0.209 (0.003 - 0.290)	0.257 (0.001 - 0.600)	0.222 (0.001-0.425)	-50.9

Note: We present our results assuming that that L_{NS}/L_S=2.31 and μ=1.5e-8, or, alternatively, L_{NS}/L_S=2.5 and μ=1.8e-8. The latter provides a direct comparison to Boyko et al. (2008).

[a]The DFE of Eyre-Walker et al. (2006) is rescaled assuming an ancestral population size of N=10,000 diploids. Additionally, this dataset is comprised of 320 autosomal genes, in contrast to the exome sequencing datasets presented here.

[b]The DFE of Li et al. (2010) is scaled by a population size of N=52,097 diploids.

Table S6. Masking singletons has little effect on inference of the DFE.

		With singletons		Singletons masked	
Dataset	**Distribution**	**Parameters**	**Log-likelihood**	**Parameters**	**Log-likelihood**
1kG EUR	gamma	α=0.186, β=875	-1450.5	α=0.186, β=866	-1444.1
	neu+gamma	p_{neu}=0.031, α=0.199, β=820.6	-1450.8	p_{neu}=0.088, α=0.232, β=662	-1444.9
ESP EUR	gamma	α=0.169, β=1327.4	-3012.6	α=0.162, β=1630.9	-2999.5
	neu+gamma	p_{neu}=0.092, α=0.207, β=1082.3	-3010.3	p_{neu}=0.0232, α=0.170, β=1530.7	-2999.8
LuCamp	gamma	α=0.215, β=562.1	-3334.7	α=0.213, β=587.1	-3327.2
	neu+gamma	p_{neu}=0.164, α=0.338, β=367.7	-3326.8	p_{neu}=0.1696, α=0.352, β=338.3	-3318.9

Notes: We assumed L_{NS}/L_S=2.31 and μ=1.5e-8 for these calculations. α denotes the shape parameter and β denotes the scale parameter. The scale parameter is scaled by the ancestral population size.

Table S7. The DFE of new mutations and the DFE of segregating sites.

Sample	α (shape) MLE	β (scale) MLE
True DFE of new mutations	0.187	0.0356
Gamma distribution fit to s (s is known)		
segregating variants, 2n=24	0.112	0.00201
segregating variants, 2n=864	0.164	0.00634
new mutations	0.187	0.0356
Gamma distribution inferred from SFS with Fit∂a∂i and multinomial likelihood		
2n=24	0.163	2.39*
2n=864	0.194	0.0248
Gamma distribution inferred from SFS with Fit∂a∂i and Poisson likelihood		
2n=24	0.166	0.0497
2n=864	0.189	0.0364

Notes: These results represent a single simulation replicate. We show that Fit∂a∂i infers the DFE of new mutations correctly for many replicates in Figure S7. These DFEs are scaled in terms of s. *The estimated scale parameter is likely inaccurate due to the small sample size as well as the fact the multinomial likelihood was used.

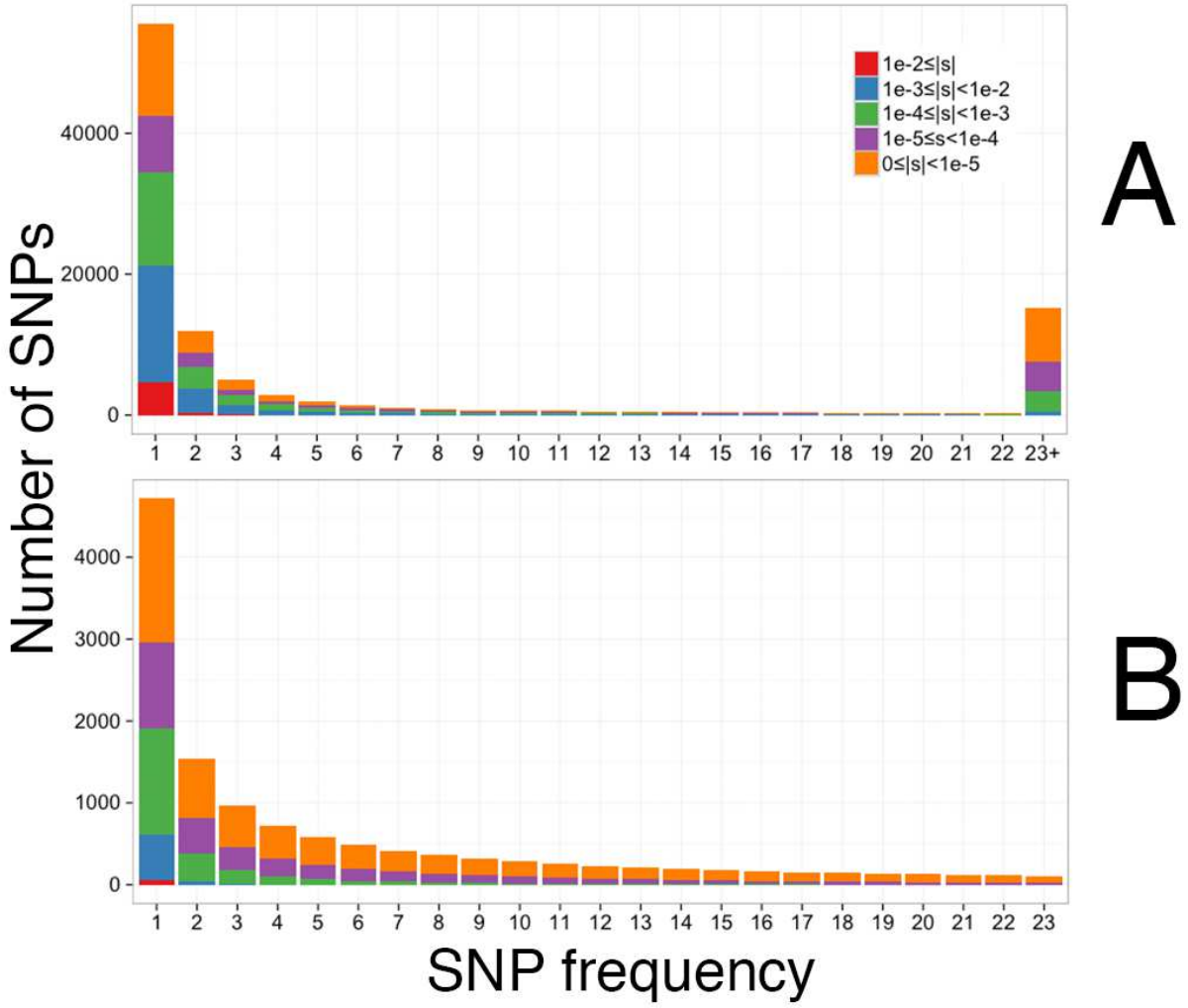

Figure S1. The expected contribution of mutations with various selective effects to the nonsynonymous SFS in (A) a sample of 2596 chromosomes and (B) a sample 24 chromosomes under the demographic model and DFE inferred from the full LuCamp dataset. Note that fewer than 1000 moderately to strongly deleterious mutations (blue and red) are segregating in the small sample, while more than 20,000 of them are predicted to be segregating in the large sample.

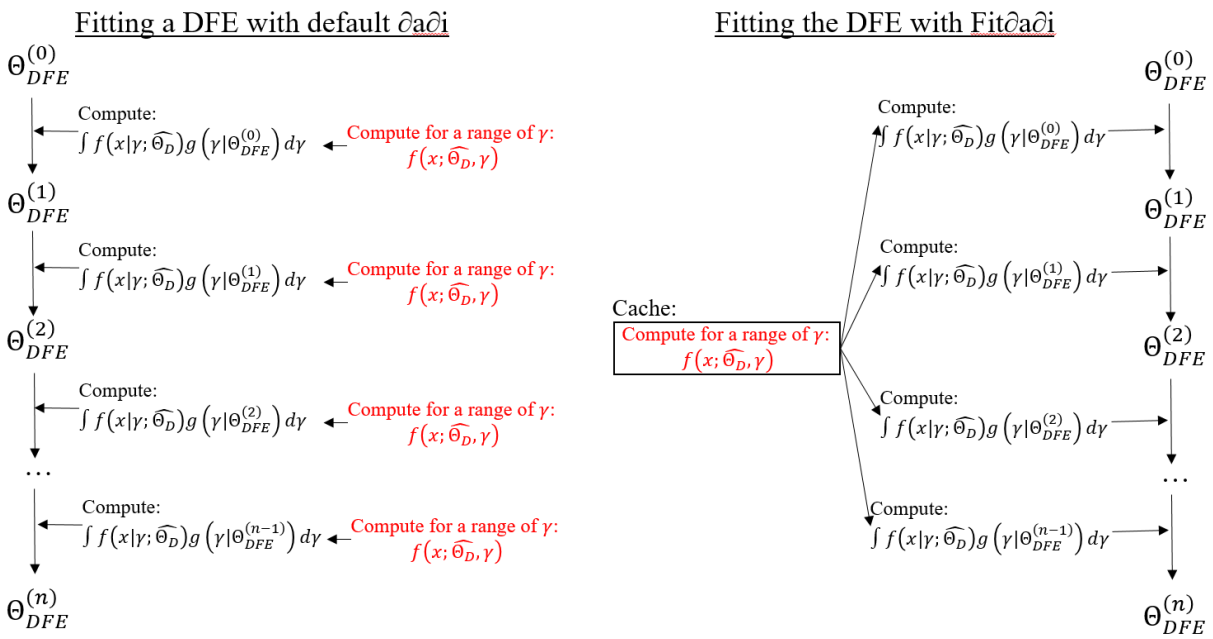

Figure S2. A flow chart depicting the efficiency of Fit∂a∂i compared to the default implementation of ∂a∂i. Fitting a DFE with the default implementation of ∂a∂i is slow because the same frequency spectra must be calculated for each step in the optimization of the DFE. We compute the spectra once and call the saved frequency spectra for each optimization step.

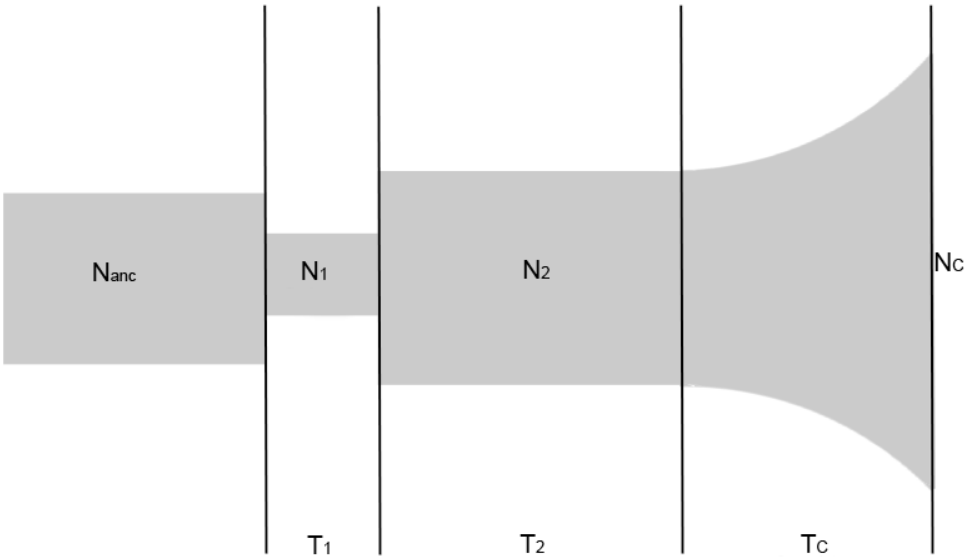

Figure S3. A pictorial representation of the demographic model fit to the SFS of synonymous sites of our datasets. The times denote the length of time of each epoch. See Table S1 for the parameter values we inferred from the data.

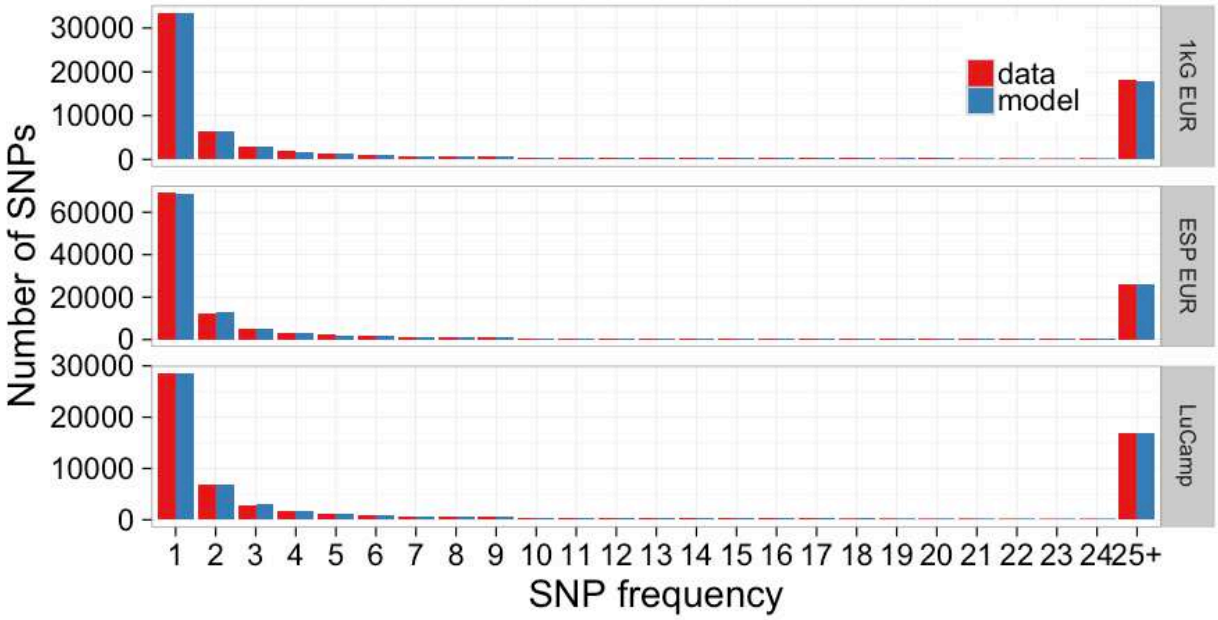

Figure S4. Comparison of the observed synonymous SFS to the SFS from the best fitting demographic model. Each SFS has been folded, then truncated to 25 entries. All the alleles of frequency 25 or greater in the folded frequency spectrum are summed into the last entry. Because these demographic models were fit using the multinomial likelihood, the model SFS has been scaled by θ_S.

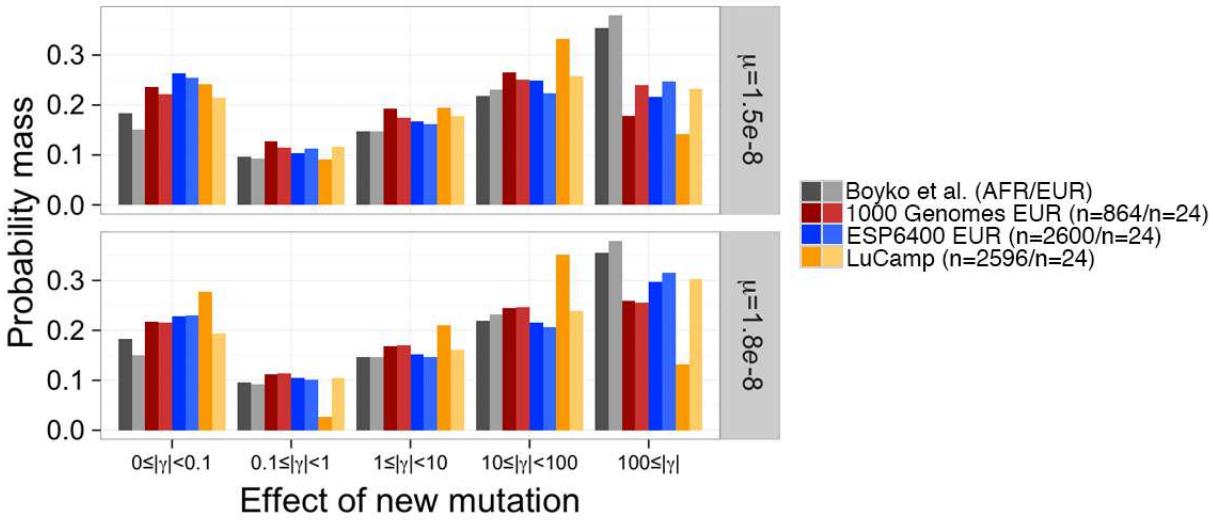

Figure S5. The distribution of population-scaled selection coefficients (*γ*=2*Ns*) of new mutations for our best-fit DFEs compared to Boyko et al. (2008). This figure is the same as **Figure 4**, except estimates are scaled by twice the ancestral population size. Results are presented for the best fit DFE to each full dataset and the best fit DFE when the data were projected down to *n*=24 chromosomes. Our DFEs predict more nearly neutral mutations (0≤|2*Ns*|<0.1) and fewer strongly deleterious mutations (100≤|2*Ns*|) than Boyko et al. (2008), regardless of the mutation rate or the manner in which selection coefficients are parameterized (**Figure 4**). The top panel denotes our favored mutation rate while the bottom panel denotes the mutation rate used by Boyko et al. (2008).

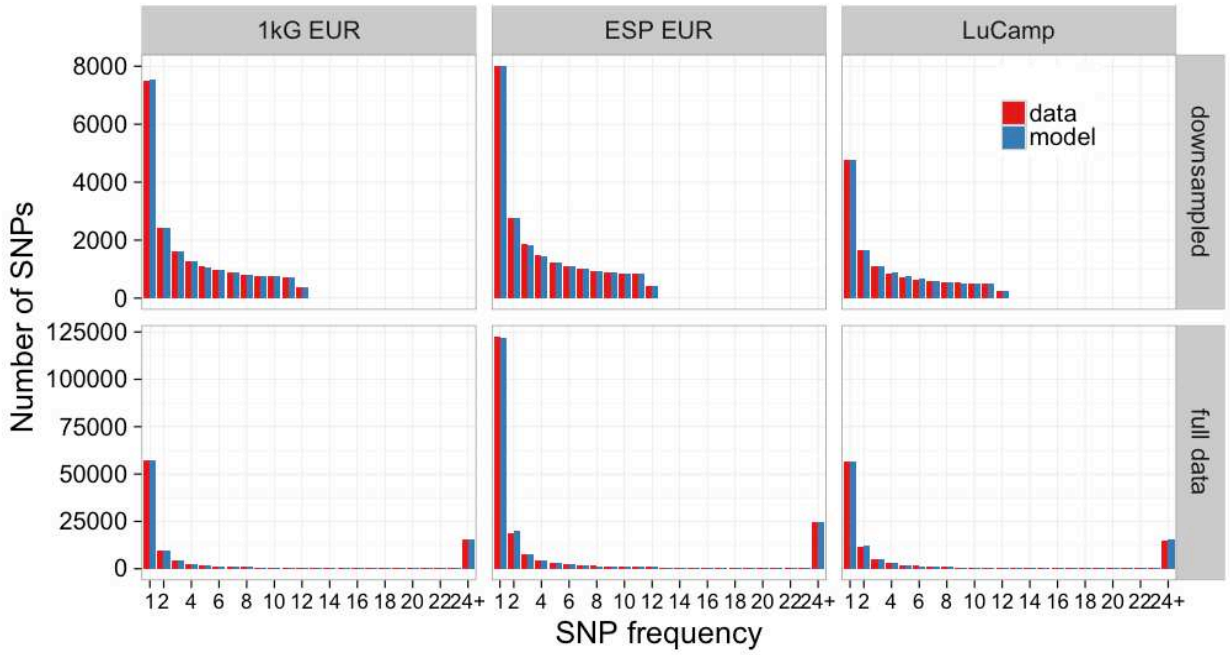

Figure S6. Comparison of the observed nonsynonymous SFS to the SFS from the best fitting demographic and selective models. For the full data, the discrete, neutral+gamma, and neutral+gamma are the best fitting DFEs to the 1000 Genomes, ESP, and LuCamp datasets, respectively. To the downsampled datasets, the neutral+gamma, gamma, and gamma DFEs are the best fitting DFEs to the 1000 Genomes, ESP, and LuCamp datasets, respectively. Each SFS has been folded, then truncated to 25 entries. All the alleles of frequency 25 or greater in the folded frequency spectrum are summed into the last entry.

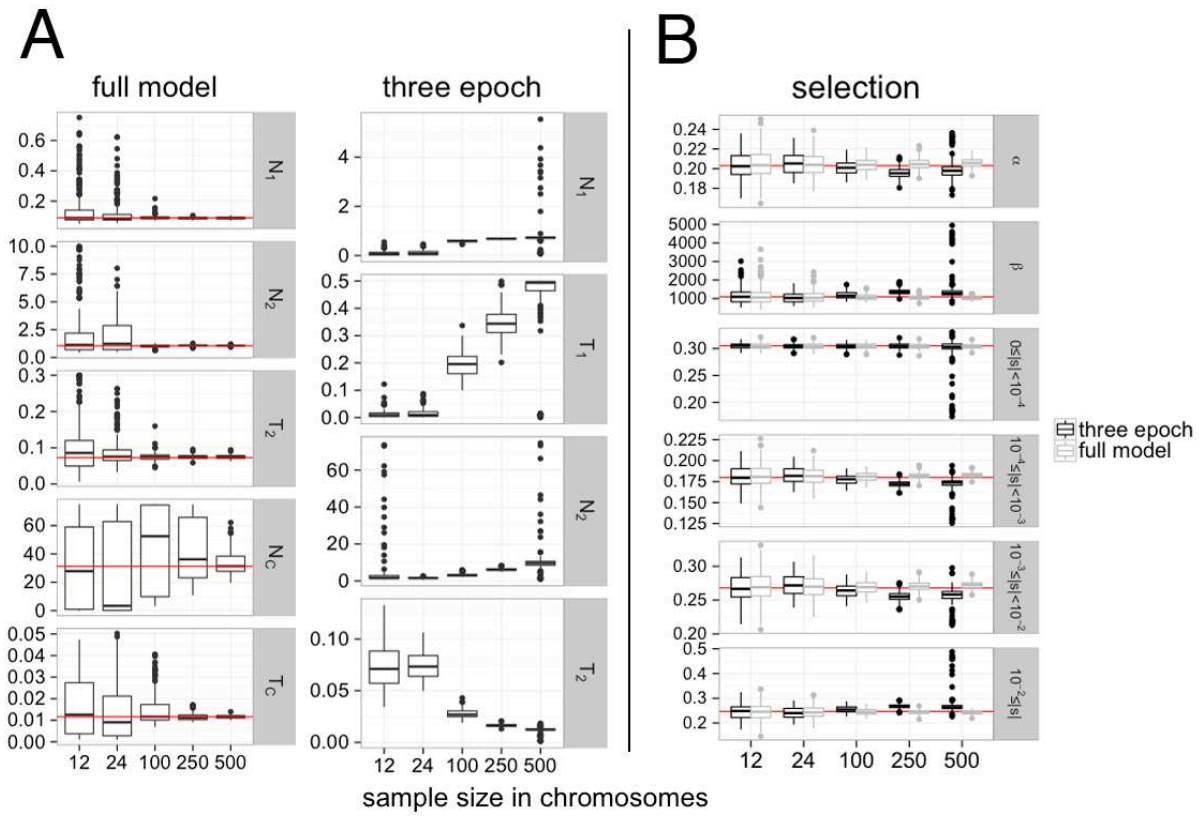

Figure S7. Mis-specification of the demographic model has little effect on selection inference from small samples. Two hundred datasets were simulated under the demographic model fit to the LuCamp synonymous SFS (Table S1, Figure S3) for a range of sample sizes. True values are depicted with a red line. (A) Parameter estimates of the correct demographic model (full model) and a three epoch model fit to the simulated data. (B) Parameter estimates of the gamma DFE as well as the proportions of each range of selective effect for a range of sample sizes assuming L_{NS}/L_S=2.5 and μ=1.8e-8 (α=0.203, β=1082.1). The shape and scale parameters are directly inferred, and the proportions are computed from the gamma distribution. For the full model, sample size does not bias inference of selection. The accuracy of the selection inference conditioned on the three epoch demography declines with increasing sample size because the simple demographic model cannot account for the excess of rare variation in the data. The boxes cover the first and third quartiles, and the band represents the median. The whiskers cover the highest and lowest datum within 1.5 times the interquartile range from the first and third quartiles. Lastly, any data outside that region are plotted as outlier points.

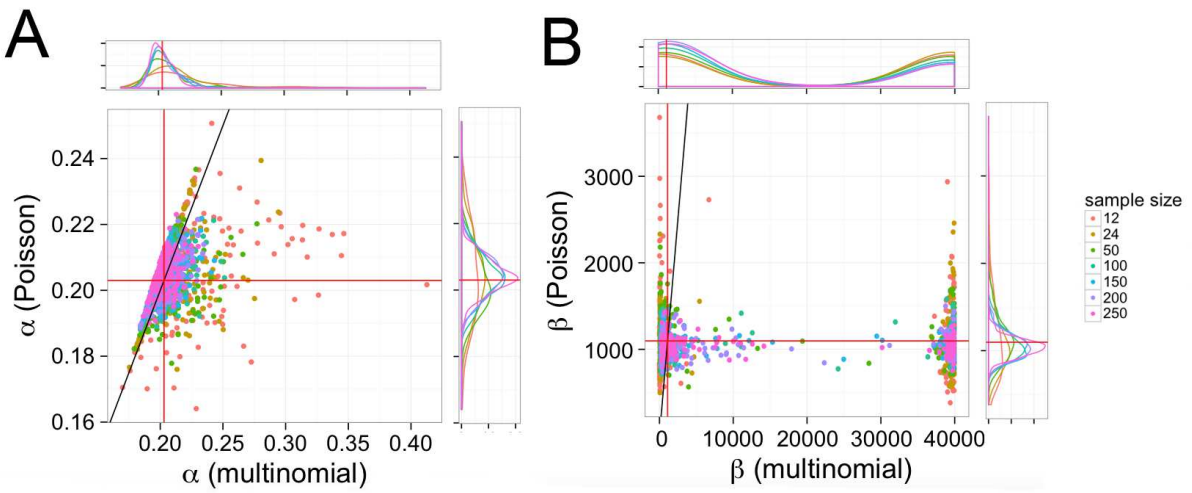

Figure S8. Fitting only the curvature of the SFS is insufficient for estimating the scale parameter of the gamma DFE, especially at small sample sizes. Depicted are the relative performances of the multinomial and Poisson likelihoods when fitting a gamma DFE to 200 simulated datasets of varying sample sizes. The multinomial likelihood only fits the curvature of the SFS while the Poisson likelihood accounts for the total number of SNPs in the data. Simulations were performed using the demographic model and gamma DFE inferred from the LuCamp dataset, assuming L_{NS}/L_S=2.5 and μ=1.8*e*-8 (α=0.203, β=1082.1). True parameter values are depicted with a red line, and the slope of the black line is 1. The density plots describe the marginal densities of the MLEs of the simulation replicates. (A) Shape parameter (α) of the gamma distribution. (B) Scale parameter (β) of gamma distribution. Note the bimodal distribution of the MLEs of the scale parameter at the optimization boundaries when the multinomial likelihood is used.

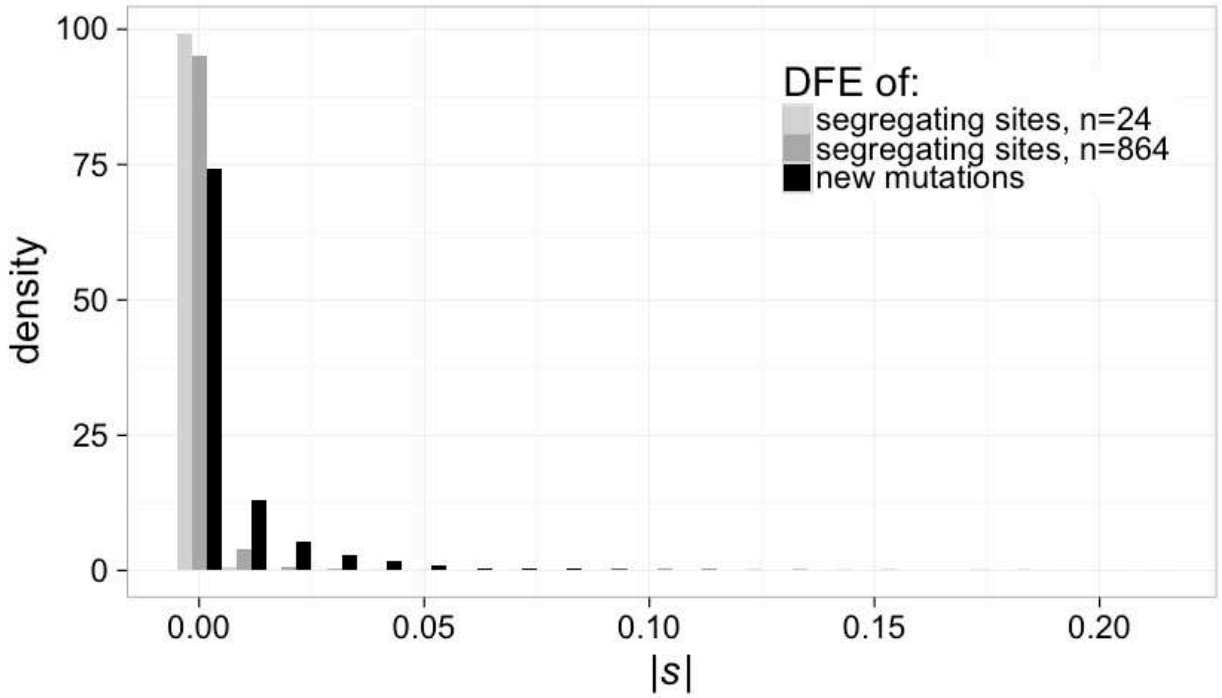

Figure S9. The DFE of segregating sites versus the DFE of new mutations. DFEs from three different simulations are shown as discretized histograms. In the first two, the selection coefficients of only the segregating variants in samples of size n=24 and n=864 chromosomes were tallied. In the third, the selection coefficients of new mutations (which could be segregating, fixed, or lost) were tallied. The DFEs of only segregating sites show a distinct skew towards neutrality, reflecting that strongly deleterious variants are less likely to be found segregating in smaller samples. Additionally, our simulations show a clear distinction between the DFE of new mutations and that of segregating variants.

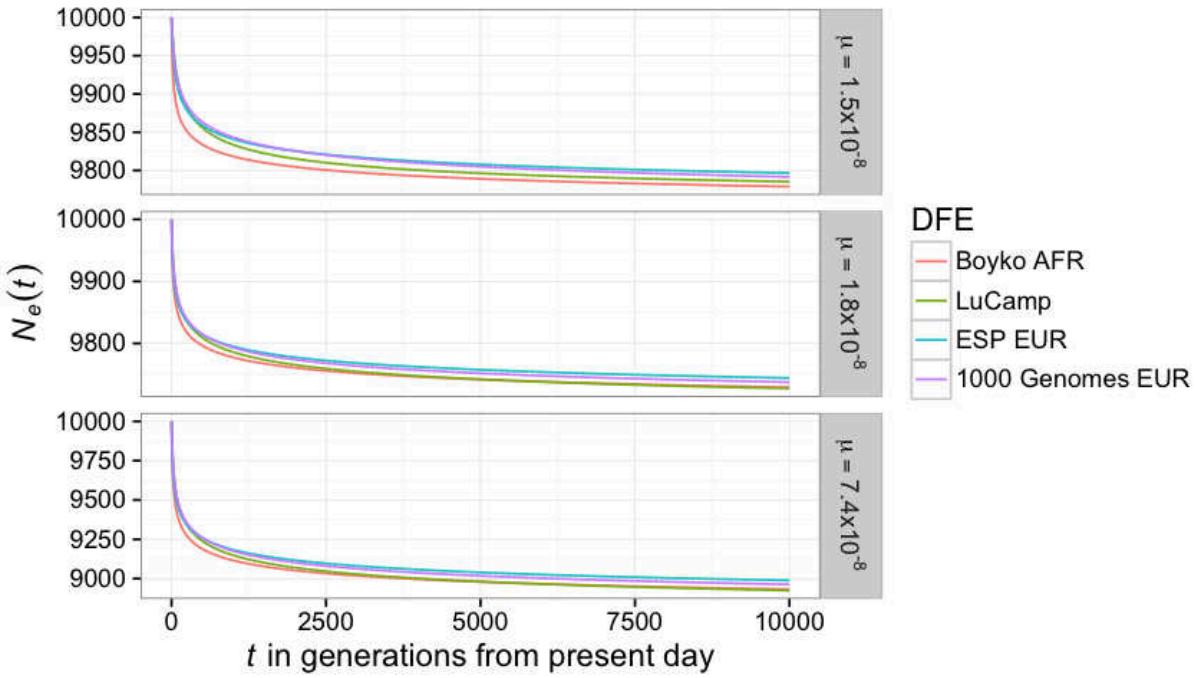

Figure S10. The reduction in genetic diversity due to background selection as a function of time. The reduction in N_e was computed using the deterministic approximation in Nicolaisen and Desai (2013), assuming: a chromosome of length 100Mb with 1.5% coding sequence; a recombination rate of $1x10^{-8}$ per bp; mutation rates similar to those used in our study as well as the deleterious mutation rate of McVicker et al. (2009) of $7.4x10^{-8}$ per bp; and the DFEs inferred in our study as well as that of Boyko et al. (2008). Except the DFE from Boyko et al. (which assumed a mutation rate of $1.8x10^{-8}$), the DFEs used to compute the reduction in diversity for the various mutation rates are the best-fitting gamma or neutral+gamma DFEs inferred using their respective mutation rates. The DFEs used for the mutation rate of $7.4x10^{-8}$ are the best-fitting DFEs inferred assuming the mutation rate was $1.8x10^{-8}$. These calculations assume a constant population size of 10,000 diploids, and values of N_e less than 10,000 are due to background selection reducing linked neutral diversity.

CHAPTER 3:

DELETERIOUS VARIATION SHAPES THE GENOMIC LANDSCAPE OF INTROGRESSION

ABSTRACT

While it is appreciated that population size changes can impact patterns of deleterious variation in natural populations, less attention has been paid to how gene flow affects and is affected by the dynamics of deleterious variation. Here we use population genetic simulations to examine how gene flow impacts deleterious variation under a variety of demographic scenarios, mating systems, dominance coefficients, and recombination rates. Our results show that admixture between populations can temporarily reduce the genetic load of smaller populations and cause increases in the frequency of introgressed ancestry, especially if deleterious mutations are recessive. Additionally, when fitness effects of new mutations are recessive, between-population differences in the sites at which deleterious variants exist creates heterosis in hybrid individuals. Together, these factors lead to an increase in introgressed ancestry, particularly when recombination rates are low. Under certain scenarios, introgressed ancestry can increase from an initial frequency of 5% to 30-75% and fix at many loci, even in the absence of beneficial mutations. Further, deleterious variation and admixture can generate correlations between the frequency of introgressed ancestry and recombination rate or exon density, even in the absence of other types of selection. The direction of these correlations is determined by the specific demography and whether mutations are additive or recessive. Therefore, it is essential that null models of admixture include both demography and deleterious variation before invoking other mechanisms to explain unusual patterns of genetic variation.

INTRODUCTION

There is tremendous interest in quantifying the effects that demographic history has had on the patterns and dynamics of deleterious variation and genetic load [1–7]. Several studies have suggested that recent human demography has had little impact on load [8,9] while others have suggested weak, but subtle differences between human populations [10–14]. All of these studies have typically focused on how population size changes, such as expansions and

bottlenecks, have affected deleterious variation. Other types of complex demography, however, have received considerably less attention.

In particular, gene flow may be important for shaping patterns of deleterious variation. Population admixture, or hybridization between closely related species, appears to be quite common in nature [15] and has had a significant role in shaping human genomes [16]. Gene flow alone can subtly change the effects of selection on deleterious variation [12], but should have notable fitness consequences if deleterious variation is distributed differently between admixing populations. For example, Neanderthals likely had a higher genetic load than coincident human populations due to the former's smaller long-term population size [17,18]. As a result, it is thought that gene flow from Neanderthals into the ancestors of modern humans could have increased the genetic load of some human populations by 0.5% [17], and that linked selection removed much of Neanderthal ancestry from humans since that time. In contrast, domesticated species likely have increased genetic load due to domestication bottlenecks and hitchhiking of deleterious alleles with artificially selected variants [19–21]. Gene flow from wild populations could alleviate the genetic load of domesticated species, and increases in the frequency of wild-population ancestry should be observed in the domesticated population [22]. Such changes in patterns of introgression are important to consider when studying how natural selection shapes the evolution of hybrid ancestry, a major goal in evolutionary biology.

Differences in the distribution of deleterious variation between hybridizing populations is one reason why natural selection may shape the evolution of hybrid ancestry. Hybridization can also decrease the fitness of a population, for instance, if the parent lineages have diverged significantly and evolved genomic incompatibilities, or if parent lineages have evolved under unique and strong selective pressures in different environments. In both cases, linked selection removes hybrid ancestry especially in regions of low recombination and high functional density [23–25]. This creates genome wide, negative correlations between the local recombination rate, or functional density, and the frequency of introgressed ancestry, a pattern that is observed in

humans [23,25,26], swordtail fish [25], and mice [27]. The similar outcomes of both these processes mean that models of selection on deleterious variation should be considered before interpreting genomic patterns of introgression as evidence of divergence and speciation.

Another complication to studying the effects of deleterious mutations on introgression is that strongly deleterious new mutations are more likely to be fully or partially recessive [28–30]. Furthermore, dominance coefficients vary between species, and can range from close to additive in humans [13] to mostly recessive in Arabidopsis [30]. If some proportion of deleterious recessive variants is private to a population, admixed populations could experience heterosis when recessive variants are masked (heterozygous) in hybrid individuals [31]. As a result, heterosis may participate in a tug-of-war on hybrid ancestry with additive variants by increasing the frequency of linked ancestry [17], increasing apparent migration rates in regions linked to selected variants [32,33], particularly when gene flow occurs in a highly structured population [34]. Heterosis should also increase the probability that introgressed ancestry will persist in an admixed population, even if the introgressed ancestry contains more deleterious alleles [17]. Given the extent to which hybridization is thought to be common to all species [15], with levels of shared polymorphism in taxa such as *Arabidopsis* motivating arguments for the bifurcating species concept to be revoked [35], it is crucial to understand the contribution of heterosis to patterns of hybrid ancestry.

Hybridization also transfers novel adaptive variants between evolutionarily distinct lineages [36]. In humans, many Neanderthal variants are thought to be adaptive [37], possibly affecting phenotypes such as skin pigmentation [38,39], the response to oxygen levels at high altitudes [40,41], and immunity to pathogens [42]. In this case, the introduction of beneficial alleles via gene flow will also oppose the effect of linked selection from deleterious variation, since introgressed ancestry would increase in frequency by hitchhiking with adaptively introgressed variants. Interestingly, North American populations of *Drosophila melanogaster* exhibit an overall enrichment for introgressed African ancestry in genomic regions of low

recombination [43,44]. The divergence time between these two *D. melanogaster* populations is small, and so selection on hybrid individuals may be driven by adaptive variants that arose over shorter time scales than genomic incompatibilities. On the other hand, no correlation between recombination rate and introgression is observed in invasive Californian sunflowers [45]. How selection against additive deleterious variation, selection for adaptive variants, and heterosis interact to determine these genomic patterns is unknown.

The objective of this study is to develop a clearer idea for null models of the dynamics of introgression in hybridizing populations while considering the effect of deleterious variants on fitness. Specifically, we aim to understand how selection on introgressed ancestry is determined by differences in the effective population size, mating system, genome structure, recombination rate, distribution of fitness effects, and distribution of dominance coefficients. Previous simulation and empirical work have shown that for at least some systems, deleterious variation is a significant modulator of gene flow [17,18,22,24,25], but few studies have investigated these questions outside of demographic models specific to a system. This study presents a series of simulations utilizing demographic models that generalize biological scenarios of interest by borrowing population genetic parameters and genomic structure from humans and *Arabidopsis thaliana*, two markedly different organisms with markedly different population genetic parameters. We include a realistic distributions of fitness effects and simulate under various models of dominance. In addition, we examine how the relationship between the genomic landscape of introgressed ancestry and recombination rates or functional content is determined by the underlying demography.

RESULTS

Forward simulations

We used SLiM 3.0 [46] in conjunction with tools from pyslim [47] and msprime [48] to simulate a series of five models of admixture in the presence of deleterious variation. Each of

the five models was based on a divergence model where an ancestral population at equilibrium splits into two subpopulations. At some time after the split, a single pulse of admixture occurs at a proportion of 5%, in one direction and for a single generation. Due to practical considerations only an initial admixture proportion of 5% was simulated. **Figure 1** provides a cartoon representation of these models and the specific model parameters can be found in **Table S1**.

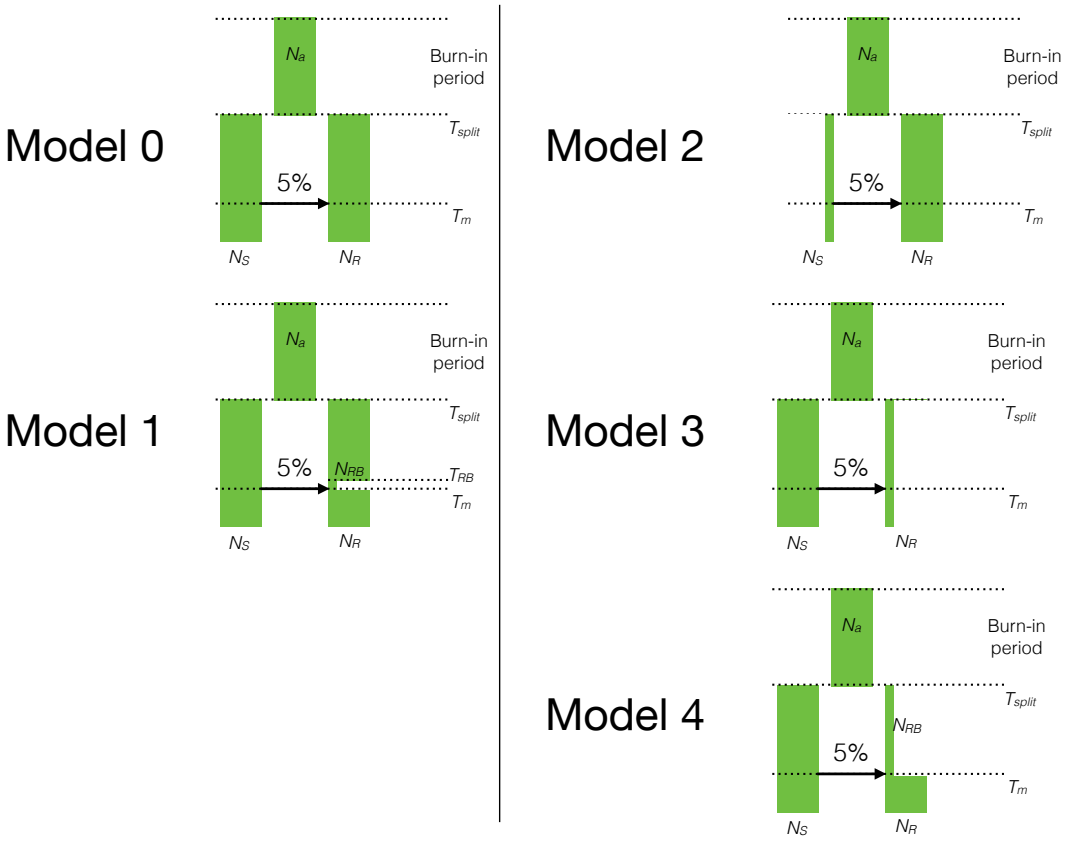

Figure 1. The demographic models used for the simulations. After a burn-in period of $10N_A$ (100,000) generations, a single population diverged into two subpopulations. The demography of the subpopulations was modified in ways that changed the distribution of deleterious variation. $2N_A$ (20,000) generations after the split, a single pulse of admixture occurred such that 5% of the ancestry of the recipient population came from the donor population. Arrows in each panel denote the direction of gene flow. The simulation was run for N_A (10,000) additional

generations after admixture. Population sizes were changed as shown for each model. See **Table S1** for specific parameter values for each model.

All simulated sequence included genic structure (exon/intron/intergenic regions), which was either randomly generated or incorporated from a reference genome as described in the following sections. Only new nonsynonymous mutations were assigned nonzero selection coefficients, which were drawn from a gamma distribution of fitness effects (DFE) with shape parameter 0.186 and E[*s*]= -0.01314833 [49] except when specified otherwise. In other words, no positively selected mutations were simulated.

Throughout, we will refer to the subpopulation that migrants originate from as the *donor* subpopulation, and the subpopulation that migrants join as the *recipient* subpopulation. Furthermore, we will refer to ancestry in the recipient subpopulation that originated in the donor subpopulation as *introgression-derived ancestry*. We use p_I to denote the total proportion of ancestry that is introgression-derived in the recipient subpopulation.

See the **Methods** for additional details on the simulations.

Demography and recombination rate create differences in load between populations

To better understand how deleterious variants shape patterns of introgressed ancestry, we first simulated small genomic segments with randomly generated genic structure, of length ~5 Mb and selection coefficients from a gamma DFE. Two hundred simulation replicates using each of the 5 demographic models in **Figure 1** (parameters in **Table S1)**, each of the per base pair recombination rates $r = 10^{-6}$, 10^{-7}, 10^{-8}, and 10^{-9}, and additive (h=0.5) or recessive (h=0.0) fitness effects were generated, for a total 8,000 independent replicates.

In the 20,000 generations between the population split and admixture event, deleterious mutations accumulate at different rates across subpopulations for each unique model (**Figure**

S1), illustrated by the relative difference in subpopulation fitness in **Figure 2**. We report subpopulation fitness while ignoring the deleterious variants that have fixed in both subpopulations, since selection will not act on globally monomorphic variants. Because some weakly deleterious variants will fix in one subpopulation yet be lost in the other, each subpopulation's fitness also steadily decreases through time.

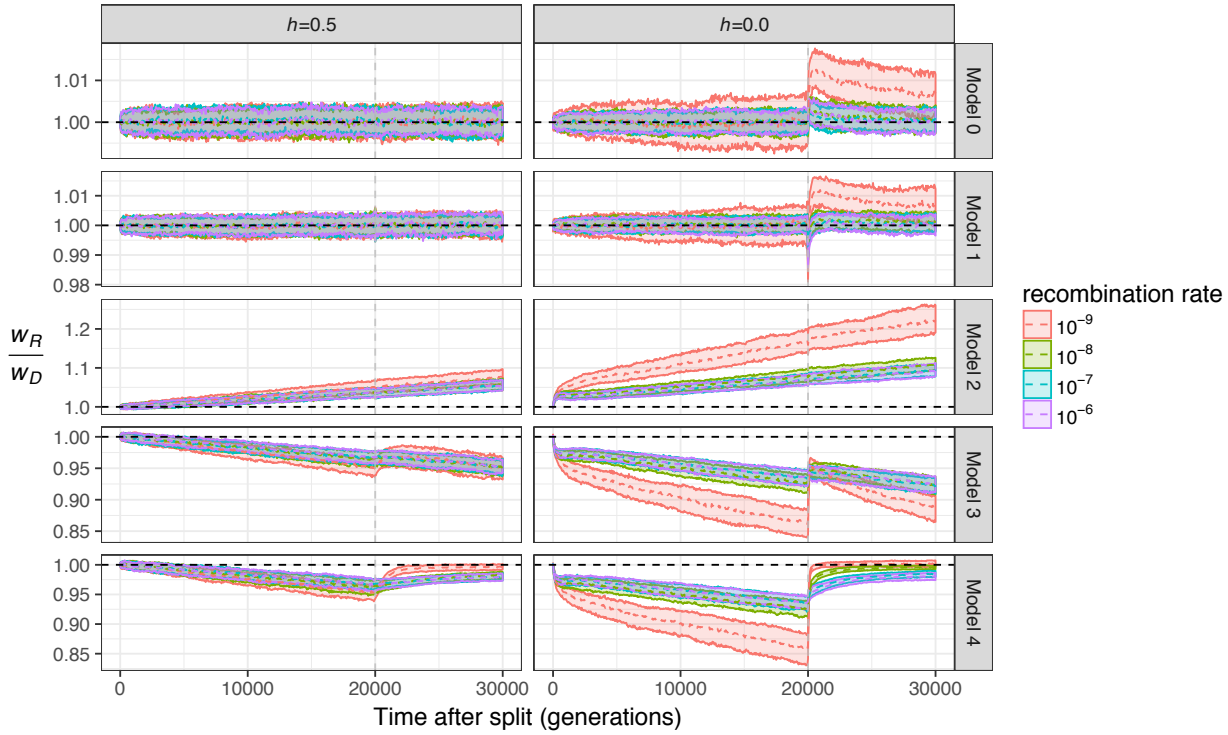

Figure 2. The change in the ratio of fitness over time due to demography. Each individual plot depicts the ratio of the mean fitness of the recipient population (w_R) to the donor population (w_D) for the demographic models shown in **Figure 1**. The mean (dotted line) and the 25th to 75th percent quantiles are shown for 200 simulation replicates. The vertical gray line depicts the time of gene flow, and the horizontal dashed black line depicts w_R/w_S=1. Different colors denote distinct recombination rates used in the simulations. Left panel denotes additive mutations (h=0.5) while the right panel shows recessive mutations (h=0).

In the additive fitness model, this relative difference in fitness is simply determined by relative differences in subpopulation size. When there are no differences in subpopulation size (Model 0), the fitness of both donor and recipient subpopulations decreases at approximately the same rate ($w_R \approx w_D$, **Figure 2**). A similar pattern is for a short bottleneck in the recipient population ($w_R \approx w_D$, Model 1, **Figure 2**), reflecting the insensitivity of additive genetic load (measured in terms of the number of deleterious variants per haplotype in **Figure S2**) to short-term changes in N_e [8]. In contrast, long-term differences in population size (Models 2-4, **Figure 2**) provide enough time for deleterious variants to drift to higher frequency in the smaller subpopulation, resulting in substantial differences (approximately 5%) in fitness between subpopulations.

When deleterious mutations are recessive, a qualitatively similar relationship between subpopulation size and subpopulation fitness is generally observed. When there are no differences in population size (Model 0), the fitness of donor and recipient subpopulations decreases at a similar rate ($w_R \approx w_D$, **Figure 2**). A short bottleneck in the recipient population (Model 1) increases the frequency of homozygous, recessive genotypes immediately post-bottleneck (**Figure S3**) which slightly decreases the recipient subpopulation's fitness immediately before admixture (**Figure 2**). Finally, similar to the additive fitness model, long-term differences in population size result in substantial differences (>10%) in relative fitness between admixing populations.

The recombination rate is a key factor in determining differences in fitness between the two subpopulations. When the recombination rate is low, the fitness of the smaller subpopulation decreases at a lower rate, reflecting the reduced efficacy of purifying selection in low recombination regions [50]. Relative subpopulation differences in fitness between high recombination (r=10^{-6}) and low recombination (r=10^{-9}) simulations are about 2% for the additive fitness model and about 8% for the recessive fitness model.

Demography and recombination determine changes fitness post-admixture

Similar to the manner in which they affect subpopulation differences in fitness, recombination rates interact with demography to determine changes in subpopulation fitness after admixture.

When fitness effects are additive, admixture is unlikely to cause immediate and large changes in fitness, while subpopulation differences in fitness lead to gradual changes in fitness over time. If admixing subpopulations have the same fitness (Models 0 and 1, **Figures 2 and S1**), admixture predictably has no impact on the recipient population's fitness. If donor haplotypes have lower fitness than the recipient (Model 2, **Figure 2**), the recipient population's fitness is negligibly decreased by admixture (**Figure S2**), specifically because relative differences in donor and recipient are relatively small (<10%) and the initial frequency of donor ancestry is always 5%. If instead the donor subpopulation has higher fitness (Models 3 and 4, **Figure 2**), recipient fitness is relatively unaffected at the time of admixture but increases over time (**Figure S2**) as the fitter haplotypes experience directional selection. The velocity and magnitude of these changes depends on the recombination rate, as variants originating from the same subpopulation are generally selected in the same direction, and these variants remain on the same haplotypes when recombination is low.

When fitness effects are recessive, admixture instead causes immediate and large changes in fitness as recessive alleles are masked in heterozygous, hybrid individuals (**Figures 2 and S1**). The qualitative patterns observed are consistent across all demographic models, but the magnitude of these changes is significantly larger in simulations where the recipient subpopulation has lower fitness. The recombination rate again plays a key role in determining fitness in the recipient subpopulation, with the largest changes in fitness occurring in simulations with low recombination. This occurs because the largest differences in pre-admixture fitness are observed when recombination is low (**Figure 2**), but also because the heterozygosity of hybrids is maximized if recombination does not occur between donor and recipient haplotypes. This

linkage effect is particularly important as most of the variants under selection should have weak effects, since selection is likely to prevent strongly deleterious variants from drifting to high frequency even in a small population.

Demography and recombination rate determine patterns of introgression

We next explore changes the frequency of introgressed ancestry (p_I) over time in the different models.

In the additive fitness case, changes in the frequency of introgression-derived ancestry are directly predictable from the differences in subpopulation fitness. When there are no differences in load ($w_R \approx w_D$, Models 0 and 1, **Figures 2 and S1**) between mixing haplotypes, selection does not favor a particular ancestry and donor subpopulation ancestry remains, on average, at the initial admixture proportion of 5% in the recipient (**Figure 3**). If donor subpopulation haplotypes have lower fitness as in Model 2 (**Figures 2 and S1**) deleterious donor ancestry is removed by selection, leading to a long-term p_I of less than 5%. If instead the donor subpopulation has higher fitness (Models 3 and 4, **Figure 2**), p_I is increased above 5% by selection. This increase is greatest (p_I = 75%) when there is an expansion after the time of admixture and in regions of low recombination (Model 4).

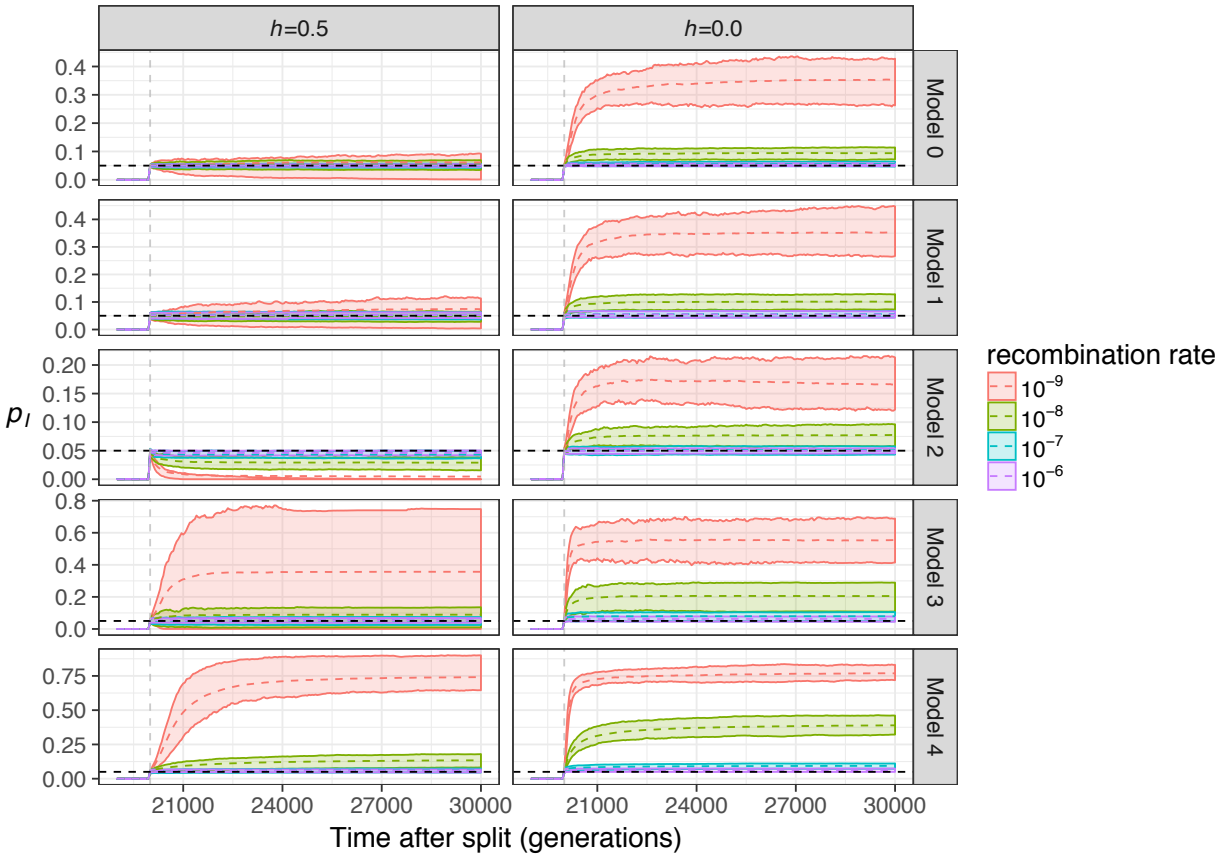

Figure 3. The frequency of introgression-derived ancestry (p_I) in each model. Earlier generations are not shown since p_I=0 prior to admixture. The mean (dotted line) and the 25th to 75th percent quantiles are shown for 200 simulation replicates. The vertical gray line depicts the time of gene flow, and the horizontal dashed black line depicts the initial admixture proportion of 0.05. Different colors denote distinct recombination rates used in the simulations. Left panel denotes additive mutations (h=0.5) while the right panel shows recessive mutations (h=0).

In a recessive fitness model, selection initially favors donor ancestry in the recipient subpopulation. In all cases (Models 0-4, **Figure 3**), the frequency of introgression-derived ancestry increases after admixture, regardless of whether the donor subpopulation's fitness is less fit or more fit than the recipient. This effect is explained by heterosis, which occurs when recessive deleterious variants are masked as heterozygotes in hybrid individuals (**Figure S3**),

particularly in the generations immediately following admixture. At this time point, recombination has had little chance to shuffle donor and recipient haplotypes and heterozygosity is maximized in admixed individuals.

Again, the recombination rate is a key parameter that determines patterns of introgressed ancestry. As described previously, variants that are selected in the same direction remain linked when recombination is low (r=10^{-9}, **Figure 3**), maximizing the effect of selection and minimizing selective interference between recombinant haplotypes. When recombination is high (r=10^{-6}), the proportion of donor ancestry is unaffected by selection post-admixture (long-term p_I=5%, **Figure 3**), as recombination quickly decouples variants under selection from their ancestry backgrounds. Importantly, when recombination rates are low (r=10^{-9}), the frequency of introgressed ancestry can increase substantially to up to 75% in the recipient population, despite the initial admixture proportion of 5%. Even with higher recombination rates, when deleterious mutations are recessive and there is a population expansion at the time of admixture (Model 4), introgressed ancestry can increase up to 25% frequency.

The impact of the population split time on heterosis

So far, we have fixed the split time before admixture at 2N generations, a substantial time for differences in deleterious variation to accumulate between subpopulations. To further examine the relationship between split time and selection on introgression-derived ancestry, we simulated with Models 0 and 4 but also varied the time between the split and admixture (t_s). For simulations with a demography analogous to Model 0, we simulated two divergent populations of equal size. For those analogous to Model 4, the recipient subpopulation's size was reduced to 1,000 diploids immediately after the split and recovered to the original size at the same time that gene flow occurred. The recombination rate was set to r=10^{-9} in these simulations.

Figure 4 depicts the long-term proportion of introgressed ancestry, p_I, 10,000 generations after the admixture event for these two sets of models. We found that across our

range of simulated t_s, the long-term frequency of introgressed ancestry increases monotonically with t_s regardless of the underlying demography. Longer split times result in more deleterious variation being unique to each subpopulation, causing heterosis after admixture as private deleterious variants are masked by introgressed ancestry (**Figure S4**). However, these differences appear to reach equilibrium after 20,000 generations (**Figure 4**), about when most deleterious variants are private to one subpopulation (**Figure S4**). We also found as a bottleneck increases in duration, differences in subpopulation fitness become a significant contributor to the increase in long-term p_I, but note the apparent equilibrium at 20,000 generations. At a split time and thus bottleneck time of >20,000 generations, heterosis and differences in load increase long-term p_I nearly 2-fold relative the model with no differences in load (compare Model 0 to Model 4 in **Figure 4**). When parametrizing the population split times in terms of the realized F_{ST} values computed from the SNPs in the simulation output, we find that even for low levels of differentiation (F_{ST}>0.04), there is a pronounced increase in introgressed ancestry (**Figure 4**). Interestingly, simulations with large long-term p_I (e.g. Model 4 at 1,000 generations or Model 0 at 5,000 generations) can have a level of differentiation of F_{ST}<0.2 at the time of admixture, suggesting that even moderate levels of differentiation between subpopulations are sufficient to drive heterosis in low recombination regions (**Figure 4**).

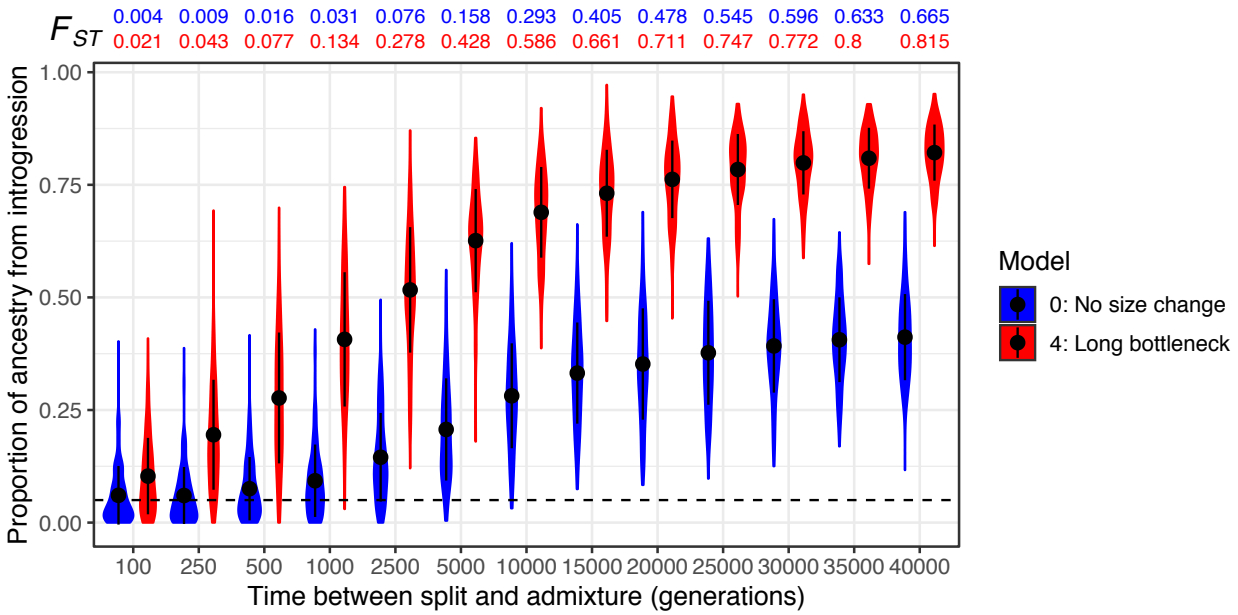

Figure 4. Population split time before admixture and population size impact the amount of introgressed ancestry when mutations are fully recessive. The proportion of ancestry that is introgression-derived, p_I, at the time of N_A (10,000) generations after admixture, is shown for 200 simulation replicates and two demographic models (Model 0 and Model 4, refer to **Figure 1**) for a range of times between subpopulation divergence and the admixture event. The recombination rate in all simulations is r=10^{-9} per base pair. Violin plots represent the density while dot and whiskers represent the mean and one standard deviation to either side. The horizontal dashed black line represents the initial admixture proportion of 0.05. Note that as the split time increases,

Human genome structure results in a heterogeneous landscape of introgression

So far, we have shown how selection on load shapes introgression-derived ancestry in a set of simple simulations. However, recombination rates and gene density are heterogeneous across actual genomes, and our simulations suggest this variation also could influence the genomic landscape of introgression.

To investigate how a realistic genomic structure affects patterns of introgression, we simulated with three of the demographic models described previously (Models 0, 2, and 4) using exon definitions and recombination map for a 100 Mb segment of human chromosome 1. We fixed the exon definitions and recombination map to be the same for all simulations. Only new nonsynonymous mutations were assigned non-zero selection coefficients drawn from a gamma DFE. In addition to simulating both additive and recessive fitness effects separately, we also simulated an inverse relationship between dominance coefficients and selection coefficients, which we will refer to as the $h(s)$ relationship, using the function estimated by Henn et al. [13]. We generated 100 simulation replicates for each of the three demographic models. At the end of each simulation, we split the simulated chromosome into non-overlapping 100kb windows and computed the frequency of introgression-derived ancestry, exon density, and the average per base pair recombination rate in each window.

The frequency of introgression-derived ancestry generally exhibited genome-wide increases after admixture when mutations were partially or fully recessive and varied in accordance with differences in N_e between subpopulations when mutations were additive. In the model with equal subpopulation sizes (Model 0), we observed no average change in the frequency of introgression-derived ancestry when mutations were additive. When new deleterious mutations were partially or fully recessive, we observed an overall genome-wide increase in the frequency of introgression-derived ancestry (**Figure 5**), with many regions reaching high frequency (>50%) in single simulation replicates (**Figure S5**). This increase in frequency is only due to selection on recessive mutations and local variation in recombination rate, since no positively selected mutations were simulated.

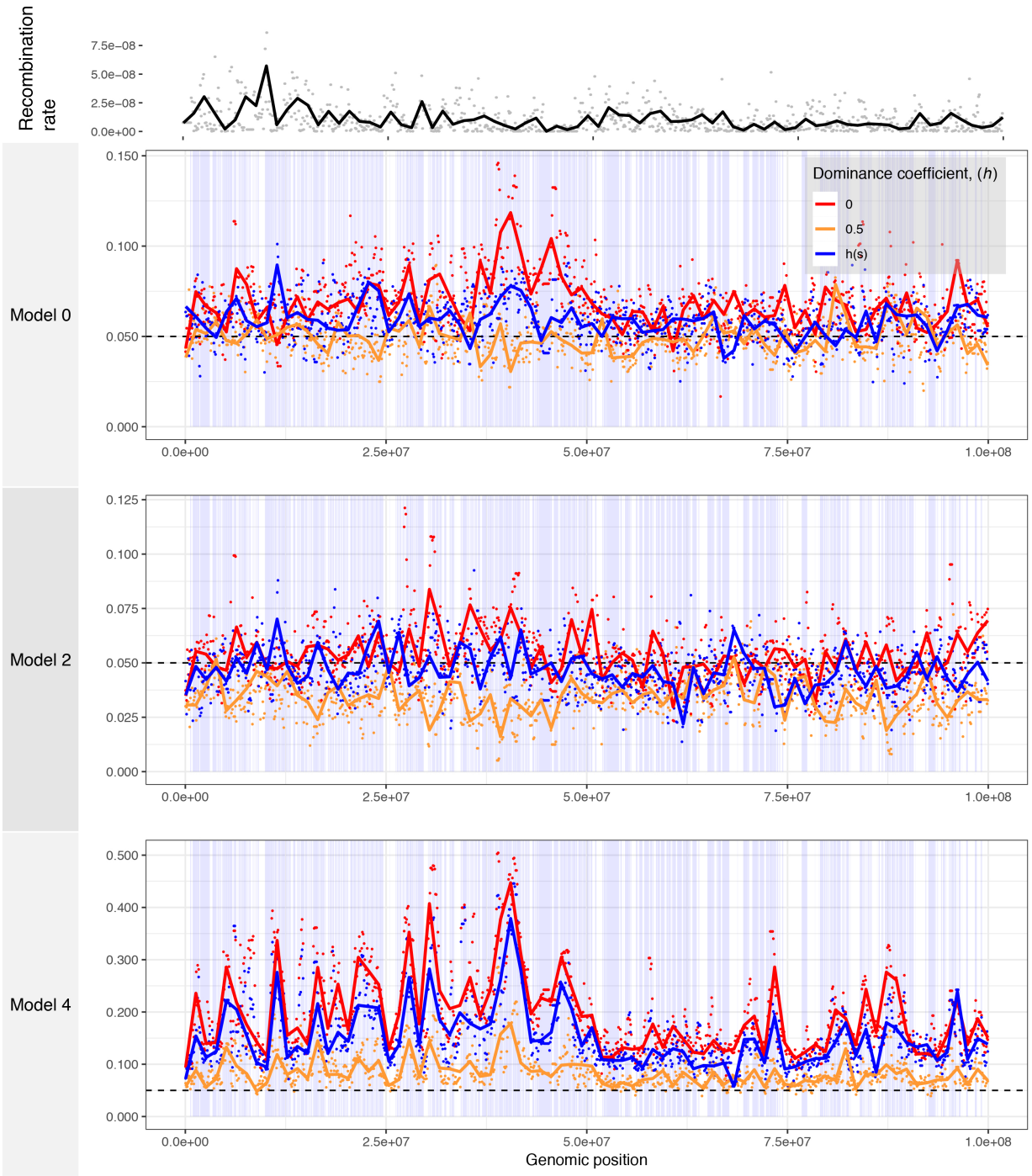

Figure 5. The average genomic landscape of introgression in simulations with human genomic structure. The frequency of ancestry that is introgression-derived is shown for non-overlapping 100 kb windows in a simulated 100 Mb region of chromosome 1. The model numbers refer to the models shown in **Figure 1**. Points represent a single value for each 100 kb window and

lines are loess curves fitted to the data. The horizontal black dashed line represents the initial frequency of introgression-derived ancestry, p_I=0.05. Vertical blue bars represent genes in which deleterious mutations can occur. Red curves denote the results for recessive mutations, orange curves show the results for additive mutations, and blue curves show the results for simulations with a $h(s)$ relationship.

In the model where introgressing haplotypes carried a larger deleterious burden (Model 2) and when deleterious mutations were not all recessive, we observed an overall depletion of introgressed ancestry consistent with the effects of purifying selection upon introgressed ancestry (**Figure 5**). However, in simulations with fully recessive mutations, the effects of heterosis were strong enough such that many genomic regions showed average increases in frequency of 1.5 to 2 times that of the initial introgression frequency of 5%. Importantly, Harris and Nielsen [17] predicted that heterosis would increase the frequency of introgressed ancestry by only a few percent, but our simulations with a similar demographic model show that larger increases in the frequency of introgressed ancestry, especially in exon-dense and low recombination regions.

Finally, when we simulated the introgression of haplotypes from a subpopulation with lower genetic load (Model 4), we observed drastic, genome-wide increases in the average frequency of introgressed ancestry in the recipient subpopulations (**Figure 5**) as well as many fixed loci in individual simulations (**Figure S5**), regardless of whether fitness effects of mutations were additive or recessive. For example, local regions of the simulated chromosome showed an average increase in introgressed ancestry from an initial frequency of 5% up to 50-60% frequency. Furthermore, peaks of introgression are highly correlated between the simulations with different models of dominance, suggesting that the interplay between exon density and recombination strongly affects the way that selection acts on introgressed ancestry in this

model. This is the type of signature that would be unlikely to be generated under neutral demographic models and could be mistakenly attributed to adaptive introgression.

It is also notable that the frequency of introgression-derived ancestry (p_I) in each window appears to be driven not only by recombination but by exon density, or the local concentration of sites at which deleterious mutations can occur. For recessive mutations, p_I is greatly increased on the left-hand side of the simulated chromosome, which tends to be more gene-rich than the right-hand side of the chromosome (**Figures 5** and **S5**). Importantly, the recombination rate was not significantly correlated with exon density (Spearman's ρ=-0.0457, p=0.149) in our simulations, showing these factors likely act independently to shape the landscape of introgression.

To more formally explore these relationships, we examine the correlations between genomic features and the average frequency of introgressed ancestry across 100 simulation replicates, measured in 100 kb windows (**Figures 6** and **7**). In the model of equal subpopulation sizes (Model 0), the frequency of introgression-derived ancestry is not significantly related to the recombination rate or exon density when mutations have additive effects, but is positively correlated to exon density when fitness effects are fully or partially recessive (**Figure 7**). The $h(s)$ relationship results in intermediate levels of introgression relative to simulations with strictly additive or fully recessive new mutations. For Model 2, the frequency of introgression-derived ancestry is positively correlated to the recombination rate and negatively correlated to exon density when fitness is additive. When fitness effects are fully recessive for this model, the frequency of introgressed ancestry is negatively correlated to recombination rate (middle panel in middle row in **Figure 6**) and positively correlated to exon density (middle panel in middle row in **Figure 6**). However, under the $h(s)$ relationship, introgression derived ancestry is not significantly correlated to the recombination rate but is correlated with exon density. Lastly, when introgressed ancestry comes from a larger subpopulation with a lower deleterious burden than the recipient subpopulation (Model 4), the frequency of introgression-derived ancestry is

always negatively correlated with recombination rate, and positively correlated with exon density. For Model 4, these correlations are observed for all models of dominance.

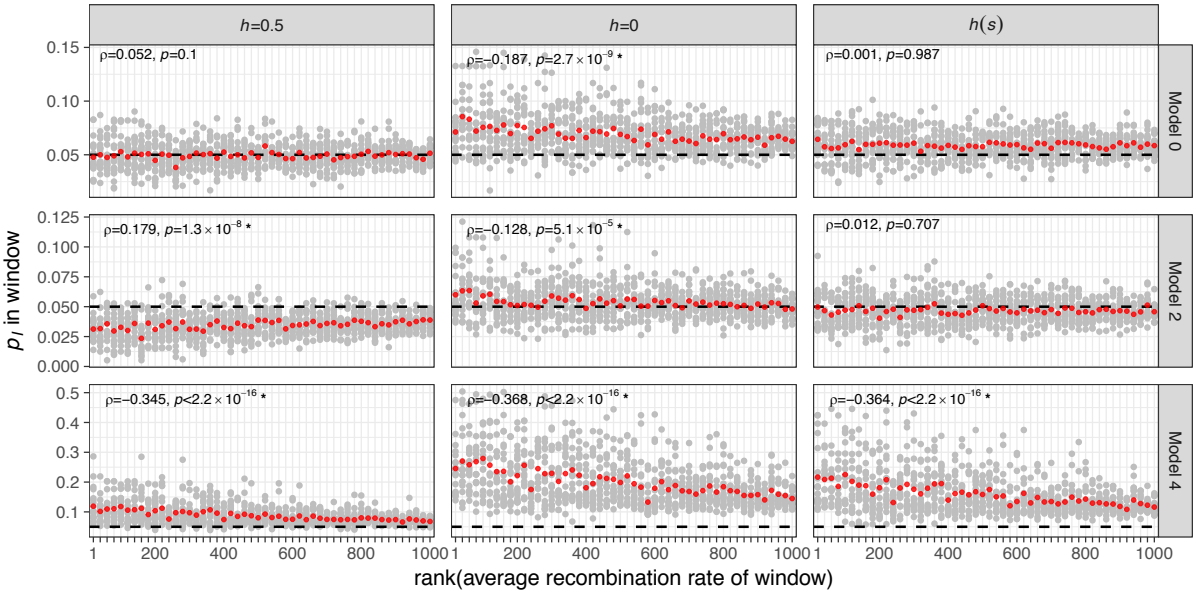

Figure 6. The relationship between recombination rate and frequency of introgressed ancestry for different demographic and selective scenarios. The frequency of introgression-derived ancestry (p_I) is plotted against the average recombination rate of non-overlapping 100 kb windows in each window at time N_A (10,000) generations after admixture. Gray dots represent the average p_I of a single window in 100 simulation replicates, while red dots represent the average p_I of 5% of windows as ordered by rank of recombination rate. Rank was randomly assigned for ties. The horizontal black line represents the initial p_I of 5%. Spearman's ρ is computed for the relationship between recombination rate and p_I in each window and p-values indicate the significance of H_1: $\rho \neq 0$. The model numbers refer to the models shown in **Figure 1**.

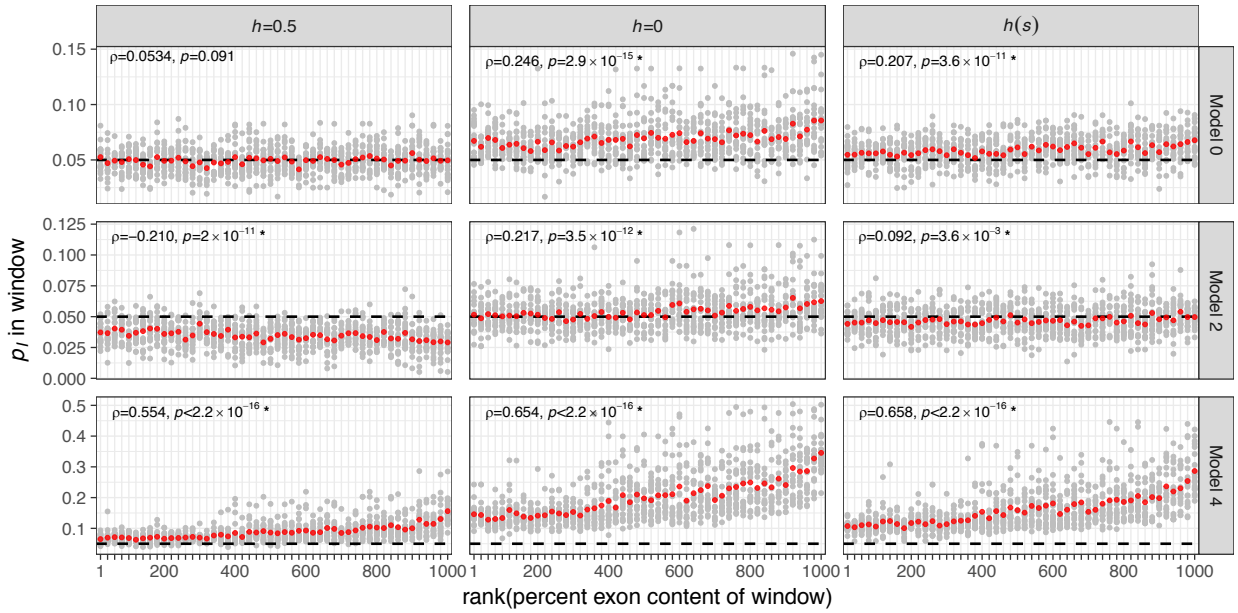

Figure 7. The relationship between exon density and frequency of introgressed ancestry for different demographic and selective scenarios. The frequency of introgression-derived ancestry (p_I) is plotted against the average exon density of non-overlapping 100 kb windows in each window at time N_A (10,000) generations after admixture. Gray dots represent the average p_I of a single window in 100 simulation replicates, while red dots represent the average p_I of 5% of windows as ordered by rank of exon density. Rank was randomly assigned for ties. The horizontal black line represents the initial p_I of 5%. Spearman's ρ is computed for the relationship between recombination rate and p_I in each window and p-values indicate the significance of H_1: $\rho \neq 0$. The model numbers refer to the models shown in **Figure 1**.

Deleterious mutations impact the length of introgression deserts

Using these same simulations, we examined how selection on deleterious variation after admixture might influence the distribution of introgression deserts, or long stretches of the genome of the recipient population devoid of introgressed ancestry (**Figure S6**). When subpopulation fitnesses are expected to be the same (Model 0), the distribution of introgression deserts for models with deleterious mutations is similar to a neutral model, suggesting that

selection does not appreciably impact the distribution of deserts. When introgression-derived ancestry is expected to be deleterious (Model 2), simulations with additive fitness are enriched for longer ancestry deserts, though only slightly so. If instead introgression-derived ancestry is less deleterious than ancestry in the recipient population (Model 4), the length distribution of introgression deserts is shifted to be shorter, with the shortest introgression deserts occurring in models with recessive mutations (h=0) where both selection on load and heterosis act synergistically to increase the frequency of introgressed ancestry.

Introgression on the X chromosome

The observation that human X chromosomes are five-fold more resistant to introgression than the human autosome has been interpreted as a signature of genomic incompatibility between Neanderthals and humans, caused by an overrepresentation of male hybrid sterility genes on the X chromosome [23]. However, the evolution of the X chromosome differs from the autosomes in a number of important aspects, particularly in the strength of selection on deleterious variants [51], which may contribute to differences in patterns of introgression [17,18]. It is additionally unclear how selection on recessive variants might contribute, or counteract, the apparent resistance of the X chromosome to introgression.

To investigate the expected patterns of introgression on the X chromosome, we modeled X chromosome admixture with the simulation framework previously described. Although we used the same DFE for all these simulations, we utilized an analogous model of fitness that accounts for dosage compensation and the hemizygous sex [51,52]. Chromosome structure, recombination rates, and the DFE were the same as the simulations of human chromosome 1. See **Methods** for additional details on the calculation of fitness in these simulations.

Our simulations show that deleterious variation alone can result in significant differences between introgression on the X and the autosomes (**Figure 8**). When fitness is additive, stronger selection occurs on the X chromosome as deleterious variants are exposed in males.

This does not affect the X to autosome introgression ratio (X/A ratio) for Model 0, since both populations carry a similar burden of deleterious variants. For Model 2, selection removes introgressed ancestry from the X more quickly (X/A < 1), and for Model 4, selection increases the frequency of introgressed ancestry more than the autosome (X/A > 1). When fitness is recessive, the effect of heterosis is weaker for the X chromosome, since the hemizygous sex cannot be heterozygous. This effect also results in less observed introgression on the X than the autosome (X/A < 1) for all considered models. Finally, under the $h(s)$ relationship, our models predict amounts of introgression that are intermediate between strictly additive or strictly recessive models.

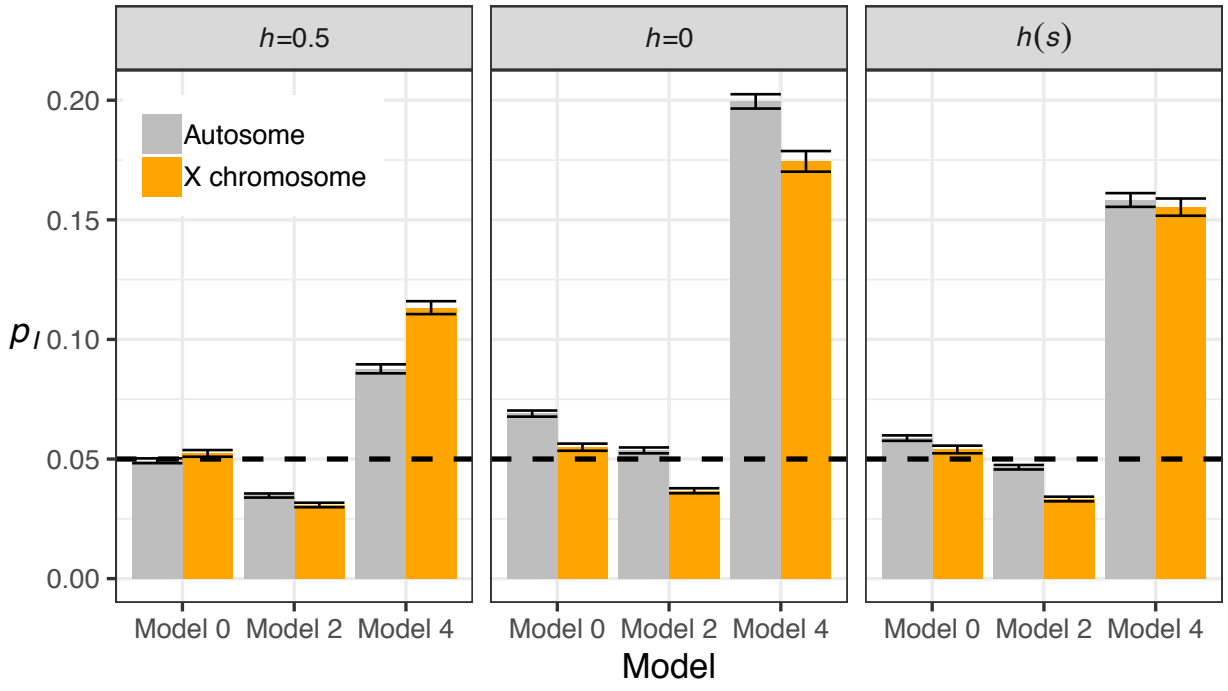

Figure 8. Differences in introgression between the X chromosome and autosomes. The average frequency of introgression-derived ancestry across the entire simulated chromosome (p_I) at time N_A (10,000) generations after admixture is shown for three demographic models and three models of fitness. Model numbers refer to the models shown in **Figure 1**. Bars represent

the mean p_I of 100 simulation replicates and error bars represent standard errors of the means. The horizontal dashed black line represents the initial p_I of 5%.

Arabidopsis genome structure results in a homogeneous landscape of introgression

Human-like demography and genomic parameters may not generalize well for the purpose of understanding introgression in other species. Functional density, recombination rates, effective population sizes, dominance, and the DFE can differ by an order of magnitude between species. To provide an alternative picture of how introgression dynamics are driven by deleterious variation in a natural system where dominance and selection have been estimated, we simulated Models 0, 2, and 4 using the genomic structure of *Arabidopsis thaliana*.

While the simulated demography was similar to the ones described previously, we used exon definitions and a recombination map of most (29.1 out of 30.4 Mb) of *A. thaliana* chromosome 1, and chromosome structure was fixed to be the same in all 100 simulation replicates. Both exon density and recombination rates are higher in *A. thaliana* (medians of 100kb windows 4.8×10^{-1} and 4.6×10^{-7}, respectively) than humans (medians of 100kb windows 1.6×10^{-2} and 8.04×10^{-9}, respectively). The ancestral population size was set to N_A=100,000 diploids, and the DFE to a gamma distribution with shape parameter 0.185 and E[s]= -0.0004866 [30]. We also assumed that dominance coefficients followed the $h(s)$ relationship estimated by that study and did not simulate scenarios with only additive or only recessive new mutations. To the best of our knowledge, this is the only estimate of the $h(s)$ relationship in a natural population other than humans. We split the simulated chromosome into non-overlapping 100kb windows and computed the frequency of introgression-derived ancestry, exon density, and the average recombination rate in each window.

The genomic landscape of introgression in our simulated *Arabidopsis* population varied little (**Figure 9**), even in a single simulation replicate of the same demographic model (**Figure S7**). For Model 0, introgressed ancestry rose quickly from an initial frequency of 5% to about 24%, N_A generations after admixture. There was little spatial variation in the frequency of introgression-derived ancestry. For example, p_I did not appear to be affected by the paucity of exons near the centromere (**Figure 9**). In Model 2, introgression-derived ancestry was quickly removed from the recipient subpopulation. This meant that p_I decreased to 0% across the whole chromosome. The converse was true for Model 4, where introgression-derived ancestry was favorable, and selection resulted in a complete replacement of recipient population ancestry (p_I=100%).

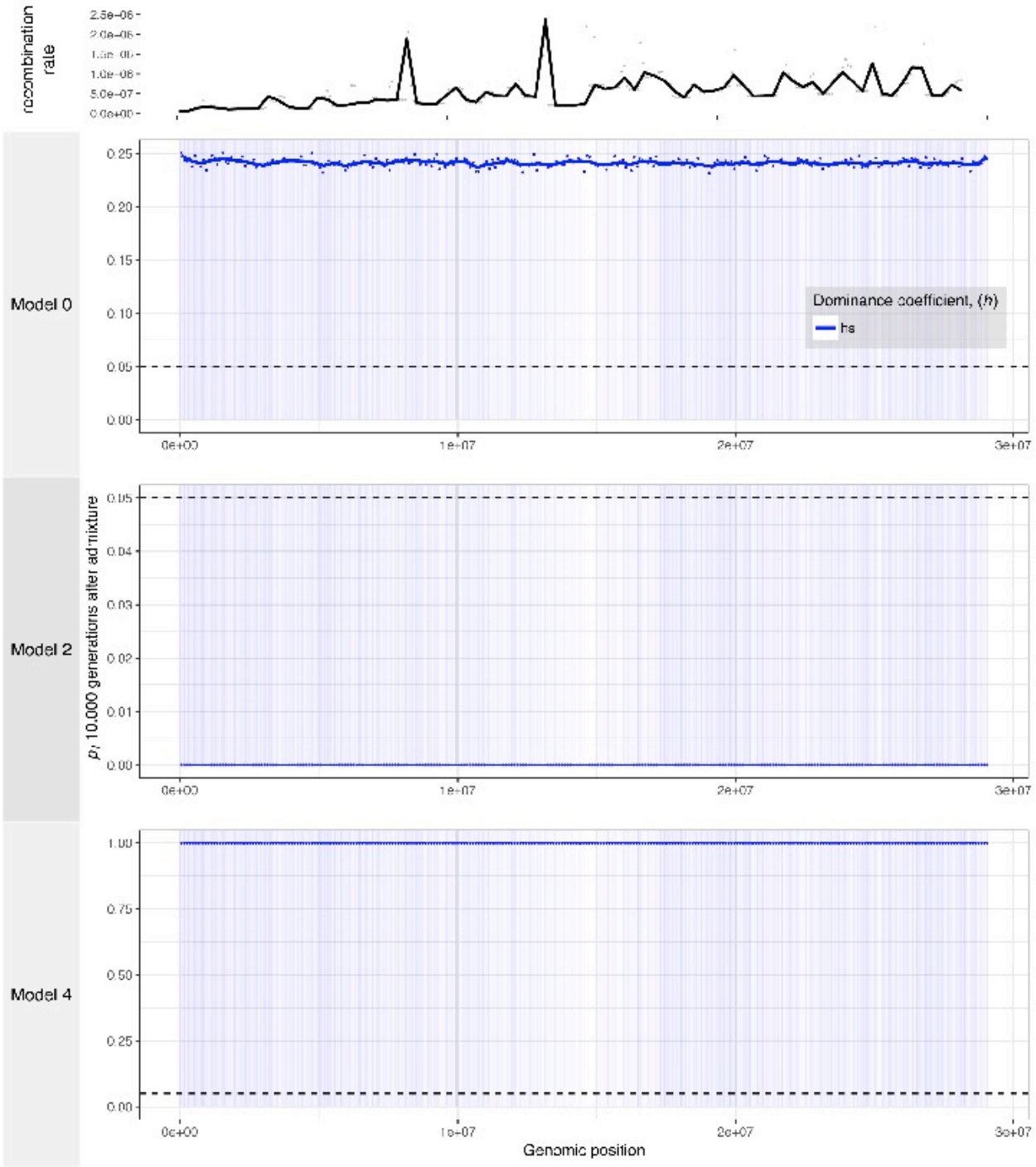

Figure 9. The average genomic landscape of introgression in simulations with *Arabidopsis* genomic structure. The frequency of ancestry that is introgression-derived is shown for non-overlapping 100 kb windows in a simulated 29.1 Mb region of chromosome 1. The model numbers refer to the models shown in **Figure 1**. Points represent a single value for each 100 kb

window and lines are loess curves fitted to the data. The horizontal black dashed line represents the initial frequency of introgression-derived ancestry, p_I=0.05. Vertical blue bars represent genes in which deleterious mutations can occur. Blue curves show the results for simulations with a $h(s)$ relationship.

Introgression is more likely in partially selfing populations than outcrossing populations

A notable life history feature distinguishing *Arabidopsis thaliana* from its congeners is the capability to self-fertilize [53]. Populations that are capable of self-fertilization may experience an overall reduced N_e leading to an accumulation of weakly deleterious variants relative to an outcrossing population, and increased levels of inbreeding depression. On the other hand, strongly deleterious recessive mutations should be purged in a selfing population [54,55]. Relative differences in the types of deleterious variation between groups with different mating systems may then initiate another kind of selective tug-of-war after admixture.

To investigate how deleterious mutations affect levels of introgression when admixture occurs between two populations with different mating systems, we simulated gene flow between a partially selfing and an outcrossing subpopulation using the same *A. thaliana* parameters as described in the previous section. We limited our simulated demographic model to Model 0 so that any differences in deleterious variation between subpopulations could be attributed to the mating system. Seven different gene flow scenarios were simulated, with selfing probabilities of 0%, 25%, 50%, and 75% in either subpopulation (**Figure 10**). Specifically, we simulated: first, with two outcrossing populations (0% to 0%); then with the outcrosser (0%) as the donor and the partial selfer (selfing probabilities of 25%, 50%, 75%) as the recipient, then the partial selfer (25%, 50%, 75%) as the donor and the outcrosser (0%) as the recipient. Self-incompatibility alleles were not simulated.

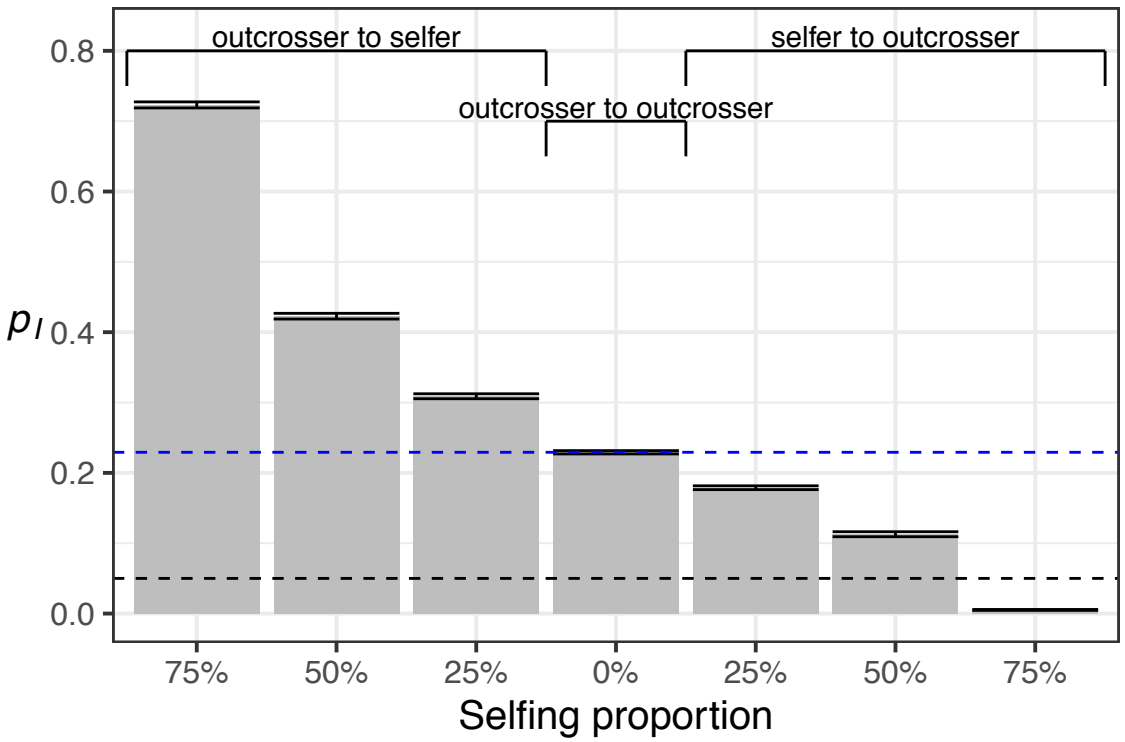

Figure 10. The impact of partial selfing on the frequency of introgression-derived ancestry in Model 0 (**Figure 1**) with *Arabidopsis* genomic structure. The frequency of introgression-derived ancestry (p_I) at time N_A (10,000) generations after admixture is plotted for seven different scenarios of admixture between a partially selfing population and an outcrossing population. Bar plots denote the average p_I of 100 simulation replicates and error bars represent standard errors of the averages. The horizontal dashed black line represents the initial p_I of 5%, and the horizontal dashed blue line represents the p_I that is expected between when both subpopulations are outcrossers. Labels on the x-axis denote the probability of selfing in the population that is partially selfing.

Our simulations show that the long-term frequency of introgression (10,000 generations after admixture) depends on the proportion of selfing individuals in the selfing subpopulation (**Figure 10**). In other words, selfing reduces N_e relative to an outcrosser, resulting in increased drift and a greater accumulation of deleterious mutations. These differences in load result in

patterns of introgression qualitatively similar to those observed previously in this study. In the simulations between two outcrossing populations, p_I increases from 5% to a long-term 20-25%, due to heterosis from the large proportion of recessive mutations predicted by the $h(s)$ relationship. This is the same result as the simulations of Model 0 in the previous section. When the outcrosser is the donor, p_I increases monotonically with the selfing probability of the recipient, this time above the fraction expected between two outcrossing populations. When the partially selfing population is the donor, long-term p_I usually increases by heterosis from the initial 5% value, although the long-term p_I monotonically decreases as the selfing probability increases. At a selfing probability of 75%, the outcrossing population is almost completely resistant to introgression. In the absence of fitness epistasis, it is likely that a combination of high recombination rates and strong initial selection from differences in deleterious mutations between populations counteracts any loss of donor ancestry from the purging of strongly deleterious recessive variants.

DISCUSSION

We have shown through simulations that deleterious variation can greatly influence the dynamics of introgression between admixing populations, in markedly different directions, magnitudes, and manners depending on the demographic model, mating system, models of selection, and genomic structure. In particular, the recombination rate is a key parameter that determines the way in which deleterious variants accumulate between populations and how selection acts on introgression-derived ancestry after admixture, ultimately determining the genomic landscape of introgression.

Our work demonstrates how demography can shape patterns of deleterious variation in different populations. Previous studies have examined the role of population size changes [1,8,12,56,57] and serial founder effect models [13,58] on deleterious variation. Interpreting how differences in the distribution of deleterious variation impact fitness has been a contentious

issue [4,5,7,9,11]. In this study, we observed that admixture can increase the fitness of the recipient population, sometimes drastically if the donor population is of larger long-term effective population size and thus carries lower genetic load. Generally, gene flow is observed to drive smaller, subtle changes in fitness. Nevertheless, the influx of new alleles can result in a rearrangement of deleterious variants in an admixed population (**Figures S2** and **S3**), and subtle changes to fitness can lead to significant shifts in the frequency of introgressed ancestry (e.g. see Model 0, *h*=0.0, in **Figure 3)**. These effects can be long lasting, persisting for thousands of generations in some of our simulations (**Figures 2, 3, S1**). If hybridization is a significant feature of a study population, studies concerning load should consider the fitness consequences of admixture as well as population size changes.

That dynamics of introgression-derived ancestry can be driven by deleterious variation is also important for the study of selection on gene flow between populations or species. Patterns of introgression between hybridizing species are often asymmetric, vary across the genome, and can be driven by demography at expansion fronts [59], dispersal processes [60], or by natural selection. However, when natural selection is implicated as driving changes in introgression-derived ancestry, processes such as genomic incompatibility or adaptive introgression are invoked to explain variation in introgression across the genome. We have shown that differences in demography and mating system create between-population differences in standing deleterious variation, and that selection upon these differences provides an alternative hypothesis to selection on alleles transplanted onto a new genomic background or new environment. To the best of our knowledge, only a few studies have considered the contribution of selection on deleterious variation to observed patterns of introgression [12,24,32], and mostly in specific systems [17,18,22,25].

Selection on deleterious variation may be particularly important for determining patterns of introgression in natural populations that are out of demographic equilibrium. Models of increased genetic drift predict accumulations of genetic load at the edges of expanding

populations [13,58] which suggests introgression into the expanding population could be driven by selection on genetic load. We have also shown that population bottlenecks can greatly affect patterns of introgression, particularly when assuming a recessive fitness model. If recessive deleterious variation also creates heterosis in admixed individuals, the effects of heterosis and population size will be synergistic, further enhancing introgression in genomic regions of low recombination. Our simulations also directly suggest heterosis may contribute to the pervasive patterns of introgression and shared polymorphism between different species in the genus *Arabidopsis* [35] even if hybridizing species have similar amounts of deleterious variation.

Because selection can alter patterns of introgression even if hybrid ancestry is not explicitly deleterious, genome-wide inferences of admixture proportions that assume neutrality are likely to be biased. For instance, our simulations predict the amount of introgression is strongly influenced by deleterious mutations in Arabidopsis, and the manner in which this occurs is dependent on the demography. Observed proportions of ancestry range from 0% for Model 2 to 100% for Model 4 (**Figures 9** and **S8**), despite the true admixture proportion of 5%. Taking the observed proportion of introgressed ancestry at face value, researchers would not infer the true initial admixture proportion of 5% accurately. Similarly, linkage disequilibrium patterns are often used to infer the timing of admixture events and to test competing demographic hypotheses about admixture [61]. If the distribution of segments of introgressed ancestry can be altered by deleterious mutations relative to what is predicted under a neutral model (e.g. Model 4 in **Figure S7**), these inferences can also be biased. To circumvent this problem, we recommend focusing on putatively neutral regions of the genome far from genes.

Likewise, our simulations may provide grounds for a plausible alternative explanation of the negative correlation between recombination rate and introgressed African ancestry observed in North American populations of *D. melanogaster* [43,44], which is the opposite of what is usually observed by other empirical studies of hybridization. Corbett-Detig and Nielsen [44] proposed that widespread adaptive introgression could bring along larger linkage blocks in

low recombination regions. If *D. melanogaster* has accumulated genetic load through the serial colonization of the world in association with humans [62,63], selection may favor introgression of the origin population (African) haplotypes in low recombination regions, similar to what we observed in Model 4 of our simulations. This could act synergistically with the effect of heterosis, which can happen in significant amounts even when divergence is low (**Figure 4**), and the divergence for which significant increases in introgressed ancestry are observed is comparable to that between populations of *D. melanogaster* [64]. Admittedly, our models bear little resemblance to the estimated demography of *D. melanogaster* (e.g. [63]). Similar to humans [8], there may be little difference in additive load between populations due to recent demography, and we have not simulated with a DFE and model of dominance estimated from *D. melanogaster*. Further study of these population genetic features is necessary to estimate the relative contribution of these processes to the genomic pattern of introgression in *D. melanogaster*.

Importantly, we do not claim that deleterious variation can explain all the patterns of introgression in any species, but rather that it is a plausible alternative explanation and therefore possible confounder that is important to consider when testing hypotheses about the nature of selection on gene flow. It is alternatively possible that colonizing populations of *D. melanogaster* experience a reduction in the rate of fixation of adaptive alleles due to reduced N_e, creating favorable conditions for the introgression of parent population haplotypes. Additionally, there is strong evidence for the role of sexual selection and fitness epistasis between the X and the autosomes in separating populations of *D. melanogaster* [65–67]. In hybridizing swordtail fish, recombination rates are positively correlated with the frequency of introgressed ancestry even when the minor parent population, analogous to the donor population in our simulations, has a larger effective population size [25]. This pattern suggests that hybrid ancestry has an overall deleterious effect, meaning that genomic incompatibility is the dominant force shaping hybrid genomes in that system. In humans, regions of high recombination rate are enriched for

introgressed Neanderthal ancestry particularly in genes that code for virus-interacting proteins [41], suggesting that in these regions putatively adaptive variants were more likely to recombine off the deleterious Neanderthal background and increase in frequency. In these two latter cases, selection on deleterious variation or heterosis may instead obscure genome-wide signals of incompatibility or adaptive introgression.

Because selection on additive and recessive variation can act in complementary or opposing directions, our study also highlights the fundamental importance of understanding the distribution of selection coefficients and their relationship to dominance coefficients in natural populations (i.e. the $h(s)$ relationship). In this study, we simulated human genomic structure, where new mutations are more likely to have additive fitness effects [13], and *Arabidopsis* genomic structure, where deleterious new mutations are likely to be more recessive [30]. In these two scenarios, we found that modes of dominance interacted with demography, recombination rates, and functional density in complex ways. Importantly, we observed an increase in introgressed ancestry as a result of the heterosis effect even when mutations were not completely recessive, that is, dominance was modeled with the $h(s)$ relationship. While the effects observed in the present study may be applicable to real populations with realistic amounts of dominance, the $h(s)$ relationship is unknown for virtually all natural systems. Therefore, we cannot easily predict the contribution of heterosis to introgression and shared polymorphism between closely related species.

Nevertheless, the underlying demographic model will determine how additive and recessive new mutations should interact after gene flow. For example, the introgression of deleterious haplotypes in Model 2 was facilitated by heterosis but impeded by additive load, leading to uncertainty about the overall contribution of the effects of deleterious variation in certain scenarios, such as Neanderthal to human admixture [17]. In other demographic models, selection on additive and recessive variants should operate in the same direction. As another example, if admixture occurs between a partially selfing and outcrossing population, our

simulations predict that selection works to remove ancestry from the selfing population, since it carries an overall larger burden of deleterious variants. It may yet be possible that strongly deleterious recessive variants, which should be purged in the selfer, play a role in preventing some introgression from the outcrossing to the selfing population. Without knowing the $h(s)$ relationship for a specific system, it is difficult to disentangle the effects of selection on additive versus recessive variation.

Our work further highlights the importance of considering deleterious variation when comparing complementary lines of evidence to make inferences about selection on hybrids. Even in the absence of fitness epistasis, our models predict an overall depletion of hybrid ancestry on the X chromosome compared to the autosomes. While the magnitude of this difference (about 1.5-fold) is far less than the 5-fold difference observed in humans [23], our results clearly show that simpler models of deleterious variation have the potential to mimic some of the signals that are considered evidence of hybrid incompatibility. Granted, we have only provided a simple model of selection on sex chromosomes to contrast to previous simulations of the autosomes, while ignoring the fact that recombination, chromosome structure, and the DFE are unlikely to be the same between the X and the autosomes. Additionally, it has been shown that sex-biased demographic processes have occurred throughout human history [68–72]. Future work should test the extent to which our results hold across more realistic population genetic models.

The recombination rate also plays a key role in determining the landscape of introgressed ancestry in the presence of deleterious variation. Models of Hill-Robertson interference [50,73] predict that deleterious mutations will not be removed as effectively in regions of the genome with low recombination rates because they may be linked to the non-deleterious alleles at other sites. We observe this effect in our simulations, where fitness declines the fastest when recombination rates are low, both pre- and post- admixture (**Figure S2**). However, we observe the opposite effect immediately after admixture. Specifically, in our

simulations, the fitness in the admixed population increased the most for the lowest recombination rates, suggesting that deleterious mutations were most effectively eliminated when recombination rates were the lowest (**Figure S2**). This occurs because selection for a haplotype will be most effective when all alleles on a haplotype tend to have weak fitness effects in the same direction [17,18,67]. For example, if introgression-derived ancestry carries fewer deleterious variants than the other haplotypes in the recipient population, selection will act to increase the frequency of the protective alleles contained within the introgressed ancestry. This applies directly to our simulations of admixture since immediately following an admixture event, all the protective or deleterious variants are found on the same haplotype. Higher rates of recombination will shuffle selected variants onto different haplotypes, creating selective interference between recombinant haplotypes.

One significant limitation of our study is that we have not considered all possible combinations of demographic, selective, and genomic parameters relevant for all species. For example, heterosis appears to stabilize long-term patterns of introgression at some frequency, but we only simulated an admixture fraction of 5%. It is possible that the magnitude or direction of observed changes may change with different major and minor parent ancestry proportions. It is therefore difficult to directly assess whether the specific conclusions seen for one combination of parameters will directly apply in a different specific system. Instead, our goal is to highlight the need to consider deleterious variation as a possible null model that should be investigated and rejected before attributing unusual patterns of introgressed ancestry to other evolutionary processes. That being said, we have observed some commonalities across models. For example, in Model 4, when mutations are either fully recessive or have an intermediate dominance coefficient assigned as a function of the selection coefficient, we observe an increase in introgressed ancestry in the recipient populations when either using simple models (**Figure 3**), models relevant for human populations (**Figure 5**) or models relevant for *A. thaliana* (**Figure 6**).

This interplay between deleterious variation and recombination has substantial implications for detecting adaptive introgression. A major objective of genomic studies of hybridization is to identify loci that are adaptively introgressed and to ascertain the overall importance of introgression to adaptive evolution [37]. Genomic regions that contain introgressed haplotypes at high frequency are considered likely candidates for adaptive introgression [37,40,74,75], but we have shown that selection on genetic load can increase the frequency of introgression-derived ancestry, even in the absence of beneficial new mutations. Thus, outlier-based approaches that compare summary statistics computed for a particular window of the genome to a null distribution that does not account for deleterious variation may be misled. Linked deleterious variants may also impede positive selection on introgressed adaptive variants, particularly if they are recessive [76]. Because recombination can move an adaptive variant off of ancestry backgrounds of varying fitness, standard models of adaptive evolution, especially ones that do not consider deleterious variation, are unlikely to accurately describe genomic patterns generated by adaptive introgression. Finally, it may be difficult to differentiate heterosis due to the masking of deleterious recessive alleles from heterozygote advantage at introgressed loci, despite the fact that these are two very different evolutionary processes with dramatically different biological interpretations.

Our results argue that new null models are needed in studies seeking to identify candidates of adaptive introgression. These new null models should include deleterious genetic variation, as well as complex demography. In order for these models to accurately capture the dynamics of deleterious variation, they should also include realistic parameters for the DFE of deleterious mutations and the relationship between dominance coefficients and selection coefficients. Lastly, the new null models should also include realistic models of the variation in recombination rate across the genome, as recombination rate is a key determinant of the dynamics of introgression (**Figure 3**). Failure to consider deleterious variation in a realistic way

in studies of admixing populations or hybridizing species can mislead inferences about the evolution of hybrids.

MATERIALS AND METHODS

Simulation details

All simulations were performed with SLiM 3.0 [46]. We chose to discard from our simulations, and therefore from calculations of fitness, mutations that were fixed in the ancestral or both subpopulations. Although fixed deleterious variants contribute to the overall genetic load of finite populations, they will have no effect on the relative differences between admixing subpopulations and no effect on the dynamics of introgression-derived ancestry. Therefore, each fitness calculation does not reflect the true fitness of each population, but rather the fitness components that are relevant during gene flow.

An admixture event in SLiM is handled by modifying the way the parents in each generation are chosen (SLiM manual 5.2.1). For example, at an admixture proportion of 5% the recipient population reproduces as follows. Five percent of the parents of the recipient population, in that generation, are chosen from the donor population, and 95% of the parents are chosen from the recipient population.

Scaling of forward simulations

We rescaled simulation parameters by a scaling constant, c, to reduce the computational burden of forward simulations. Population sizes were scaled to be N/c, times to t/c, selection coefficients to sc, and the mutation rate to μc. Recombination rates were scaled as $0.5(1-(1-2^r)^c)$, which is approximately rc for small r and small c. The total length of simulated sequence was not changed in scaled simulations. Note, the simulation parameters we reference in this paper are always unscaled. The manner in which we scaled simulations follows Algorithm 1 in Uricchio and Hernandez [77] and is similar to how Lange and Pool [78] simulated populations of

Drosophila melanogaster, although the primary features of interest in our simulations are related to the dynamics of introgression-derived ancestry through time.

Because scaled simulations may not exactly recapitulate the dynamics of unscaled simulations, we used a set of test simulations to choose c=5 for most simulations. The dynamics of p_I for scaled simulations (c=2, 5, and 10) were compared to an unscaled simulation (c=1), using the demography of Model 4, a gamma DFE, and additive fitness (h=0.5). Per base pair recombination rates of r=10^{-7} and 10^{-8} were simulated separately. Although all scaled simulations exhibit slight differences from the unscaled simulations, a scaling factor of c=5 provided a reasonably accurate representation of the unscaled dynamics of p_I (**Figure S8**) while keeping simulation run times within reasonable limits. We additionally note that our intent in this study is to understand qualitative patterns of introgression rather than to obtain accurate qualitative estimates from a particular system, and the qualitative patterns are consistent irrespective of the scaling factor.

Tracking introgression

The proportion of admixture that is introgression-derived (p_I) was tracked in one of two ways: by placing marker mutations at a fixed interval or by tracking the tree sequences (genealogies) across the simulated genome. In the former case, p_I was estimated by placing marker mutations in the donor population immediately before the admixture event. These mutations were spaced at 500 base pair intervals over the genome of every individual. After admixture, p_I was estimated in the recipient population by taking the averaged allele frequency of marker mutations per window, or throughout the whole simulated chromosome. In the latter case, the true ancestry proportions were calculated, since the information on start/end coordinates and the lineages that trace their ancestry back through donor and recipient populations is preserved. Although tracking tree sequences provides the most accurate

estimate of p_I, marker mutation tracking was used for computational efficiency in some simulations.

Simulations with randomly generated chromosomal structure

The sequences from simulations with randomly generated chromosome structure were approximately 5Mb in length, and contained intergenic, intronic, and exonic regions, but only nonsynonymous new mutations experienced natural selection. The per base pair mutation rate was constant and set to $\mu=1.5\times10^{-8}$ and we set nonsynonymous and synonymous mutations to occur at a ratio of 2.31:1 [79]. The selection coefficients (s) of new nonsynonymous mutations were drawn from a gamma-distributed DFE with shape parameter 0.186 and expected selection coefficient E[s] = -0.01314833 [49] for both additive and recessive dominance models.

The chromosomal structure of each simulation was randomly generated by drawing exon lengths from $Lognormal(\mu = log(50), \sigma^2 = log(2))$, intron lengths from $Lognormal(\mu = log(100), \sigma^2 = log(1.5))$, and the length of noncoding regions from $Unif(100,5000)$, following the specification in the SLiM 3.0 manual (7.3), which is modeled after the distribution of intron and exon lengths in Deutsch and Long [80]. The per base pair per chromosome recombination rate (r) was fixed in each simulation, but we varied r between different sets of simulations where $r\in\{10^{-6},10^{-7},10^{-8},10^{-9}\}$. Lastly, we simulated 200 replicates for each set of simulations, of each specific h and r.

Chromosome-wide F_{ST} was calculated for all variants from exons, introns, and intergenic regions by calculating F_{ST} at individual sites following Hudson et al. [81] and by combining F_{ST} across sites following Bhatia et al. [82].

Simulations of human genomic structure

In simulations of fixed chromosome structure from the human genome, we fixed the structure to 100 Mb randomly chosen from human genome build GRCh37, chromosome 1

(chr1:5,005,669-105,005,669). The exon ranges were defined by the GENCODE v14 annotations [83] and the sex-averaged recombination map by Kong et al. [84], averaged over a 10 kb scale. The per base pair mutation rate was constant and set to $\mu=1.5\times10^{-8}$ and we set nonsynonymous and synonymous mutations to occur at a ratio of 2.31:1 [79]. The selection coefficients (s) of new nonsynonymous mutations were drawn from a gamma-distributed DFE with shape parameter 0.186 and expected selection coefficient E[s] = -0.01314833 [49] for all models of dominance. All other new mutations were neutral. We simulated additive fitness (h=0.5), recessive fitness (h=0), and the $h(s)=0.5/(1-7071.07s)$ relationship [13] separately, using the same DFE for s for each simulation. All simulations were scaled by a factor of c=5.

Simulations of Arabidopsis genomic structure

In simulations of fixed chromosome structure from the genome of *Arabidopsis thaliana*, we fixed the structure to 29.1 Mb from chromosome 1 (chr1:488,426-29,588,426). The exon ranges were defined using the Araport11 annotations [85] and the recombination map using Salomé et al. [86]. The per base pair mutation rate was constant and set to $\mu=7\times10^{-9}$ and we again set nonsynonymous and synonymous mutations to occur at a ratio of 2.31:1. The selection coefficients (s) of new nonsynonymous mutations were drawn from the gamma distribution estimated by Huber et al. [30] (shape parameter 0.185 and E[s] = -0.00048655). We simulated dominance with the $h(s)$ relationship estimated by that study: $h(s)=1((1/0.987)-39547s)$. Simulations were scaled at c=100, but we note that we could not test the difference between c=100 and smaller scaling factors (e.g. c=50) due to limits on computational time.

Avoiding heterosis in the additive fitness model

Computing fitness as additive (h=0.5) within a locus but multiplicative across loci was problematic for our simulations because it created heterosis in admixed individuals. This occurred because the product of a fitness decrease reduces fitness less than the sum of a

fitness decrease. As a simple example, imagine two additive deleterious alleles are in a single individual, each with selection coefficient *s* where *s* is the absolute value of the selection coefficient. If they are found as a single homozygous site, the fitness decrease is usually computed as 1-*s*. If they are found as two heterozygous sites, the fitness would be computed as $(1\text{-}0.5s)^2=1\text{-}s+0.25s^2$. The fitness of the heterozygous individual is larger than the homozygous individual by $0.25s^2$, despite carrying the same number of deleterious variants. Because admixed individuals are more likely to carry deleterious alleles as heterozygotes than non-admixed individuals, the fitness of the admixed individuals was always higher than a non-admixed individual in the above computation of fitness, even when the number of deleterious variants per individual was the same.

Our intent was to examine the contribution of genetic load to selection on introgressed ancestry, but we identified an inherent advantage of heterozygosity in the additive model that biased the direction of selection to favor introgressed ancestry. To address this, we computed heterozygote fitness at a locus as 1-*hs* and homozygote fitness as $(1\text{-}0.5s)^2$, and the fitness across loci was computed multiplicatively. In the additive case (h=0.5), an individual's fitness was then multiplicative across all deleterious variants, such that an individual *j* carrying *i* variants each with selection coefficient s_i had fitness w_i:

$$w_j = \prod_i (1 + 0.5s_i)$$

Fitness is then monotonically related to the number of deleterious variants regardless of zygosity and is approximately equivalent to additive fitness. This computation in essence created a slight underdominance-like effect, but importantly this effect was caused by the difference in homozygous fitness rather than a difference in heterozygote fitness (*i.e.* the dominance coefficient). In practice, the difference in homozygous fitness is negligible for weakly deleterious alleles and strongly deleterious alleles are unlikely to be found as homozygotes. Therefore, the overall underdominance effect should be minimal. To confirm this, we simulated

100Mb of human chromosome 1 in an equilibrium population, with selection coefficients drawn from a gamma DFE with the two fitness models. The frequency spectrum was unaffected by our calculation of fitness (**Figure S9**), suggesting our simulations approximate the standard additive model well.

We used the same calculation for additive and partially recessive fitness models for consistency when simulating the $h(s)$ relationship. Completely recessive fitness (h=0) was computed the standard way, that is, as $1-s_i$ when homozygous for the deleterious allele and as 1 otherwise.

Selection on the X chromosome

We modeled fitness of the sex chromosomes following the framework described by Charlesworth et al. [52] and Veeramah et al. [51], with a slight modification to preserve the multiplicative fitness scenario described for the autosome. The specific fitness models for each dominance scenario – additive, recessive, and with the $h(s)$ function – are presented in **Table S2**. Importantly, the fitness of females that are homozygous and males that have the selected allele are the same, and, in the additive model, heterozygous females have an intermediate fitness. This models dosage compensation in females, assuming levels of gene expression map to the same fitness values for males and females.

Data availability

All scripts necessary for reproducing the simulations, parsing the simulation output, and creating the figures displayed in this manuscript are available at: https://github.com/LohmuellerLab/admixture_load_scripts.

Supplemental Results

Table S1. Demographic parameters of the simulated models shown in Figure 1.

Model	N_S	N_{RB}	N_R	T_m	T_{RB}	T_{split}
Model 0	1	--	1	0.5	--	1.5
Model 1	1	0.1	1	0.5	0.5025	1.5
Model 2	0.1	--	1	0.5	--	1.5
Model 3	1	--	0.1	0.5	--	1.5
Model 4	1	0.1	1	0.5	--	1.5

NOTE.—All population sizes are relative to the ancestral population size (N_A), where N_A=10,000 diploids unless specified otherwise. All times are in units of generations/($2N_A$) from the present day. Parameters are defined as follows. N_S: size of the source subpopulation, N_{RB}: size of the bottleneck in the recipient population, N_R: size of the recipient population, T_m: time of migration, T_{RB}: time at which the bottleneck in the recipient population began, T_{split}: time at which the subpopulations diverged

Table S2. Fitness model for sex chromosomes.

	genotype		
dominance	XX heterozygote	XX homozygote	XY
Additive (h=0.5)	$(1+hs)$	$(1+hs)^2$	$(1+hs)^2$
Recessive	1	$(1+0.5s)^2$	$(1+0.5s)^2$
$h(s)$	$(1+hs)$	$(1+0.5s)^2$	$(1+0.5s)^2$

NOTE.—h denotes the dominance coefficient and $(1+0.5s)^2$ is approximately equal to $(1-s)$.

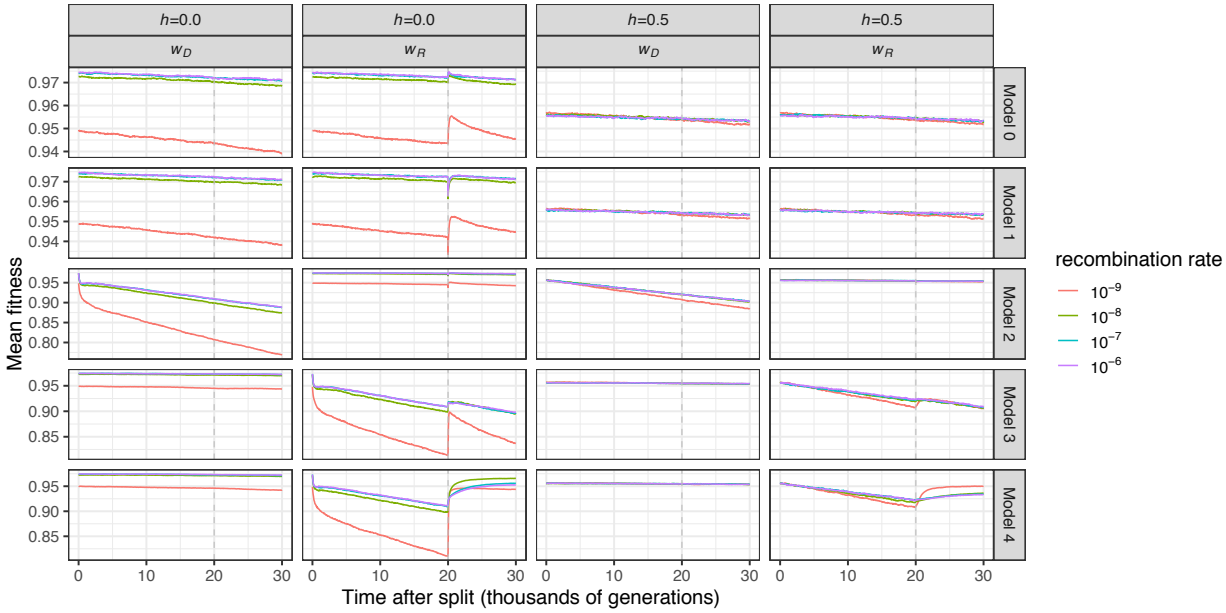

Figure S1. The change in the mean fitness of the donor and recipient subpopulation in each model. The mean (solid line) is shown for 200 simulation replicates. The vertical grey line depicts the time of gene flow. Different colors denote distinct recombination rates used in the simulations. The left two panels depict simulations with recessive mutations (h=0) while the right two panels show simulations with additive mutations (h=0.5). Variants that are fixed in both subpopulations are not considered in the calculation of fitness. The model numbers refer to the models shown in **Figure 1**.

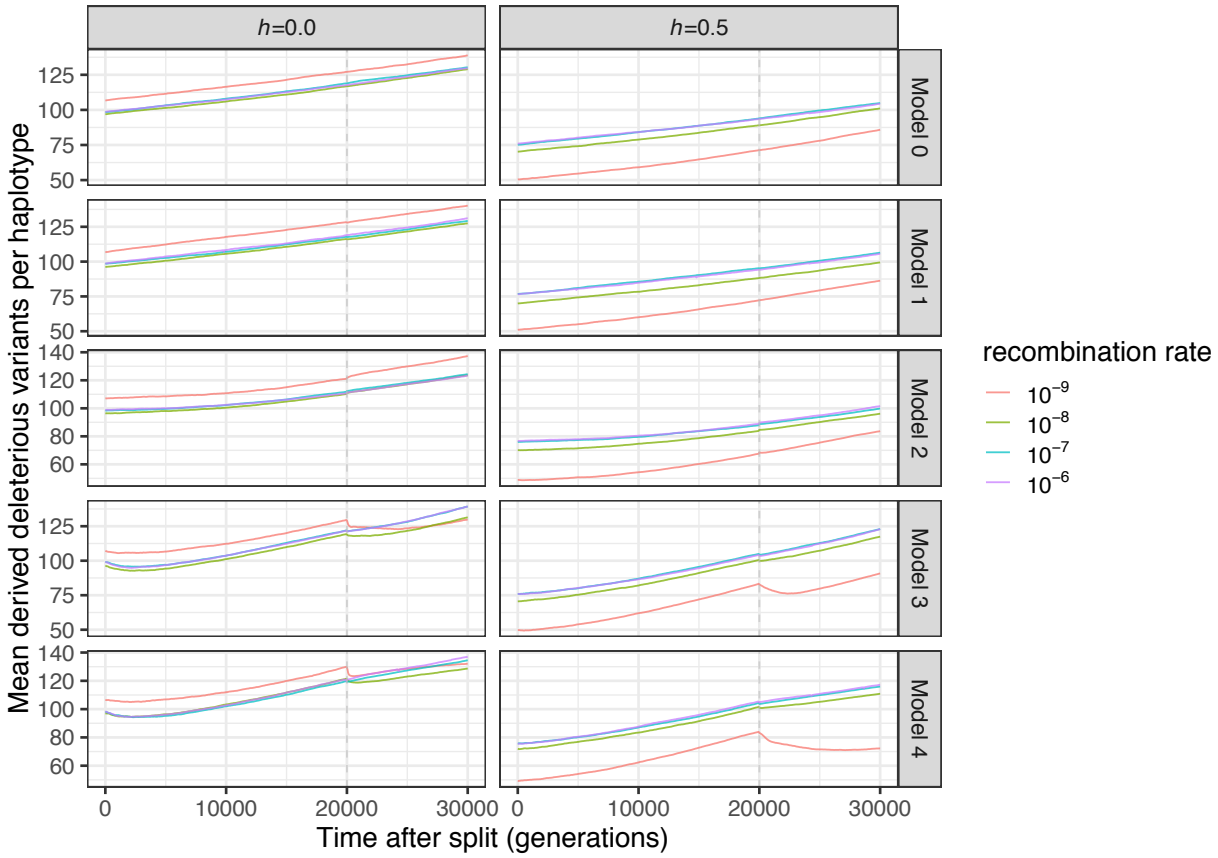

Figure S2. The change in the mean number of derived deleterious sites (s<0) per haplotype in each model in the recipient subpopulation. The mean (solid line) is shown for 200 simulation replicates. The vertical gray line depicts the time of gene flow. Different colors denote distinct recombination rates used in the simulations. The left panel shows simulations with recessive mutations (h=0) while the right panel shows simulations with additive mutations (h=0.5). Variants that are fixed in both subpopulations are not counted. The model numbers refer to the models shown in **Figure 1**.

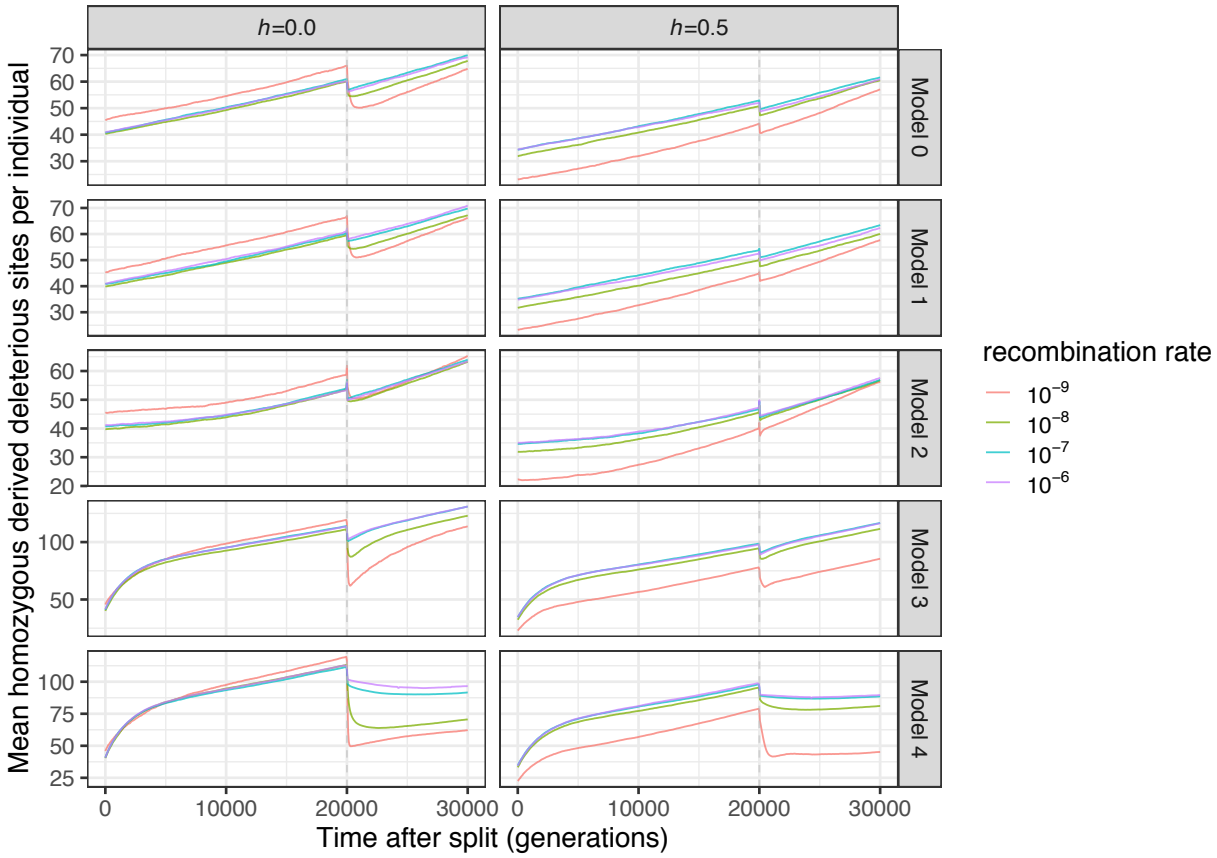

Figure S3. The change in the mean number of homozygous derived deleterious sites per individual in the recipient subpopulation. The mean (solid line) is shown for 200 simulation replicates. The vertical gray line depicts the time of gene flow. Different colors denote distinct recombination rates used in the simulations. The left panel shows simulations with recessive mutations (h=0) while the right panel shows simulations with additive mutations (h=0.5). Variants that are fixed in both subpopulations are not counted. The model numbers refer to the models shown in **Figure 1**.

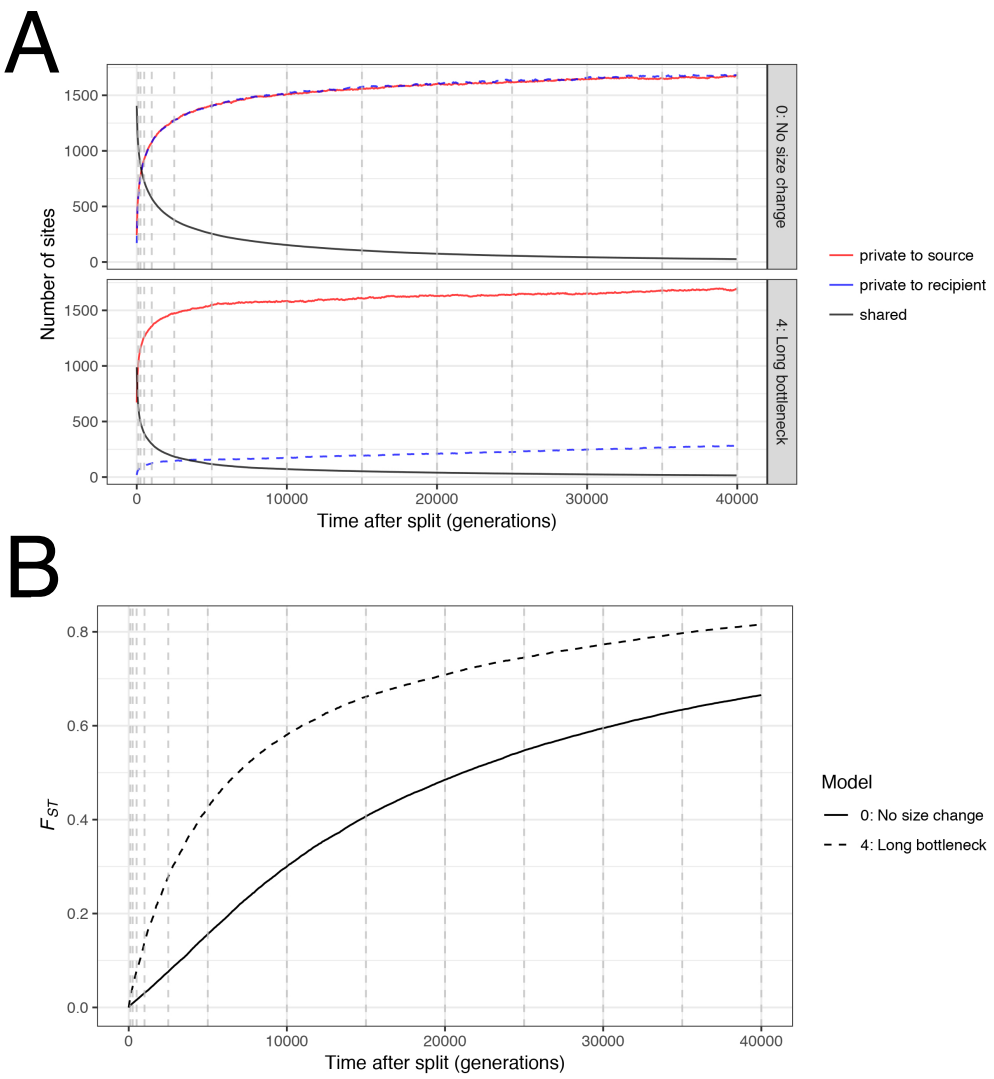

Figure S4. The relationship of the split time to measures of divergence between subpopulations in Models 0 and 4 (refer to **Figure 1**). The vertical gray lines represent the time between population divergence and admixture (100, 250, 500, 1,000, 2,500, 5,000, 10,000, 20,000, 25,000, 30,000, 35,000, and 40,000 generations) in the demographic models as depicted in **Figure 4**. (**A**) Population split time and population size impact the number of sites private to each subpopulation at the time of admixture. The numbers of sites that are private to the donor and recipient subpopulations, or shared between subpopulations, are shown for 200 simulation replicates and two demographic models. (**B**) F_{ST} increases continuously in Models 0 and 4 after the split. Increased drift in Model 4 drives larger increases in F_{ST}.

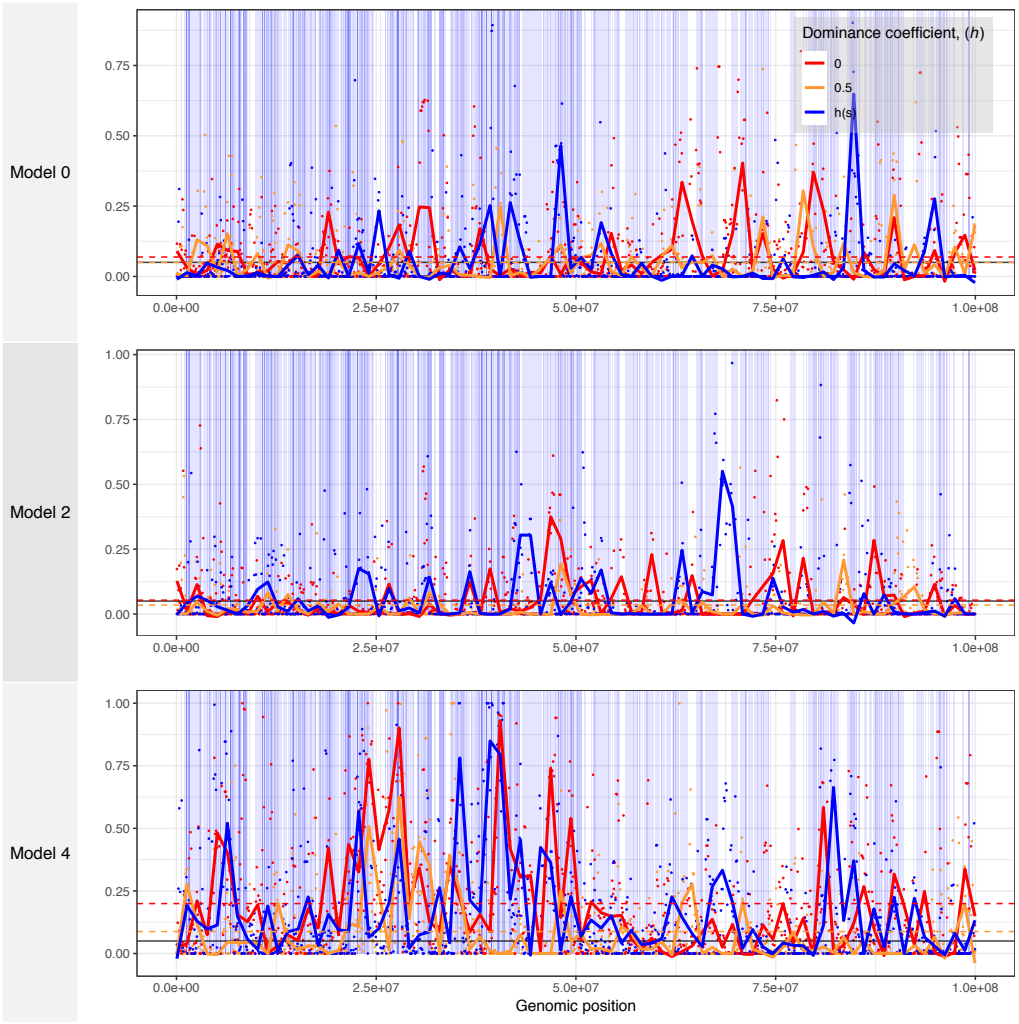

Figure S5. The genomic landscape of introgression in one replicate of simulations with human genomic structure. The frequency of ancestry that is introgression-derived is shown for non-overlapping 100 kb windows in a simulated 100 Mb region of chromosome 1. The model numbers refer to the models shown in **Figure 1**. Points represent a single value for each 100 kb window and lines are loess curves fitted to the data. The horizontal dashed black dashed line represents the initial frequency of introgression-derived ancestry, p_I=0.05. Vertical blue bars represent genes in which deleterious mutations can occur. Red curves denote the results for recessive mutations, orange curves show the results for additive mutations, and blue curves show the results for simulations with a $h(s)$ relationship.

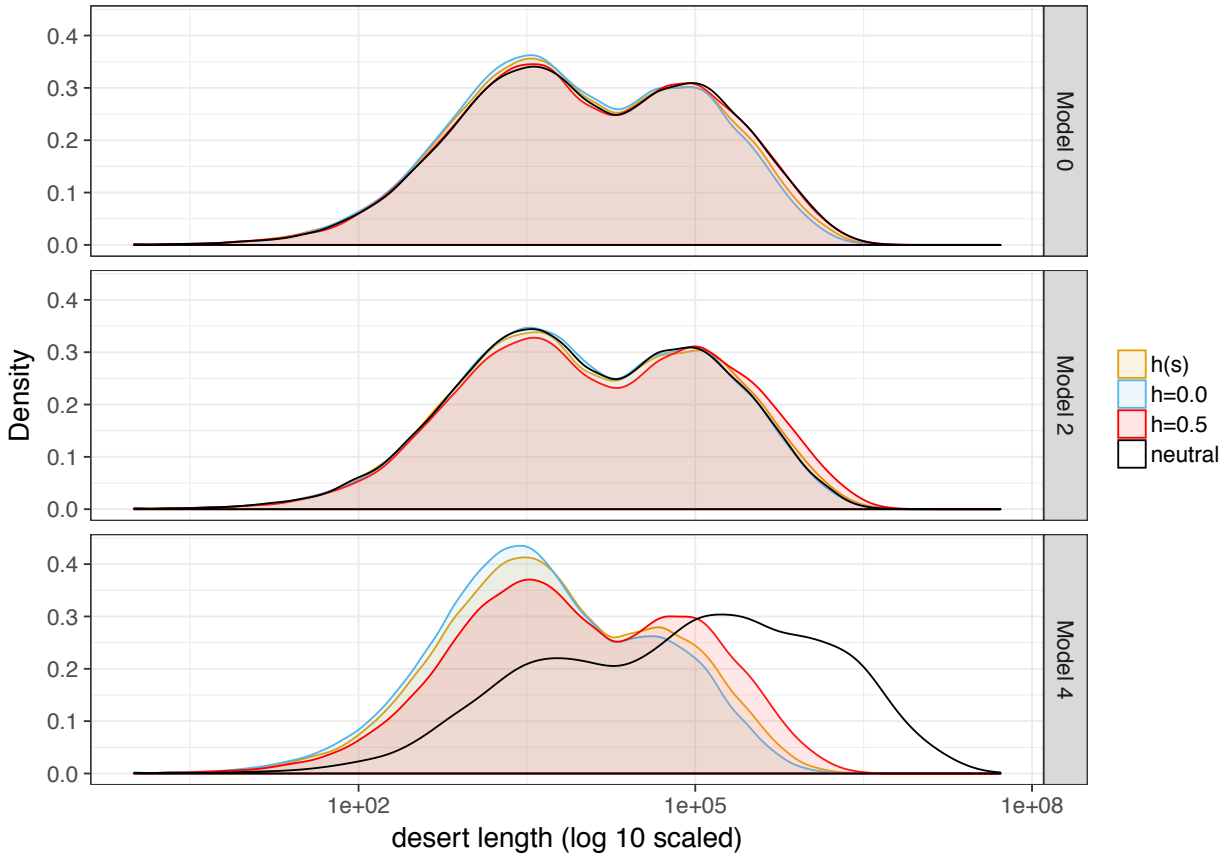

Figure S6. The density distribution of introgression desert lengths for simulations of human chromosome 1. Introgression deserts are segments without any hybrid ancestry. Model numbers refer to the models shown in **Figure 1**.

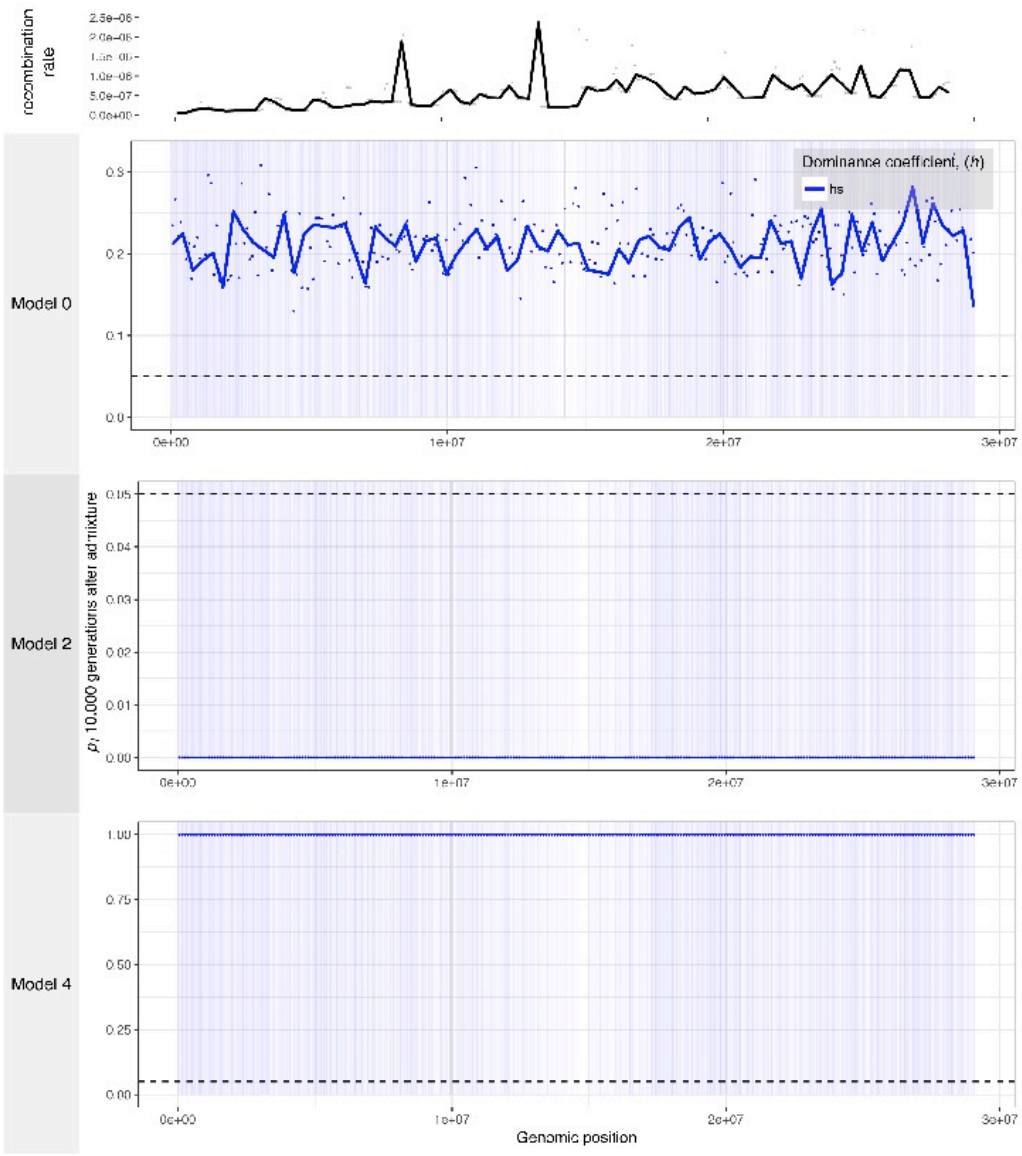

Figure S7. The genomic landscape of introgression in one replicate of simulations with *Arabidopsis* genomic structure. The frequency of ancestry that is introgression-derived is shown for non-overlapping 100 kb windows in a simulated 100 Mb region of chromosome 1. The model numbers refer to the models shown in **Figure 1**. Points represent a single value for each 100 kb window and lines are loess curves fitted to the data. The horizontal dashed black line represents the initial frequency of introgression-derived ancestry, p_I=0.05. Vertical blue bars represent genes in which deleterious mutations can occur. Blue curves show the results for simulations with a $h(s)$ relationship.

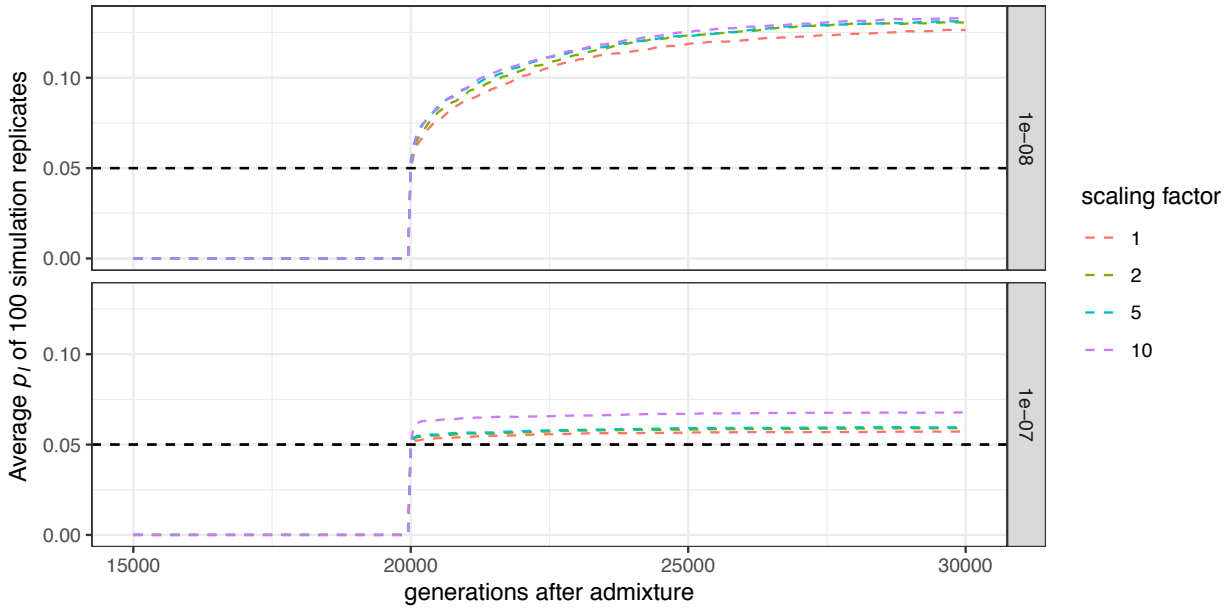

Figure S8. Scaled simulations accurately reproduce introgression in simulations with no scaling. The average frequency of introgressed ancestry (p_I) of 100 simulation replicates of Model 4 and additive fitness (h=0.5) is plotted through time. The average p_I for four different scaling factors (c=1, 2, 5, and 10) is shown. The simulations in this study use c=5 unless mentioned otherwise. Details are provided in the **Methods**.

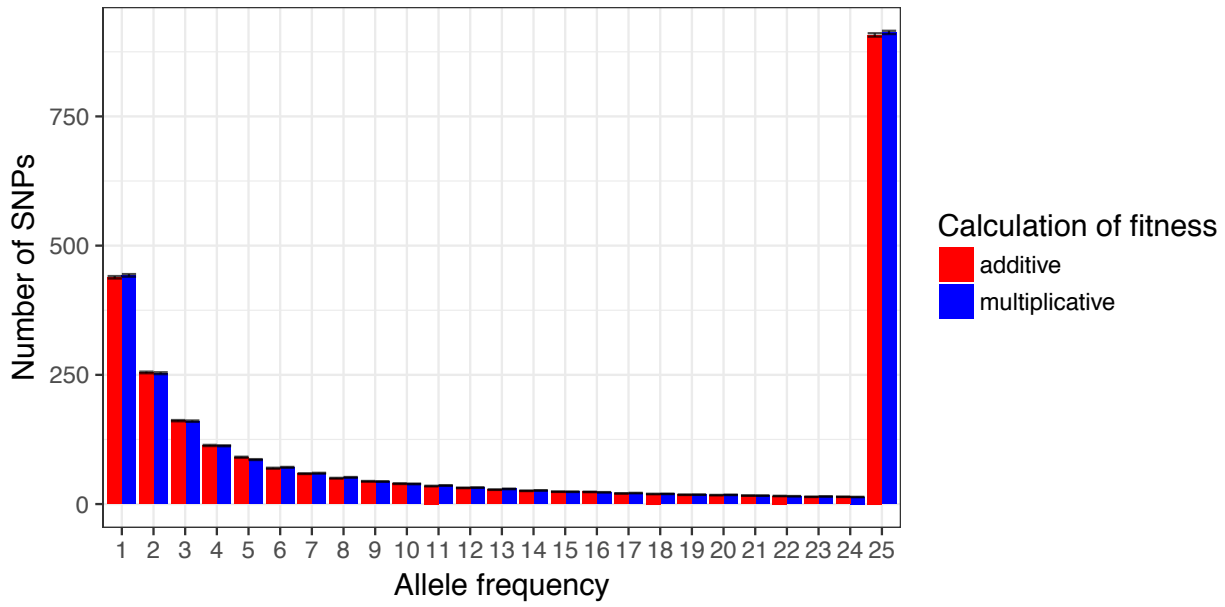

Figure S9. The site frequency spectrum of nonsynonymous variants is the same when fitness is calculated as multiplicative within a locus as it is when fitness is additive within a locus. Simulations were of an equilibrium population with 100 Mb of human genomic structure in a sample of size *n*=2,000 chromosomes. Confidence intervals represent standard errors computed from 100 simulation replicates. All variants at frequency ≥25 are summed together in the last entry of the SFS.

REFERENCES

1. Kirkpatrick M, Jarne P. The effects of a bottleneck on inbreeding depression and the genetic load. Am Nat. 2000;155: 154–167. doi:10.1086/303312

2. Gazave E, Chang D, Clark AG, Keinan A. Population growth inflates the per-individual number of deleterious mutations and reduces their mean effect. Genetics. 2013;195: 969–978. doi:10.1534/genetics.113.153973

3. Lohmueller KE. The impact of population demography and selection on the genetic architecture of complex traits. PLOS Genet. 2014;10: e1004379. doi:10.1371/journal.pgen.1004379

4. Lohmueller KE. The distribution of deleterious genetic variation in human populations. Curr Opin Genet Dev. 2014;29: 139–146. doi:10.1016/j.gde.2014.09.005

5. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. Nat Rev Genet. 2015;16: 333–343. doi:10.1038/nrg3931

6. Brandvain Y, Wright SI. The limits of natural selection in a nonequilibrium world. Trends Genet. 2016;32: 201–210. doi:10.1016/j.tig.2016.01.004

7. Simons YB, Sella G. The impact of recent population history on the deleterious mutation load in humans and close evolutionary relatives. Curr Opin Genet Dev. 2016;41: 150–158. doi:10.1016/j.gde.2016.09.006

8. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. Nat Genet. 2014;46: 220–224. doi:10.1038/ng.2896

9. Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. Nat Genet. 2015;47: 126–131. doi:10.1038/ng.3186

10. Casals F, Hodgkinson A, Hussin J, Idaghdour Y, Bruat V, Maillard T de, et al. Whole-exome sequencing reveals a rapid change in the frequency of rare functional variants in a founding population of humans. PLOS Genet. 2013;9: e1003815. doi:10.1371/journal.pgen.1003815

11. Fu W, Gittelman RM, Bamshad MJ, Akey JM. Characteristics of Neutral and deleterious protein-coding variation among individuals and populations. Am J Hum Genet. 2014;95: 421–436. doi:10.1016/j.ajhg.2014.09.006

12. Gravel S. When is selection effective? Genetics. 2016;203: 451–462. doi:10.1534/genetics.115.184630

13. Henn BM, Botigué LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, et al. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc Natl Acad Sci. 2016;113: E440–E449. doi:10.1073/pnas.1510805112

14. Pedersen C-ET, Lohmueller KE, Grarup N, Bjerregaard P, Hansen T, Siegismund HR, et al. The effect of an extreme and prolonged population bottleneck on patterns of deleterious variation: insights from the Greenlandic Inuit. Genetics. 2017;205: 787–801. doi:10.1534/genetics.116.193821

15. Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. Mol Ecol. 2016;25: 2337–2360. doi:10.1111/mec.13557

16. Wall JD, Yoshihara Caldeira Brandt D. Archaic admixture in human history. Curr Opin Genet Dev. 2016;41: 93–97. doi:10.1016/j.gde.2016.07.002

17. Harris K, Nielsen R. The genetic cost of Neanderthal introgression. Genetics. 2016;203: 881–891. doi:10.1534/genetics.116.186890

18. Juric I, Aeschbacher S, Coop G. The strength of selection against Neanderthal introgression. PLOS Genet. 2016;12: e1006340. doi:10.1371/journal.pgen.1006340

19. Marsden CD, Vecchyo DO-D, O'Brien DP, Taylor JF, Ramirez O, Vilà C, et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. Proc Natl Acad Sci. 2016;113: 152–157. doi:10.1073/pnas.1512501113

20. Liu Q, Zhou Y, Morrell PL, Gaut BS. Deleterious variants in Asian rice and the potential cost of domestication. Mol Biol Evol. 2017;34: 908–924. doi:10.1093/molbev/msw296

21. Moyers BT, Morrell PL, McKay JK. Genetic costs of domestication and improvement. J Hered. 2017; doi:10.1093/jhered/esx069

22. Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford M. The interplay of demography and selection during maize domestication and expansion. bioRxiv. 2017; 114579. doi:10.1101/114579

23. Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, et al. The genomic landscape of Neanderthal ancestry in present-day humans. Nature. 2014;507: 354–357. doi:10.1038/nature12961

24. Aeschbacher S, Selby JP, Willis JH, Coop G. Population-genomic inference of the strength and timing of selection against gene flow. Proc Natl Acad Sci. 2017;114: 7061–7066. doi:10.1073/pnas.1616755114

25. Schumer M, Xu C, Powell DL, Durvasula A, Skov L, Holland C, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018; eaar3684. doi:10.1126/science.aar3684

26. Sankararaman S, Mallick S, Patterson N, Reich D. The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. Curr Biol. 2016;26: 1241–1247. doi:10.1016/j.cub.2016.03.037

27. Janoušek V, Munclinger P, Wang L, Teeter KC, Tucker PK. Functional organization of the genome may shape the species boundary in the house mouse. Mol Biol Evol. 2015;32: 1208–1220. doi:10.1093/molbev/msv011

28. Simmons MJ, Crow JF. Mutations affecting fitness in Drosophila populations. Annu Rev Genet. 1977;11: 49–78. doi:10.1146/annurev.ge.11.120177.000405

29. Agrawal AF, Whitlock MC. Inferences about the distribution of dominance drawn from yeast gene knockout data. Genetics. 2011;187: 553–566. doi:10.1534/genetics.110.124560

30. Huber CD, Durvasula A, Hancock AM, Lohmueller KE. Gene expression drives the evolution of dominance. Nat Commun. 2018;9: 2750. doi:10.1038/s41467-018-05281-7

31. Crow JF. Alternative hypotheses of hybrid vigor. Genetics. 1948;33: 477–487.

32. Ingvarsson PK, Whitlock MC. Heterosis increases the effective migration rate. Proc R Soc Lond B Biol Sci. 2000;267: 1321–1326. doi:10.1098/rspb.2000.1145

33. Bierne N, Lenormand T, Bonhomme F, David P. Deleterious mutations in a hybrid zone: can mutational load decrease the barrier to gene flow? Genet Res. 2002;80: 197–204. doi:10.1017/S001667230200592X

34. Whitlock MC, Ingvarsson PK, Hatfield T. Local drift load and the heterosis of interconnected populations. Heredity. 2000;84 (Pt 4): 452–457.

35. Novikova PY, Hohmann N, Nizhynska V, Tsuchimatsu T, Ali J, Muir G, et al. Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. Nat Genet. 2016;48: 1077–1082. doi:10.1038/ng.3617

36. Hedrick Philip W. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol. 2013;22: 4606–4618. doi:10.1111/mec.12415

37. Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E. Evidence for archaic adaptive introgression in humans. Nat Rev Genet. 2015;16: 359–371. doi:10.1038/nrg3936

38. Ding Q, Hu Y, Xu S, Wang C-C, Li H, Zhang R, et al. Neanderthal ORIGIN OF THE HAPLOTYPES CARRYING THE FUNCTIONAL VARIANT Val92Met in the *MC1R* in modern humans. Mol Biol Evol. 2014;31: 1994–2003. doi:10.1093/molbev/msu180

39. Ding Q, Hu Y, Xu S, Wang J, Jin L. Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in East Asians. Mol Biol Evol. 2014;31: 683–695. doi:10.1093/molbev/mst260

40. Huerta-Sánchez E, Jin X, Asan null, Bianba Z, Peter BM, Vinckenbosch N, et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 2014;512: 194–197. doi:10.1038/nature13408

41. Enard D, Petrov DA. RNA viruses drove adaptive introgressions between Neanderthals and modern humans. bioRxiv. 2017; 120477. doi:10.1101/120477

42. Abi-Rached L, Jobin MJ, Kulkarni S, McWhinnie A, Dalva K, Gragert L, et al. The shaping of modern human immune systems by multiregional admixture with archaic humans. Science. 2011;334: 89–94. doi:10.1126/science.1209202

43. Pool JE. The mosaic ancestry of the Drosophila genetic reference panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. Mol Biol Evol. 2015;32: 3236–3251. doi:10.1093/molbev/msv194

44. Corbett-Detig R, Nielsen R. A hidden Markov Model approach for simultaneously estimating local ancestry and admixture time using next generation sequence data in samples of arbitrary ploidy. PLOS Genet. 2017;13: e1006529. doi:10.1371/journal.pgen.1006529

45. Owens GL, Baute GJ, Rieseberg LH. Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. Mol Ecol. 2016;25: 2630–2643. doi:10.1111/mec.13569

46. Haller BC, Messer PW. SLiM 2: Flexible, interactive forward genetic simulations. Mol Biol Evol. 2017;34: 230–240. doi:10.1093/molbev/msw211

47. Kelleher J, Thornton K, Ashander J, Ralph P. Efficient pedigree recording for fast population genetics simulation. bioRxiv. 2018; 248500. doi:10.1101/248500

48. Kelleher J, Etheridge AM, McVean G. Efficient coalescent simulation and genealogical analysis for large sample sizes. PLOS Comput Biol. 2016;12: e1004842. doi:10.1371/journal.pcbi.1004842

49. Kim BY, Huber CD, Lohmueller KE. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics. 2017;206: 345–361. doi:10.1534/genetics.116.197145

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

51. Veeramah KR, Gutenkunst RN, Woerner AE, Watkins JC, Hammer MF. Evidence for increased levels of positive and negative selection on the X chromosome versus autosomes in humans. Mol Biol Evol. 2014;31: 2267–2282. doi:10.1093/molbev/msu166

52. Charlesworth B, Coyne JA, Barton NH. The relative rates of evolution of sex chromosomes and autosomes. Am Nat. 1987;130: 113–146.

53. Charlesworth D, Vekemans X. How and when did *Arabidopsis thaliana* become highly self-fertilising. BioEssays. 2005;27: 472–476. doi:10.1002/bies.20231

54. Lande R, Schemske DW. The evolution of self-fertilization and inbreeding depression in plants. I. genetic models. Evolution. 1985;39: 24–40. doi:10.2307/2408514

55. Charlesworth D, Charlesworth B. Inbreeding depression and its evolutionary consequences. Annu Rev Ecol Syst. 1987;18: 237–268. doi:10.1146/annurev.es.18.110187.001321

56. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, et al. Proportionally more deleterious genetic variation in European than in African populations. Nature. 2008;451: 994–997. doi:10.1038/nature06611

57. Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR. Dominance of deleterious alleles controls the response to a population bottleneck. PLOS Genet. 2015;11: e1005436. doi:10.1371/journal.pgen.1005436

58. Peischl S, Excoffier L. Expansion load: recessive mutations and the role of standing genetic variation. Mol Ecol. 2015;24: 2084–2094. doi:10.1111/mec.13154

59. Currat M, Ruedi M, Petit RJ, Excoffier L. The hidden side of invasions: massive introgression by local genes. Evol Int J Org Evol. 2008;62: 1908–1920. doi:10.1111/j.1558-5646.2008.00413.x

60. Amorim CEG, Hofer T, Ray N, Foll M, Ruiz-Linares A, Excoffier L. Long-distance dispersal suppresses introgression of local alleles during range expansions. Heredity. 2017;118: 135–142. doi:10.1038/hdy.2016.68

61. Sankararaman S, Patterson N, Li H, Pääbo S, Reich D. The date of interbreeding between Neandertals and modern humans. PLOS Genet. 2012;8: e1002947. doi:10.1371/journal.pgen.1002947

62. Stephan W, Li H. The recent demographic and adaptive history of *Drosophila melanogaster*. Heredity. 2006;98: 65–68. doi:10.1038/sj.hdy.6800901

63. Duchen P, Živković D, Hutter S, Stephan W, Laurent S. Demographic inference reveals African and European admixture in the North American *Drosophila melanogaster* population. Genetics. 2013;193: 291–301. doi:10.1534/genetics.112.145912

64. Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D. Population genomic analysis uncovers African and European admixture in *Drosophila melanogaster* populations from

the south-eastern United States and Caribbean Islands. Mol Ecol. 2015;24: 1499–1509. doi:10.1111/mec.13137

65. Ting C-T, Takahashi A, Wu C-I. Incipient speciation by sexual isolation in Drosophila: Concurrent evolution at multiple loci. Proc Natl Acad Sci. 2001;98: 6709–6713. doi:10.1073/pnas.121418898

66. Lachance J, True JR. X-autosome incompatibilities in *Drosophila melanogaster*: Tests of Haldane's rule and geographic patterns within species. Evolution. 2010;64: 3035–3046. doi:10.1111/j.1558-5646.2010.01028.x

67. Kao JY, Lymer S, Hwang SH, Sung A, Nuzhdin SV. Postmating reproductive barriers contribute to the incipient sexual isolation of the United States and Caribbean *Drosophila melanogaster.* Ecol Evol. 2015;5: 3171–3182. doi:10.1002/ece3.1596

68. Keinan A, Mullikin JC, Patterson N, Reich D. Accelerated genetic drift on chromosome X during the human dispersal out of Africa. Nat Genet. 2009;41: 66–70. doi:10.1038/ng.303

69. Hammer MF, Mendez FL, Cox MP, Woerner AE, Wall JD. Sex-biased evolutionary forces shape genomic patterns of human diversity. PLOS Genet. 2008;4: e1000202. doi:10.1371/journal.pgen.1000202

70. Emery LS, Felsenstein J, Akey JM. Estimators of the human effective sex ratio detect sex biases on different timescales. Am J Hum Genet. 2010;87: 848–856. doi:10.1016/j.ajhg.2010.10.021

71. Arbiza L, Gottipati S, Siepel A, Keinan A. Contrasting X-linked and autosomal diversity across 14 human populations. Am J Hum Genet. 2014;94: 827–844. doi:10.1016/j.ajhg.2014.04.011

72. Lohmueller KE, Degenhardt JD, Keinan A. Sex-averaged recombination and mutation rates on the X chromosome: A comment on Labuda et al. Am J Hum Genet. 2010;86: 978–980. doi:10.1016/j.ajhg.2010.03.021

73. Keightley PD, Otto SP. Interference among deleterious mutations favours sex and recombination in finite populations. Nature. 2006;443: 89–92. doi:10.1038/nature05049

74. Gittelman RM, Schraiber JG, Vernot B, Mikacenic C, Wurfel MM, Akey JM. Archaic hominin admixture facilitated adaptation to out-of-Africa environments. Curr Biol. 2016;26: 3375–3382. doi:10.1016/j.cub.2016.10.041

75. Richards EJ, Martin CH. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. PLOS Genet. 2017;13: e1006919. doi:10.1371/journal.pgen.1006919

76. Assaf ZJ, Petrov DA, Blundell JR. Obstruction of adaptation in diploids by recessive, strongly deleterious alleles. Proc Natl Acad Sci. 2015;112: E2658–E2666. doi:10.1073/pnas.1424949112

77. Uricchio LH, Hernandez RD. Robust forward simulations of recurrent hitchhiking. Genetics. 2014;197: 221–236. doi:10.1534/genetics.113.156935

78. Lange JD, Pool JE. Impacts of recurrent hitchhiking on divergence and demographic inference in Drosophila. Genome Biol Evol. 2018;10: 1882–1891. doi:10.1093/gbe/evy142

79. Huber CD, Kim BY, Marsden CD, Lohmueller KE. Determining the factors driving selective effects of new nonsynonymous mutations. Proc Natl Acad Sci. 2017;114: 4465–4470. doi:10.1073/pnas.1619508114

80. Deutsch M, Long M. Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res. 1999;27: 3219–3228.

81. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA sequence data. Genetics. 1992;132: 583–589.

82. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting F_{ST}: The impact of rare variants. Genome Res. 2013;23: 1514–1521. doi:10.1101/gr.154831.113

83. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: The reference human genome annotation for The ENCODE Project. Genome Res. 2012;22: 1760–1774. doi:10.1101/gr.135350.111

84. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010;467: 1099–1103. doi:10.1038/nature09525

85. Baerenfaller K, Grossmann J, Grobei MA, Hull R, Hirsch-Hoffmann M, Yalovsky S, et al. Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. Science. 2008;320: 938–941. doi:10.1126/science.1157956

86. Salomé PA, Bomblies K, Fitz J, Laitinen R a. E, Warthmann N, Yant L, et al. The recombination landscape in *Arabidopsis thaliana* F_2 populations. Heredity. 2012;108: 447–455. doi:10.1038/hdy.2011.95