Broad-scale variation in human genetic diversity levels is predicted by purifying selection on
coding and non-coding elements

David A. Murphy

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

© 2021

David A. Murphy

All Rights Reserved

Abstract

Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements

David A. Murphy

Genome-wide neutral diversity levels are shaped by both positive and purifying selection on linked sites. In humans like most species, the relative importance of these types of selection in shaping patterns of neutral diversity remains an open question. We can infer their relative contribution from observed patterns of neutral diversity by using information about recombination rates and targets of natural selection. To this end, I fit a joint model of the effects of positive selection (selective sweeps) and purifying selection (background selection) to genetic polymorphism data from the 1000 Genomes Project. I show that a model of the effects of background selection provides a good fit to patterns in diversity data and that incorporating the effects of selective sweeps does not improve the fit. Using my approach, the effects of background selection explain up to 60% of the variation in neutral diversity levels on the 1Mb scale and account for patterns in the data for which positive selection via selective sweeps had been invoked as explanations. I find that over 80% of the selected regions affecting neutral diversity levels are located outside of exons and that phylogenetic conservation is the best predictor of the source of selection in these regions. My results show that the genome-wide effects of background selection are pervasive, with measurable reductions in neutral diversity throughout almost the entirety of the autosomes. I provide maps of the effects of background

selection and software for making similar inferences, which should provide important tools for future research that relies on interpreting patterns in neutral diversity levels.

Table of Contents

<u>List of Figures</u>	<u>v</u>
<u>Acknowledgments.....</u>	<u>x</u>
<u>Introduction.....</u>	<u>1</u>
<u>Chapter 1: Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements</u>	<u>10</u>
1.1 Abstract	10
1.2 Introduction.....	10
1.3 Model and inference	12
1.4 Background selection.....	14
1.5 Selective sweeps	15
1.6 The source of background selection	18
1.7 The deleterious mutation rate.....	20
1.8 Background selection in autosomes.....	22
1.9 Conclusion	24
<u>Chapter 2: Perspective and future directions</u>	<u>25</u>
<u>References.....</u>	<u>31</u>
<u>Appendix 1: Model and inference method.....</u>	<u>37</u>
1.1 Model and inference problem	38
1.2 Calculating lookup tables.....	44

1.3 Binning neutral sites	49
1.4 Optimization	50
1.5 Thresholding	57
1.6 Software	63
<u>Appendix 2: Data sources and filters</u>	<u>65</u>
2.1 Polymorphism data	65
2.2 Multiple species alignment data.....	66
2.3 Genetic map	66
2.4 Human gene annotations.....	66
2.5 CADD scores	67
2.6 ENCODE cCRE annotations	68
2.7 Substitutions in the human lineage	68
2.8 Covariates of B	69
<u>Appendix 3: Choice of exogenous parameters</u>	<u>70</u>
3.1 Choosing putatively neutral sites based on phylogenetic conservation.....	70
3.2 Removing sites at the telomeric ends of chromosomes	73
3.3 Estimating local variation in mutation rates	74
<u>Appendix 4: Fitting models with different targets of selection</u>	<u>77</u>
4.1. Background selection model based on phylogenetic conservation.....	77

4.2 Background selection model based on genic annotations.....	85
4.3 Background selection models separating conserved exonic and non-exonic sites	87
4.4 Background selection models based on other annotations	92
4.5 Models with selective sweeps.....	100
4.6 Comparison with previous work by McVicker and colleagues	104
<u>Appendix 5: Assessing estimates of the deleterious mutation rate.....</u>	<u>108</u>
5.1 Estimates of the total mutation rate per site.....	108
5.2 Estimating the proportion of deleterious mutations at putatively selected sites.....	109
5.3 Interpreting the relationship between the two estimates.....	112
<u>Appendix 6: Statistics</u>	<u>113</u>
6.1 Estimates of explained variance.....	113
6.2 Comparing the fit of different maps.....	115
6.3 Sampling error in parameter estimates	117
<u>Appendix 7: Results for other human populations</u>	<u>118</u>
<u>Appendix 8: Diversity levels where background selection is weakest ($B \approx 1$).....</u>	<u>124</u>
8.1 Covariates of B	125
8.2 Mutational spectrum and biased gene conversion	129
8.3 A footprint of archaic introgression?	136

<u>Appendix 9: Additional Figures.....</u>	<u>140</u>
<u>References to appendices.....</u>	<u>144</u>

List of Figures

Chapter 1 Figures.

Figure 1. Modeling and inferring the effects of linked selection in humans.	13
Figure 2. Comparison of diversity levels predicted by our best-fitting maps of background selection effects with observations.	15
Figure 3. A background selection model predicts neutral diversity levels observed around human-specific nonsynonymous (NS) substitutions.....	17
Figure 4. Estimates of the proportion of deleterious mutations at selected sites for our best-fitting background selection models.....	21
Figure 5. Observed vs. predicted neutral diversity levels across the autosomes.	23

Appendix Figures.

Figure A1. Modeling and inferring the effects of linked selection in humans.	39
Figure A2. Distribution of relative errors in predictions before and after modifying <code>calc_bkgd</code> . 47	
Figure A3. Illustration of the two-step algorithm.	53
Figure A4. Comparison of inferred and ground-truth parameters for datasets simulated ‘deterministically’ under the best-fitting background selection model.	54
Figure A5. Comparison of inferred and ground-truth parameters for datasets simulated with noise under the best-fitting background selection model.....	55
Figure A6. Comparison of inferred and ground-truth parameters for datasets simulated with noise under a joint model of background selection and selective sweeps.....	57
Figure A7. Comparison of inference results with and without thresholding.....	60

Figure A8. The proportion of variance explained in 10 kb, 100 kb and 1 Mb windows using a range of B thresholds applied to lookup tables ('lookup') or during maximization ('maximization').	61
Figure A9. Predicted and observed diversity levels along chromosome 1 in the YRI sample.	62
Figure A10. The distribution of phastCons scores across autosomes for varying phylogenetic distances from humans.	72
Figure A11. The variance in diversity levels explained by our two best-fitting models using different choices of putatively neutral sites.	73
Figure A12. The effect of removing putatively neutral sites near telomeres on model fit and parameter estimates.	74
Figure A13. Choosing the parameters used in estimating the relative mutation rate at putatively neutral sites.	75
Figure A14. Comparison of background selection models based on phastCons conservation scores in phylogenies of difference depths.	81
Figure A15. The spatial distribution of putatively selected sites targets remains similar when we vary the phylogenetic depth of the alignment used to infer conservation (shown in a), and the proportion of sites with the highest conservation scores included (in b).	82
Figure A16. Comparison of background selection models based on phastCons scores using different proportions of autosomal sites as selection targets.	84
Figure A17. The background selection model based on simple genic annotations fits worse than our best-fitting phastCons-based model.	86
Figure A18. The relationship between simple genic annotations and our main measures of constraint.	87

Figure A19. Dividing conserved sites into exonic and non-exonic sets leads to different estimates of selection parameters in each, but to little improvement in fit compared to the model based on conservation alone.	90
Figure A20. Comparison of background models using exonic, non-exonic and all conserved sites as targets of selection.	91
Figure A21. spatial correlations of exonic, non-exonic and all conserved sites for varying window sizes ('con _e ', 'con _n ' and 'con _a ', respectively).	92
Figure A22. The model based on the ENCODE annotations of cCRE fit the data substantially worse than our best-fitting phastCons-based model using conservation in the 99-vertebrate alignment (conserved).....	95
Figure A23. The relationship between ENCODE cCRE annotations and our main measures of constraint.....	95
Figure A24. The model based on CADD scores offer little improvement over the model based on phastCons scores (based on the 99-vertebrate alignment).....	97
Figure A25. The spatial correlation between the 6% of sites with the highest CADD and phastCons scores.	97
Figure A26. Comparison of background selection models based on CADD scores using different proportions of autosomal sites as selection targets.	100
Figure A27. Models with sweeps alone fit substantially worse than models with background selection alone.....	103
Figure A28. The spatial correlation between targets of selection in sweep models and in our best-fitting phastCons-based background selection model.....	104

Figure A29. Our maps of the effects of background selection fit the data much better than the maps from McVicker et al. ² .	106
Figure A30. Different estimates of the deleterious mutation rate at putatively selected sites, measured relative to total mutation rates per site (u_0).	111
Figure A31. The relative difference between R^2 estimates using all the data and the exclusion approach described in the text, for our two best-fitting models.	114
Figure A32. Assessing the differences in fit between our best-fitting models on three spatial scales.	115
Figure A33. Significance level of differences in fit between our best-fitting models and variations on these models, on the 1 Mb spatial scale.	116
Figure A34. Estimates of the sampling errors of parameter estimates for our two best-fitting models.	118
Figure A35. The maps and parameter estimates for different populations are remarkably similar.	119
Figure A36. proportion of variance in diversity levels explained (R^2) in different populations, using the population specific map vs. the YRI map.	120
Figure A37. Differences among populations in the variance in diversity levels explained by our map of the effects background selection.	123
Figure A38. Observed vs. predicted neutral diversity levels across the autosomes (similar to Fig. 5).	125
Figure A39. Average recombination rate (a) and functional density per bp (b) and per cM (c) as a function of predicted B .	126

Figure A40. GC content (a), CpG sites in CpG islands (b), proportion of methylated CpGs in neutral sites (c), and proportion of neutral sites in C>G hypermutable regions (d) as a function of predicted <i>B</i>	128
Figure A41. The relationship between chromosomal position and predicted <i>B</i>	129
Figure A42. of different types of substitutions as a function of predicted <i>B</i>	130
Figure A43. Contribution of different types of substitutions to diversity levels (a) and number of substitutions per site in the 8-primate phylogeny (b) as a function of predicted <i>B</i>	131
Figure A44. Observed vs. predicted neutral diversity levels for different types of substitutions.	133
Figure A45. Observed vs. predicted neutral diversity levels for different types of substitutions after removing C>G hypermutable regions (from both the inference and observations).	135
Figure A46. Comparison of the best-fitting CADD-based models with and without C>G hypermutable regions.	136
Figure A47. Estimated proportion of Neanderthal (NE) ancestry as a function of predicted <i>B</i> in Europeans (CEU) (a) and East-Asians (CHB/CHS) (b).	138
Figure A48. A background selection model predicts neutral diversity levels around different genomic features.	140
Figure A49. Predicted and observed neutral diversity levels along chromosome 1 based on data from representative continental populations.	141
Figure A50. A background selection model predicts neutral diversity levels around human-specific nonsynonymous (NS) substitutions in representative continental populations.	142
Figure A51. Observed vs. predicted neutral diversity levels across the autosomes in representative continental populations.	143

Acknowledgments

I would like to thank my advisor, Guy Sella, for guiding me through the many twists and turns of this project. Guy gave me the opportunity to jump into an entirely new field at the beginning of my PhD and devoted a significant amount of his time to nurturing my understanding of population genetics as well as providing me the opportunity to develop a whole new set of skills in programming and statistics. Guy's extremely high standard of quality when it has come to this project and science in general, while difficult at times, has made the completion of the journey all the more satisfying.

I would also like to thank the members of my committee, Molly Przeworski and Itsik Pe'er, for their helpful comments during the course of the project and for lending their expertise. I am grateful to Eyal Elyashiv, from whom I inherited the groundwork for my project, and who guided me through the early stages. I am deeply grateful to Guy Amster for being an encouraging mentor while he was at Columbia, as well as a very good friend. I also thank Guy for his substantial contributions to this project and for being a limitless source of math and coding knowledge. I would also like to thank Priya Moorjani, Eduardo Amorim and Ipsita Agarwal for helpful discussions over the course of this project.

Finally, I would like to express my gratitude to the members of the Przeworski, Andolfatto and Sella labs for many friendly and helpful conversations over the years.

Introduction

We can learn a lot about evolutionary forces by studying patterns in neutral diversity levels throughout the genome. In particular, natural selection favoring adaptive alleles and disfavoring deleterious ones can affect diversity patterns at linked neutral sites by processes collectively referred to as linked selection¹. When selection favors a new, highly beneficial mutation and drives it to fixation in the population, a process referred to as a “hard” selective sweep, neutral alleles present on the haplotype on which the adaptive mutation hitchhike their way to fixation as well, reducing neutral diversity levels near the selected site^{2,3}. “Soft sweeps” in which an allele segregates at low frequency before becoming beneficial and sweeping to fixation and “partial sweeps”, in which a beneficial mutation is driven to intermediate frequencies, also reduce neutral diversity levels near the selected sites⁴⁻⁹. In functional regions of the genome, deleterious mutations continuously arise and are removed by purifying selection, and neutral alleles sharing haplotypes with these deleterious mutations are removed with them (reducing local neutral diversity levels) in a process referred to as background selection¹⁰⁻¹⁵. Each of these processes reduces diversity levels near targets of selection to a degree that depends on the density of selection targets, the strength of selection and the rate of recombination. We can use the patterns of reduction in genome-wide diversity levels due to linked selection to learn about the mode and frequency of natural selection, as well as the targets of selection and its extent throughout the genome. Characterizing the effects of linked selection also provides an important baseline for other types of population genetics inferences based on neutral diversity data, such as inferences about demography and scans for recent adaptive substitutions¹⁶⁻²⁰.

The effects of linked selection have been observed across taxa. For example, levels of diversity are lower in regions with a greater density of amino acid substitutions in both *Drosophila melanogaster* and humans²¹⁻²³. Similarly, diversity levels are reduced in regions of low recombination and greater functional density in *Drosophila melanogaster*, humans, and across other taxa²³⁻²⁸. With the advent of large genomic datasets, including high resolution genetic maps, genome-wide polymorphism data, multiple sequence alignments and improved functional annotations, it has become possible to infer genome-wide maps of the effects of linked selection along with parameters of selection that best explain diversity patterns.

Pioneering work by McVicker and colleagues showed that a simple model of background selection on conserved exonic regions does surprisingly well at explaining genome-wide patterns in neutral diversity levels in apes and in particular humans²⁹. Despite the impressive fit to patterns in diversity data, their map of the expected effects of background selection provided a poor prediction of patterns around certain genomic features, such as the reduction in diversity levels around amino acid substitutions. Moreover, the inferred deleterious mutation rate for their background selection model was unrealistically high²⁹. These results suggested that their model may have absorbed the effects of selective sweeps or purifying selection on unannotated functional elements²⁹. Subsequent work has suggested that selective sweeps have had little effect on diversity levels in humans, with no more of a reduction of diversity around amino acid substitutions than around synonymous ones^{30,31}. The poor quantitative fit of the McVicker et al. predictions to diversity levels around synonymous and nonsynonymous substitutions was used to argue that the effects of background selection could be more pronounced around synonymous than nonsynonymous substitutions, thereby masking the effects of selective sweeps³². Thus, the

relative contributions of different kinds of linked selection to diversity patterns in the human genome remain controversial.

A natural approach to resolving these problems is to model the effects of background selection and selective sweeps jointly. Using such an approach with data from *Drosophila melanogaster*, Elyashiv et al. characterized the effects of selective sweeps and background selection genome-wide, showing that both modes of linked selection substantially affect neutral diversity levels and that the effects of each mode of selection can be clearly differentiated³³. My thesis builds on the approach of Elyashiv et al.³³, using recent high quality genomic data from humans to disentangle the relative contributions of background selection and selective sweeps. Chapter 1 presents an overview of my main results. I show that a model of background selection driven by purifying selection on phylogenetically conserved regions explains most of the variance in autosomal neutral diversity levels in a Yoruba population on the megabase scale. When I consider models that combine background selection with selective sweeps and make a variety of different assumptions about the targets of adaptation, I consistently infer the fraction of adaptive substitution resulting in sweeps to be approximately zero. My inferred maps of the effects of background selection and selection parameters resolve the outstanding issues in McVicker et al. using a model of background selection alone: First, these maps are able to explain the reduction in diversity levels around amino acid substitutions (and other genomic features). Second, I infer a plausible deleterious mutation rate for the background selection model, which is below estimates of the total point mutation rate and in line with an independent estimate that I calculate using evolutionary substitution rates. In contrast to McVicker et al.²⁹, I find that the majority of the deleterious mutations driving the effects of background selection occur outside of exons,

underscoring the functional significance of non-coding elements. I show that background selection reduces neutral diversity levels throughout most of the autosomes, and my maps of the reduction in diversity are well calibrated for almost the entire range of predicted selection effects. The exception is for ~2% of neutral sites where I predict the least reduction in diversity from background selection, which I interpret as possible evidence of residual archaic ancestry from ancient introgression events.

In Chapter 2, I provide an overview of my results in the context of earlier work and results from other species. This section includes a perspective on what we have learned about linked selection up to now and what our results mean for future studies. I also discuss a variety of applications for our maps of the effects of background selection in other population genetics inferences that rely on patterns in diversity levels that I show to be strongly affected by background selection.

The remainder of the thesis provides a more detailed and technical account of the methods described in chapter 1, as well as additional results and analyses for different models and datasets that I considered. These results are organized into a set of 9 appendices. Appendix 1 describes the models I rely on and the inference method used to generate maps of the effects of linked selection. This method builds on elements of the work of Elyashiv et al.³³ and McVicker et al.²⁹, with the introduction of several important improvements. First, I introduce measures to control the precision of my estimates of the effects of both selective sweeps and background selection. I introduce additional steps for compressing large amounts of human polymorphism data in order to make inferences computationally feasible. I then describe an optimization algorithm that I developed for inferring selection parameters via composite likelihood and detail tests of

optimization robustness using simulated data. I find that misspecification of selection targets can lead to unrealistically large reductions in diversity levels in certain regions of the genome. To address this potential problem, I introduce a threshold on predictions of diversity reduction, which improves the predictions with minimal loss of specificity in my maps. Finally, this section includes a description of the software and datasets that I make publicly available to enable further studies.

In Appendices 2 and 3, I describe the sources of data and exogenous parameters used for making inferences. First, I describe the set of polymorphic and monomorphic sites that I used to measure neutral diversity levels throughout the autosomes. I discuss how I classify sites as neutral, as well as a method to correct for the effects of variation in the neutral substitution rate on neutral polymorphism levels. I describe the annotations used as putative targets of selection for the models of background selection and selective sweeps, as well as the choice of genetic maps.

In Appendix 4, I present results for all of the different models and selection targets that I considered. Using various metrics of fit, I show that, among all choices, background selection due to purifying selection on phylogenetically conserved regions yields the best fit to observed patterns in neutral diversity data. Notably, annotations based on functions, such as coding and UTR segments or functional proxies such as open chromatin, yield significantly worse fits to the data. The only annotation on par with phylogenetic conservation (but no better) is the combined annotation dependent depletion (CADD) scores³⁴, are based heavily on conservation. I find that models of selective sweeps alone fit the data much worse than models of background selection, and when I consider background selection and selective sweeps jointly, the inferred rates of

adaptive substitutions are essentially zero. I also compare my maps of the effects of background selection to the maps from McVicker et al.²⁹ and show that my predictions fit the data much better in addition to requiring a much lower, more realistic deleterious mutation rate.

In Appendix 5, I assess the deleterious mutation rate that I infer using my best fitting models of background selection by deriving an upper bound for the *total* mutation rate using current estimates of different kinds of mutations (i.e., point mutations, indels and mobile element insertions)³⁵⁻³⁷. In contrast to McVicker et al.²⁹, my best estimates of the deleterious mutation rate are well below this upper bound. I also compare evolutionary substitution rates between neutral and conserved regions (based on the annotations described in Appendices 3 and 4) as an independent measure of the deleterious mutation rate in these regions. Specifically, the ratio of observed substitutions in conserved vs. neutral regions can provide a rough estimate of what fraction of mutations in conserved regions are deleterious. Comparing this ratio to the inferred deleterious mutation rate divided by the *total* mutation rate, I find that the two methods are within less than a factor of two; for a subset of the best fitting models, the two estimates overlap.

Appendix 6 deals with assessments of uncertainty in my estimates and measures of statistical significance when comparing the goodness of fit across models. Because the composite likelihood-based method does not lend itself to standard measures of uncertainty, I use a combination of techniques to assess fits. To compare different models, I divide diversity data and predictions into 1Mb blocks, in which diversity levels are approximately independent, and create permutations of paired maps from these blocks of data and predictions, which serves as a null distribution for the difference in fit between maps. I compute the proportion of permutations for

which the difference between models is greater than the difference for the full, non-permuted maps, finding that among the best fitting models, the difference in fit is not significant. I use a leave-one-out approach to assess issues of overfitting, finding that the relative difference between models containing all of the data and those with 2Mb blocks of data removed are extremely small. I use the inferred parameters from the 2Mb block leave-one-out analysis to assess the standard errors in my parameter estimates with jackknifing, finding that they are extremely small.

In Appendix 7, I apply the best fitting models to diversity data from all 26 populations of the 1000 Genomes Project and compare the resulting maps. I show that the inferred selection parameters and resulting maps of the effects of background selection are extremely similar across populations, but the variance in diversity levels that the maps explain differ among populations because of contributions to the variance from differences in demographic history¹⁷. I introduce a simple model that decomposes the variance in diversity levels into components explained by background selection, demography and an interaction between selection and demography. Using this model, I show that, in general, the more recently a population is inferred to have passed through a bottleneck, the greater the contribution of demography to the total variance in diversity levels (although the picture is somewhat more complicated for recently admixed populations). I also find evidence of an interaction between demographic effects and background selection³⁸⁻⁴¹, which I propose as an avenue for further study.

In Appendix 8, I discuss outliers in neutral diversity data, which occur in the ~2% of neutral autosomal sites where the effects of background selection are predicted to be negligible. In these

regions, diversity levels are found to be 10-15% greater than the predicted neutral level, depending on the model used. I describe a number of analyses exploring possible confounders that may explain this pattern, such as hypermutable regions, biased gene conversion and relative distance from putative origins of replication. While many of these factors are highly correlated with the predicted strength of background selection, none of them seem to explain the outlier neutral diversity levels. I then outline a hypothesis that explains the elevated diversity in regions of weak background selection as remnants of ancient admixture between archaic hominins and humans, based on the idea that archaic alleles were mildly deleterious⁴²⁻⁴⁶. Under this scenario, purifying selection purged archaic ancestry throughout the genome, except in regions subject to the weakest selection, where I find elevated diversity levels. In support of this hypothesis, I use high-resolution maps of inferred archaic ancestry in European and East Asian populations⁴⁷ and show that the predicted level of archaic ancestry in both lines up with the predicted effects of background selection.

Finally, Appendix 9 includes additional figures referred to throughout the thesis, including results based on diversity data from other populations and observed vs. predicted diversity levels near different genomic features.

In summary, my thesis shows that a fairly simple model of background selection goes a long way in explaining a broad set of patterns in neutral diversity data in the human genome. Specifically, I show how this model explains most of the variation in diversity levels on the 1Mb scale and also explains the reduction in diversity levels around genomic features like amino acid substitutions. The latter result, combined with the deleterious mutation rate that I infer, resolve

outstanding controversies about the modes of linked selection prevalent in the human genome. I also show that background selection reduces diversity levels throughout most of the autosomes and is primarily driven by purifying selection on non-coding conserved regions that are dispersed throughout the genome. My maps of the effects of background selection provide a valuable tool for studies that rely on patterns of neutral diversity levels in humans to draw conclusions about demography and ancient introgression events. Importantly, this method is applicable to any model system where a basic minimum of data is available, and as the genomic era proceeds this set will include a broad range of taxa.

Chapter 1: Broad-scale variation in human genetic diversity levels is predicted by purifying selection on coding and non-coding elements

1.1 Abstract

Analyses of genetic variation in many taxa have established that neutral genetic diversity is shaped by natural selection at linked sites. Whether the source of selection is primarily the fixation of strongly beneficial alleles (selective sweeps) or purifying selection on deleterious mutations (background selection) remains unknown, however. We address this question in humans by fitting a model of the joint effects of selective sweeps and background selection to autosomal polymorphism data from the 1000 Genomes Project. After controlling for variation in mutation rates along the genome, a model of background selection alone explains ~60% of the variance in diversity levels at the megabase scale. Adding the effects of selective sweeps driven by adaptive substitutions to the model does not improve the fit, and when both modes of selection are considered jointly, selective sweeps are estimated to have had little or no effect on linked neutral diversity. The regions under purifying selection are best predicted by phylogenetic conservation, with ~80% of the deleterious mutations affecting neutral diversity occurring in non-exonic regions. Thus, background selection is the dominant mode of linked selection in humans, with marked effects on diversity levels throughout autosomes.

1.2 Introduction

Selection at a given locus in the genome affects diversity levels at sites linked to it^{2-15,24,48-59}. In particular, when a new, strongly beneficial mutation increases in frequency to fixation in the population, it carries with it the haplotype on which it arose, thus reducing levels of neutral

diversity nearby, in what is sometimes called a “hard selective sweep”^{2,3}. “Soft sweeps”, particularly those in which an allele segregates at low frequency before becoming beneficial and sweeping to fixation, and “partial sweeps”, in which a beneficial mutation rapidly increases to intermediate frequencies, also reduce neutral diversity levels near the selected sites⁴⁻⁹. Similarly, when deleterious mutations are eliminated from the population by selection, so are the haplotypes on which they lie. This process too reduces diversity levels near selected sites, in a phenomenon known as “background selection”^{10-15,52,58,59}. Because the lengths of the haplotypes associated with selected alleles depend on the recombination rate, linked selection processes cause a greater reduction in levels of neutral genetic diversity in regions with lower rates of recombination or a greater density of selected sites. These predicted relationships have been observed in numerous taxa, including plants, *Drosophila*, rodents and primates, establishing the effects of linked selection in a wide range of species^{21-27,53,54,57,60,61}.

More recently, the advent of large genomic datasets and detailed functional annotations have made it possible to build maps of the effects of linked selection that predict levels of diversity along the genome^{29,33}. The first effort predated the availability of genome-wide resequencing data, relying instead on information about incomplete lineage sorting among human, chimpanzee and gorilla, which reflects variation in diversity levels along the genome in the common ancestor of humans and chimpanzees²⁹. This pioneering paper showed that a model of background selection fits variation in human-chimpanzee divergence levels along the genome remarkably well, with only a few parameters. Nonetheless, the estimate of the deleterious mutation rate was unrealistically high, much greater than estimates of the total mutation rate per site in humans^{35,36} (Appendix 5), raising the possibility that it was soaking up effects of other modes of selection,

notably those of selective sweeps. Subsequent work suggested that selective sweeps have had little effect on diversity levels in humans, with no more of a reduction of diversity around nonsynonymous substitutions than around synonymous ones^{30,62}, but the interpretation of these findings was contested. Notably, it was suggested that the effects of background selection are more pronounced around synonymous substitutions than nonsynonymous substitutions, thereby masking the effects of selective sweeps³². Thus, we still lack an understanding of the relative contribution of adaptation and purifying selection⁵¹, as well as a map of their effects on human diversity levels.

1.3 Model and inference

Here, we resolve these issues by considering the effects of background selection and selective sweeps on diversity levels jointly (Fig. 1 and Appendix 1). To this end, we model the effects of background selection on neutral diversity levels as a function of genetic distance from regions that may be under purifying selection (following ¹¹ and ⁶³), where the deleterious mutation rate per site and distribution of selection effects in a given type of region (e.g., exons) are parameters to be estimated (Fig. 1a). In turn, the effects of sweeps are modeled as a function of genetic distance from selected substitutions on the human lineage (following ⁶⁴ and ⁶⁵), where the fraction of substitutions of a given type (e.g., nonsynonymous) that were beneficial and their distribution of selection effects are again parameters to be estimated (Fig. 1b). Importantly, our model should capture the effects of any kind of sweeps, be they hard, partial or soft, so long as they eventually resulted in a substitution and affected diversity levels nearby (see SOM Section D in ³³). Given the positions of different types of selected regions and substitutions, their corresponding selection parameters, and a fine-scale genetic map, the model allows us to

calculate the marginal probability that any given neutral site in the genome is polymorphic in a sample (Fig. 1c). Provided measurements of polymorphism at neutral positions throughout the genome, we combine information across sites and samples to calculate the composite likelihood of selection parameters, and find the parameter values that maximize this likelihood (Fig. 1 and Appendix 1). In addition to parameter estimation, this approach yields a map of the expected neutral diversity levels along the genome (Fig. 1c).

Figure 1. Modeling and inferring the effects of linked selection in humans. Given the targets of selection and corresponding selection parameters (a and b), we calculate the expected neutral diversity levels along the genome (c). We infer the selection parameters by maximizing their composite likelihood given observed diversity levels (c). Based on these parameter estimates, we calculate a map of the expected effects of selection on linked diversity levels.

To infer the effects of background selection and selective sweeps on human diversity levels, we analyze autosomal polymorphism data from 26 human populations, collected in Phase III of the 1000 Genomes Project⁶⁶. Here, we focus on data from 108 genomes sampled from the Yoruba

population (YRI), but in Appendices 7 and 9, we show similar results for other populations. To estimate diversity levels at neutral sites, we focus on non-genic autosomal sites that are the least conserved in a multiple sequence alignment of 25 supra-primates (see Appendix 3.1). To control for variation in the mutation rates across neutral sites, we use estimates of the relative mutation rate for contiguous, non-overlapping blocks of 6000 putatively neutral sites, obtained from substitution rates in an eight-primate phylogeny (see Appendix 3.3). To rely on a high-resolution genetic map built that is least dependent on diversity levels, we use the map inferred from ancestry switches in African-Americans, which is highly correlated with other maps⁶⁷.

1.4 Background selection

We first focus on two of our best-fitting models of the effects of background selection (see below and Appendix 4). In both cases, we take as targets of purifying selection the 6% of autosomal sites estimated as most likely to be under selection constraint. In one, we choose these sites using phastCons conservation scores obtained for a 99-vertebrate phylogeny that excludes humans⁶⁸. In the other, we rely on Combined Annotation-Dependent Depletion (CADD) scores, which are based primarily on phylogenetic conservation (excluding humans) but also on information from functional genomic assays^{34,69} (to avoid circularity, we use scores that were generated without using the McVicker et al. *B*-map²⁹ as input; see Appendix 2.5). From these models, we obtain a map of predicted diversity levels (accounting for variation in mutation rates), which we can then compare to observed data (Figs. 2a and A24). As a measure of the precision of our

Figure 2. Comparison of diversity levels predicted by our best-fitting maps of background selection effects with observations. (a) Predicted and observed diversity levels along chromosome 1 in the YRI sample. Diversity levels are measured in 1 Mb windows, with a 0.5 Mb overlap, with the autosomal mean set to 1. **(b)** The proportion of variance in YRI diversity levels explained by background selection models on different spatial scales. Shown are the results for four choices of putative targets of selection: all sites with the highest 6% of CADD or phastCons scores (denoted CADD and phastCons, respectively) and the subset of these sites that are exonic (denoted CADD_e and phastCons_e, respectively). See Appendix 4 for similar graphs with other choices, and Appendices 7 and 9 for other populations. We show that our estimates are not inflated by over-fitting in Appendix 6.

predictions, we consider the variance in diversity levels explained in non-overlapping autosomal windows (Fig. 2b). As can be seen, our predictions explain a large proportion of the variance across spatial scales: at the 1 Mb scale using CADD scores, they account for 59.9% of the variance in diversity levels, ~87% more than explained previously²⁹ (see Appendix 4.6).

1.5 Selective sweeps

Next, we examine whether incorporating selective sweeps alongside background selection improves our predictions. Our inference should be able to tease apart the effects of selective sweeps, primarily because their effects, unlike those of background selection, should be centered around the locations of substitutions. Moreover, as noted, we expect to capture the effects of selective sweeps, be they hard, partial or soft²⁻⁸, so long as they resulted in substitutions and

substantially affected diversity levels (see SOM Section D in ³³). Indeed, previous work that applied a similar methodology to data from *Drosophila melanogaster* was able to identify distinct effects of background selection and sweeps³³. To examine whether we can identify such effects in humans, we consider several choices of putatively selected substitutions along the human lineage, including any nonsynonymous substitutions or any nonsynonymous and non-coding substitutions in constrained regions, allowing each type to have its own selection parameters and considering different measures of constraint (see Appendix 4.5). For no choice does incorporating sweeps improve our fit. In fact, in all cases our estimates of the proportion of substitutions resulting in sweeps with discernable effects on neutral diversity is approximately 0.

Moreover, in contrast to previous attempts^{29,30}, our model of background selection alone provides good quantitative fits to the diversity levels observed around different genomic features and in particular around nonsynonymous and synonymous substitutions (Figs. 3 and A48). Together, these results refute the hypothesis that reduced diversity levels around nonsynonymous substitutions in humans reflect “masked” effects of selective sweeps³²; more generally, they indicate that selective sweeps resulting in substitutions had little effect on diversity levels in contemporary humans.

Figure 3. A background selection model predicts neutral diversity levels observed around human-specific nonsynonymous (NS) substitutions. Shown are the results for putatively neutral sites as a function of their genetic distance to the nearest nonsynonymous substitution (in 160 bins, each spanning 0.005cM). For observed values, we average diversity levels within each bin. For predicted values, we average predicted diversity levels for the best-fitting CADD-based model (using the mean mutation rate) within each bin, and then multiply by an estimate of relative mutation rates in that bin compared to the average (obtained from substitution data; see Appendix 3.3). Both observed and predicted diversity levels are plotted relative to the autosomal mean. See Figs S48 and A50 for similar graphs for other genomic features and using data from other populations.

The lack of a footprint of sweeps in neutral diversity data does not imply that adaptation was rare in recent human evolution. Instead, much of it may have been driven by selection on genetically complex traits^{30,62,70-72}, i.e., traits with heritable variation arising from many segregating loci.

The response to a shift in selection pressures on complex traits is expected to be highly polygenic, with short-term, rapid but tiny changes to allele frequencies at many loci that segregated before the shift, and a longer term, small excess in the fixation of alleles that change the trait in the direction favored by selection⁷³. Importantly, polygenic adaptation is unlikely to result in sweeps and is likely to introduce only minor perturbations to the allele trajectories expected when selection pressures on traits remain constant⁷³, implying that its effects on neutral diversity levels should be minor¹. In contrast, ongoing stabilizing selection on complex traits, i.e., selection acting to maintain traits near an optimal value, is thought to be common⁷⁴⁻⁷⁷ and

could have a substantial effect on neutral diversity levels⁷³. Stabilizing selection induces purifying selection against minor alleles that affect complex traits^{75,78,79}, and purifying selection on these alleles could be a major source of background selection⁷³. In other words, if much of the selection in humans is driven by ongoing and changing selection pressures on complex traits, we may expect background selection to be the dominant mode of linked selection, as our results indicate.

1.6 The source of background selection

Focusing then on models of background selection alone, we ask which genomic annotations appear to be the sources of purifying selection. Previous work found selection on non-exonic regions to contribute little, to the extent that removing conserved non-exonic sites from a model of background selection had little effect on predicted diversity levels²⁹. In contrast, when we include only conserved exonic regions in our inference, our predictive ability is considerably diminished (Fig. 2b).

Moreover, in models that include separate selection parameters for conserved exonic and non-exonic regions, purifying selection on non-exonic regions accounts for most of the reduction in linked neutral diversity (Appendix 4.3). Our estimates suggest that ~80% of deleterious mutations affecting neutral diversity occur in non-exonic regions (e.g., in the model with the top 6% of phastCons scores, ~84% of selected sites and ~76% of mutations are non-exonic; with the top 6% of CADD scores, ~83% of selected sites and ~85% of deleterious mutations are non-exonic; see Appendices 4.3 and 4.6). Our estimates of the average strength of selection differ between exonic and non-exonic regions, but because the total reduction in diversity levels caused

by background selection is fairly insensitive to the strength of selection (with the reduction being more localized for weakly selected mutations than for strong ones), the proportions of deleterious mutations that occur in these regions approximate their relative effects on neutral diversity levels⁵⁰ (see Appendices 4.3, 4.4 and 4.6). Thus, our estimates suggest that purifying selection on non-exonic regions accounts for ~80% of the reduction in linked neutral diversity. Moreover, including separate selection parameters for conserved exonic and non-exonic regions does not appear to improve our predictions (Appendix 4.3 and Fig. A19).

Incorporating additional functional genomic information also does little to improve our predictions (Appendices 4.2 and 4.4). Notably, including separate selection parameters for coding regions and for each of the Encyclopedia of DNA Elements (ENCODE) classes of candidate cis-regulatory elements (cCRE)⁸⁰, considerably diminishes our predictive ability (Appendix 4.4). Indeed, using CADD scores^{34,69}, which augment information on phylogenetic conservation with functional genomic information, offers little improvement over relying on conservation alone (e.g., explaining 59.9% compared to 59.7% of the variance in diversity levels in 1Mb windows, a difference that is not statistically significant; Appendix 6). Thus, at present, functional annotations that do not incorporate phylogenetic conservation appear to provide poorer predictions of the effects of linked selection and those that do offer little improvement over conservation alone (see Appendices 4.1- 4).

In turn, our predictions are fairly insensitive to the phylogenetic depth of the alignments used to infer conservation levels, although we appear to do slightly better using a 99-vertebrate alignment (excluding humans) compared to its monophyletic subsets (e.g., Figs. A14c and A33a

and Appendix 6.2). Our best-fitting models based on a variety of metrics are obtained using 5-7% of sites with the top CADD or phastCons scores as selection targets (Fig. A16 and A26), a percentage in good accordance with more direct estimates of the proportion of the human genome subject to functional constraint⁸¹⁻⁸⁴.

1.7 The deleterious mutation rate

Reassuringly, the deleterious mutation rates that we estimate for our best-fitting models are plausible (Fig. 4). Current estimates of the average mutation rate per site per generation in humans, including point mutations^{35,36}, indels³⁶, mobile element insertions³⁷, and structural mutations^{85,86} lie in the range of $1.29 \times 10^{-8} - 1.38 \times 10^{-8}$ per base pair per generation (Appendix 5). Further accounting for the length of deletions³⁶ (as a deletion that starts at a neutral site and includes selected sites should be included but a deletion that affects several selected sites should only be counted once), we find that the upper bound on the rate of deleterious mutations falls in the range of $1.29 \times 10^{-8} - 1.51 \times 10^{-8}$ per base pair per generation (Appendix 5). The estimates for all of our best-fitting models fall well below this bound (Fig. 4). This is what we would expect, because not every mutation at putatively selected sites will be deleterious: some sites are misclassified as constrained and some mutations at selected sites are selectively neutral.

Figure 4. Estimates of the proportion of deleterious mutations at selected sites for our best-fitting background selection models. Shown are the results for models using 5-7% of sites with the highest phastCons scores (a) and CADD scores (b) as selection targets. For estimates based on fitting the models, we divide our estimates of the deleterious mutation rate per selected site by the estimate of the total mutation rate per site, where the ranges correspond to the range of estimates of the total rate, i.e., $1.29 \times 10^{-8} - 1.51 \times 10^{-8}$ per site per generation (Appendix 5.1). For estimates based on evolutionary rates, we divide the estimated rates per site on the human lineage (from the common ancestor of humans and chimpanzees) at selection targets by the rates at matched sets of putatively neutral sites (see text and Appendix 5.2 for details).

To test whether the estimated proportion of deleterious mutations is plausible, we compare it with an independent estimate based on the relative reduction in evolutionary rates at constrained vs. putatively neutral sites along the human lineage (these sets of sites were identified from an alignment that excludes humans; Appendices 3.1, 4.1 and 4.4). Because deleterious mutations at selected sites rarely fix in the population whereas neutral mutations fix with a much higher probability (which is the same at conserved and neutral sites), this relative reduction allows us to estimate the proportion of deleterious mutations⁸⁷. To this end, we were also mindful of potential differences in point mutation rates between conserved and neutral sites, calculating rates separately for sites inferred to be GC versus AT in the human-chimpanzee ancestor and checking that other aspects known to affect mutation or substitution rates (i.e., the triplet context, CpG mutations and biased gene conversion) have little effect on our estimates (Appendix 5). By this

approach, our estimates are closer to (and even overlap) those obtained from fitting models of background selection based on CADD scores compared to those based on phastCons scores (Fig. 4); this is expected, given that CADD scores are much better than phastCons scores at identifying constraint on a single site resolution^{34,69}, which markedly influences evolutionary rates at putatively selected sites (but not the predictions of background selection effects). We expect the two estimates to be similar but not identical, both because weak selection has a larger effect on evolutionary rates than on linked diversity levels^{13-15,52,56} and because estimates based on the effects of background selection may absorb the deleterious mutation rate at selected sites that were not included in our sets but are closely linked to sites in them (Appendix 5). In summary, given the fit to data and plausible estimates of the deleterious rates, it is natural to interpret our maps as reflecting the effects of background selection, i.e., as maps of B (defined as the ratio of expected diversity levels with background selection, π , and in its absence, π_0)¹⁰.

1.8 Background selection in autosomes

Our maps are also well calibrated (Fig. 5). When we stratify diversity levels at putatively neutral sites by our predictions, predicted and observed diversity levels are similar throughout nearly the entire range of predicted values (e.g., $R^2 = 0.96$ when sites are in predicted percentile bins). One exception is for ~4% of sites in which background selection is predicted to be the strongest (i.e., with the lowest B), where our predictions are imprecise. This behavior is due to a technical approximation we employ in fitting the models (see Appendix 1.5). The other exception is for ~2% of sites in which background selection is predicted to be the weakest (i.e., with B near 1), where observed diversity levels are markedly greater than expected. We observe similar behavior in all the human populations examined (Fig. A51), and we cannot fully explain it by known

mutational and recombination effects (e.g., of base composition and biased gene conversion; Appendix 8). This behavior could reflect ancient introgression of archaic human DNA into ancestors of contemporary humans (Appendix 8.3), indicated also in other population genetic signatures^{42,43,47,88-95}. Such introgressed regions are expected to increase genetic diversity and persist the longest in regions with low functional density and high recombination, corresponding to weak background selection effects^{44,96,97}.

Figure 5. Observed vs. predicted neutral diversity levels across the autosomes. Shown are the results for the best-fitting CADD-based model. Light orange scatter plot: we divide putatively neutral sites into 100 equally sized bins based on the predicted B . For predicted values (x-axis), we average the predicted B in each bin. For observed values (y-axis), we divide the average diversity level by the estimate of the average relative mutation rate (obtained from substitution data; see Appendix 3.3) in each bin, and normalize by the autosomal average of π_0 (estimated from fitting the model; see Appendix 1.1). Dark orange curve: the LOESS fit for a similarly defined scatter plot but with 2000 rather than 100 bins (with span=0.1). For similar graphs corresponding to other models and using data from other populations, see Appendices 4 and 9, respectively.

Setting these outlier regions aside, we can use the maps to characterize the distribution of background selection effects in human autosomes. We note that background selection effects that are not captured by our models would cause us to underestimate the range and extent of background selection effects³³. We find that diversity levels throughout almost all of the autosomes are affected by background selection, with a $\sim 37\%$ reduction in the 10% most affected sites, a non-zero ($\sim 2.1\%$) reduction even in the 10% least affected (after excluding outliers in the top 2% of bins; see Fig. 5), and a mean reduction of $\sim 17\%$. These conclusions are robust across our best-fitting maps and populations (Appendix 4 and Figs. A35 and A51). An important implication is that our maps of the effects of background selection provide a more accurate null model than currently used for other population genetic inferences that rely on diversity levels, notably inferences about demographic history^{16,17,98-100}.

1.9 Conclusion

Our results indicate that background selection is the dominant mode of linked selection in human autosomes and the major determinant of neutral diversity levels on the Mb scale. They further reveal that background selection effects arise primarily from purifying selection at non-coding regions of the genome. Non-coding regions are known to exhibit substantial functional turnover on evolutionary timescales⁸¹⁻⁸⁴, and yet we find phylogenetic conservation to be the best predictor of selected regions, and that, at least at present, augmenting measures of conservation with functional genomic information in humans offers little improvement. It therefore remains unclear how much our maps can still be improved. Even without these potential refinements, our findings demonstrate that a simple model of background selection, conceived nearly three

decades ago¹⁰, is able to provide a reliable quantitative prediction of genetic diversity levels throughout human autosomes.

Chapter 2: Perspective and future directions

In this thesis I have shown that background selection is the dominant mode of linked selection in humans and that its effects are pervasive throughout the human genome. The relatively simple model of background selection that I use, with a minimal set of external parameters (i.e., the 6% most conserved segments in a 99-vertebrate alignment^{68,101} and the African American genetic map⁶⁷) is able to explain up to 60% of the variation in diversity levels at the megabase scale (depending on the demographic history of the population from which polymorphism data is gathered^{17,66}). The rate of deleterious mutations that we infer to be driving the effects of background selection is plausible: well below biological upper limits that we estimate by the combining rates across mutation types from large pedigree studies and concordant with evolutionary rates in conserved regions³⁵⁻³⁷. We find that the majority of deleterious mutations driving background selection occur outside of exons, consistent with ~85% of conserved regions being non-exonic and underscoring the important evolutionary role of non-exonic conserved elements whose precise function remain unclear. Our ability to explain patterns in diversity levels around amino acid substitutions supports previous studies that suggested that selective sweeps were rare in recent human evolution³⁰. More generally, we show the reduction in neutral diversity levels due to background selection are pervasive, with diversity levels below the neutral expectation throughout much of the genome.

As noted, our results are based on a relatively simple set of inputs, despite our efforts to learn more about targets of selection using recent functional and bioinformatics-based annotations^{34,69,80}. Although at present we find that conservation is the most informative annotation, we expect that better understanding of functional regions throughout the genome will improve our ability to pinpoint regions driving background selection, though the gains beyond what can be inferred based on conservation may be marginal. The bulk of conserved regions lying outside of exons are of unknown function and as we learn more about these regions it may allow us to identify analogous features that are not well defined by conservation scores due to high turnover, yet subject to purifying selection. Similarly, we expect that higher resolution population-specific genetic maps based on large-scale pedigree studies may also improve our predictions. Although these and other improvements on the data side (e.g., larger sample sizes for human polymorphism data) are expected to improve our results, we expect these will all be incremental improvements subject to diminishing returns.

With presently available resources, a natural extension of our method would be to map the effects of linked selection on the human X chromosome compare the effects of linked selection on the X to the autosomes. Because of X hemizyosity in males, selection on both adaptive alleles and deleterious mutations is more intense than in autosomes and our inferences about the effects of linked selection on diversity levels in the X should allow us to gauge the extent to which these differences match our expectations. The question of the relative contributions of selective sweeps and background selection on the X remains a controversial point as in autosomes¹⁰²; while our results largely rule out pervasive sweeps in autosomes, it is possible that, due to the different selection dynamics on the X, the contribution from sweeps may significant

and discernable using our method. Several features of the diversity patterns across the X have also been suggested as evidence for adaptations and ancient admixture¹⁰²⁻¹⁰⁴. Mapping the effects of linked selection on the X chromosome diversity levels should therefore offer clarification to this area of inquiry.

Another notable features of my results about linked selection in humans is their strong contrast to the relative contributions of selective sweeps and background selection in *Drosophila Melanogaster* that were inferred using the same joint modeling approach we use here³³. In *Drosophila*, it was found that the effects of linked selection cause a far greater reduction in genome-wide diversity levels – 73% on average – and that classic sweeps on amino acid substitutions are quite common and a major contributor to the reduction in diversity³³. It is not clear why selective sweeps would be a much more common mode of adaptation in *Drosophila* compared to humans, where polygenic adaptation is the more predominant mode. At least some adaptations in *Drosophila* appear to be oligogenic^{105,106}, so perhaps the architecture of adaptive traits is more amenable to recurrent selective sweeps. Setting sweeps aside, the degree of inferred reduction in diversity levels in *Drosophila* due to background selection alone is substantially greater than what we infer in humans. This may be in part due to misspecification in the model for *Drosophila*, where the inferred deleterious mutation rate was somewhat unrealistically high (see SOM Sections H-J in ³³). In addition, the synonymous polymorphism levels used to measure neutral polymorphism levels in this study are considerably more closely linked to selection targets in the *Drosophila* genome, which could inflate the overall reduction in diversity levels relative to what is seen in the human case, where putatively neutral sites are more widely dispersed and less tightly linked to the conserved segments under purifying selection. With

future studies of the effects of linked selection in different taxa, it will be interesting to see whether the contributions of different kinds of linked selection line up with what we report in humans versus results from *Drosophila*, or something entirely different.

The increasing availability of detailed genomic data now makes it possible to study the effects of linked selection in other species to better understand variation in selection across taxa¹⁰⁷⁻¹⁰⁹.

Early work in model species, including mice and *C. elegans*, have demonstrated hallmarks of linked selection^{25,28}. By applying my method, it will be possible to create a richer picture of the parameters of selection in these species and compare the relative contributions of selective sweeps and background selection as well as the overall reduction in diversity levels to our results in humans and previous results in *Drosophila*. Additionally, population-scale sequencing projects and newly available genetic maps for great apes have made it possible to pose the same questions about modes of linked selection in our nearest relatives¹⁰⁷⁻¹⁰⁹. In the case of the great apes, the set of conserved elements is expected to be quite similar to those we use as selection targets in humans⁶⁸. Comparing conserved elements defined from a common multiple species alignment and inferred to be under purifying selection across apes should allow determine the degree of overlap in conservation between humans and our nearest relatives and to better understand the rate of turnover in these regions. We can also learn about whether or not selective sweeps appear to have been more frequent in the recent evolutionary history of these near-relatives of humans, or if the pattern of background selection dominating is common among apes. We can use the same resource for ancestor reconstruction as we used for the human-chimp ancestor to identify other lineage-specific substitutions among the apes that may have led to sweeps¹¹⁰.

Our map of the effects of background selection appear to point to patterns in the data that could provide the basis for further studies, such as the weak interaction between background selection effects and demography (see Appendix 7). Our map of the effects background selection across populations is remarkably stable despite the heterogeneity in the coalescent rates introduced by different demographic histories, although variance in diversity levels due to demography appears to interact with background selection at least weakly³⁸⁻⁴¹. In fact, with the heterogeneity in the coalescent across the autosomes is surprising that the method works well even in a single population. Further theoretical work considering the interaction of background selection and demography may further shed light on this result.

The effect of background selection on diversity, which can be thought of as local reductions to the local effective population within segments of the genome, has important ramifications for scans for positive selection based on outlier allele frequencies. Because the reduction in the local effective population size can greatly affect the rate of genetic drift, the variance in allele frequencies is larger and outliers relative to average diversity levels may be explained by the local reduction in the effective population size due to background selection³¹. Since the targets of background selection generally overlap putative adaptive substitutions and high-frequency derived SNPs identified near genes, our map is an important resource to control for this effect^{30,32}. As the targets of background selection that we identify appear to be stable across populations and in most cases subject to very ancient purifying selection, this should be true not only for studies of selection in modern populations but also for scans for selection on more ancient human samples from the last 10-20kya¹¹¹.

The evidence we show that the unexplained peak in diversity levels around $B = 1$ may be due to archaic introgression is perhaps merely suggestive given the resolution of current archaic ancestry mapping⁴⁷. At the moment there are a number of possibly confounding factors, such as the fact that at least some maps of archaic introgression that have been generated use the older McVicker B statistic as an input, and the fact that identifying archaic introgression becomes more difficult with greater recombination rates using some of these methods^{88,89}. There may be other patterns due to differences in mutation rates in the telomeric ends of the chromosomes (where $B = 1$ sites cluster) that we have not controlled for, as even the relatively blunt filter removing 250Mb of C>G hypermutable regions appears to diminish – though not remove – this peak¹¹². Although multiple lines of evidence support the idea that archaic ancestry appears to decrease in regions under greater purifying selection⁴²⁻⁴⁴, the specific peak around $B = 1$ may just coincide with this pattern. By using the peak in diversity levels around $B = 1$ to make specific predictions about the timing of a putative introgression event (based on assumptions about the strength of selection against archaic alleles and information about recombination rates), it may be possible to assess whether our speculation about archaic introgression is plausible.

In sum, the predictions about natural selection in humans and the method for inference that I have developed in my thesis provide a substantial resource for future study, both in humans and other species. This work represents an important first step towards filling in gaps in our knowledge about the mode and extent of natural selection across the living world as genomic era accelerates.

References

1. Barton, N.H. Genetic hitchhiking. *Philos Trans R Soc Lond B Biol Sci* **355**, 1553-62 (2000).
2. Maynard Smith, J.M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet Res* **23**, 23-35 (1974).
3. Kaplan, N.L., Hudson, R.R. & Langley, C.H. The "hitchhiking effect" revisited. *Genetics* **123**, 887-99 (1989).
4. Przeworski, M., Coop, G. & Wall, J.D. The signature of positive selection on standing genetic variation. *Evolution Int J Org Evolution* **59**, 2312-23 (2005).
5. Pennings, P.S. & Hermisson, J. Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. *Mol Biol Evol* **23**, 1076-84 (2006).
6. Pennings, P.S. & Hermisson, J. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet* **2**, e186 (2006).
7. Hermisson, J. & Pennings, P.S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335-52 (2005).
8. Coop, G. & Ralph, P. Patterns of neutral diversity under general models of selective sweeps. *Genetics* **192**, 205-24 (2012).
9. Berg, J.J. & Coop, G. A Coalescent Model for a Sweep of a Unique Standing Variant. *Genetics* **201**, 707-25 (2015).
10. Charlesworth, B., Morgan, M.T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-303 (1993).
11. Hudson, R.R. & Kaplan, N.L. Deleterious background selection with recombination. *Genetics* **141**, 1605-17 (1995).
12. Nordborg, M., Charlesworth, B. & Charlesworth, D. The effect of recombination on background selection. *Genet Res* **67**, 159-74 (1996).
13. McVean, G.A. & Charlesworth, B. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929-44 (2000).
14. Gordo, I., Navarro, A. & Charlesworth, B. Muller's ratchet and the pattern of variation at a neutral locus. *Genetics* **161**, 835-48 (2002).
15. Good, B.H., Walczak, A.M., Neher, R.A. & Desai, M.M. Genetic diversity in the interference selection limit. *PLoS Genet* **10**, e1004222 (2014).
16. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-6 (2011).
17. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet* **46**, 919-25 (2014).
18. Sabeti, P.C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832-7 (2002).
19. Nielsen, R. *et al.* Genomic scans for selective sweeps using SNP data. *Genome Res* **15**, 1566-75 (2005).
20. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72 (2006).

21. Macpherson, J.M., Sella, G., Davis, J.C. & Petrov, D.A. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in *Drosophila*. *Genetics* **177**, 2083-99 (2007).
22. Andolfatto, P. Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Res* **17**, 1755-62 (2007).
23. Cai, J.J., Macpherson, J.M., Sella, G. & Petrov, D.A. Pervasive hitchhiking at coding and regulatory sites in humans. *PLoS Genet* **5**, e1000336 (2009).
24. Begun, D.J. & Aquadro, C.F. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519-20 (1992).
25. Nachman, M.W. Patterns of DNA variability at X-linked loci in *Mus domesticus*. *Genetics* **147**, 1303-16 (1997).
26. Payseur, B.A. & Nachman, M.W. Gene Density and Human Nucleotide Polymorphism. *Molecular Biology and Evolution* **19**, 336-340 (2002).
27. Wright, S.I. *et al.* Testing for effects of recombination rate on nucleotide diversity in natural populations of *Arabidopsis lyrata*. *Genetics* **174**, 1421-30 (2006).
28. Cutter, A.D. Selection at Linked Sites in the Partial Selfer *Caenorhabditis elegans*. *Molecular Biology and Evolution* **20**, 665-673 (2003).
29. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet* **5**, e1000471 (2009).
30. Hernandez, R.D. *et al.* Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920-4 (2011).
31. Coop, G. *et al.* The role of geography in human adaptation. *PLoS Genet* **5**, e1000500 (2009).
32. Enard, D., Messer, P.W. & Petrov, D.A. Genome-wide signals of positive selection in human evolution. *Genome Res* **24**, 885-95 (2014).
33. Elyashiv, E. *et al.* A Genomic Map of the Effects of Linked Selection in *Drosophila*. *PLoS Genet* **12**, e1006130 (2016).
34. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
35. Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471-5 (2012).
36. Besenbacher, S. *et al.* Multi-nucleotide de novo Mutations in Humans. *PLoS Genet* **12**, e1006315 (2016).
37. Gardner, E.J. *et al.* Contribution of retrotransposition to developmental disorders. *Nat Commun* **10**, 4630 (2019).
38. Comeron, J.M. Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philos Trans R Soc Lond B Biol Sci* **372**(2017).
39. Torres, R., Stetter, M.G., Hernandez, R.D. & Ross-Ibarra, J. The Temporal Dynamics of Background Selection in Nonequilibrium Populations. *Genetics* **214**, 1019-1030 (2020).
40. Torres, R., Szpiech, Z.A. & Hernandez, R.D. Human demographic history has amplified the effects of background selection across the genome. *PLoS Genet* **14**, e1007387 (2018).
41. Wang, L. *et al.* The interplay of demography and selection during maize domestication and expansion. *Genome Biol* **18**, 215 (2017).
42. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-7 (2014).

43. Vernot, B. & Akey, J.M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017-21 (2014).
44. Juric, I., Aeschbacher, S. & Coop, G. The Strength of Selection against Neanderthal Introgression. *PLoS Genet* **12**, e1006340 (2016).
45. Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Current Biology* **26**, 1241-1247 (2016).
46. Harris, K. & Nielsen, R. The Genetic Cost of Neanderthal Introgression. *Genetics* **203**, 881-91 (2016).
47. Steinrucken, M., Spence, J.P., Kamm, J.A., Wiczorek, E. & Song, Y.S. Model-based detection and analysis of introgressed Neanderthal ancestry in modern humans. *Mol Ecol* **27**, 3873-3888 (2018).
48. Hill, W.G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet Res* **8**, 269-94 (1966).
49. Wiehe, T.H. & Stephan, W. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol Biol Evol* **10**, 842-54 (1993).
50. Hudson, R.R. How can the low levels of DNA sequence variation in regions of the drosophila genome with low recombination rates be explained? *Proc Natl Acad Sci U S A* **91**, 6815-8 (1994).
51. Kim, Y. & Stephan, W. Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* **155**, 1415-27 (2000).
52. Comeron, J.M. & Kreitman, M. Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**, 389-410 (2002).
53. Wright, S.I. & Andolfatto, P. The impact of natural selection on the genome: emerging patterns in *Drosophila* and *Arabidopsis*. *Annual Review of Ecology and Systematics* **39**, 193-213 (2008).
54. Sella, G., Petrov, D.A., Przeworski, M. & Andolfatto, P. Pervasive natural selection in the *Drosophila* genome? *PLoS Genet* **5**, e1000495 (2009).
55. Stephan, W. Genetic hitchhiking versus background selection: the controversy and its implications. *Philos Trans R Soc Lond B Biol Sci* **365**, 1245-53 (2010).
56. Charlesworth, B. Background selection 20 years on: the Wilhelmine E. Key 2012 invitational lecture. *J Hered* **104**, 161-71 (2013).
57. Cutter, A.D. & Payseur, B.A. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* **14**, 262-74 (2013).
58. Nicolaisen, L.E. & Desai, M.M. Distortions in Genealogies due to Purifying Selection and Recombination. *Genetics* **195**, 221-230 (2013).
59. Cvijović, I., Good, B.H. & Desai, M.M. The Effect of Strong Purifying Selection on Genetic Diversity. *Genetics* **209**, 1235-1278 (2018).
60. Nordborg, M. *et al.* The pattern of polymorphism in *Arabidopsis thaliana*. *PLoS Biol* **3**, e196 (2005).
61. Begun, D.J. *et al.* Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol* **5**, e310 (2007).
62. Coop, G. *et al.* The Role of Geography in Human Adaptation. *Plos Genetics* **5**(2009).

63. Nordborg, M., Charlesworth, B. & Charlesworth, D. The effect of recombination on background selection. *Genetics Research* **67**, 159-174 (1996).
64. Barton, N.H. The effect of hitch-hiking on neutral genealogies. *Genetical Research* **72**, 123-133 (1998).
65. Gillespie, J.H. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**, 909-19 (2000).
66. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
67. Hinch, A.G. *et al.* The landscape of recombination in African Americans. *Nature* **476**, 170-5 (2011).
68. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-50 (2005).
69. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* **47**, D886-D894 (2019).
70. Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J.K. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**, 1411-23 (2010).
71. Pritchard, J.K. & Di Rienzo, A. Adaptation - not by sweeps alone. *Nat Rev Genet* **11**, 665-7 (2010).
72. Sella, G. & Barton, N.H. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annu Rev Genomics Hum Genet* (2019).
73. Hayward, L.K. & Sella, G. Polygenic adaptation after a sudden change in environment. (Cold Spring Harbor Laboratory, 2019).
74. Wright, S. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *Journal of Genetics* **30**, 243-256 (1935).
75. Robertson, A. A mathematical model of the culling process in dairy cattle. *Animal Science* **8**, 95-108 (1966).
76. Sella, G. & Barton, N.H. Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics* **20**, 461-493 (2019).
77. Walsh, B. & Lynch, M. *Evolution and Selection of Quantitative Traits*, (Oxford University Press, 2018).
78. Simons, Y.B., Bullaughey, K., Hudson, R.R. & Sella, G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLOS Biology* **16**, e2002985 (2018).
79. Wright, S. Evolution in Mendelian Populations. *Genetics* **16**, 97-159 (1931).
80. Consortium, E.P. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).
81. Ward, L.D. & Kellis, M. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* **337**, 1675-8 (2012).
82. Rands, C.M., Meader, S., Ponting, C.P. & Lunter, G. 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *PLoS Genetics* **10**, e1004525 (2014).
83. Kellis, M. *et al.* Defining functional DNA elements in the human genome. *Proc Natl Acad Sci U S A* **111**, 6131-8 (2014).

84. Ponting, C.P. & Hardison, R.C. What fraction of the human genome is functional? *Genome Res* **21**, 1769-76 (2011).
85. Sudmant, P.H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81 (2015).
86. Belyeu, J.R. *et al.* De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *Am J Hum Genet* **108**, 597-607 (2021).
87. Kimura, M. & Crow, J.F. The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics* **49**, 725-38 (1964).
88. Skov, L. *et al.* Detecting archaic introgression using an unadmixed outgroup. *PLoS Genet* **14**, e1007641 (2018).
89. Durvasula, A. & Sankararaman, S. Recovering signals of ghost archaic introgression in African populations. *Sci Adv* **6**, eaax5097 (2020).
90. Wall, J.D. & Hammer, M.F. Archaic admixture in the human genome. *Curr Opin Genet Dev* **16**, 606-10 (2006).
91. Wall, J.D., Lohmueller, K.E. & Plagnol, V. Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol Biol Evol* **26**, 1823-7 (2009).
92. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).
93. Plagnol, V. & Wall, J.D. Possible ancestral structure in human populations. *PLoS Genet* **2**, e105 (2006).
94. Green, R.E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).
95. Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sánchez, E. Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics* **16**, 359-371 (2015).
96. Harris, K. & Nielsen, R. The Genetic Cost of Neanderthal Introgression. *Genetics* **203**, 881-891 (2016).
97. Schumer, M. *et al.* Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656-660 (2018).
98. Terhorst, J., Kamm, J.A. & Song, Y.S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics* **49**, 303-309 (2017).
99. Steinrücken, M., Kamm, J., Spence, J.P. & Song, Y.S. Inference of complex population histories using whole-genome sequences from multiple populations. *Proceedings of the National Academy of Sciences* **116**, 17115-17120 (2019).
100. Spence, J.P., Steinrücken, M., Terhorst, J. & Song, Y.S. Inference of population history using coalescent HMMs: review and outlook. *Current Opinion in Genetics & Development* **53**, 70-76 (2018).
101. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**, 708-15 (2004).
102. Nam, K. *et al.* Extreme selective sweeps independently targeted the X chromosomes of the great apes. *Proceedings of the National Academy of Sciences* **112**, 6413-6418 (2015).
103. Patterson, N. *et al.* Ancient Admixture in Human History. *Genetics* **192**, 1065-1093 (2012).

104. Lipson, M. & Reich, D. A working model of the deep relationships of diverse modern human genetic lineages outside of Africa. *Molecular Biology and Evolution*, msw293 (2017).
105. Bastide, H., Lange, J.D., Lack, J.B., Yassin, A. & Pool, J.E. A Variable Genetic Architecture of Melanic Evolution in *Drosophila melanogaster*. *Genetics* **204**, 1307-1319 (2016).
106. Bell, G. The Oligogenic View of Adaptation. *Cold Spring Harbor Symposia on Quantitative Biology* **74**, 139-144 (2009).
107. Prado-Martinez, J. *et al.* Great ape genetic diversity and population history. *Nature* **499**, 471-475 (2013).
108. Robinson, J.A. *et al.* Analysis of 100 high-coverage genomes from a pedigreed captive baboon colony. *Genome Research* **29**, 848-856 (2019).
109. Stevison, L.S. *et al.* The Time Scale of Recombination Rate Evolution in Great Apes. *Molecular Biology and Evolution* **33**, 928-945 (2016).
110. Paten, B., Herrero, J., Beal, K., Fitzgerald, S. & Birney, E. Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res* **18**, 1814-28 (2008).
111. Slatkin, M. & Racimo, F. Ancient DNA and human history. *Proceedings of the National Academy of Sciences* **113**, 6380-6387 (2016).
112. Jonsson, H. *et al.* Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519-522 (2017).

Appendix 1: Model and inference method

Here we detail the model and inference method used in this study. In Appendix 1.1, we describe our model for the effects of background selection and selective sweeps and our approach to inferring the parameters of these models. This section is adapted from Elyashiv and colleagues¹, who applied a similar approach to data from *Drosophila melanogaster*; we reproduce it here for completeness. In Appendix 1.2, we describe how we calculate lookup tables for the effects of background selection and sweeps, which our inference relies upon. We introduce several changes to the methods used in previous studies^{1,2}, which allow us to better control the precision of maps of the effects of linked selection. In Appendix 1.3, we describe how we represent neutral polymorphism data and maps of the effects of linked selection in our calculations in order to increase computational tractability. In Appendix 1.4, we describe the optimization algorithm that we use to find the selection parameters that maximize our models composite-likelihood, and we apply the optimization to simulated datasets in order to demonstrate its efficacy and robustness. In Appendix 1.5, we introduce a thresholding approach that contends with biases in our optimization that arise from model misspecification, and we investigate how this thresholding affects our inferences. Finally, in Appendix 1.6, we provide an overview of the software that we use for inference and for other key analyses in the paper. The software, its documentation, and maps of the effects of linked selection are available for download at (github.com/sellalab/HumanLinkedSelectionMaps).

1.1 Model and inference problem

We model the effects of background selection and selective sweeps on neutral heterozygosity levels (i.e., the probability of observing different alleles in a sample size of two), π , at an autosomal position x . In a coalescent framework, the model takes the form

$$\pi(x) = \frac{2u(x)}{2u(x) + 1/(2N_e B(x)) + S(x)}, \quad (1)$$

where $u(x)$ is the local mutation rate, N_e is the effective population size without linked selection, $B(x)$ is the local (multiplicative) reduction in the effective population size due to background selection and $S(x)$ is the local coalescence rate caused by selective sweeps^{1,3}. This approximation can be derived by considering the probability that a mutation occurs (at a rate $2u(x)$ per generation) before the pair of lineages coalesces, owing either to genetic drift ($1/2N_e B(x)$), which includes the effect of background selection, or to a selective sweep ($S(x)$). While we consider autosomes, the model can be extended to sex chromosomes with straightforward modifications.

The model for the effects of background selection, $B(x)$, follows Hudson & Kaplan⁴ and Nordborg et al.⁵ (Fig. A1a). We assume a set of distinct annotations $i_B = 1, \dots, I_B$ under purifying selection (e.g., conserved exonic and non-exonic regions) and positions in the genome $A_B = \{a_B(i_B) | i_B = 1, \dots, I_B\}$, where $a_B(i_B)$ denotes the set of genomic positions with annotation i_B . The selection parameters at these annotations are given by $\Theta_B = \{(u_d(i_B), f(t|i_B)) | i_B = 1, \dots, I_B\}$, where u_d is the rate of deleterious mutations and $f(t)$ is the distribution of selection coefficients in heterozygotes for a deleterious mutation. The reduction in the effective population size is then

$$B(x|A_B, \Theta_B, R) = \text{Exp} \left(- \sum_{i_B} \sum_{y \in a_B(i_B)} \int \frac{u_d(i_B)}{t(1 + r(x, y)(1 - t)/t)^2} f(t|i_B) dt \right), \quad (2)$$

where R is the genetic map and $r(x, y)$ is the genetic distance between the focal position x and positions y (only positions on the same chromosome are considered). The integrand reflects the effect that a site under purifying selection at position y exerts on a neutral site at position x . This expression and its combination across sites provide a good approximation to the effect of background selection so long as selection is sufficiently strong (i.e., when $2N_e t \gg 1$).

Figure A1. Modeling and inferring the effects of linked selection in humans. Given the targets of selection and corresponding selection parameters (a and b), we calculate the expected neutral diversity levels along the genome (c). We infer the selection parameters by maximizing their composite-likelihood given observed diversity levels (c). Based on these parameter estimates, we calculate a map of the expected effects of linked selection on diversity levels.

In turn, the model for the effect of selective sweeps follows from an approximation used by Barton⁶ and Gillespie⁷, among others (Fig. A1b). Similarly to the model for background selection, we assume a set of distinct annotations $i_S = 1, \dots, I_S$ subject to sweeps, but here the specific positions at which substitutions have occurred are known, $A_S = \{a_S(i_S) | i_S = 1, \dots, I_S\}$ with $a_S(i_S)$ denoting the set of substitution positions with annotation i_S . The selection parameters at these annotations are $\Theta_S = \{(\alpha(i_S), g(s|i_S)) | i_S = 1, \dots, I_S\}$, where α is the fraction of substitutions that are beneficial and $g(s)$ is the distribution of their additive selection coefficients. For autosomes, the expected rate of coalescence per generations at position x due to sweeps is then approximated by

$$S(x|A_S, \Theta_S, R, \bar{N}_e, T) = \frac{1}{T} \sum_{i_S} \alpha(i_S) \sum_{y \in a(i_S)} \int \text{Exp}(-r(x, y)\tau(s, \bar{N}_e)) g(s|i_S) ds, \quad (3)$$

where T is the length of the lineage (in generations) over which substitutions occurred, the positions of substitutions y are summed over the chromosome with the focal site, $\bar{N}_e$ is the average effective population size and $\tau(s, \bar{N}_e)$ is the expected time to fixation of a beneficial substitution with selection coefficient s and given an effective population size $\bar{N}_e$. We use the diffusion approximation for the fixation time

$$\tau(s, N_e) = \frac{2(\ln(4N_e s) + \gamma - (4N_e s)^{-1})}{s}, \quad (4)$$

where γ is the Euler constant⁸. This model relies on several simplifying assumptions and approximations. In particular, the term $1/T$ relies on an assumption of one substitution per site per lineage and neglects variation in the length of lineages across loci. In combining the effects over substitutions, we further assume that the timings of beneficial substitutions are independent and uniformly distributed along the lineage, and that they are infrequent enough such that we can

ignore interference among them⁹. The exponent approximates the probability of coalescence of two samples due to a classic sweep with additive selection coefficient s (where $2N_e s \gg 1$) in a panmictic population of constant effective size $\bar{N}_e$. (For the relationships between these expressions and other kinds of sweeps see SOM Section D in Elyashiv et al.¹). In principle, we should use the local N_e incorporating the effects of background selection but given the logarithmic dependence of Equation (3) on N_e , we simply use the average $\bar{N}_e$.

To infer the selection parameters Θ_B and Θ_S , we use a composite-likelihood approach across sites and samples¹⁰ (Fig. A1). We denote the positions of neutral sites by X and the set of samples by I . We then summarize the observations by a set of indicator variables across sites and all pairs of samples $O = \{O_{i,j}(x) | x \in X, i \neq j \in I\}$, where $O_{i,j}(x) = 1$ indicates that samples i and j ($j \neq i$) differ at position x and $O_{i,j}(x) = 0$ indicates that they are the same. In these terms, the composite log-likelihood takes the form

$$\log(L) = \sum_{x \in X} \sum_{i \neq j \in I} \log(\Pr\{O_{i,j}(x) | \Theta_B, \Theta_S\}), \quad (5)$$

where

$$\Pr\{O_{i,j}(x) | \Theta_B, \Theta_S\} = \begin{cases} \pi(x | \Theta_B, \Theta_S) & O_{i,j}(x) = 1 \\ 1 - \pi(x | \Theta_B, \Theta_S) & O_{i,j}(x) = 0 \end{cases}$$

Using composite-likelihood circumvents the complications of considering linkage disequilibrium (LD) and of coalescent models for larger sample sizes. Importantly, maximizing this composite-likelihood should yield unbiased point estimates^{11,12}. Beyond losing the information in LD patterns and in the site frequency spectrum, the main cost of this approach is the difficulty in

assessing uncertainty in parameter estimates (as standard asymptotic results do not apply). We therefore use other ways to assess the reliability of our inferences.

To make the composite-likelihood calculations (i.e., the calculation of $\pi(x|\Theta_B, \Theta_S)$) feasible genome-wide, we discretize the distribution of selection coefficients on a fixed grid. Given a grid of negative and positive selection coefficients, t_g and s_k , $g = 1, \dots, G$ and $k = 1, \dots, K$, the distribution of selection coefficients for each annotation becomes a set of weights on this grid, $w(t_g|i_B)$ and $w(s_k|i_S)$. (In principle, the grid could also be annotation-specific.) For background selection, these weights reflect the rate of deleterious mutations with a given selection coefficient and their sum should therefore be bound by the maximal deleterious mutation rate per site. For sweeps, the weights reflect the fraction of beneficial substitutions with a given selection coefficient and their sum should be bound by 1. In these terms, the effect of background selection takes the form

$$B(x|\Theta_B) = \text{Exp}\left(-\sum_{i_B} \sum_{g=1}^G w(t_g|i_B) b(x|t_g, i_B)\right), \quad (6)$$

where $\text{Exp}(-b(x|t_g, i_B))$ is the proportional reduction in the effective population size induced by having one deleterious mutation per generation per site with selection coefficient t_g at all the positions in annotation i_B . By the same token, the effects of sweeps take the form

$$S(x|\Theta_S) = \frac{1}{T} \sum_{i_S} \sum_{k=1}^K w(s_k|i_S) s(x|s_k, i_S), \quad (7)$$

where $\frac{1}{T} s(x|s_k, i_S)$ is the probability of coalescence per generation induced by sweeps in annotation i_S , if all the substitutions in this annotation are beneficial with selection coefficient

s_k . By using a grid, we can calculate a lookup table of $b(x|t_g, i_B)$ and $s(x|s_k, i_S)$ once and then use it repeatedly to calculate the likelihood of different sets of weights. Moreover, the interpretation of estimated distributions on a grid is arguably simpler than that of the continuous parametric distributions commonly used (e.g., gamma and exponential), which impose rigid interdependencies between the densities associated with different selection coefficients with little justification and while the data is only informative about a subset of the domain. In the next section, we describe additional simplifications in the calculation of $b(x|t_g, i_B)$ and $s(x|s_k, i_S)$.

Other parameters are estimated as follows. Consider equation (1) rewritten as

$$\pi(x) = \frac{\pi_0 \cdot (u(x)/\bar{u})}{\pi_0 \cdot (u(x)/\bar{u}) + 1/B(x) + S(x; \bar{N}_e, T)}, \quad (8)$$

to clearly specify all the additional parameters required for inference. $\pi_0 \equiv 4N_e \bar{u}$ is (approximately) the average neutral heterozygosity, given the effective population size in the absence of linked selection and the average mutation rate per site ($\bar{u}$); π_0 is estimated through the likelihood maximization. The local variation in mutation rate $u(x)/\bar{u}$ is estimated based on substitution rates at putatively neutral sites in an eight-primate phylogeny (excluding humans) in nonoverlapping windows, with a window size chosen to balance true variation in mutation rates and measurement error (see Appendix 3.3). Finally, $\bar{N}_e$ is estimated based on the average genome-wide heterozygosity at putatively neutral sites, after dividing out by a direct estimate of the spontaneous point mutation rate of 1.2×10^{-8} per site per generation¹³, and $T/2\bar{N}_e$ is estimated by $(\bar{K}/2)/\pi_0$, where $\bar{K}$ is the average number of point substitutions per putatively neutral site on the human lineage (see Appendix 2.7).

1.2 Calculating lookup tables

Here we describe how we calculate the lookup tables for

$$s(x|s_k, i_S) \equiv \sum_{y \in a(i_S)} \text{Exp}(-r(x, y)\tau(s_k, \bar{N}_e)) \quad (9)$$

and

$$b(x|t_g, i_B) \equiv \sum_{y \in a_B(i_B)} \frac{1}{t_g(1 + r(x, y)(1 - t_g)/t_g)^2} \quad (10)$$

at all putatively neutral autosomal positions (x), given annotations (i_B and i_S) and selection coefficients (t_g and s_k). We focus on one annotation and selection coefficient at a time and therefore simplify the notation to $b(x)$ and $s(x)$, and omit the variables in τ and the subscripts of the selection coefficients. When we refer to accuracy in this section, we assume that there is no model misspecification (e.g., that putatively neutral sites are neutral, that sets of selected sites and selection parameter values are accurate, that genetic maps are accurate, etc.); once we control the accuracy in this sense, the main sources of error in our predictions will be due to model misspecification.

Our general approach is to calculate $b(x)$ and $s(x)$ with high accuracy at a subset of positions and to use linear interpolation between them. The distances between these positions are chosen such that maps built using the lookup tables maintain a preset level of accuracy ϵ . Specifically, we require that our approximation $\tilde{s}$ and $\tilde{b}$ at any position x satisfy

$$\left| \frac{\tilde{s}(x) - s(x)}{s(x)} \right| < \epsilon \text{ and } \left| \frac{\text{Exp}(-u_M \cdot \tilde{b}(x)) - \text{Exp}(-u_M \cdot b(x))}{\text{Exp}(-u_M \cdot b(x))} \right| < \epsilon,$$

where u_M is an upper bound on the deleterious mutation rate per site per generation. When these conditions are met one can show (based on Eqs. 6 and 7) that the relative accuracy of S and B , and consequently of the expected neutral diversity level π (based on Eq. 1), are also bound by ϵ .

Sweeps. Assume that we have calculated s accurately at position x and consider the distance Δ at which the relative change in s is bound by ϵ , i.e., where

$$\left| \frac{s(x + \Delta) - s(x)}{s(x + \Delta)} \right| \leq \epsilon. \quad (11)$$

From Eq. 11, we find that

$$\begin{aligned} |s(x + \Delta) - s(x)| &\leq \sum_y |\text{Exp}(-r(x + \Delta, y) \cdot \tau) - \text{Exp}(-r(x, y) \cdot \tau)| \\ &= \sum_y \text{Exp}(-r(x + \Delta, y) \cdot \tau) \cdot \left| 1 - \text{Exp}((r(x + \Delta, y) - r(x, y)) \cdot \tau) \right| \\ &\approx \sum_y \text{Exp}(-r(x + \Delta, y) \cdot \tau) \cdot |1 - \text{Exp}(r(x + \Delta, x) \cdot \tau)| \\ &\approx s(x + \Delta) \cdot (r(x + \Delta, x) \cdot \tau), \end{aligned}$$

where the approximations assume $r(x + \Delta, x) \ll 1$. Consequently, by solving for Δ such that

$$r(x + \Delta, x) = \epsilon / \tau \quad (12)$$

we assure that the relative accuracy between x and $x + \Delta$ is bound by ϵ . We therefore calculate s at the selected set of positions on a chromosome beginning at one end and choosing our step sizes according to Eq. (12) until we reach the other end.

Background selection. Our calculation for background selection is based on the algorithm developed by McVicker and colleagues² (their `calc_bkgd` program) with several important

modifications (Fig. A2). The problems that require these modifications are most pronounced for small selection coefficients, whose background selection effects are localized at short genetic distances from selected segments where they can be quite strong. First, McVicker et al. used an additional lookup table to integrate over the effects of background selection exerted by a contiguous selected segment (SI of McVicker et al.²). This lookup table had poor resolution for small selection coefficients at short genetic distances from selected segments, and we have increased the resolution accordingly to fix the problem. Second, the algorithm for choosing the step size Δ is designed to control the absolute error, such that

$$\left| \text{Exp}\left(-u_M \cdot \tilde{b}(x)\right) - \text{Exp}\left(-u_M \cdot b(x)\right) \right| < \epsilon,$$

rather than the relative error (Eq. 13), which results in large relative errors when background selection effects are the strongest (which is with small selection coefficients). Third, the choice of step size Δ is based on the local behavior of background selection at the previous position, and consequently it sometimes skips over selected segments largely ignoring their highly localized effects (which are due to small selection coefficients). We describe how we resolve the last two problems in turn.

Figure A2. Distribution of relative errors in predictions before and after modifying `calc_bkgd`. We consider the model in which autosomal sites with the top 6% of CADD scores are chosen as selection targets, the deleterious mutation rate is $u_d = 7.4 \cdot 10^{-8}$ per bp per generation and the selection coefficient is the lowest in our grid ($t = 10^{-4.5}$), because this is the case most prone to errors (see text). We calculate B -values accurately (using Eq. 10) at a million positions picked randomly from the 22 autosomes and use these values to calculate the relative errors based on the McVicker et al. algorithm (a) and on our modified algorithm (b). The side panel shows the proportion of sites in which the error exceeds ϵ (below), as well as its breakdown in multiples of ϵ .

Assume that we have calculated b accurately at position x and consider the distance Δ at which the relative change in $\text{Exp}(-u_M \cdot b)$ is bound by ϵ (see Eq. 13), i.e., where

$$\left| \frac{\text{Exp}(-u_M \cdot b(x + \Delta)) - \text{Exp}(-u_M \cdot b(x))}{\text{Exp}(-u_M \cdot b(x + \Delta))} \right| \leq \epsilon. \quad (13)$$

Rearranging the left-hand side, we find that

$$|1 - \text{Exp}(-u_M \cdot (b(x) - b(x + \Delta)))| \leq \epsilon,$$

and assuming that $|u_M \cdot (b(x) - b(x + \Delta))| \ll 1$ we find that this requirement is well approximated by

$$|b(x + \Delta) - b(x)| \approx |b'(x) \cdot \Delta + b''(x) \cdot \Delta^2/2| \leq \epsilon/u_M.$$

As our putative step size, we therefore take the (smallest) solution of the quadratic

$$|b'(x) \cdot \Delta + b''(x) \cdot \Delta^2 / 2| = \epsilon / u_M. \quad (14)$$

As in the case of sweeps, we calculate b at a selected set of positions on a chromosome, beginning on one end and choosing our step sizes in a way that maintains the preset relative accuracy ϵ until we reach the other end. Assuming that we have calculated b accurately at position x , our algorithm for choosing the step size consists of the following steps:

1. If x is at the end of the chromosome, stop.
2. Calculate a candidate step size Δ^* by solving Eq. 14.
3. If Δ^* is greater than a preset maximal step size Δ_{max} then set $\Delta^* = \Delta_{max}$.
4. If there is a selected segment between positions x and $x + \Delta^*$ then set Δ^* such that $x + \Delta^*$ is the midpoint between x and the beginning of the (closest) selected segment. This step assures that we do not ‘skip’ selected segments.
5. Convert Δ^* from Morgans to base-pairs, rounding downwards. But if the step ≤ 1 bp then set it to 1 bp, calculate $b(x + \Delta^*)$, set x to $x + \Delta^*$, and return to step 1.
6. Calculate $b(x + \Delta^*)$. If $|b(x + \Delta^*) - b(x)| > \epsilon / u_M$ then set the step size in Morgans to $\Delta^* / 2$ and return to step 4. Otherwise, set x to $x + \Delta^*$ and return to step 1.

Interpolation and representation of lookup tables. We calculate $b(x)$ or $s(x)$ at every autosomal position x (for a given selection coefficient and selected annotation) by linear interpolation between adjacent positions at which we calculated s and b accurately. We then discretize the values of $b(x)$ or $s(x)$ on a linear grid of values corresponding to the preset accuracy ϵ , and group together contiguous autosomal segments with the same discrete value. We

intersect these segments with our list of putatively neutral sites (Appendix 3.1) to obtain lookup tables consisting of contiguous segments of putatively neutral sites with the same coarse-grained s and b values for our sets of selected annotations and selection coefficients.

1.3 Binning neutral sites

A direct calculation of the composite log-likelihood function for given sets of selected annotations and selection coefficients and parameters (Eq. 5) requires that we store and access lookup tables and calculate the log-likelihood function at $\sim 6.5 \times 10^8$ putatively neutral autosomal sites (see Appendix 2.1). Doing so would entail high computation and memory demands in the search for selection parameters that maximize the composite-likelihood. For example, our best-fitting models of background selection (see Main Text) with a grid of 6 selection coefficients would require storing and repeatedly accessing lookup tables that amount to $6.5 \times 10^8 \times 8 \times 6 \approx 32\text{GB}$ (given a precision of $\epsilon = 0.01$), and models involving multiple annotations for background selection and sweeps push the memory requirement to hundreds of GBs.

We reduce the computational and memory demands by dividing the set of putatively neutral sites into bins in which all the effects of background selection and sweeps predicted by the lookup tables and our estimates of the local (relative) mutation rate ($u(x)/\bar{u}$ in Eq. 8; Appendix 3.3) are identical. The composite log-likelihood function can then be calculated by summing over log-likelihood functions corresponding to bins, where the calculation per bin requires only the bin-specific parameters and bin-specific summaries of polymorphism. The number and identity of bins varies with the sets of selected annotations and selection coefficients and parameters and

with the precision (ϵ). For our best-fitting models, the average number of sites per bin is ~ 100 , implying a ~ 100 -fold reduction in demands on memory and in the number of log-likelihood calculations. For our most complex selection models (Appendix 4), the binning reduces memory and computational demands tenfold.

1.4 Optimization

Here we describe how we developed and tested the algorithm we use in order to find the selection parameters that maximize the composite-likelihood of our different models. The high dimensional parameter space (including up to 55 parameters in the most complex model in Appendix 4) potentially makes this optimization problem non-trivial.

One step optimization. First, we tested the performance of standard optimization algorithms from the SciPy minimization toolkit¹⁴. To this end, we generated polymorphism datasets based on our best-fitting model of background selection based on phastCons conservation scores, as follows:

- 1) We fixed the total deleterious mutation rate to $u_d = 10^{-8}$ per base per generation, and randomly divided it among the 6 selection coefficients of the model by sampling from a Dirichlet distribution (with $\alpha = 1$). We set the expected neutral diversity level in the absence of background selection to $\pi_0 = \pi_{YRI}$, where π_{YRI} is a value of π_0 from an iteration of our best-fitting phastCons-based model using polymorphism data from the Yoruba (YRI) population (Appendix 2.1).

- 2) We generated the map of expected neutral diversity levels in autosomes given the chosen parameters. The map was represented in terms of the expected levels at each bin of putatively neutral sites (see Appendix 1.3).
- 3) We generated a polymorphism dataset corresponding to a sample size $n = 108$ pairs of (haploid) autosomes by picking the number of pairwise differences in each bin such that the average diversity level in it most closely matched the level predicted by the map. The discretization step introduces small differences between average and expected diversity levels in bins.

We tested each algorithm by applying it to *10 simulated datasets*, with *3 sets of initial conditions* for each dataset, corresponding to weak, intermediate and strong background selection (with $u_d = 5 \times 10^{-10}$, 5×10^{-9} , and 5×10^{-8} per base per generation, respectively), and *5 randomly chosen initial conditions* in each set (with the total rate divided among the 6 selection coefficients by sampling from a Dirichlet distribution with $\alpha = 1$) amounting to *150 runs*. The initial value of π_0 was always set to the average diversity level in the dataset $\bar{\pi}$.

None of the algorithms closely converged to the ground truth parameters in all cases. Nelder-Mead downhill simplex minimization¹⁵ (NM) and Constrained Trust Region minimization¹⁶ (CTR) performed the best overall, closely recovering the true parameters in $\sim 2/3$ of cases. While CTR was slightly more reliable, it was also up to ten times slower than NM. We therefore decide to combine them in order to leverage the relative strengths.

Two-step minimization algorithm. After some experimentation we converged on the following two-step algorithm (Fig. A3):

- 1) We apply NM with multiple initial conditions. For models of background selection with a single selected annotation we generate 3 sets of initial conditions, with 5 randomly chosen initial conditions per set, as we described above. For models of sweeps with a single annotation we generate the initial conditions analogously. Namely, we generate 3 sets of initial conditions corresponding to a low, intermediate and high proportion of beneficial substitutions (with $\alpha = 0.0125, 0.125$ and 1 , respectively) with 5 randomly chosen initial conditions per set (with the total proportion divided among selection coefficients by sampling from a Dirichlet distribution with $\alpha = 1$). For models with background selection and sweeps and/or multiple annotations, we generate 3 sets of initial conditions, corresponding to the weak/low, intermediate, and strong/high categories, with 5 random initial conditions per set that are chosen similarly for each mode and annotation. In all cases, the initial value of π_0 is set to the average diversity level in the dataset $\bar{\pi}$.
- 2) We apply the CTR algorithm with a single initial condition that is chosen based on the output of the previous step. Specifically, we focus on the sets of selection parameters inferred in the 3 out of 15 initial runs that yielded the highest composite-likelihood, and use their average as our initial condition.

Figure A3. Illustration of the two-step algorithm. In this example, the optimization is applied to a model of background selection with a single selected annotation and a grid of 6 selection coefficients. See text for details.

We tested the two-step algorithm under a variety of scenarios. When we applied it to the aforementioned ‘deterministically’ simulated datasets corresponding to the best-fitting model of background selection, it always closely recovered the ground truth parameters (Fig. A4). The tiny differences between predicted and simulated diversity levels introduced by discretizing sometimes caused tiny differences between the inferred and ground-truth parameter values (see e.g., Fig. A4c), but the composite log-likelihood of the inferred parameters was always higher, indicating that the algorithm is working well. Moreover, the runtime of the CTR algorithm in step 2 was typically short, presumably because its initial conditions were close to the true maximum.

Figure A4. Comparison of inferred and ground-truth parameters for datasets simulated ‘deterministically’ under the best-fitting background selection model. Panels a-d correspond to different simulated datasets. Boxed region in (c) highlights the small differences between inferred and ground truth parameters introduced by discretization.

We also tested the algorithm on simulated datasets that include substantial noise in diversity levels. We generated the datasets for a sample size $n = 2$ by sampling the number of pairwise differences in a bin of neutral sites from a Binomial distribution with a probability of success that equals the predicted diversity level (replacing step 3 in the simulations described above). The parameters inferred by our optimization algorithm were always similar to those used in the corresponding simulations, but with noticeable differences (Fig. A5). In all cases, however, the composite-likelihood of the inferred parameters was greater than that of the ground-truth parameters indicating that the differences were due to overfitting (which is expected given the noise we introduced in the simulations) rather than a problem in the optimization.

Figure A5. Comparison of inferred and ground-truth parameters for datasets simulated with noise under the best-fitting background selection model. Panels a-d correspond to different simulated datasets.

Lastly, we tested the optimization algorithm on datasets simulated under a joint model of background selection and selective sweeps. We modeled the effects of sweeps driven by

nonsynonymous substitutions, assuming that they made up $\alpha = 0.25$ of the nonsynonymous substitutions on the human lineage since divergence from the common ancestor with chimpanzees (see Appendix 2.7), and randomly dividing this proportion among 6 selection coefficients of by sampling from a Dirichlet distribution (with $\alpha = 1$). We modeled background selection as we detailed above, and generated the dataset using the ‘noisy’ simulation scheme corresponding to a sample size of $n = 2$. The parameters inferred by our optimization algorithm were always similar to those used in the simulations, with greater composite-likelihood of inferred than of ground-truth parameters indicative of overfitting (Fig. A6) as we observed in the case with background selection alone. We obtained similar results when we simulated datasets under a variety of scenarios corresponding to the combinations weak, intermediate and strong background selection ($u_d = 5 \times 10^{-10}$, 5×10^{-9} and 5×10^{-8} per base per generation, respectively) with low, intermediate, and high proportions of beneficial substitutions ($\alpha = 0.0125, 0.125$ and 1 , respectively).

Figure A6. Comparison of inferred and ground-truth parameters for datasets simulated with noise under a joint model of background selection and selective sweeps. Panels a and b correspond to different simulated datasets.

1.5 Thresholding

Our inference is strongly affected by forms of model misspecification that cause erroneous predictions of strong background selection effects (i.e., low values of B) and thus of low diversity levels at a relatively small proportion of neutral sites in our dataset. (We refer to neutral rather than putatively neutral sites for brevity and because low error in the identification of neutral sites is irrelevant to the problem at hand). These kinds of erroneous predictions can occur, for example, at neutral sites near regions that are incorrectly annotated as conserved or that are truly conserved but have proportionally fewer weakly deleterious mutations than most similarly annotated regions (because weakly deleterious mutations have strong localized effects on diversity levels). Even when neutral sites near such regions make up a small proportion of the dataset, having more of them be polymorphic than predicted can substantially reduce the

composite-likelihood of models that may otherwise fit the data well (see Eq. 5), potentially biasing our inference. Here we present evidence for this problem, show how we modify our inference to solve it – by imposing a lower threshold for the value of B in the lookup tables or in the optimization, and address the consequences of this modification.

In Figs. A7-A9, we compare the results of our inference with and without thresholding for our best-fitting CADD-based model (the results for other models are qualitatively similar). Under the aforementioned forms of model misspecification, we might expect excess neutral diversity in regions where background selection is predicted to be strongest. Accordingly, when we apply the inference with little or no thresholding and focus on 1% of neutral sites where background selection is predicted to be the strongest, we find that observed diversity levels are up to twofold higher than our predictions (Fig. A7a). Additionally, we expect this form of model misspecification to bias the inferred distribution of selection effects toward larger selection coefficients, because smaller selection effects cause a more localized reduction in diversity levels and are therefore expected to be heavily penalized by having even relatively few misspecified regions. Accordingly, we find that the inferred distribution without thresholding is shifted toward greater selection coefficients ($t \geq 10^{-2.5}$) compared to the distributions with thresholding (Fig. A7c(i)).

Importantly, the map of background selection effects generated without thresholding fits the data more poorly than the maps with thresholding. Notably, when we compare observed and predicted diversity levels around nonsynonymous substitutions, we find that the predictions generated without thresholding underestimate the reduction in diversity levels near

nonsynonymous substitutions (inset in Fig. A7b). This can be explained by the bias toward larger selection coefficients, which causes the inference without thresholding to underestimate the reduction in diversity levels near conserved regions that are specified correctly (in order to avoid the reduction in diversity levels near misspecified regions). Additionally, when we compare the fit of maps with and without thresholding, we find that without thresholding the composite-likelihood is lower (Fig. A7c(iv)), the variance in diversity levels explained throughout the range of window sizes is lower (Figs. A7d and A8) and the calibration of our predictions is poorer (Fig. A7a; this remains the case when we exclude the top and bottom 5% of our predicted values, such that the predictions with and without thresholding span the same ranges of values; e.g., Pearson R^2 of 0.99 and 0.97 with a threshold of $B = 0.6$ and without thresholding, respectively).

Figure A7. Comparison of inference results with and without thresholding. The results shown correspond to our best-fitting CADD-based model (see Main Text), with threshold values of $B = 0$ (without threshold, labeled ‘none’), 0.2, 0.5 and 0.6 applied in the lookup tables. a) Observed vs. predicted neutral diversity levels across the autosomes. The graph was generated as detailed in Fig. 5. Note that the division of neutral sites among bins varies with the choices of thresholds because it is based on corresponding maps. b) Observed vs. predicted neutral diversity levels as a function of genetic distance from human-specific nonsynonymous (NS) substitutions. The graph was generated as detailed in Fig. 3, using a narrower range of genetic distances to NS substitutions to highlight differences among thresholds. c) Parameter estimates and summaries of the inferences. From left to right: i) The estimated distribution of fitness effects, described in terms of the rate of mutation per generation with a given selection coefficient. Mutation rates (throughout) are measured relative to the estimate of the total mutation rate in humans, $u_0 = 1.4 \cdot 10^{-8}$ per bp per generation (see Appendix 5). ii) The total deleterious mutation rate (u_d) measured in units of u_0 . iii) Our prediction of the mean reduction in neutral diversity level due to background selection, measured as the ratio of the average predicted level across the genome, $\bar{\pi}$, to the predicted level in the absence of selection at linked sites, π_0 . iv) The reduction in composite log-likelihood (CLL) per site relative to the model with the highest CLL. Differences in CLL should be interpreted with caution, as this measure does not account for linkage disequilibrium. d) The proportion of variance in diversity levels explained (R^2) on different spatial scales (measured in non-overlapping windows).

We considered two ways of thresholding, where in both we set any value of B that is below the threshold to the threshold value: 1) applying the threshold in the lookup tables, i.e., *before* the composite-likelihood maximization step, and 2) applying the threshold *at each step* of the maximization, when B values are calculated for a given distribution of selection effects (see Eq. 6). The two approaches yield similar improvements in fit at equivalent threshold levels, and even

applying a relatively low threshold improves fits markedly compared to B -maps without thresholding (Fig. A8). Based on our metrics of fit, we find that applying a threshold of $B = 0.6$ in the lookup tables yields the best fits (Fig. A7-9), although thresholds within the range $0.45 \leq B \leq 0.65$ yield comparable results. Nonetheless, lower thresholds yield better fits to data in regions of the genome where selection is particularly strong (e.g., Fig. A7a-b, red box in Fig. A9). It may therefore be useful to use a lower threshold when considering regions of the genome that are subject to especially strong background selection. We provide B -maps for a range of thresholds that can be downloaded at github.com/sellalab/HumanLinkedSelectionMaps (in addition to the ‘best-fitting B -maps’ presented in the Main Text).

Figure A8. The proportion of variance explained in 10 kb, 100 kb and 1 Mb windows using a range of B thresholds applied to lookup tables (‘lookup’) or during maximization (‘maximization’).

Figure A9. Predicted and observed diversity levels along chromosome 1 in the YRI sample. Diversity levels are measured in 1 Mb windows, with a 0.5 Mb overlap, with the autosomal mean set to 1. Thresholds were applied in the lookup tables. Lower thresholds yield better predictions in regions with low diversity levels, e.g., near 50 Mb (red box).

While thresholding largely resolves the aforementioned problem of model misspecification, it also introduces some problems. First, as we already noted, it leads to an underestimation of background selection effects at ~4% of the genome in which background selection effects is predicted to be the strongest. Second, thresholding potentially biases our estimates of the distribution of selection effects. While this bias is probably smaller than the bias without thresholding, its form and magnitude are not obvious. This is why we decided not to report the inferred distributions of selection effects in the Main Text. We are working on more principled ways of resolving the problems introduced by model misspecification, but these fall beyond the scope of the current paper.

1.6 Software

We provide a set of Python programs to download and format the genomic data that we use (see Appendix 2), infer maps of the effects of linked selection and reproduce all of the analyses and figures described in this study (github.com/sellalab/HumanLinkedSelectionMaps). We rely on publicly available software for some steps, including the PHAST package^{17,18}, which we use to identify conserved regions and to estimate substitution rates (see Appendices 3 and 5), and a modified version of the `calc_bkgd` program from McVicker and colleagues², which we use to generate lookup tables of the effects of background selection (see Appendix 1.2).

Running the inference pipeline. The inference pipeline is controlled by a data structure called `RunStruct`, which is initialized with information about input/output file paths used, model parameters and other control variables, such as the precision ϵ of lookup tables (see Appendix 1.2) and the B threshold (see Appendix 1.5). Once `RunStruct` has been initialized, the pipeline proceeds through the following steps:

- 1) Download and organize input files (annotations, genetic maps, etc.).
- 2) Create lookup tables of the effects of background selection and/or selective sweeps (Appendix 1.2) for the given set of selected annotations and grid of selection coefficients.
- 3) Organize polymorphism dataset that includes polymorphism data at putatively neutral sites (Appendix 2.1), corresponding estimates of substitution rates (Appendix 3.3) and corresponding values of lookup table into our compressed bins format (Appendix 1.3).
- 4) Run the two-step optimization algorithm to obtain estimates of model parameters, a map of the predicted effects of linked selection, and summary statistics including, e.g., the estimated

deleterious mutation rate (u_d) and proportion of beneficial substitutions (α) associated with different annotations and the average reduction in diversity levels ($\bar{\pi}/\pi_0$).

Parallelization and runtimes. The composite-likelihood calculations during optimization can be partitioned into sums over subsets of bins of neutral sites, which in turn allows us to parallelize the optimization. The number of processing cores used in optimization is controlled by RunStruct. For our best-fitting models of background selection, loading lookup tables and neutral polymorphism data and running the two-step optimization requires ~ 1 GB of memory for each of the 15 processes in step 1 and the single process in step 2. Running each process on a single core takes ~ 12 -24 hours or ~ 200 -400 CPU $\times$ GB hours. The computing cluster we used allows up to 12 cores per process and thus using parallelization we were able to run the optimization for the best-fitting models in 1-2 hours. Our most complex models (see Appendix 4) required up to 10GB of memory per process and took up to 60 hours with using 12 cores (i.e., $\sim 10^4$ CPU $\times$ GB hours).

Appendix 2: Data sources and filters

2.1 Polymorphism data

We download 1000 Genomes Project phase 3 VCF files for all 26 populations from across the world¹⁹. Unless otherwise noted, results in the Main Text and Supplementary Online Materials are based on autosomal data from Yoruba (YRI); the results for other populations are reported in Appendices 7 and 9 of this Supplementary Online Materials.

We apply several filters to these data. First, we restrict our analysis to bases that pass all filters, denoted ‘P’ in the 1000 Genomes Project strictMask accessibility mask^{19,20}. In addition, we remove low-complexity, simple repeats, duplications, and hg19 build gaps using repeatMasker files downloaded from UCSC²¹. For each population, we restrict polymorphic sites to that population’s subset of biallelic SNPs from VCF files, excluding indels and other variants using VCFTools²². Remaining sites are treated as monomorphic.

We apply additional filters to restrict our analyses to putatively neutral sites. First, we remove the union of genic regions, as detailed in section 2.4. Second, we remove all remaining sites with phastCons conservation scores greater than 0.001 as described in section 3.1. Third, we remove putatively neutral sites at the telomeric ends of autosomes, near the edges of our genetic maps (Appendix 2.3), as detailed in Appendix 3.2. Accessibility and repeat masks remove ~33.3% of all autosomal sites; excluding genic regions removes an additional ~3.3%; filtering based on phastCons scores removes another ~40.5%; and filtering sites at the telomeric ends removes

~1.2% more. We are left with a set of ~653M putatively neutral sites, which correspond to ~23% of autosomal sites (based on hg19 build).

2.2 Multiple species alignment data

We rely on multiple sequence alignments to identify phylogenetically conserved and non-conserved regions of the genome, as well as for estimating local variation in neutral substitution rates (see Appendices 3 and 4). To this end, we download mutation annotation format (MAF) files containing 99 vertebrate genomes aligned to the human genome (build hg19), using the Multiz software from UCSC²³.

2.3 Genetic map

We use the Hinch et al.²⁹ genetic map, which was inferred from ancestry switches in African-Americans. At the >10 kb scale, it is highly correlated to other fine-scale maps (e.g., Halldorsson et al.³⁰, International HapMap Consortium³¹). Among high-resolution genetic maps in humans, however, this one is likely the least confounded by diversity levels along the genome.

2.4 Human gene annotations

We use genic annotations from the UCSC knownGene database³² to identify putative targets of selection as well as regions that should be removed from our set of putative neutral sites. To this end, we rely on exon coordinates from knownGene transcripts to identify four kinds of annotations: 1) upstream/downstream regions, defined as 1 kb upstream of a transcript start and 1 kb downstream of a transcript end; 2) untranslated region (UTR), both 5' and 3'; 3) protein

coding sequences (CDSs); and 4) splice regions, defined as 200 bp from the start and end of each intron.

For the purpose of identifying putative targets of selection, we rely on a non-overlapping subset of these four annotations. For genes with multiple splice variants, we keep only the set of exons within the longest isoform. In rare cases of two overlapping gene predictions, we retain the gene with the longer exonic sequence. For the purpose of removing putatively functional regions from our set of putative neutral sites, however, we remove the union of all four annotations for all gene transcripts.

2.5 CADD scores

We use CADD scores^{27,28} in order to annotate putative targets of selection in a couple of models (Appendices 4.4 and 4.5). The standard CADD scores rely on the map of background selection effects generated by McVicker et al. (2009) as one of their inputs. While this input has minor effects on CADD scores (i.e., the top 1-10% of scores; see Supplementary Table 3 in Kircher et al.²⁷), in order to avoid any measure of circularity we approached the Kircher Lab (Martin Kircher, Lusiné Nazaretyan, Philip Rentzsch and Max Schubach), who manage the development of CADD scores, and who kindly agreed to generate and share a version of CADD score without the background selection map as input (this set of CADD scores is available on request from either the Kircher or Sella labs). For each site in the genome, we retain the highest of the three CADD scores (corresponding to the three possible point mutations). We use the distribution of scores across the autosomes to determine cutoffs for our annotations (e.g., sites within the top 6% of scores) and use sites with scores that exceed these cutoffs as putative targets of selection

(sometimes in conjunction with another annotation, e.g., exons).

2.6 ENCODE cCRE annotations

In two of our models (Appendix 4.4), we consider regulatory elements identified by the ENCODE project as putative targets of selection²⁶. To this end, we download ENCODE candidate cis-regulatory elements (cCREs) from the Tier 1a group of biosamples, which include experimental support from all relevant assays used to define elements: high DNase signal and high H3K4me3, H3K27ac or CTCF signal²⁶. The resulting cCREs are categorized as 1) enhancer-like signatures (ELS), 2) promoter-like signatures (PLS), 3) CTCF-bound (CTCF) and 4) poised elements marked by DNase and H3K4me3 (H3K4me3). cCRE annotations were downloaded for each individual Tier 1a biosample using the SCREEN tool²⁶ and lifted over from hg38 to hg19 coordinates.

2.7 Substitutions in the human lineage

We rely on an estimate of the human-chimpanzee ancestor inferred using the Enredo-Pecan-Ortheus (EPO) 6-species alignment pipeline²⁴ to identify likely substitutions on the human lineage. We use subsets of these substitutions that arose in putative targets of positive selection as candidate substitutions resulting in classic sweeps (Appendix 4.5). We derive sets of likely substitutions in a couple of different ways. First, we compare the reconstructed ancestral genome with the human hg19 reference, taking the differences as putative substitutions. In this case and others, we do not differentiate between low and high confidence calls (lower and upper case, respectively) in the estimated ancestor. Because the hg19 reference genome is a composite of

genomes with different ancestries²⁵, we also consider population-specific inferences of substitutions for YRI and CEU. To this end, we compare the reconstructed ancestral genome with the polymorphism data collected in the 1000 Genome Project for a given population. If a site is monomorphic in the population and differs from the HC ancestor, we include the site in our set of substitutions. For biallelic sites where one of the two alleles is ancestral, we randomly choose one of the alleles with probabilities that are weighted by allele frequency; if the chosen allele is the derived one, the site is considered a substitution. We generate two such samples for a given population to see whether different choices of substitutions affect our results. In practice, each of these sets differs from the set based on the hg19 reference at fewer than 1% of sites, the differences between the two samples for a given population are even smaller, and the results of our inference end up being insensitive to these differences (Appendix 4.6).

2.8 Covariates of B

In Appendix 8, we ask whether genomic features that covary with B could account for the divergence between observed and predicted diversity levels in the ~2% of sites in which background selection is predicted to be the weakest. In addition to annotations of features whose sources were already mentioned, we also use the following datasets: 1) BED files of CpG islands downloaded from the UCSC Table Browser²¹; 2) BED files of testis CpG methylation levels in downloaded from the GEO database (GEO accession: GSM1127119)³³; 3) coordinates of C>G hypermutable regions, given at 1 Mb resolution, taken from the Supplemental Information of Jonsson et al.³⁴; 4) coordinates of centromeres and telomeres taken from the hg19 gaps track in the UCSC Table Browser²¹; 5) inferred proportions of archaic ancestry in European (CEU) and East-Asian (CHB/CHS) populations based on estimates from Steinrücken et al.³⁵.

Appendix 3: Choice of exogenous parameters

Fitting our model to data requires several choices beyond those of datasets and filters. Here, we describe how we chose our set of putatively neutral sites and estimate the substitution rate at these sites. In Appendix 4, we describe how our results depend on the choice of targets of selection.

3.1 Choosing putatively neutral sites based on phylogenetic conservation

Our main source of information for choosing the set of putatively neutral sites is the degree of conservation in multiple species alignments. To this end, we rely on running phastCons¹⁷ on subsets of the 99-vertebrate alignment (from which we exclude the human genome). PhastCons fits a phylogenetic hidden Markov Model (phylo-HMM) with two states, neutral and conserved, to multiple species alignments of contiguous sites along the genome using the relative substitution rates in the alignment columns to infer conservation. The phastCons score is the posterior probability that any given site is conserved. In principle, including more species in the alignment increases the power to distinguish between conserved and neutral sites (Fig. A10a). However, as the phylogenetic distance from humans increases, sequence conservation might become less informative about conservation in humans because of functional turnover^{36,37}. In practice, the latter effect is ameliorated by the fact that phastCons only uses information at aligned sites and the proportion of the genome that aligns to the human reference decreases with phylogenetic distance (Fig. A10b), especially in regions with considerable turnover.

In relying on phastCons scores to identify a set of putatively neutral sites, we need to choose two parameters: the phylogenetic depth of species included in the alignment and the cutoff conservation score below which a site will be considered neutral. In both cases, we pick the parameter values that maximize the variance in diversity levels explained by our best-fitting models (Fig. A11). Given these criteria, we chose to base our set of neutral sites on the alignment of supra-primates (Fig. A11a), and use the 35% of sites (in the set remaining after filters and removing genic regions; see Appendices 2.1 and 2.4) with the lowest phastCons scores in this alignment, which includes sites with scores ≤ 0.001 (Fig. A11b). These choices are robust to the phylogenetic depth used to specify the selection targets (see Appendix 4) and to the window size in which we measure the variance explained by our model (we show the results for windows of 1 Mb in Fig. A11).

Figure A10. The distribution of phastCons scores across autosomes for varying phylogenetic distances from humans. a) The number of species included at each phylogenetic depth is noted in the caption. As the number of species in the alignment increases, the ability to distinguish between conserved and neutral sites increases. b) The proportion of a species' sequenced genome that aligns to the human reference (hg19) decreases with their phylogenetic distance from humans. The decrease is not monotonic because of other factors, e.g., the quality of the sequencing. The proportion is not 1 for humans because of missing information in the reference genome.

Figure A11. The variance in diversity levels explained by our two best-fitting models using different choices of putatively neutral sites. In (a) we vary the phylogenetic depth of the multi-species alignment (i.e., the maximal phylogenetic distance from humans to any/all of the other species) and in (b) we vary the cutoff phastCons score for the least conserved sites included in our set. The best fit corresponds to the least conserved 35% of sites (phastCons scores ≤ 0.001) in the supra-primate alignment (euar).

3.2 Removing sites at the telomeric ends of chromosomes

The Hinch et al. genetic map²⁹ does not include recombination rate estimates for ~0.5-1 Mb at the 5' and 3' ends of autosomes. Consequently, we are unable to describe background selection effects of putatively selected regions that lie in these telomeric regions, and our inferences and predictions at putatively neutral sites near the telomeres are less accurate. We therefore exclude putatively neutral sites in telomeric regions not covered by the genetic map. Similar to our approach in the previous section, we choose the map size of the region to remove based on how the choice affects the model fit to diversity levels across autosomes (Fig. A12a). We find that filtering putatively neutral sites in 0.1 cM from the edge of the genetic map, which amounts to ~0.8% of neutral sites, largely removes this 'edge effect'. This genetic distance makes sense, as it is roughly one at which background selection effects of deleterious mutations with $s = 10^{-3}$ – the strongest selection effects inferred to contribute substantially (Fig. A12b) – become

negligible. Moreover, our estimates of model parameters are fairly insensitive to the removal of larger regions (Fig. A12b).

Figure A12. The effect of removing putatively neutral sites near telomeres on model fit and parameter estimates. We show the result for our best-fitting CADD-based model; results for phastCons scores are highly similar (not shown). **a**) The proportion of variance in diversity levels explained for different window sizes, as a function of the size of the removed region (in cM). **b**) (i-iii) Estimates of model parameters as a function of the size of the removed region (in cM).

3.3 Estimating local variation in mutation rates

We rely on estimates of substitution rates at putatively neutral sites along the genome to control for the effect of variation in mutation rates on neutral diversity levels (see Eq. 1 in Appendix 1.1). To this end, we use phyloFit¹⁸ to estimate the substitution rate in a phylogeny, in windows

of putatively neutral sites across the genome. We choose the species to include in the phylogeny based on the following considerations. The number of substitutions in a given window can be approximated by a Poisson random variable with expectation λ , which is proportional to the total branch length of the phylogeny, T , and the number of putatively neutral sites in the window, n . Consequently, the precision of our estimates of the relative mutation rate increase with $\lambda \propto n \cdot T$. Including more species in the phylogeny increase T but reduces n , because it reduces the fraction of putatively neutral sites that align to the human reference in all the species included. Fig. A13a shows the trade-off between the two factors, for all subsets of 9 primate species included in the 99-vertebrate alignment (see Appendix 2.2). We chose the subset that maximizes $n \cdot T$, which includes 8 of the 9 species (gibbon is removed) with an average of ~ 0.135 substitutions per putatively neutral site.

Figure A13. Choosing the parameters used in estimating the relative mutation rate at putatively neutral sites. a) The trade-off between the fraction of aligned sites and total branch length for subsets of the primate phylogeny. The fraction of aligned sites is estimated

for our set of putatively neutral sites, and the total branch length is measured in terms of the average number of substitutions per site on the phylogeny, estimated by phyloFit. The maximum product of the fraction and branch length is attained by including all primates included in the 99-vertebrate alignment other than gibbon. b) The variance in diversity levels explained by our best-fitting models across 1 Mb windows, for different choices of window sizes (i.e., the number of putatively neutral sites) used to control for variation in mutation rates at putatively neutral sites.

We estimate relative mutation rates along the genome based on the estimated substitution rates in the 8-primate phylogeny in windows with a fixed number of contiguous putatively neutral sites. Using windows with a greater number of sites decreases the sampling error but reduces the spatial resolution of our estimates. We use the variance in diversity levels explained by our best-fitting models as a criterion for choosing the window size, finding that a window with 6000 putatively neutral sites performs best among the options we examined (Fig. A13b). This choice corresponds to mean physical window sizes of 26,454 bp (with a S.D. of 18,455 bp) and to a mean relative error of ~3.3% in our estimates of the relative mutation rate per window. We also examined other ways of estimating the relative mutation rate, including using windows of fixed physical length and sliding windows with varying degrees of overlap, but none of these approaches yielded better results.

In the analyses in which we bin neutral sites, either by their distance to genomic elements (e.g., Fig. 3) or by predicted B (e.g., Fig. 5), we estimate the relative mutation rate in each bin. To this end, we use phyloFit¹⁸ to estimate the substitution rate in the 8-primate phylogeny on all sites in that bin jointly and then normalize this estimate by the average across bins.

Appendix 4: Fitting models with different targets of selection

Our framework allows us to fit models of background selection, selective sweeps, or both, based on different choices of putative targets of negative and/or positive selection. Here we detail the analysis of the models and choices that are described in the Main Text. We use several criteria to evaluate how well the models fit the data; these indicate that models of background selection alone in which the targets of selection are chosen based on constrained elements annotated by either phastCons or CADD scores are best supported by the data. We also compare the predictions of these models with those of McVicker and colleagues².

4.1. Background selection model based on phylogenetic conservation

We first consider a model of background selection in which targets of selection are chosen based on phylogenetic conservation. We identify conserved genomic elements using phastCons scores¹⁷ calculated on monophyletic subsets of the 99-vertebrate alignment to the human genome²³, all of which exclude the human genome itself (see Appendix 2.2). We vary the phylogenetic depth of the subset of species considered (i.e., the maximal distance from humans). For a given depth, we obtain targets of selection by specifying a proportion of selected sites (i.e., of the total autosomal length in hg19) and choosing those sites that have the highest phastCons scores in the alignment (after excluding some sites, e.g., from up to 5% of the four-ape alignment to less than 0.1% of the 99-vertebrate alignment, that are in our putatively neutral set). As we have done for previous choices (e.g., Appendix 3.1), we examine how our choices of phylogenetic depth and of proportion of selected sites affect the models' fit to autosomal diversity levels.

We find the fit to be largely insensitive to the choice of phylogenetic depth, with models based on conservation in the full 99-vertebrate alignment fitting slightly better than other choices of depth (Fig. A14). Notably, the explained variance in diversity levels (in windows of different sizes) is similar across depths, increasing slightly with the number of species included, other than for the four-ape phylogeny (Fig. A14b and c). The fits of predicted diversity levels along the genome (e.g., Fig. A14d) and around genomic features (e.g., Fig. A14e) are similar, with none of the choices of depth clearly outperforming others. Moreover, for all choices, the predicted diversity levels are well calibrated (Fig. A14f), with the exception of regions in which background selection is predicted to be very weak, i.e., $B \approx 1$ (see Appendix 8). When we restrict each annotation to the top 6% of scores in sites for which all phylogenetic depths include phastCons scores (~98% of sites satisfy this criterion), our results are unchanged.

Distantly related species, such as those added when we move from supra-primates (25) to vertebrates out to lamprey (99), have little effect on phastCons scores and thus on our models, because only a small proportion of their genomes align with humans (Fig. A10b). This can be seen in the high correlations between the number of conserved sites based on different depths across windows of different sizes (Fig. A15a). The spatial distribution of conserved sites is even fairly insensitive to varying the species included from four apes to 99 vertebrates (Fig. A15a). Interestingly, we later show that the improvement in fit across 1 Mb windows of the model based on conservation in 99 vertebrates compared with models based on conservation in shallower phylogenies is statistically significant, except for the model based on four-apes (Fig. A33), whereas the spatial distributions of conserved sites in the 99-vertebrate and four-ape models are

the least correlated (Fig. A15). The v-shaped dependence on phylogenetic depth may reflect a tradeoff in which phastCons scores based on deeper alignments have greater power to identify long-lived selected regions (see, e.g., Fig. A10a), whereas those based on apes are better at identifying regions that are selected in humans but exhibited functional turnover in the deeper phylogeny³⁶ (see also Appendix 6.2).

The model fit is also fairly insensitive to the cutoff conservation score used in choosing selection targets, although choosing 5-7% of autosomal sites as targets does appear to yield slightly better fits than other choices (Fig. A16). Notably, the variance explained for different window sizes is maximized between 5-7% (Fig. A16b and c); at the higher end of the range of cutoffs from 2-9%, the fits of diversity levels along the genome (e.g., Fig. A16d) and around genomic features (e.g., Fig. A16e) appear to be slightly worse, and the stratification of observed values by predicted ones spans a smaller range (Fig. A16f). Among comparisons between models based on 6% and all other cutoffs in the range of 2-9%, only 8 and 9% lead to a statistically significant reduction of fit in windows of 1 Mb (Fig. A33). Based on these analyses, we use the model with the 6% of autosomal sites with the highest phastCons scores based on the 99-vertebrate alignment in many of our analyses, and refer to this as our *best-fitting phastCons-based model* in both the Main Text and throughout the Supplementary Online Materials.

Figure A14. Comparison of background selection models based on phastCons conservation scores in phylogenies of difference depths. Shown are results of models based on conservation in four apes, eight primates, 12 prosimians, 25 supra-primates, 50 laurasiatherians, 61 mammals and 99 vertebrates extending out to lamprey. In all cases, we take the 6% of autosomal sites with the highest phastCons scores (excluding putatively neutral sites) as our targets of selection. Throughout the Supplementary Online Materials, with the exception of Appendix 7, we show results using data from YRI. *The panels describe:*

a) Parameters and summaries of models (from left to right): i) Estimated distribution of fitness effects, described in terms of the rate of mutations with given selection coefficients. Mutation rates throughout are measured relative to the estimate of the estimated average mutation rate per bp per generation in humans, $u_0 = 1.4 \cdot 10^{-8}$ (see Appendix 5). As detailed in Appendix 1.5, the inferred distribution of selection coefficients should be interpreted with caution. ii) Estimated total deleterious mutation rate per selected site (u_d) measured in units of u_0 . iii) Estimated autosomal average fold-reduction in neutral diversity levels due to selection at linked sites, i.e., the ratio of average predicted heterozygosity, $\bar{\pi}$, to average predicted heterozygosity in the absence of selection at linked sites, π_0 . iv) The reduction in composite log-likelihood (CLL) per site relative to the model with the highest CLL. Differences in CLL should be interpreted with caution, as diversity levels at putatively neutral sites are not independent. b) The proportion of variance in diversity levels explained (R^2) on different spatial scales (measured in non-overlapping contiguous windows). c) Close-up on the variance explained for several window sizes. d) Predicted and observed diversity

levels along chromosome 1. Diversity levels are measured in 1 Mb windows, with 0.5 Mb overlap, and are normalized by the mean level (as detailed in Fig. 2). The results here and in subsequent panels are shown for a subset of depths, including four apes, 25 supra-primates and 99 vertebrates. e) Predicted and observed diversity levels as a function of genetic distance to the nearest human-specific nonsynonymous (NS) substitutions. The plot was generated as detailed in Fig. 3. Inset shows closeup between -0.05 and 0.05 cM. f) Observed vs. predicted neutral diversity levels across the autosomes. The plot was generated as detailed in Fig. 5.

Figure A15. The spatial distribution of putatively selected sites targets remains similar when we vary the phylogenetic depth of the alignment used to infer conservation (shown in a), and the proportion of sites with the highest conservation scores included (in b). We compare two choices of selection targets at a time, and show the Pearson correlations (ρ) between the numbers of putatively selected sites among windows of different genetic lengths (measured in Morgans). The range of window sizes roughly corresponds to the spatial scales over which selection affects linked neutral diversity for the estimated range of selection effects. When we vary the phylogenetic depth, we use the 6% of autosomal sites with the highest phastCons scores, and when we vary the conservation cutoff, we use phastCons scores based on the 99-vertebrate alignment. c) The deleterious mutation rate per gamete per generation inferred as a function of assumed proportion of selected sites in autosomes.

Figure A16. Comparison of background selection models based on phastCons scores using different proportions of autosomal sites as selection targets. In all cases considered, we rely on conservation in 99 vertebrates. Otherwise, all panels are as described in Fig. A14.

The insensitivity of our fits to varying the conservation cutoff can be understood as follows. phastCons estimates the probability that runs of sites belong to conserved segments¹⁷. When we reduce the conservation cutoff, shorter segments with high scores tend to expand to include adjacent, lower scoring sites. This results in a high spatial correlation between the conserved sites corresponding to different cutoffs (Fig. A15b). Given a lower conservation cutoff and longer ‘selected’ segments, we infer a lower deleterious mutation rate per site (Fig. A16a(ii)) but a similar deleterious mutation rate per segment (see, e.g., Fig. A15c), thereby producing similar troughs in diversity around such segments and similar fits overall.

4.2 Background selection model based on genic annotations

Next, we consider a model of background selection in which selection targets are chosen based on simple genic annotations, i.e., the exons divided into UTRs and protein coding sequences (CDSs), as well as regions in the immediate vicinity of these sequences controlling transcript regulation: regions 1 kb up- and downstream of transcript start/end, and splice regions 200 bp at the start and end of introns^{38,39} (see Appendix 2.4 for details). We allow selection parameters to vary among annotations, but find that in the best-fitting model only protein coding and splice regions have non-negligible deleterious mutation rates (for other annotations, $u_d/u_0 < 10^{-6}$).

We also find that this model fits much worse than our best-fitting phastCons-based model (Fig. A17): the variance in diversity levels it explains is substantially lower across different window sizes (Fig. A17b), its fit to diversity levels along the genome is discernably worse (e.g., Fig. A17c), and when observed diversity levels are stratified by the model's predictions, they are less calibrated (Fig. A17e). The genic model does do reasonably well at predicting how diversity levels drop with genetic distance around nonsynonymous

Figure A17. The background selection model based on simple genic annotations fits worse than our best-fitting phastCons-based model. All the panels are as described in Fig. A14 (but with the hatch-marked blue bars in a(i) and (ii) corresponding to different annotations of the genic model).

substitutions (e.g., Fig. A17d). The generally poorer fit as well as the reasonably good fit around nonsynonymous substitutions can be understood in terms of the overlap between our simple genic annotations and direct measures of constraint (Fig. A18). Namely, the genic annotations miss most constrained sites, which are intronic or intergenic (Fig. A18b), but most protein coding regions (CDSs) are constrained (Fig. A18a) explaining why models including them as an annotation perform well near them.

Figure A18. The relationship between simple genic annotations and our main measures of constraint. Specifically, we examine the overlap of the 6% of autosomal sites with the highest phastCons or CADD scores with the genic annotation detailed in the text; we added intronic (INTRON) and intergenic (INTERG) annotations for completeness. a) The fraction of each genic annotation within the 6% most constrained sites. b) The fraction of the 6% most constrained within each genic annotation. c) Enrichment of genic annotations in the 6% most constrained sites, i.e., the ratio of their proportion among constrained and all autosomal sites.

4.3 Background selection models separating conserved exonic and non-exonic sites

While background selection models based on simple genic annotations do worse than those based on phylogenetic conservation, using such annotations in conjunction with conservation could allow for improved fits. Notably, it is often argued that purifying selection in protein coding regions is stronger than in functional non-coding regions^{36,40}; if this were true, then allowing them to have different selection parameters could result in better fits. To examine this possibility, we fit a model with two types of selection target: exonic (i.e., segments combining CDSs and UTRs) and non-exonic conserved sites (see details in Appendix 2.4).

We infer a higher deleterious mutation rate and stronger selection in exonic compared to non-exonic sites (Fig. A19a), although we note that our estimates of selection parameters could be

affected by thresholding (see Appendix 1.5). The total deleterious mutation rate per gamete is similar in models with and without the exonic/non-exonic division ($U = 1.6$ and $U = 1.73$ per gamete per generation, respectively), but the (weighted) average selection effect is greater in the model with the division ($\bar{s} = 1.71 \times 10^{-3}$ vs. $\bar{s} = 6.8 \times 10^{-4}$ for the models with and without division, respectively), primarily due to stronger selection in conserved exonic sites. Overall, despite affording additional parameters, dividing conserved sites into exonic and non-exonic has little effect on our fits (Fig. A19b-e).

Regardless of whether we separate exonic and non-exonic conserved sites, most of the reduction in diversity levels is caused by selection in non-exonic regions. Weakly selected mutations cause a large reduction in neutral diversity levels over short genetic distances, whereas strongly selected mutations cause a weak reduction over long genetic distances; but the integral reduction in diversity levels due to weak and strong selection on a given set of deleterious mutations end up roughly equivalent⁴¹. This property allows us to use estimates of the total deleterious mutation rates in conserved exonic and non-exonic regions as a rough measure of their proportional effects on neutral diversity levels, despite differences in selection effects in these regions. These estimates suggest that ~80% of deleterious mutations occur in non-exonic regions, indicating that they account for most of the reduction in linked neutral diversity (e.g., in the model with the top 6% of phastCons scores, ~84% of selected sites and ~76% of deleterious mutations are non-exonic; with the top 6% of CADD scores, ~83% of selected sites and ~85% of deleterious mutations are non-exonic; also see discussion in Appendix 4.6).

Given that the bulk of deleterious mutations exerting background selection occur in non-exonic

regions, it is not surprising that a model including only conserved non-exonic sites fits the data only slightly worse than a model including all conserved sites as targets of selection (Fig. A20). By the same token, it is not surprising that a model including only conserved exonic sites fits the data substantially worse than models with either conserved non-exonic or all conserved sites as targets of selection (Fig. A20). Moreover, the estimate of the deleterious mutation rate per site in the exonic model is much higher than in the other two (Fig. A20a(ii)).

Figure A19. Dividing conserved sites into exonic and non-exonic sets leads to different estimates of selection parameters in each, but to little improvement in fit compared to the model based on conservation alone. Our set of conserved sites consists of the 6% of sites with the highest phastCons scores in the 99-vertebrate alignment (see Appendix 4.1). All panels are as described in Fig. A14. Because of thresholding (Appendix 1.5), the model based on conservation alone is not formally nested in the one with the division into exonic and non-exonic sets, explaining how its maximum composite-likelihood can be slightly greater.

Figure A20. Comparison of background models using exonic, non-exonic and all conserved sites as targets of selection. Our set of conserved sites consists of the 6% of sites with the highest phastCons scores in the 99-vertebrate alignment (see Appendix 4.1). All panels are as described in Fig. A14.

It is somewhat surprising that the model based on conserved exonic sites alone fits the data as well as it does (Fig. A20b and c). This can be understood by noting that the spatial distribution of conserved exonic sites and of all conserved sites are fairly highly correlated (Fig. A21). Given similar spatial distributions of selected sites, the distribution of background selection effects in the model with all conserved sites can be approximated by having a higher deleterious mutation rate per site at the fewer selected sites in the exonic model. These considerations explain why we infer a similar (albeit lower) average reduction in diversity levels but a substantially higher deleterious mutation rate in the exonic model (Fig. A20a(ii) and (iii)). They also help to explain

differences between our inferences and those of McVicker and colleagues², notably their implausibly high estimate of the deleterious mutation rate given that their main model assumes selection only at conserved exonic sites (see Main Text and Appendix 4.6).

Figure A21. spatial correlations of exonic, non-exonic and all conserved sites for varying window sizes ('con_e', 'con_n' and 'con_a', respectively).

4.4 Background selection models based on other annotations

We consider two additional widely-used functional annotations as putative background selection targets. First, we rely on the expanded encyclopedias of DNA elements (ENCODE) annotations of candidate cis-regulatory elements (cCREs), including enhancer-like signatures (ELS), promoter-like signatures (PLS), CTCF-bound (CTCF) and poised/DNase-hypersensitive (H3K4me3) assayed in 25 Tier 1a biosamples²⁶, alongside protein coding sequences (CDSs) (see Appendices 2.4 and 2.6 for data sources and definition of elements). ENCODE cCREs attempt to capture the diverse repertoire of regulatory elements across cell types that control gene expression in different cellular and biological contexts. They are based on a large set of

epigenomic assays, including ChIP-seq measuring the occupancy of histone marks associated with both activation and repression of gene expression, pulldown of DNA-bound transcription factors, and DNA accessibility measured in terms of DNase sensitivity. Since we infer the majority of autosomal sites under purifying selection to be non-exonic (see Appendix 4.3), we reason that some combination of cCREs may substantially overlap these sites. Importantly, cCRE annotations may allow us to better partition non-exonic regions into sub-classes of sites experiencing different selection strengths. We define our choices of selection targets (other than CDSs) by grouping cCRE in two alternative ways. In one, we take the union of cCREs of a given type over all 25 biosamples. In the other, we divide cCREs of a given type into those identified in few ($\leq$ median number) or in many ($>$ median number) biosamples (in practice, most cCREs included in the first set are cell-type specific whereas most of those in the second are found in a few to all cell-types). The model in which cCREs of a given type are split performs slightly better, presumably because of the additional degrees of freedom. Both models, however, fit the data substantially worse than either of our best-fitting models (Fig. A22). The poor fit accords with the modest overlap between cCREs and our estimates of constraint sites (Fig. A23). Moreover, PLSSs, the cCREs that are most highly enriched in constrained sites (Fig. A23a and c) are inferred to have a negligible deleterious mutation rate.

Next, we consider Combined Annotation-Dependent Depletion (CADD) scores^{27,28}. CADD scores predict the ‘deleteriousness’ of every point mutation in the genome. They are generated by using machine learning to integrate information from a diverse set of annotations (122 annotations in version 1.6), such as measures of phylogenetic conservation (including phastCons scores based on the 99-vertebrate alignment), predictions of regulatory elements (including many

of the assays used for constructing the ENCODE cCREs), genic annotations (including those described in Appendices 2.7 and 4.2) and predicted functional consequences of variants in protein coding sequences. The algorithm is trained using the depletion of 14.7 million high-frequency (>95%) derived alleles (based on 1000 Genomes Data) relative to 14.7 million simulated variants with the same genomic distribution as the criterion for ‘deleteriousness’. While the standard CADD scores (version 1.6) incorporate the McVicker et al. map² of background selection effects as one of the annotations, we use a version in which this annotation was excluded in order to avoid circularity (see Appendix 2.5). We use the maximal score at each site (corresponding to the

Figure A22. The model based on the ENCODE annotations of cCRE fit the data substantially worse than our best-fitting phastCons-based model using conservation in the 99-vertebrate alignment (conserved). The results shown correspond to the model in which we split each type of cCREs into those that occur in few (subscript 1) and many biosamples (subscript 2). We infer a non-negligible deleterious mutation rate (i.e., $u_d/u_0 > 0.01$) in 2 of the 8 cCRE-based putative selection targets: enhancer like sequences and CTCF binding sites identified in few biosamples, ELS_1 and $CTCF_1$ respectively, as well as in protein coding regions (CDS). All the panels are as described in Fig. A14.

Figure A23. The relationship between ENCODE cCRE annotations and our main measures of constraint. Specifically, we examine the overlap of the 6% of autosomal sites with the highest phastCons or CADD scores with promoter like sequences (PLS), enhancer like sequences (ELS), CTCF-bound (CTCF), poised/DNase-hypersensitive (H3K4me3), as well as sites that are not in any of these annotations (NONE). a) The fraction of each cCRE annotation within the 6% most constrained sites. b) The fraction of the 6% most constrained within each cCRE annotation. c) Enrichment of cCRE annotations in the 6% most constrained sites, i.e., the ratio of their proportion among constraint and all autosomal sites.

most deleterious of three possible point mutations), and, for comparison with our best-fitting phastCons-based model (Appendix 4.1), we begin by considering the 6% of autosomal sites with the highest CADD scores (excluding putatively neutral sites) as targets of selection.

Despite incorporating many sources of information beyond phylogenetic conservation, and doing better than phastCons scores at predicting functional consequences of variants at a single site resolution²⁷, the model based on CADD scores offers only a minor improvement over our best-fitting phastCons-based model (Fig. A24). For example, the model based on CADD scores explains 59.9% of the variance in diversity levels in 1 Mb windows compared to 59.7% for the model based on phastCons scores, although this difference and differences in other window sizes are not statistically significant (see Fig. A33 and Appendix 6.2). The little improvement is not that surprising, given that phylogenetic conservation is the annotation most correlated with CADD scores genome-wide²⁷, and that the spatial distributions of sites with top CADD and phastCons scores are highly correlated on the spatial scales that impact background selection effects (Fig. A25).

Figure A24. The model based on CADD scores offer little improvement over the model based on phastCons scores (based on the 99-vertebrate alignment). In both cases, we take the 6% of sites with the highest scores. All the panels are as described in Fig. A14.

Figure A25. The spatial correlation between the 6% of sites with the highest CADD and phastCons scores.

The fit of models based on CADD scores is fairly insensitive to the proportion of sites included as selection targets, with proportions of 5-7% yielding slightly better fits than other choices (Fig. A26). This insensitivity and the increase in estimates of the deleterious mutation rate per site with decreasing proportion of sites used as selection targets (Fig. A26a(ii)) can be explained in the same way that we explained similar observations for models based on phastCons scores (Appendix 4.1).

Based on the analyses in Figs. A24 and A26, we refer to the model with the 6% of autosomal sites with the highest CADD scores as our *best-fitting CADD-based model*, and use it in most of our analyses here and in the Main Text. While the differences in fit of our best-fitting CADD-based and phastCons-based models are minor, the improved predictions of CADD compared to phastCons scores at the single site resolution substantially affects our estimates of the deleterious mutation rate based on evolutionary rates and thus their agreement with estimates based on the effects of background selection (see Main Text and Appendix 5).

Figure A26. Comparison of background selection models based on CADD scores using different proportions of autosomal sites as selection targets. All panels are as described in Fig. A14.

4.5 Models with selective sweeps

Next, we examine whether models that include both background selection and selective sweeps fit the data better than models with background selection alone. Our inference should be able to tease apart the effects of sweeps, primarily because these effects, unlike those of background selection, are centered around the locations of substitutions. This feature should hold true for hard, partial or soft sweeps^{8,42-46}, so long as they result in substitutions and have a substantial effect on diversity levels (SOM Section D in Elyashiv et al.¹). Indeed, previous work that applied

a similar methodology to data from *Drosophila melanogaster* was able to identify and quantify distinct signatures of background selection and sweeps alongside one another¹.

We consider a variety of models characterized by different sets of putatively selected sites. For background selection, we consider the two sets used in our best-fitting models based on phastCons and CADD scores. We also consider several choices for targets of *positive selection*, i.e., for sweeps, corresponding to different kinds of substitutions that we infer to have occurred on the human lineage from the common ancestor with chimpanzees (see Appendix 2.7). Notably, we consider models that include the set of all nonsynonymous substitutions paired with either of the two sets for background selection. We also consider models with the substitutions that have occurred at sites with the top 2%, 3%, ..., 9% of phastCons or CADD scores, where in each case we separate substitutions into sets of nonsynonymous and other, and pair that choice with the corresponding set for background selection (i.e., based on phastCons or CADD scores). For each of these choices, we infer the set of substitutions on the human lineage in two ways, either comparing the estimated human-chimpanzee ancestral genome²⁴ with the human reference genome (hg19) or with a population (YRI or CEU) sample of human genomes (see Appendix 2.7). We perform the inference for all of these models (18 in total) using the same grid of selection coefficients for each of the sets of selected sites, and data from either YRI or CEU. *In all cases, our estimate of the fraction of beneficial substitutions, α , is essentially 0 ($< 10^{-9}$).* We do not show the results because they are indistinguishable from those for the corresponding models with background selection alone (i.e., see Figs. A16 and A26).

We also consider models with sweeps alone. Fig. A27 shows the results for a subset of these

models, including the best-fitting one (e.g., based on variance explained). These models fit the data substantially worse than those with background selection alone, as seen by each of our measures (interestingly, even when considering the reduction in diversity levels around nonsynonymous substitutions; Fig. A27d). Sweep models do account for substantial variance in diversity levels, but given that they add nothing to a model of background selection alone yet fit much worse, this is plausibly because they approximate some of the effects of background selection. Notably, both background selection and sweeps cause reductions in diversity levels near selected sites, and the densities of sites that give rise to background selection and sweeps in the corresponding models are spatially correlated along the genome (Fig. A28). Moreover, the sweep models that fit the data best are those that rely on substitutions whose spatial distributions are the most highly correlated with the distributions of selection targets in our best-fitting background selection models (e.g., compare the fits and correlations for the models based on substitutions in the most conserved 2% and 9% of sites in Figs. A27 and A28). Taken together, the evidence presented here supports previous studies^{47,48} indicating that sweeps had little effect on current diversity levels and that background selection is the dominant mode of linked selection in humans.

Figure A27. Models with sweeps alone fit substantially worse than models with background selection alone. Shown are the results for sweep models based on either: all nonsynonymous substitutions (NS); nonsynonymous and other substitutions at sites within the top 9% of phastCons scores (9%: NS/other); or nonsynonymous and other substitutions at sites within the top 2% of phastCons scores (2%: NS/other). For comparison, we also show the results of our best-fitting phastCons-based background selection model (conserved). The panels are as described in Fig. A14, with the exception of the bottom halves of panels a(i) and

(ii), which show the proportions of substitutions that are estimated to be adaptive for a given selection coefficient (i) or in total (ii), for different annotations and sweeps models.

Figure A28. The spatial correlation between targets of selection in sweep models and in our best-fitting phastCons-based background selection model. Results shown for sweep models based on human-specific substitutions at sites within the top 2% and 9% of phastCons scores (see text for details).

4.6 Comparison with previous work by McVicker and colleagues

For completeness, we conclude by comparing our inferences about the effects of background selection with those of McVicker and colleagues². The McVicker et al. study was done more than a decade ago, before genome-wide resequencing polymorphism data were available.

Instead, they ingeniously used a five-primate alignment of ~4.7 million putatively neutral sites, relying on incomplete lineage sorting between human, chimpanzee and gorilla in order to learn about variation in the effective population size along the genome of the common ancestor of humans and chimpanzees. We rely on diversity levels in samples of 108 individuals at ~653 million putatively neutral sites (Appendix 2.1). Similar to this study, they relied on conservation scores and estimates of neutral substitution rates based on multiple sequence alignments, but they

based themselves on the genomes of 15 placental mammals when we have 99 aligned vertebrate genomes at our disposal (Appendix 2.2). Lastly, they used a genetic map based on LD patterns⁴⁹, whereas we rely on genetic maps based on ancestry switches in African Americans²⁹. The McVicker study also differed in several aspects of the methodology. Notably, McVicker et al. did not incorporate selective sweeps into their models, and were therefore unable to exclude the possibility that sweeps had made a substantial contribution to their inferred effects of background selection². Also, McVicker et al. assumed that selection coefficients are distributed exponentially, whereas we assumed a more flexible (non-parametric) distribution on a grid. Despite limitations, the McVicker et al. maps of the effects of background selection capture substantial variation in diversity levels along the human genome (Fig. A29a and Fig. 7 in McVicker et al.²).

Figure A29. Our maps of the effects of background selection fit the data much better than the maps from McVicker et al.². Shown are the results for our best-fitting CADD-based model. All panels are as described for the corresponding ones in Fig. A14.

Nonetheless, our maps of the effects of background selection fit the data substantially better than the map from McVicker et al., both quantitatively and qualitatively (Fig. A29). They explain considerably greater proportions of the variance in diversity levels across window sizes (Fig. A29c); for example, they explain ~60% compared to ~32% of the variance on the 1 Mb scale. Our predictions are well calibrated, whereas those of McVicker et al. are not (Fig. A29d). Our

predictions also do substantially better at capturing diversity patterns near specific genomic features, as illustrated by the fit to diversity levels around nonsynonymous substitutions (Fig. A29b). The relatively poor quantitative fit of the McVicker et al. predictions around synonymous and nonsynonymous substitutions⁴⁷ was used to argue that the effects of background selection could be more pronounced around synonymous than nonsynonymous substitutions, thereby masking the effects of selective sweeps⁵⁰. In this regard, the close fit of our predictions helps to refute one of two arguments for a residual, important role of selective sweeps.

We turn to the second argument, regarding estimates of the deleterious mutation rate, next. Our work and that of McVicker et al. differ markedly in our inferences about the rate and genomic distribution of deleterious mutations causing background selection in humans. In fact, the main problem in interpreting the McVicker et al. findings in terms of background selection alone is that they are based on an estimated deleterious mutation rate of 7.4×10^{-8} per generation at their ‘conserved exonic’ sites (defined as sites within the top 5.3% of conservation scores in segments that overlap exons, accounting for ~1.1% of euchromatic autosomal sites) – more than fivefold higher than current estimates of the total mutation rate per site (see next Appendix). In contrast, as we detail in the next section, our estimates of the deleterious mutation rate per selected site are quite plausible (1.00×10^{-8} per generation for both of our best-fitting models based on phastCons and CADD scores; Fig. 4 in Main Text). The results of McVicker et al. further suggest that background selection arises predominantly from deleterious mutations in the ‘conserved exonic’ regions covering ~1.1% of euchromatic autosomal sites (i.e., they estimate ~2.3 mutations per gamete per generation in such regions in exons compared to ~0.1 elsewhere). In contrast, our results suggest that background selection arises mostly from deleterious

mutations at non-exonic sites (i.e., from ~ 1.22 and ~ 1.27 mutations per gamete per generation in non-exonic compared to ~ 0.38 and ~ 0.23 mutations in exonic sites in the models based on phastCons and CADD scores, respectively). Notably, in our best-fitting models, these deleterious mutations occur in 6% of autosomal sites as opposed to only $\sim 1\%$ in the McVicker et al. model. Having the effects of background selection arise from deleterious mutations in a substantially greater fraction of the genome largely explains why our estimates of the deleterious mutation rate are much lower and much more plausible (Fig. 4 and A30).

Appendix 5: Assessing estimates of the deleterious mutation rate

Here, we consider the plausibility of the deleterious mutation rate that we estimated by fitting models of background selection. First, we consider the total mutation rate per site in humans, which provides an upper bound on the deleterious mutation rate. Second, we rely on the reduction in substitution rates at our selection targets relative to putative neutral sites to obtain estimates of the proportion of mutations at selected sites that are deleterious. These estimates should be largely independent of those that we obtained by fitting background selection models, and can therefore be used to evaluate the plausibility of the latter. Lastly, we briefly consider to what extent we should expect the two kinds of estimates to line up.

5.1 Estimates of the total mutation rate per site

The total mutation rate per site includes contributions from point mutations, indels, mobile element insertion (MEIs) and copy variants such as inversions. Current estimates of mutation

rates per site per generation in humans are $1.2 \times 10^{-8} - 1.29 \times 10^{-8}$ for point mutations^{13,51}, $8.79 \times 10^{-10} - 9.82 \times 10^{-10}$ for indels⁵¹, whereas the rate for MEIs and other structural variants (including inversions and duplications) are more than two orders of magnitude lower than the point mutation rate⁵²⁻⁵⁴, making their contribution to our calculations below negligible. Adding up point mutation and indel rates results in a per site per generation estimate of $1.29 \times 10^{-8} - 1.38 \times 10^{-8}$. In estimating an upper bound on the rate of deleterious mutations at selected sites, we may consider weighting deletions by their length. For instance, we would like to count a deletion that begins at a neutral site but includes a selected site yet avoid counting one that includes multiple selected sites more than once. Counting deletions, which account for ~ 0.725 of indels, between once and up to their mean size of $\sim 2.88\text{bp}$ ⁵¹, yields estimates of the total mutation rate in the range of $1.29 \times 10^{-8} - 1.51 \times 10^{-8}$ per site per generation. Throughout the paper, we use the middle of this range, i.e., 1.4×10^{-8} per site per generation, as our estimate for the total mutation rate (u_0). The estimates of the deleterious rate per putatively selected site for our best-fitting models fall well below the estimated total mutation rate (Fig. 4 and Appendices 4.1 and 4.4), as one would hope.

5.2 Estimating the proportion of deleterious mutations at putatively selected sites

Next, we estimate the proportional reduction of the substitution rate at selection targets relative to that at putatively neutral sites. We apply phyloFit¹⁸ to the human-chimp-gorilla-orangutan (HCGO) alignment (based on the HCGO sequences from the 99-vertebrate alignment described

in Appendix 2.2) in order to estimate the substitution rate per site on the human lineage from the ancestor with chimpanzee, for sets of selected and neutral sites (Appendix 3.1). To control for differences in base composition between the two sets, we estimate the reduction in substitution rates separately for each type of ancestral nucleotide (e.g., substitutions from G>X), and weight the proportional reductions by the proportions of each nucleotide in the set of selected sites. Controlling for the composition of triplets rather than single nucleotides produces similar estimates. Note that in choosing our sets of neutral and selected sites based on phylogenetic conservation (Appendices 3.1 and 4.1), we excluded the human genome from the alignments and therefore our estimates of the reduction in substitution rates on the human branch should be minimally confounded with the choice of sites. Similarly, the conservation scores that serve as input for calculating CADD scores are based on the same 99-vertebrate alignment excluding the human reference genome (see Supplementary Table 1 in Kircher et al.²⁷).

Figure A30. Different estimates of the deleterious mutation rate at putatively selected sites, measured relative to total mutation rates per site (u_0). Estimates based on evolutionary rates are shown for sets of selected sites chosen based on either: (a) the top 4-8% of phastCons scores for the 99-vertebrate alignment, (b) the top 4-8% of CADD scores, or (c) the top 6% of phastCons scores for alignments of varying phylogenetic depths. As expected, (see Appendix 4.1), the estimates are fairly insensitive to the phylogenetic depth (c). d and e) Estimates based on evolutionary rates vs. those based on the effects of background selection, for sets of putatively selected sites based on phastCons scores for the 99-vertebrate alignment (d) and on CADD scores (e). f) Estimates for different sets of putatively selected sites based on evolutionary rates (ER) and background selection effects (BS). The range of estimates based on background selection effects (in d-f) is due to the uncertainty about the total mutation rate per site (Appendix 5.1).

The estimates of the proportion of mutations that are deleterious are shown in Fig. A30 (and Fig. 4), along with their comparison to estimates from background selection models. Expectedly, estimates based on substitution rates decline slightly as the cutoff phastCons or CADD score decreases (i.e., as the percentage of sites included in the selected set increases) (Fig. A30a and b). Importantly, estimates based on substitution rates are substantially greater for the sets chosen based on CADD than on phastCons scores (Fig. A30a and b), whereas estimates based on background selection effects are similar in both cases (Fig. A30d and e). We interpret this finding as reflecting the greater ability of CADD scores to identify selection on a single site resolution²⁷, plausibly because CADD scores incorporate measures of phylogenetic conservation based on one site at a time (e.g., phyloP, GERP^{55,56}) in addition to measures that rely on runs of sites, such as phastCons scores. Consequently, the two estimates of the deleterious mutation rate are within a factor of 2 for our best-fitting phastCons-based model whereas they overlap for our best-fitting CADD-based model, while the estimates based on background selection effects are similar in both cases (Fig. A30d and e).

5.3 Interpreting the relationship between the two estimates

We would expect the two estimators of the deleterious mutation rate to yield similar but not identical answers. For one, the range of selection coefficients that cause a substantial reduction in evolutionary rates, e.g., $4N_e s \gtrsim 3^{57}$, is greater than the range of effects that cause a substantial reduction in diversity levels via classic background selection, e.g., $4N_e s \gtrsim 10^{58-60}$. This consideration suggests that estimates based on evolutionary rates should be greater than those based on the effects of background selection (although non-equilibrium demographic history, notably changes in population size, might complicate quantitative expectations). On the other

hand, we cannot expect to identify all selected sites and only those by our criteria. Estimates based on the effects of background selection plausibly soak up much of the contribution of missing selected sites, because their spatial distribution is likely to be highly correlated with sites that are included in our sets (see Appendix 4.1). In contrast, estimates based on evolutionary rates are affected only by the sites in our sets and would be biased downwards by the accidental inclusion of effectively neutral sites. For these reasons, we do not expect the two estimates of the deleterious mutation rate to align perfectly. Nonetheless, it is encouraging that when we rely on selected sites that amount to current estimates of the proportion the human genome under selection, i.e., ~5-9%^{36,40}, our two estimates of the deleterious mutation rate are quite similar. Moreover, the similarity is highest when we use CADD scores, which are better than phastCons scores at identifying selection on a single site resolution²⁷. Thus, our results resolve the issues raised by the substantial overestimation of the deleterious mutation rate in past work².

Appendix 6: Statistics

6.1 Estimates of explained variance

Our main quantitative measure for the fit of our models is the variance in diversity levels explained by our predictions, R^2 , for different window sizes. A concern in using R^2 as a measure of fit is that it not be inflated by overfitting. To avoid this problem, we exclude the data in a given window from the inference used to predict diversity levels in that window. Specifically, we divide the autosomal polymorphism data into contiguous non-overlapping blocks of 2 Mb, and repeat the inference using the data excluding one block at a time. As the autosomes cover just over 2.88 Gb, this amounts to repeating our inference 1,440 times. When we calculate the

contribution of a given window (of size ≤ 2 Mb) to R^2 , we use the prediction based on excluding the 2 Mb block containing that window. In Appendix 6.3, we use the same datasets and inferences to calculate jackknife estimates for the sampling error of our estimates of model parameters.

Fig. A31 shows the relative difference between our R^2 estimates using all the data and this exclusion approach for our two best-fitting models. These estimates suggest that overfitting has a tiny effect, which is not surprising given the large amounts of data used in our inferences. Given the negligible effect and computational burden of this analysis, we do not repeat it for each of the models we examine, and use the R^2 estimates based on predictions using all the data instead.

Figure A31. The relative difference between R^2 estimates using all the data and the exclusion approach described in the text, for our two best-fitting models.

6.2 Comparing the fit of different maps

We use permutations of paired maps to test whether differences in R^2 between two maps are statistically significant. Assume without loss of generality that R^2 for a given window size is greater for map I than for map II. We divide autosomes into 2881 contiguous non-overlapping blocks of 1 Mb and generate a new map (map A) by picking each 1 Mb block from map I or II at random; we generate the complementary map (map B) by picking the alternative 1 Mb blocks throughout. This way, we generate n paired maps and calculate the difference in R^2 between each pair, ΔR^2 , to obtain a distribution for the expected differences in R^2 between maps I and II under the null hypothesis that their fit to polymorphism data is roughly equivalent. Having r denote the number of permutations with ΔR^2 greater than or equal to the observed difference ΔR_0^2 , we estimate the p-value for ΔR_0^2 under the null by $p = (r + 1)/(n + 1)$.

Figure A32. Assessing the differences in fit between our best-fitting models on three spatial scales. We show the distribution of differences in explained variance (ΔR^2) for 10,000 paired permutations of the best-fitting CADD and phastCons based maps. The part of the distributions with $\Delta R^2 \geq \Delta R_0^2$ is in red, and the corresponding p-value is shown above.

Figure A33. Significance level of differences in fit between our best-fitting models and variations on these models, on the 1 Mb spatial scale. a and b) Comparison of our best-fitting phastCons-based model with phastCons-based models with alternative phylogenetic depths (a) and conservation thresholds (b). c) Comparisons of our best-fitting CADD-based model with CADD-based models with alternative thresholds

We illustrate this procedure by comparing our two best-fitting models (Fig. A32). The fit of these models is very similar, with, e.g., $R^2 = 0.599$ and 0.597 at the 1 Mb scale for the models based on CADD and phastCons scores respectively. We find that the difference between the fits is not statistically significant, supporting our claim that the functional annotations incorporated in CADD offer little or no improvement in predictive power (see Main Text and Appendix 4.4). Using this procedure to compare our best-fitting phastCons-based model with phastCons-based models with alternative phylogenetic depths or conservation thresholds, we only find significant differences at the 1 Mb scale in a small subset of cases (Fig. A33a and b). The same is true for comparisons of our best-fitting CADD-based model with CADD-based models with alternative thresholds (Fig. A33c). We note that even when the fits are significantly worse, they are still far closer to the fits of our best-fitting models than any of the models based on other choices of selection targets discussed in Appendix 4.

Interestingly, the difference in fit between the model based on conservation in 99-vertebrates and four-apes is the only non-significant comparison across phylogenetic depths (Fig. A33a), despite the fact that other phylogenetic depths have R^2 values closer to the 99-vertebrate result across various window sizes (Fig. A14c). We believe this is due to the fact that the four-ape alignment may actually better capture some recent targets of selection, but with greater noise, reflecting a tradeoff in the power to detect conservation vs. functional turnover. As a result, a non-negligible subset of 1 Mb windows in the four-ape yield better fits to the data than the 99-vertebrate map. In contrast, maps from deeper in the phylogeny are essentially highly correlated to the 99-vertebrate map, and the small differences in fit are uniformly biased in favor of the 99-vertebrate map across 1 Mb windows due to its greater power to resolve the boundaries of conserved elements.

6.3 Sampling error in parameter estimates

We use a jackknife resampling approach to estimate the sampling errors of our parameter estimates (see, e.g., Patterson et al.⁶¹). To this end, we perform the inference on datasets excluding 2 Mb blocks as described in Appendix 6.1. Specifically, denoting the parameter of interest by θ , and the estimate based on the data set excluding block $i = 1, \dots, n$ by $\hat{\theta}_i$, our jackknife estimates of the mean and variance are $\bar{\theta} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i$ and $V(\bar{\theta}) = \frac{n-1}{n} \sum_{i=1}^n (\hat{\theta}_i - \bar{\theta})^2$ respectively. We use the standard deviation $\sqrt{V(\bar{\theta})}$ as our measure of sampling error (SE). Fig. A34 shows the SEs for the parameter estimates of our two best-fitting models. As these examples illustrate, these errors are quite small and do not affect the conclusions of our analyses.

Consequently, and given the computational cost of obtaining them, we do not calculate these SEs for most models.

Figure A34. Estimates of the sampling errors of parameter estimates for our two best-fitting models. The bars denote $\pm SE$ estimated using jackknife as described in the text.

Appendix 7: Results for other human populations

Here, we examine whether the maps of the effects of background selection that we infer and evaluate using polymorphism data from YRI provide a good fit to data from other populations. To this end, we use data from each of the other 25 populations collected in Phase III of the 1000 genomes project¹⁹, which span a wide geographic range and have had different demographic histories¹⁹, to infer the maps corresponding to our best-fitting models. The population-specific maps can be found at github.com/sellalab/HumanLinkedSelectionMaps.

Overall, we find that the maps and main parameters inferred in different populations are remarkably similar (Fig. A35 and A49-S51). When we compare the predictions of relative diversity levels along autosomes (i.e., relative to the mean in each population) we find nearly

perfect correlations across window sizes (Fig. A35a and b). The distributions of selection effects of deleterious mutations, estimates of the total deleterious mutation rate per selected site, and the mean reduction in diversity levels, are all quite close among populations (Fig. A35c). The similarity among maps implies that we can use the maps of relative diversity levels inferred in YRI to predict diversity levels in other populations without loss of accuracy. Specifically, we multiply predictions based on YRI by a constant, chosen such that the predicted and observed mean diversity level in the focal population match. Fig. A36 illustrates that the adjusted YRI maps predict diversity levels as well as the population specific ones.

Figure A35. The maps and parameter estimates for different populations are remarkably similar. Shown are the results for our best-fitting models using data for one of 1000 Genomes Project populations from each continental group: Africa – Yoruba (YRI), Europe – North-Western European (CEU), South Asia – Gujrati Indian (GIH), East Asia – Japanese (JPT), and Americas – Mexican (MXL). a and b) The Pearson correlations between

predictions of relative diversity levels (compared to the population mean) in YRI vs. the other populations for the models based on phastCons (a) and CADD (b) scores. c) Comparison of parameter estimates using data from these populations (panels c (i-iii) as described in panels a (i-iii) in Fig. A14).

Figure A36. proportion of variance in diversity levels explained (R^2) in different populations, using the population specific map vs. the YRI map. We show the results for three window sizes (10 kb, 100 kb and 1 Mb) based on our best-fitting CADD-based model. Each point corresponds to one of the 26 populations sampled in the 1000 genomes project and is colored based on continental origin, i.e., African (AFR), European (EUR), South Asian (SAS), East Asian (EAS) and American (AMR).

While the maps inferred in different populations are highly similar, the proportion of variance explained differs substantially among populations (Fig. A36). These differences can be explained by the effects of different demographic histories (e.g., historical changes in effective population sizes) on variation in diversity levels across the genome. To make this more concrete, we consider a simple model for the variance in neutral diversity levels in non-overlapping windows of a given size; for simplicity, we ignore variation in mutation rates across windows. We denote the relative (average) diversity level in window i by $y_i = \pi_i / \bar{\pi}$, the predicted relative diversity level in that window by $f_i = B_i / \bar{B}$, and the corresponding residual by $e_i = y_i - f_i$, where $\bar{y} = \bar{f} = 1$ and $\bar{e} = 0$. We can now decompose the variance in relative diversity levels across windows, as

$$V(y) = V(f) + V(e) + 2Cov(e, f), \quad (15)$$

where $V(f)$ corresponds to the variance due background selection; $V(e)$, the variance of residuals, can be thought of as reflecting the effects of drift and demographic history; and the covariance, $Cov(e, f)$, can be thought of as reflecting the interaction between background selection and demographic history. Recasting the proportion of variance explained in these terms, we find that

$$R^2 \equiv 1 - \frac{V(e)}{V(y)} = \frac{V(f)}{V(y)} (1 + \beta), \quad (16)$$

where $\beta = 2 \cdot Cov(e, f)/V(f)$ is the slope of the linear regression of the residuals against the predictions, which reflects the effects of interactions between background selection and demographic history on diversity levels.

Given that we found the predicted effects of background selection to be highly similar across populations, this modeling exercise sets up a testable prediction: if the interaction terms were nil, the difference in R^2 among populations should come from the total variance in the denominator, $V(y) = V(f) + V(e)$, and specifically from the contribution of demographic history to this variance, $V(e)$. Fig. A37a-c suggest that while most of the differences in R^2 among populations are indeed explained by differences in total variance due to demographic history, the interaction terms (the β s) are non-zero. To examine these interactions further, we look at the relationship between residuals, e_i , and predictions, f_i , in several populations (Fig. A37d-h). We find a strong apparent dependency at the low and high ends of our predicted range, presumably reflecting artifacts due to thresholding at the low end (see Appendix 1.5) and possibly the effects of ancient introgression at the high end (see Appendix 8); removing 1.5% of the windows at each of these extremes appears to largely remove these effects. In the rest of the range, we find a weak

negative correlation between residuals and predictions (which becomes somewhat stronger when we remove the ends). We also find that this correlation varies substantially among populations, e.g., -0.06 to -0.1 for 1 Mb windows, which is what we would expect given differences in demographic history. Thus, our analysis suggests that interactions between demographic history and background selection also contribute to the differences in R^2 among populations.

In summary, our findings suggest that the effects of background selection are similar across human populations, and that differences among populations in the proportion of variance in diversity levels that our predictions explain are likely due to differences in population demographic history. Interestingly, there appears to be an interaction between the effects of background selection and demography on diversity levels, which varies among populations, as recently suggested by several studies⁶²⁻⁶⁵. Our maps of the effects of background selection have enabled us to identify evidence for these interaction effects and should facilitate a better understanding of these effects in the future.

Figure A37. Differences among populations in the variance in diversity levels explained by our map of the effects background selection. (a-c) The variance explained (R^2) as a function of $1/V(y)$ (where $V(y)$ is the total variance), for three choices of window size. If the interaction terms (β) were 0, we would expect populations to fall on the dashed line $R^2 = V(f) \cdot 1/V(y)$, with slope $V(f)$ and differences in $V(y)$ due to demographic history. The distances from the dashed line reflect the interaction terms (specifically, $R^2 - V(e)/V(y) = \beta \cdot V(e)/V(y)$; Eq. 15). The points correspond to the 26 populations sampled in the 1000 genomes project¹⁹ and are colored by continental origin as described in Fig. A36. Here we base our predictions on our best-fitting CADD-based map in YRI, but using other population-specific maps yields almost identical results. (d-h) The relationship between the residuals (e_i) and predictions (f_i) in representative populations (same as in Fig. A36) on the 1 Mb scale. The

1.5% of windows with lowest and highest predicted values, where our predictions are likely off for various reasons (see Appendices 1.5 and 8), are marked in blue. The β s for each population, with and without extreme points, are shown on the graph (denoted 'a' for 'all' and 't' for 'trimmed', respectively). As above, we use the predictions in YRI, but other population maps yield qualitatively similar results (β s obtained using the corresponding population specific maps are shown in parenthesis).

Appendix 8: Diversity levels where background selection is weakest

$$(B \approx 1)$$

Our maps of background selection effects are well calibrated throughout the range of predicted effects, with two exceptions. One is in the ~4% of sites in which background selection is predicted to be strongest, where predictions are imprecise; this arises from the thresholding approximation we apply in fitting, and is discussed in Appendix 1.5. The other exception is for sites in which background selection is predicted to be the weakest, where observed diversity levels are markedly greater than expected (Figs. 5 and A38). A close up on this region shows that observed values depart from predictions in the ~2% of sites where $B \gtrsim 0.98$ (Fig. A38b). Similar behavior is seen in all 26 populations sampled in the 1000 Genomes Project (Fig. A51). We consider possible explanations for it here.

First, we characterize the main covariates of the strength of background selection (quantified by B), such as recombination rate, base composition, and chromosomal position. Beyond the inherent interest in these covariates, they point toward processes that may explain the departure from our predictions near $B = 1$. Second, we investigate whether differences in rates of different kinds of mutations and of biased gene conversion associated with the covariates of B can explain

the departure near $B = 1$; our analysis suggests that they cannot. Third, we argue that the residual effect of ancient introgression between archaic humans and ancestors of extant humans may contribute to this departure.

Figure A38. Observed vs. predicted neutral diversity levels across the autosomes (similar to Fig. 5). a) Light orange scatter plot: We divide putatively neutral sites into 100 equally sized bins based on the predicted effect of background selection, B , from the best-fitting CADD-based model. For predicted values (x-axis), we average the predicted B in each bin. For observed values (y-axis), we divide the average diversity level in each bin by the average predicted diversity level in the absence of background selection, π_0 , after scaling each in bin by its estimated local (relative) mutation rate ($u(x)/\bar{u}$ in Eq. 8; Appendix 3.3). Dark orange curve: the LOESS curve for a similarly defined scatter plot but with 2000 rather than 100 bins (with $\text{span}=0.1$). **b)** A close-up near $B = 1$ corresponding to the boxed region in (a). Here, the LOESS curve has $\text{span}=0.033$ and the scatter plot corresponds to 2000 bins (showing the top 30%).

8.1 Covariates of B

We expect the effects of background selection to be strongest in regions with low rates of recombination and high densities of functional sites, because neutral variation in such regions

will be linked to more deleterious mutations. In line with these expectations, recombination rates increase with greater predicted B (Fig. A39a); in particular, they increase sharply between the 99th and 100th percentile of predicted B to >10 cM/Mb, which is tenfold the autosomal average. In addition, as expected, the average level of conservation around neutral sites decreases as B increases (Fig. A39b and c).

Figure A39. Average recombination rate (a) and functional density per bp (b) and per cM (c) as a function of predicted B . The predicted B , binning of putatively neutral sites and LOESS fitting are as described in Fig. A38a. We calculate the average recombination rate in each bin based on the African-American admixture map from Hinch et al.²⁹ (Appendix 2.3). We measure functional density by calculating the mean phastCons score (based on the 99-vertebrate alignment) in a radius of 50 kb (b) or 0.05 cM (c) around each putatively neutral site, and averaging these means over sites in each bin.

Next, we consider base composition and other factors that are known to affect rates of mutation and biased gene conversion (BGC). GC content has a J-shaped dependence on predicted B (Fig. A40a). The greater peak in GC content, in regions under weak background selection (B near 1), is plausibly largely driven by the long-term effects of BGC due to higher rates of recombination in these regions^{66,67} (Fig. A39a). Both peaks (for low and high B) are associated with an increase in the proportion of GC sites in CpG islands but this proportion is small throughout (<1%) (Fig.

A40b), suggesting that it has little effect on GC content and of mutation rates. In contrast, methylated CpG content increase sharply as predicted B approaches 1 (Fig. A40c), suggesting a corresponding increase in the rate of C>T transitions. The proportion of sites in C>G hypermutable regions also increases with predicted B (Fig. A40d).

Figure A40. GC content (a), CpG sites in CpG islands (b), proportion of methylated CpGs in neutral sites (c), and proportion of neutral sites in C>G hypermutable regions (d) as a function of predicted B . Proportions and other quantities are measured for putatively neutral sites, whose type (i.e., GC and CpG) is defined based on the inferred state in the human-chimpanzee ancestor (Appendix 2.7). Data sources are detailed in Appendix 2.8. The predicted B , binning of putatively neutral sites and LOESS fitting are as described for Fig. A38a.

Lastly, predicted B is associated with chromosomal position, with regions under weak background selection (B near 1) clustered near telomeres (Fig. A41). In turn, regions near telomeres are early replicating, and replication timing is known to affect mutational patterns⁶⁸.

Figure A41. The relationship between chromosomal position and predicted B . a) The the distance to telomeres is measured on the same chromosome (see Appendix 2.8). b) The distribution of putatively neutral sites by relative chromosomal position, defined as the ratio of a site's distance to the centromere and the distance between centromere and telomeres on that chromosomal arm (see Appendix 2.8); relative distances corresponding to the shorter chromosome arm are shown on the left (in $[-1, 0]$) and to the longer arm on the right (in $[0, 1]$). The binning of putatively neutral sites by predicted B is as described for Fig. A38a.

8.2 Mutational spectrum and biased gene conversion

As we noted, the covariates of B are associated with mutational processes and with biased gene conversion that affect diversity and divergence levels (Fig. A42). Specifically, we see the footprints of the following processes:

- Increased rates of **A>C/T>G** and **A>G/T>C** substitutions near $B = 1$ (Fig. A42a), due to higher rates of biased gene conversion that are associated with the higher rates of recombination^{66,67} (Fig. A39a). Biased gene conversion also reduces the rates of **C>A/G>T** and **C>T/G>A** substitutions, but this is not clearly visible (Fig. A42b), presumably because of other processes affecting these substitutions (see below).

- Increased rate of **C>T mutations** near $B = 1$ (Fig. A42b), associated with the higher methylated CpG content³³ (Fig. A40c).
- Reduced rates of **C>A/G>T** and **A>T/T>A** mutations near $B = 1$ (Fig. A42 a and b), associated with improved repair of these types of mutations near origins of replication, which tend to be near telomeres⁶⁸ (Fig. A41).
- Increased rates of **C>G/G>C** mutations near $B = 1$ (Fig. A42b), due to the enrichment of C>G hypermutable regions³⁴ (Fig. A40d).

Consequently, the rates of substitutions between any two bases covary with predicted B (Fig. A42). However, the different types of substitutions have markedly different contributions to levels of diversity and divergence (Fig. A43), and different dependencies on predicted B (Fig. A42).

Figure A42. of different types of substitutions as a function of predicted B . We bin putatively neutral sites by predicted B as described in Fig. A38a. We calculate the rate of $X>Y$ substitutions in a bin by dividing the estimate of the number of $X>Y$ substitutions at its sites in an 8-primate phylogeny by the estimated number of its sites with state X in the ancestor of

that phylogeny (see Appendix 3.3). We obtain the relative rate by dividing the rate in a bin by the average rate across bins. We show the rates of substitutions with ancestral state AT in (a) and GC in (b).

Figure A43. Contribution of different types of substitutions to diversity levels (a) and number of substitutions per site in the 8-primate phylogeny (b) as a function of predicted B .

We bin putatively neutral sites by predicted B as described in Fig. A38a. a) We define the ancestral state, i.e., AT or GC, as the inferred state in the human-chimpanzee ancestor (see Appendix 2.7). We define the contribution of each type of substitution to the diversity level in a bin as the ratio of the number of pairwise differences of that type and the total number of pairwise comparisons in a bin; this way, the sum over types equals the observed diversity level in a bin ($\hat{\pi}$). b) We calculate the number of substitutions of each type in a bin as described in Fig. A42, but here we normalize it by the number of sites, such that the sum over types equals the observed number of substitutions per site in the 8-primate phylogeny.

To investigate whether these processes can explain the departure from our predictions near $B = 1$, we break up the observed diversity levels by types of substitutions (Fig. A44). We reason that if all types behave similarly near $B = 1$, the differential processes affecting them cannot explain the departure from predictions (at least not fully). Note that, up to a multiplicative constant, our

observations are ratios of diversity levels and substitution rates (on the 8-primate phylogeny), implying that the signatures of processes that affect diversity and divergence similarly should cancel out. Conversely, for a process to affect our observations it must have noticeably different effects on diversity and divergence. We find that the observations associated with different types of substitutions largely align with each other and with the observations that include all types jointly (Figs. A38 and A44). Specifically, they align with predictions throughout nearly the entire range of predicted B , and are markedly higher than predictions near $B = 1$.

Figure A44. Observed vs. predicted neutral diversity levels for different types of substitutions. Diversity levels are presented as in Fig. A38, but here, we calculate diversity levels and substitution rates for sites with a given ancestral state, i.e., AT (a and c) or CG (b and d), and for each of the three types of substitutions from the ancestral state, as in Fig. A43.

Nonetheless, the observed levels associated with C>G mutations near $B = 1$ are markedly higher than for other types of substitutions (Fig. A44b and d). This effect contributes negligibly to the

total observed levels near $B = 1$, however, because C>G substitutions have a minor contribution to total diversity and divergence levels (Fig. A43). The higher levels of C>G substitutions are associated with the enrichment of C>G hypermutable regions near $B = 1$ (Fig. A40d). Notably, when we remove these regions, the observed levels associated with C>G substitutions are no longer higher than for other types of substitutions (Fig. A45). Removing C>G hypermutable regions also affects the magnitude of the departure from predictions for other types of substitutions (compare Fig. A45c and d with Fig. A44c and d), because C>G is not the only type of mutation whose rate is higher in these regions. These hypermutable regions plausibly affect diversity more than divergence (and thus our observations) given that their effects are stronger in recent human evolution than in the more distant past and in the lineages of closely related species (Ipsita Agarwal and Molly Przeworski, personal communication). This may reflect the fact that these regions were identified in extant humans and/or a dependence of their effects on life history^{34,69}. Setting the causes aside, even when we remove these regions from the set of putatively neutral sites used in our inference, observed levels of all types are still markedly higher than the revised predictions near $B = 1$ (Fig. A46).

In summary, while we cannot rule out that there are other mutational processes that contribute to the departure from predictions near $B = 1$, our analysis suggests that known mutational processes and biased gene conversion fall short of explaining these departures.

Figure A45. Observed vs. predicted neutral diversity levels for different types of substitutions after removing $C > G$ hypermutable regions (from both the inference and observations). Other than removing $\sim 12\%$ of putatively neutral sites in these regions, the details are as in Fig. A44. We note that while a greater proportion of sites is removed from bins near $B = 1$ ($\sim 20\%$ for the 100th percentile), this in itself has a minor effect on the departures from predictions in these bins.

Figure A46. Comparison of the best-fitting CADD-based models with and without C>G hypermutable regions. All panels are as described for the corresponding ones in Fig. A14.

8.3 A footprint of archaic introgression?

Next, we consider whether the excess diversity observed near $B = 1$ could reflect a residual signal of archaic introgression. The presence of archaic alleles at a locus increases diversity because their coalescence with modern human alleles traces back to the ancestors of modern humans and the archaic hominin from which they originated. Archaic introgression could help to

explain the excess diversity in regions with $B \approx 1$, if archaic alleles were more common in these regions. As we argue below, there are good reasons to believe this to be the case.

Aside from evidence for positive selection on introgressed alleles in a few cases⁷⁰⁻⁷², the pattern of Neanderthal and Denisovan introgression in contemporary human populations appears to be dominated by purifying selection to remove archaic ancestry from the human genome, as evidenced by the depletion of archaic introgression in and around genes^{70,71,73-75}. The causes for this purifying selection are still being deliberated^{73,75,76}. One hypothesis is that selection acts against introgressed alleles that are incompatible with the genetic background in modern humans, e.g., alleles that are part of Dobzhansky-Muller incompatibilities between archaic hominins and modern humans^{70,76}. Another hypothesis is that selection acts against alleles that were deleterious in both archaic hominins and modern humans, which were more common in archaic hominins because their long-term effective population sizes were smaller than in modern humans^{35,73,75}.

Regardless of its cause, we expect purifying selection to remove archaic alleles, including neutral variants, more rapidly in genomic regions under stronger background selection^{73,76,77}. This is because these regions harbor more selected sites (Fig. A39b) in which archaic alleles could be selected against, and because they have lower rates of recombination (Fig. A39a) causing selection against archaic alleles to remove larger archaic segments. Conversely, we expect the highest, residual proportion of archaic neutral variants in regions with $B \approx 1$ —precisely where we observe a 10-15% excess of diversity above our predictions (Fig. A38).

In order to test this expectation, we use fine-scale maps of archaic introgression inferred for European (CEU) and East-Asian (CHB/CHS) individuals from the 1000 Genomes Project³⁵. These maps assign a probability of Neanderthal ancestry to contiguous 500 bp segments tiling individual genomes based on the high-coverage Altai Neanderthal genome⁷⁸. We use them to estimate the average proportion of archaic ancestry per putatively neutral site in bins of predicted B . As expected, we find that the estimated proportion of archaic alleles increases with predicted B (Fig. A47). The power to identify introgressed segments using this and other methods decreases substantially in regions with $B \approx 1$, because higher recombination rate in these regions results in much shorter archaic segments^{35,79}. We therefore expect that the actual proportion of archaic ancestry increases more sharply near $B = 1$ than our analysis suggests, and may therefore better trace the sharp increase in diversity relative to predictions near $B = 1$.

Figure A47. Estimated proportion of Neanderthal (NE) ancestry as a function of predicted B in Europeans (CEU) (a) and East-Asians (CHB/CHS) (b). See text for the estimation procedure. The bins and LOESS curves were calculated as in Fig. A38.

Current inferences about archaic introgression are divided into those that incorporate sequenced Neanderthal and Denisovan genomes^{35,70,71,80,81}, such as the maps we used in Fig. A47, and those that are based only on patterns of variation in contemporary humans^{79,82-84}. When we repeat the analysis in Fig. A47 using ancestry-maps based on the latter approach in both Africans and non-Africans^{79,83}, we find that levels of archaic ancestry either increase and level off at intermediate values of B , or peak at intermediate values and decrease as B approaches 1. We believe that this departure from our expectation reflects a decrease in the power of these methods near $B = 1$ (due to higher rates of recombination), which is greater than the decrease for methods based on sequenced archaic genomes. An additional caveat is that the evidence for the contribution of archaic introgression to the African gene pool is based solely on patterns in contemporary genetic variation⁸²⁻⁸⁵ and remain more speculative in lieu of more direct evidence. Thus, while it seems plausible that the greater retention of neutral archaic variants in regions with the highest ~2% of predicted B values contributes substantially to the departure from our predictions in both African and non-African populations, at present, the evidence for such a contribution remains equivocal.

Appendix 9: Additional Figures

Figure A48. A background selection model predicts neutral diversity levels around different genomic features. Here we use our best-fitting CADD-based model and show diversity levels around: a) human-specific synonymous substitutions; b) human-specific substitutions in conserved regions; c) exons; and d) conserved exonic regions. The inference of human-specific substitutions is described in Appendix 2.7. Conserved regions are based on autosomal sites with the top 6% phastCons scores in the 99-vertebrate alignment (Appendix 4.1). The set of exons is described in Appendix 2.4. The genetic distance to the nearest element (e.g., exon) is measured to its closest edge. Other details are similar to Fig. 3.

Figure A49. Predicted and observed neutral diversity levels along chromosome 1 based on data from representative continental populations. Plots are generated as detailed in Fig. 2A.

Figure A50. A background selection model predicts neutral diversity levels around human-specific nonsynonymous (NS) substitutions in representative continental populations. Plots are constructed as detailed in Fig. 3, using polymorphism data from each population for both inferences and observations.

Figure A51. Observed vs. predicted neutral diversity levels across the autosomes in representative continental populations. Plots are constructed as detailed in Fig. 5, using polymorphism data from each population for both inferences and observations.

References to appendices

1. Elyashiv, E. *et al.* A Genomic Map of the Effects of Linked Selection in *Drosophila*. *PLoS Genet.* **12**, e1006130 (2016).
2. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
3. Wiehe, T.H. & Stephan, W. Analysis of a genetic hitchhiking model, and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**, 842-54 (1993).
4. Hudson, R.R. & Kaplan, N.L. Deleterious background selection with recombination. *Genetics* **141**, 1605-17 (1995).
5. Nordborg, M., Charlesworth, B. & Charlesworth, D. The effect of recombination on background selection. *Genet. Res.* **67**, 159-174 (1996).
6. Barton, N.H. The effect of hitch-hiking on neutral genealogies. *Genet. Res.* **72**, 123-133 (1998).
7. Gillespie, J.H. Genetic drift in an infinite population. The pseudohitchhiking model. *Genetics* **155**, 909-19 (2000).
8. Hermisson, J. & Pennings, P.S. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* **169**, 2335-52 (2005).
9. Kim, Y. & Stephan, W. Selective sweeps in the presence of interference among partially linked loci. *Genetics* **164**, 389-398 (2003).
10. Hudson, R.R. Two-locus sampling distributions and their application. *Genetics* **159**, 1805-17 (2001).
11. Fearnhead, P. Consistency of estimators of the population-scaled recombination rate. *Theor. Popul. Biol.* **64**, 67-79 (2003).
12. Wiuf, C. Consistency of estimators of population scaled parameters using composite likelihood. *J. Math. Biol.* **53**, 821-41 (2006).
13. Kong, A. *et al.* Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**, 471-5 (2012).
14. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272 (2020).
15. Nelder, J.A. & Mead, R. A Simplex Method for Function Minimization. *The Computer Journal* **7**, 308-313 (1965).
16. Conn, A.R., Gould, N.I.M. & Toint, P.L. *Trust-region methods* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2000).
17. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034-50 (2005).
18. Siepel, A. & Haussler, D. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* **21**, 468-88 (2004).
19. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
20. Altshuler, D.M. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).

21. Karolchik, D. The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, 493D-496 (2004).
22. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011).
23. Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708-15 (2004).
24. Paten, B. *et al.* Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* **18**, 1829-43 (2008).
25. Church, D.M. *et al.* Modernizing Reference Genome Assemblies. *PLoS Biol.* **9**, e1001091 (2011).
26. Moore, J.E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699-710 (2020).
27. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310-5 (2014).
28. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886-D894 (2019).
29. Hinch, A.G. *et al.* The landscape of recombination in African Americans. *Nature* **476**, 170-5 (2011).
30. Halldorsson, B.V. *et al.* Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363**, eaau1043 (2019).
31. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
32. Hsu, F. *et al.* The UCSC Known Genes. *Bioinformatics* **22**, 1036-46 (2006).
33. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41**, D991-5 (2013).
34. Jonsson, H. *et al.* Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519-522 (2017).
35. Steinrucken, M., Spence, J.P., Kamm, J.A., Wiczorek, E. & Song, Y.S. Model-based detection and analysis of introgressed Neanderthal ancestry in modern humans. *Mol. Ecol.* **27**, 3873-3888 (2018).
36. Rands, C.M., Meader, S., Ponting, C.P. & Lunter, G. 8.2% of the Human Genome Is Constrained: Variation in Rates of Turnover across Functional Element Classes in the Human Lineage. *PLoS Genet.* **10**, e1004525 (2014).
37. Ward, L.D. & Kellis, M. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* **337**, 1675-8 (2012).
38. Black, D.L. Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annu. Rev. Biochem.* **72**, 291-336 (2003).
39. Kim, T.H. *et al.* A high-resolution map of active promoters in the human genome. *Nature* **436**, 876-880 (2005).
40. Kellis, M. *et al.* Defining functional DNA elements in the human genome. *Proc. Natl. Acad. Sci. USA* **111**, 6131-8 (2014).
41. Hudson, R.R. How can the low levels of DNA sequence variation in regions of the drosophila genome with low recombination rates be explained? *Proc. Natl. Acad. Sci. USA* **91**, 6815-8 (1994).

42. Berg, J.J. & Coop, G. A Coalescent Model for a Sweep of a Unique Standing Variant. *Genetics* **201**, 707-25 (2015).
43. Coop, G. & Ralph, P. Patterns of neutral diversity under general models of selective sweeps. *Genetics* **192**, 205-24 (2012).
44. Pennings, P.S. & Hermisson, J. Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genet.* **2**, e186 (2006).
45. Pennings, P.S. & Hermisson, J. Soft sweeps II--molecular population genetics of adaptation from recurrent mutation or migration. *Mol. Biol. Evol.* **23**, 1076-84 (2006).
46. Przeworski, M., Coop, G. & Wall, J.D. The signature of positive selection on standing genetic variation. *Evolution Int. J. Org. Evolution* **59**, 2312-23 (2005).
47. Hernandez, R.D. *et al.* Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920-4 (2011).
48. Coop, G. *et al.* The role of geography in human adaptation. *PLoS Genet.* **5**, e1000500 (2009).
49. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. *Science* **310**, 321-4 (2005).
50. Enard, D., Messer, P.W. & Petrov, D.A. Genome-wide signals of positive selection in human evolution. *Genome Res.* **24**, 885-95 (2014).
51. Besenbacher, S. *et al.* Multi-nucleotide de novo Mutations in Humans. *PLoS Genet.* **12**, e1006315 (2016).
52. Gardner, E.J. *et al.* Contribution of retrotransposition to developmental disorders. *Nat. Commun.* **10**, 4630 (2019).
53. Belyeu, J.R. *et al.* De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *Am. J. Hum. Genet.* **108**, 597-607 (2021).
54. Sudmant, P.H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81 (2015).
55. Siepel, A., Pollard, K.S. & Haussler, D. New Methods for Detecting Lineage-Specific Selection. 190-205 (Springer Berlin Heidelberg, 2006).
56. Cooper, G.M. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901-913 (2005).
57. Kimura, M. & Crow, J.F. The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics* **49**, 725-38 (1964).
58. McVean, G.A. & Charlesworth, B. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929-44 (2000).
59. Gordo, I. & Charlesworth, B. The speed of Muller's ratchet with background selection, and the degeneration of Y chromosomes. *Genet. Res.* **78**, 149-61 (2001).
60. Charlesworth, B., Morgan, M.T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-303 (1993).
61. Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-93 (2012).
62. Comeron, J.M. Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **372** (2017).
63. Torres, R., Stetter, M.G., Hernandez, R.D. & Ross-Ibarra, J. The Temporal Dynamics of Background Selection in Nonequilibrium Populations. *Genetics* **214**, 1019-1030 (2020).

64. Torres, R., Szpiech, Z.A. & Hernandez, R.D. Human demographic history has amplified the effects of background selection across the genome. *PLoS Genet.* **14**, e1007387 (2018).
65. Wang, L. *et al.* The interplay of demography and selection during maize domestication and expansion. *Genome Biol.* **18**, 215 (2017).
66. Li, R. *et al.* A high-resolution map of non-crossover events reveals impacts of genetic diversity on mammalian meiotic recombination. *Nat. Commun.* **10**, 3900 (2019).
67. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10**, 285-311 (2009).
68. Stamatoyannopoulos, J.A. *et al.* Human mutation rate associated with DNA replication timing. *Nat. Genet.* **41**, 393-5 (2009).
69. Gao, Z. *et al.* Overlooked roles of DNA damage and maternal age in generating human germline mutations. *Proc. Natl. Acad. Sci. USA* **116**, 9491-9500 (2019).
70. Sankararaman, S. *et al.* The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-7 (2014).
71. Vernot, B. & Akey, J.M. Resurrecting surviving Neandertal lineages from modern human genomes. *Science* **343**, 1017-21 (2014).
72. Racimo, F., Sankararaman, S., Nielsen, R. & Huerta-Sánchez, E. Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* **16**, 359-371 (2015).
73. Juric, I., Aeschbacher, S. & Coop, G. The Strength of Selection against Neanderthal Introgression. *PLoS Genet.* **12**, e1006340 (2016).
74. Sankararaman, S., Mallick, S., Patterson, N. & Reich, D. The Combined Landscape of Denisovan and Neanderthal Ancestry in Present-Day Humans. *Curr. Biol.* **26**, 1241-1247 (2016).
75. Harris, K. & Nielsen, R. The Genetic Cost of Neanderthal Introgression. *Genetics* **203**, 881-91 (2016).
76. Schumer, M. *et al.* Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656-660 (2018).
77. Harris, K. & Nielsen, R. The Genetic Cost of Neanderthal Introgression. *Genetics* **203**, 881-891 (2016).
78. Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-9 (2014).
79. Skov, L. *et al.* Detecting archaic introgression using an unadmixed outgroup. *PLoS Genet.* **14**, e1007641 (2018).
80. Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060 (2010).
81. Green, R.E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710-722 (2010).
82. Plagnol, V. & Wall, J.D. Possible ancestral structure in human populations. *PLoS Genet.* **2**, e105 (2006).
83. Durvasula, A. & Sankararaman, S. Recovering signals of ghost archaic introgression in African populations. *Sci. Adv.* **6**, eaax5097 (2020).
84. Wall, J.D., Lohmueller, K.E. & Plagnol, V. Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol. Biol. Evol.* **26**, 1823-7 (2009).

85. Wall, J.D. & Hammer, M.F. Archaic admixture in the human genome. *Curr. Opin. Genet. Dev.* **16**, 606-10 (2006).