Determining the Effect of Natural Selection on Linked Neutral Divergence across Species

Tanya N. Phung[1], Christian D. Huber[2], Kirk E. Lohmueller[1,2,3] *

1 Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, California, United States of America, **2** Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, California, United States of America, **3** Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California, United States of America

* klohmueller@ucla.edu

Abstract

A major goal in evolutionary biology is to understand how natural selection has shaped patterns of genetic variation across genomes. Studies in a variety of species have shown that neutral genetic diversity (intra-species differences) has been reduced at sites linked to those under direct selection. However, the effect of linked selection on neutral sequence divergence (inter-species differences) remains ambiguous. While empirical studies have reported correlations between divergence and recombination, which is interpreted as evidence for natural selection reducing linked neutral divergence, theory argues otherwise, especially for species that have diverged long ago. Here we address these outstanding issues by examining whether natural selection can affect divergence between both closely and distantly related species. We show that neutral divergence between closely related species (e.g. human-primate) is negatively correlated with functional content and positively correlated with human recombination rate. We also find that neutral divergence between distantly related species (e.g. human-rodent) is negatively correlated with functional content and positively correlated with estimates of background selection from primates. These patterns persist after accounting for the confounding factors of hypermutable CpG sites, GC content, and biased gene conversion. Coalescent models indicate that even when the contribution of ancestral polymorphism to divergence is small, background selection in the ancestral population can still explain a large proportion of the variance in divergence across the genome, generating the observed correlations. Our findings reveal that, contrary to previous intuition, natural selection can indirectly affect linked neutral divergence between both closely and distantly related species. Though we cannot formally exclude the possibility that the direct effects of purifying selection drive some of these patterns, such a scenario would be possible only if more of the genome is under purifying selection than currently believed. Our work has implications for understanding the evolution of genomes and interpreting patterns of genetic variation.

Author Summary

Genetic variation at neutral sites can be reduced through linkage to nearby selected sites. This pattern has been used to show the widespread effects of natural selection at shaping patterns of genetic diversity across genomes from a variety of species. However, it is not entirely clear whether natural selection has an effect on neutral divergence between species. Here we show that putatively neutral divergence between closely related species (human and chimp) and between distantly related pairs of species (humans and mice) show signatures consistent with having been affected by linkage to selected sites. Further, our theoretical models and simulations show that natural selection indirectly affecting linked neutral sites can generate these patterns. Unless substantially more of the genome is under the direct effects of purifying selection than currently believed, our results argue that natural selection has played an important role in shaping variation in levels of putatively neutral sequence divergence across the genome. Our findings further suggest that divergence-based estimates of neutral mutation rate variation across the genome as well as certain estimators of population history may be confounded by linkage to selected sites.

Introduction

Determining the evolutionary forces affecting genetic variation has been a central goal in population genetics over the past several decades. A large body of empirical and theoretical work has suggested that neutral genetic variation within a species (diversity) can be influenced by nearby genetic variants that are affected by natural selection (reviewed in [1]). This can occur via two mechanisms. In a selective sweep, a neutral allele linked to a beneficial mutation will reach high frequency [2,3]. Selective sweeps reduce neutral genetic variation near regions of the genome that are directly affected by natural selection. The second process, background selection, also reduces neutral genetic variation [4–7]. Here, purifying selection that eliminates deleterious mutations also removes nearby neutral genetic variation. Many empirical studies have found strong evidence for the effects of background selection and selective sweeps affecting patterns of neutral genetic diversity (intra-species DNA differences) across the human genome. For example, several studies have reported a correlation between genetic diversity and recombination rate [8–13]. This correlation can be driven by selective sweeps and background selection because these processes affect a larger number of base pairs in areas of the genome with a low recombination rate than with a high recombination rate. Additionally, other studies found reduced neutral genetic diversity surrounding genes [12–17], which is consistent with the idea that there is more selection occurring near functional elements of the genome.

While the evidence for natural selection reducing genetic diversity at linked neutral sites is unequivocal, the effect of natural selection on linked neutral divergence between species (inter-species DNA differences) is less clear. Elegant theoretical arguments have suggested selection does not affect the substitution rate at linked neutral sites [18,19]. However, these theoretical arguments do not include mutations that arose in the common ancestral population, the population that existed prior to the split and formation of two descendant lineages. Such ancestral polymorphism has been shown to be a significant confounder in estimating population divergence times [20]. When also including ancestral polymorphism, it becomes less clear whether selection affects divergence at linked neutral sites.

Based on coalescent arguments, neutral polymorphism in the ancestral population will be affected by linkage to selected sites the same way as genetic diversity within a population (Fig 1). Presumably, neutral divergence between closely related species, with lots of ancestral
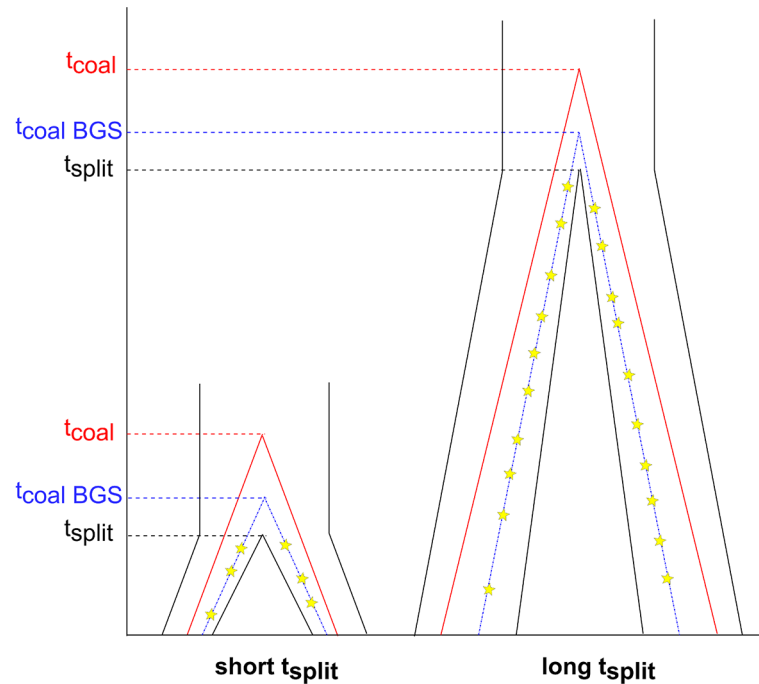
Fig 1. Models of how genealogies are affected by selection at linked neutral sites. The genealogies on the left represent species with a short split time such as human and chimpanzee. The genealogies on the right represent species with a long split time such as human and mouse. Red lines represent two lineages and their coalescent time. Blue lines represent two lineages and their coalescent time when there is selection at linked neutral sites in the ancestral population (abbreviated BGS). Yellow stars denote mutations accumulating on each of the two lineages after they split. Note that with the longer split time, the proportion of the genealogy attributed to the ancestral population decreases.

polymorphism, could be affected by selection. Indeed, McVicker et al. [15] demonstrated that background selection could explain the variation in human-chimp neutral divergence across the genome. Additionally, Cruickshank and Hahn [21] found that divergence between recently separated species pairs was reduced in regions of low recombination and in "islands of speciation". They attributed at least some of these patterns to selection affecting linked neutral sites.

However, the reduction in neutral diversity in the ancestral population is thought to have a negligible effect and/or be undetectable when considering neutral divergence from species with a very long divergence time [9,18] because there would be many opportunities for mutations to occur after the two lineages split (Fig 1). These neutral mutations that occur after the split would not be influenced by selection at linked neutral sites [18] and would dilute the signal from the ancestral polymorphism. Thus, it is generally believed that selection at linked neutral sites should not affect divergence between distantly related species. An example of this argument was presented by Hellmann et al. [9]. They argued that the positive correlation between human-baboon divergence and human recombination was due to mutagenic recombination, rather than selection affecting linked neutral sites, because of the long split time between humans and baboons (>20 million years). Reed et al. [22] suggested that though it is unlikely background selection by itself could explain the entire correlation observed by Hellmann et al., background selection may still contribute to divergence. However, there has been little quantitative investigation of the effect that selection has on divergence at linked neutral sites among distantly divergent species when including ancestral polymorphism.

In addition to conflicting conceptual predictions about the expected effect of selection on divergence at linked neutral sites, empirical studies also have been ambiguous. While some studies found no evidence for a correlation between divergence and recombination such as in *Drosophila* [23,24] or in yeast [25], other studies have reported correlations between divergence and recombination in *Drosophila* [26,27]. Further, positive correlations between human-chimpanzee divergence and human recombination rate [10,12,13], human-macaque divergence and human female recombination rate [28], or human-baboon divergence and human recombination rate [9] have been reported. Finally, even though there was evidence for a strong reduction in human-chimpanzee divergence and human-macaque divergence surrounding genes [15,28], McVicker et al. [15] attributed the reductions seen for human-dog divergence to variation in mutation rates. Thus, the degree to which divergence is affected by selection across species with different split times remains elusive.

Determining whether and how selection affects linked neutral divergence is critical to understanding the evolutionary forces influencing genetic variation and mutational processes. If selection in the ancestral population only has a limited effect on divergence, it would suggest correlations between recombination and divergence to be evidence of mutagenic recombination. This may further suggest the need to consider recombination rates when modeling variation in mutation rates across the genome [9,29–32]. Because mutations rates have been difficult to estimate reliably in humans [33,34], understanding the biological factors influencing them will be of paramount importance for obtaining improved estimates. If, on the other hand, selection can affect linked neutral divergence, reductions of linked neutral divergence surrounding genes would suggest an abundance of selection affecting linked neutral sites [35]. Selection affecting linked neutral diversity and divergence is at odds with the neutral and nearly neutral theories [36–38], which have been the prevailing views in molecular population genetics for the last several decades. It would also suggest the need to consider the effects of selection when estimating mutation rates from neutral divergence.

Here we aim to examine the effects of selection on linked neutral divergence for pairs of species with a range of split times. We first present evidence that neutral divergence is reduced at putatively neutral sites close to selected sites across a wide range of taxa, including those with split times as long as 75 million years ago. Factors such as hypermutable CpG sites, GC content, or biased gene conversion by themselves cannot explain these results. We then use coalescent simulations to explore whether models incorporating background selection in the ancestral population could generate the empirical patterns. We also present a theoretical argument as to how background selection can affect variation in neutral divergence across the genome, even for species with a long split time such as human and mouse. Finally, we show that purifying selection directly reducing divergence at putatively neutral sites cannot explain these findings unless a large fraction of the genome is directly under selection, or there is a substantial number of sites under selection in the human or mouse lineage that are not conserved across species. Even though we cannot formally reject the direct effects of purifying selection from driving some of these correlations, our empirical and simulation-based findings indicate that natural selection can indirectly affect neutral genetic divergence. In sum, the view that selection does not affect divergence at linked neutral sites between distantly diverged species should be reconsidered.

Results

Obtaining putatively neutral divergence

We wished to test whether the genetic divergence at a linked neutral site is influenced by the indirect effects of natural selection. As such, we set out to obtain putatively neutral sites by

removing sites that were potentially functional and under the direct effects of purifying selection. In particular, a site was considered putatively neutral if it was (1) located at least 5kb from the starting or ending position of an exon, (2) not located within a phastCons element that was calculated over different phylogenic scopes, (3) not alignable between human and zebrafish, and (4) not found within the top 10% of most conserved Genomic Evolutionary Rate Profiling (GERP) scores [39]. Criteria 2 and 3 remove sites that are likely to be conserved across species and therefore not neutral. We chose these filtering criteria following previous studies [12,13,40]. Additionally, we chose to remove the top 10% of sites having the most extreme GERP scores because previous work suggests <10% of the genome was under the direct effect of selection [39,41–49]. The putatively neutral sites close to genes show comparable levels of divergence to four-fold degenerate sites (S1 Fig, S1 Table). As four-fold degenerate sites are often used as a neutral standard in molecular evolution, the fact that they show similar levels of divergence as our putatively neutral noncoding sites argues that our putatively neutral sites are unlikely to be under additional direct effects of selection.

Effects on putatively neutral divergence between humans and primates

To understand the evolutionary factors affecting linked neutral divergence between closely related species, we examined human-primate divergence, particularly human-chimp divergence and human-orangutan divergence. First, we explored the relationship between neutral human-primate divergence and functional content, defined as the proportion of sites within a 100kb-window that overlapped with an exon or a phastCons region. We hypothesized that if natural selection contributes to the reduction of divergence at linked neutral sites, its effect would be more pronounced at regions with greater functional content [14]. This hypothesis predicts a negative correlation between functional content and neutral divergence. To test this, we divided the human genome into non-overlapping windows of 100kb and obtained putatively neutral divergence for each window as described above. We found a negative correlation between functional content and neutral divergence between pairs of closely related species (Spearman's $\rho_{human-chimp}$ = -0.235, $P < 10^{-16}$, Spearman's $\rho_{human-orang}$ = -0.204, $P < 10^{-16}$, Fig 2A and 2B, S2 Table).

We next examined the relationship between human-primate neutral divergence and broad-scale human recombination rate which we obtained from the deCODE genetic map [50]. While recombination has not been conserved throughout evolutionary history, the recombination rate at the broad-scale level (i.e. 100kb) was shown to be correlated between human and chimp [51,52]. We found a positive correlation between neutral human-primate divergence and human recombination rate (Spearman's $\rho_{human-chimp}$ = 0.234, $P < 10^{-16}$, Spearman's $\rho_{human-orang}$ = 0.249, $P < 10^{-16}$, Fig 2C and 2D, S3 Table), which indicates that neutral human-primate divergence is reduced in regions of low recombination rate. Additionally, when we stratified windows into those that were near genes and those that were far from genes based on the proportion of sites in each window that overlapped with a RefSeq transcript, we found that the correlation between divergence and recombination is stronger for windows with a higher overlap with RefSeq transcripts (S2 Fig). These observations indicate that neutral divergence is reduced at sites that are more tightly linked to those under the direct effect of selection, consistent with the hypothesis that natural selection indirectly reduces linked neutral divergence.

These two correlations are robust to the presence of multiple confounding factors. First, the correlations are robust to the choice of window size used for analysis as they persisted when using 50 kb windows (S2 and S3 Tables). Second, some features of the genome such as hyper-mutable CpG sites or GC content are known to correlate with genic content [10,12,13]. To test whether these features confounded the correlations found in our data, we repeated our analyses
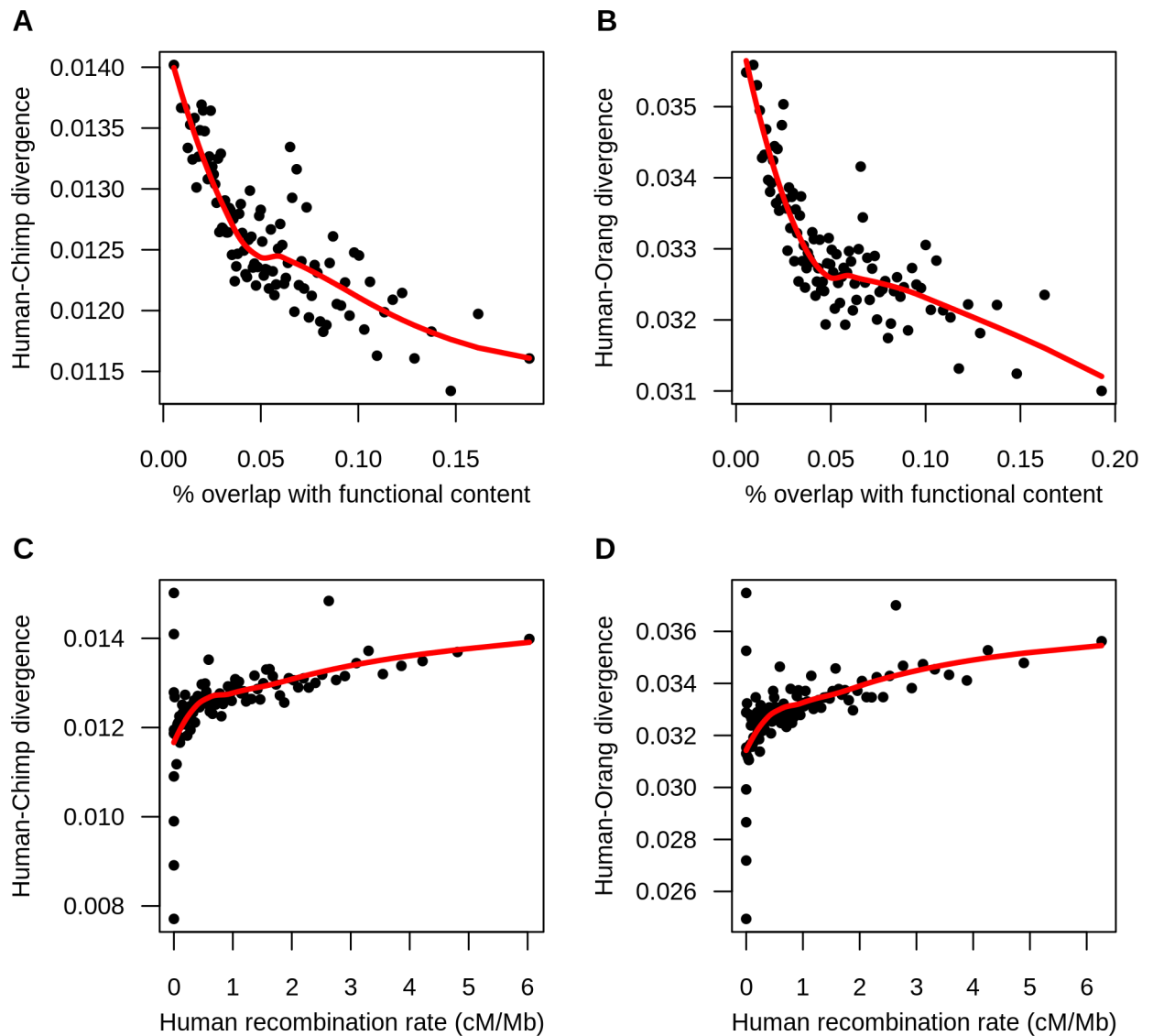
Fig 2. Human-primate divergence is reduced at putatively neutral sites near selected sites. (A) Neutral human-chimp divergence is negatively correlated with functional content. (B) Neutral human-orang divergence is negatively correlated with functional content. (C) Neutral human-chimp divergence is positively correlated with human recombination rate. (D) Neutral human-orang divergence is positively correlated with human recombination rate. Each point represents the mean divergence and functional content (A and B) or recombination rate (C and D) in 1% of the 100kb windows binned by functional content or recombination rate. Red lines indicate the loess curves fit to divergence and functional content (A and B) and divergence and recombination rate (C and D). The high variance of divergence at regions of low recombination rate is expected since the variance of divergence is inversely proportional to the recombination rate. Note that the last bin containing less than 1% of the windows was omitted from the plot. While the graph presents binned data, the correlations reported in the text are from the unbinned data.

doi:10.1371/journal.pgen.1006199.g002

removing potential CpG sites by omitting sites preceding a G or following a C [15]. The correlations persisted after filtering out CpG sites (S2 and S3 Tables). We next computed partial correlations controlling for GC content. Similarly, we found that the correlations persisted (S2 and S3 Tables).

Biased gene conversion is an additional evolutionary force that has been shown to influence patterns of divergence [53,54]. In this process, double-strand breaks in the DNA in individuals heterozygous for AT/GC variants will be preferentially repaired with the GC allele, resulting in

AT → GC substitutions occurring at a higher rate than GC → AT substitutions [54–56]. To control for the effects of biased gene conversion on this analysis, we filtered out sites that could be affected by removing any AT → GC substitutions genome-wide. The negative correlation between human-primate divergence and functional content did not change after controlling for biased gene conversion (S2 Table). Though the positive correlation between human-primate divergence and human recombination decreased after this filter (from 0.234 to 0.108), it still remained significant (S3 Table). Thus, the observed correlations are unlikely to be driven solely by choice of window size or mutational properties based on sequence composition. Because biased gene conversion appears to contribute to some of the correlation between divergence and recombination rate, subsequent analyses of this correlation use the divergence dataset filtered for biased gene conversion.

Effects on putatively neutral divergence between humans and rodents

We next explored the evolutionary forces affecting divergence between more distantly related pairs of species, specifically human-mouse and human-rat. These species were predicted to have diverged approximately 75 million years ago [41] and, as such, current thinking would predict that natural selection would not affect linked neutral sites. Similar to what was seen for the closely related species, functional content is negatively correlated with neutral human-rodent divergence (Spearman's $\rho_{human-mouse}$ = -0.184, $P < 10^{-16}$, Spearman's $\rho_{human-rat}$ = -0.149, $P < 10^{-16}$, Fig 3A and 3B, S4 Table). This negative correlation persisted when using 50kb windows and also after accounting for the confounding factors of hypermutable CpG sites, GC content, and GC-biased gene conversion (S4 Table).

Since the broad-scale recombination rate at 100kb appears to have changed over the course of evolution of the species [57], we looked for other potential signatures of whether natural selection has affected linked neutral divergence. In particular, we examined the relationship between human-rodent divergence and the strength of background selection across the genome inferred from divergence within primates [15]. This strength of background selection is captured by the B-value, which represents the degree to which neutral variation at a given position is reduced by selection relative to neutral expectations. While McVicker et al. [15] concluded that divergence between primates was indeed reduced due to background selection, they did not consider human-mouse divergence in their analyses and did not model background selection within the human-dog ancestor. As such, there is no *a priori* reason why the B-values of McVicker et al. [15] should be related to human-mouse divergence.

Nevertheless, we found a positive correlation between human-rodent divergence and the B-values from McVicker et al. [15] (Spearman's $\rho_{human-mouse}$ = 0.445, $P < 10^{-16}$, Fig 3C, S5 Table, Spearman's $\rho_{human-rat}$ = 0.402, $P < 10^{-16}$, Fig 3D, S5 Table). The positive correlation between human-rodent divergence and B-values remained significant even after accounting for the confounding factors of CpG sites, GC content, and GC-biased gene conversion. Similarly, these correlations remained when using 50kb windows (S5 Table). Taken together, the empirical correlations are consistent with the hypothesis that natural selection has contributed to reducing neutral divergence at linked sites even between species with a long split time such as human and mouse.

Models incorporating background selection in the ancestral population can generate the empirical correlations

To test whether a model including background selection in the ancestral population can explain the empirical observations regarding neutral human-primate divergence and neutral human-rodent divergence, we used a coalescent simulation approach. To a first approximation,
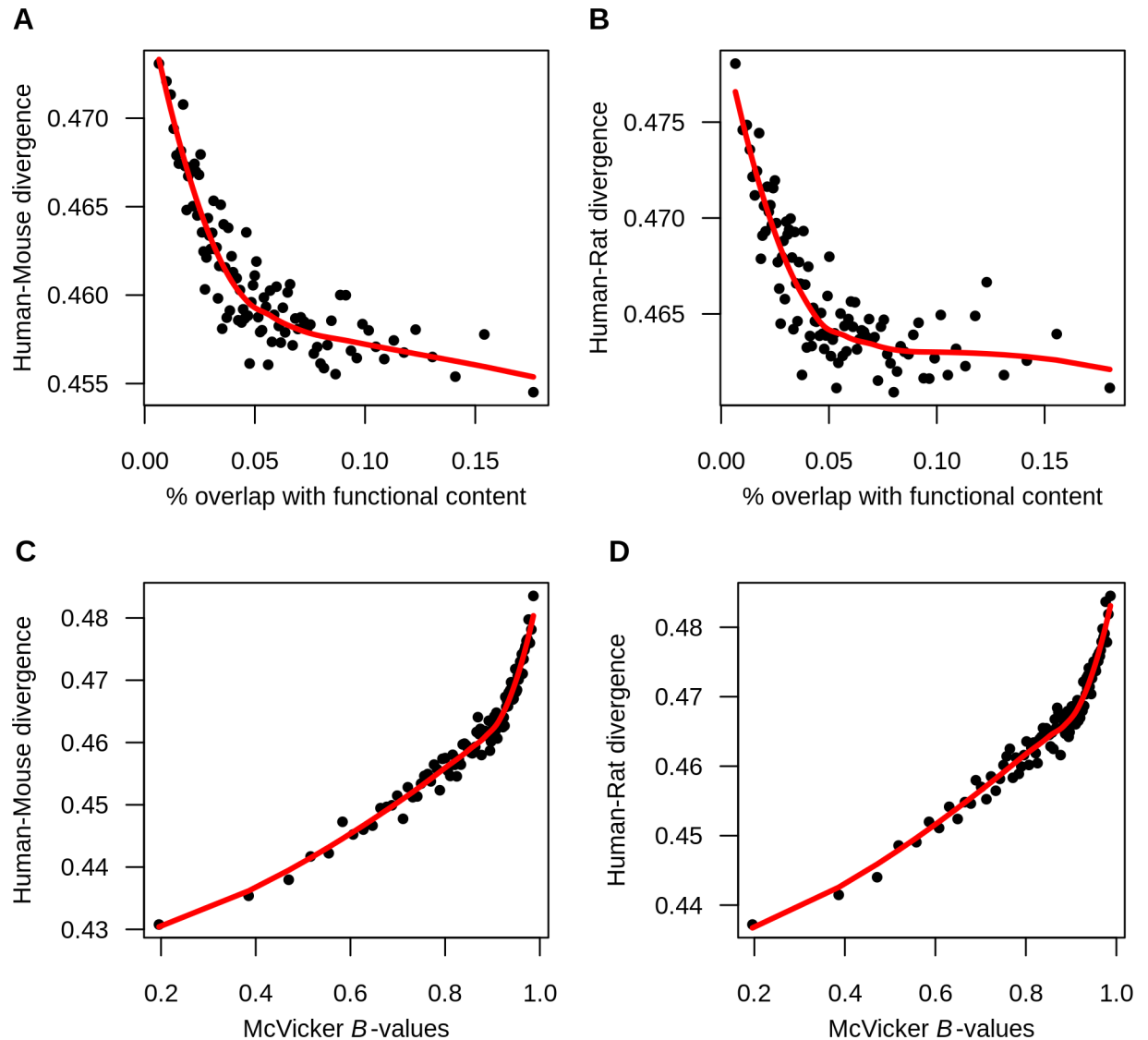
Fig 3. Human-rodent divergence is reduced at putatively neutral sites near selected sites. (A) Neutral human-mouse divergence is negatively correlated with functional content. (B) Neutral human-rat divergence is negatively correlated with functional content. (C) Neutral human-mouse divergence is positively correlated with McVicker's *B*-values. (D) Neutral human-rat divergence is positively correlated with McVicker's *B*-values. Each point represents the mean divergence and functional content (A and B) or *B*-values (C and D) in 1% of the 100kb windows binned by functional content or *B*-values. Red lines indicate the loess curves fit to divergence and functional content (A and B) and divergence and *B*-values (C and D). Note that the last bin containing less than 1% of the windows was omitted from the plot. While the graph presents binned data, the correlations reported in the text are from the unbinned data.

doi:10.1371/journal.pgen.1006199.g003

the effect of background selection in a sample size of two chromosomes can be accounted for by scaling the ancestral population size by the strength of background selection [4,5,7,15,58–62]. Thus, we modeled the effect of background selection as a reduction in the ancestral population size using the *B*-values estimated in McVicker et al. [15]. Briefly, we first used *ms* [63] to generate genetic variation in the ancestral population where the ancestral population has size $N_a B$. Then we simulated mutations that accumulated since the split between two species using a Poisson process. The total divergence was the sum of the mutations in the ancestral population and mutations accumulated since the split (see Methods). We modeled mutation rate

variation by drawing a mutation rate for each window from a gamma distribution. We chose values for the parameters of the gamma distribution as well as the ancestral population size (N_a) such that the mean and standard deviation of the simulated divergence across the genome and the correlation coefficients between divergence and other functional properties were similar to those seen empirically (S3 Fig, S6 and S7 Tables, Methods).

We first examined which models could generate the observed correlation between recombination and human-chimp divergence. Here we use the value of Spearman's ρ estimated from the data after filtering out sites that could be affected by biased gene conversion (ρ = 0.108). When considering a model without background selection (i.e. $B = 1$ for all windows), the average value of Spearman's ρ between human-chimp divergence and recombination rate was 0.042, and none of the 500 simulation replicates approached the value of Spearman's ρ seen empirically (Fig 4A, **white histogram**). On the other hand, when modeling background selection using the McVicker B-values, the average Spearman's ρ was 0.107 which was comparable to the Spearman's ρ computed from empirical human-chimp divergence with human recombination after accounting for biased gene conversion (Fig 4A, **gray histogram**).

We then tested whether a model incorporating background selection could generate a positive correlation between neutral human-rodent divergence and B-values as observed empirically. We modified our simulation approach to account for the difference in generation time
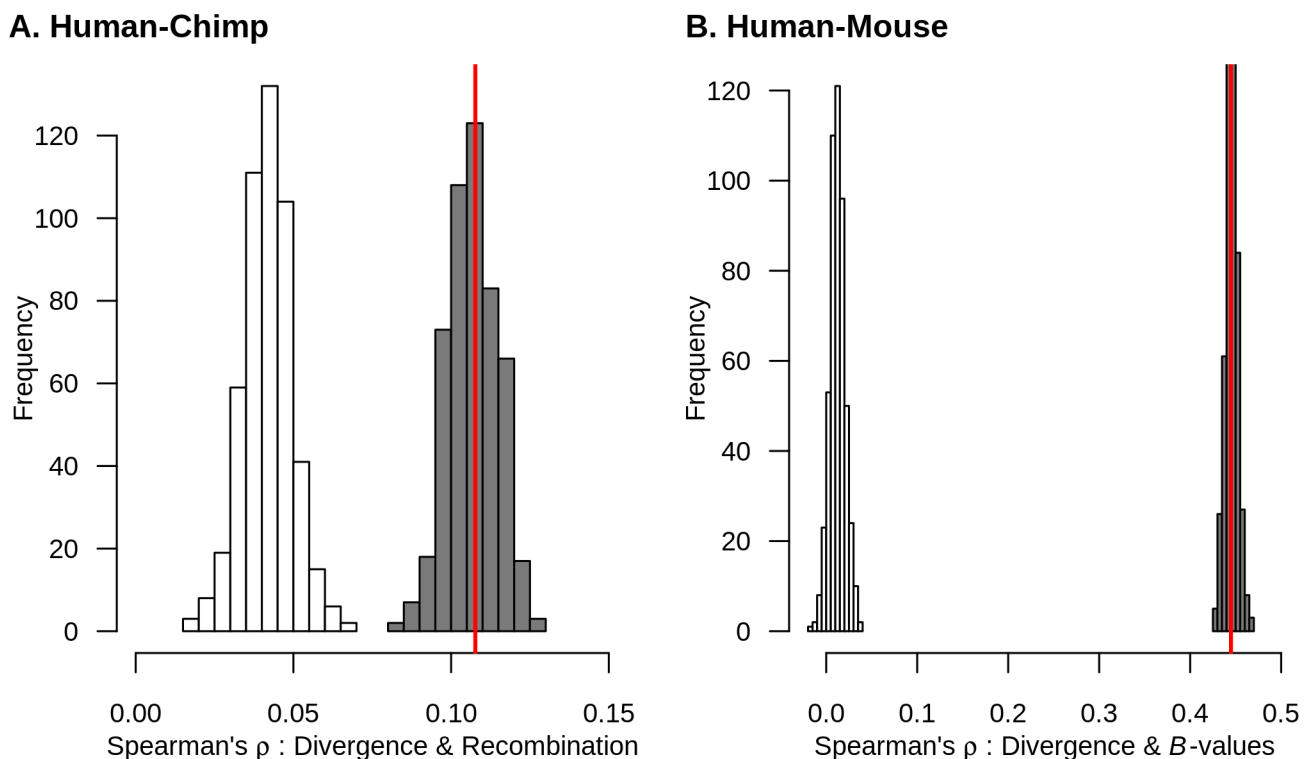

A. Human-Chimp

B. Human-Mouse

Fig 4. Models incorporating background selection can generate patterns of neutral divergence that recapitulate the empirical correlations. (A) Models of background selection predict a positive correlation between neutral human-chimp divergence and human recombination. Because our model does not include biased gene conversion, the empirical correlation was calculated omitting AT to GC sequence differences. (B) Models of background selection predict a positive correlation between neutral human-mouse divergence and McVicker's B-values. White histogram denotes 500 simulations not including background selection. Gray histogram denotes 500 simulations incorporating background selection (see text). Red line represents the correlation computed from empirical data. Thus, plausible levels of background selection can match the observed correlations while neutral simulations cannot.

doi:10.1371/journal.pgen.1006199.g004

between human and mouse (see Methods). When considering models without background selection (i.e. $B = 1$ for all windows), the average value of Spearman's ρ was 0.012, and none of the 500 simulation replicates approached the value of Spearman's ρ seen empirically (Fig 4B, **white histogram**). However, when modeling background selection using the McVicker B-values, the average Spearman's ρ was 0.446 which was comparable to the Spearman's ρ computed from empirical human-mouse divergence and McVicker's B-values (Fig 4B, **gray histogram**).

In sum, our results suggest that for a given set of parameters, a model with background selection in the ancestral population can generate the correlations observed in the empirical data (i.e. a positive correlation between neutral human-primate divergence and human recombination and a positive correlation between neutral human-rodent divergence and B-values) whereas neutral coalescent models cannot.

Intuition for why background selection is a plausible explanation for the empirical correlations

Current thinking argues that natural selection affecting linked neutral sites is not a plausible explanation for the reduction in neutral divergence between pairs of species with a long split time such as human-mouse or human-rat. Here, we outline a theoretical analysis of a simple two-locus model to gain intuition about how the mutation rate (μ), strength of background selection (B), and ancestral population size (N_a) affect the degree to which background selection can affect divergence (Fig 5A).

If background selection has any effect on the variation in neutral divergence across the genome, this can only be due to its effect on divergence in the ancestral population, since deleterious mutations do not affect the fixation rate at linked neutral sites [18]. Recombination in the ancestral population results in a distribution of coalescent times within each locus, with an average coalescent time of $\bar{t}$. We assumed that the recombination rate within each locus is large enough, such that there is no variation in $\bar{t}$ for a fixed value of B, i.e. $\text{Var}[\bar{t}|\text{B}] \approx 0$. This is a reasonable assumption as long as the window size and recombination rate are not too small. Recombination events cause the sequence to be broken into independent segments, such that for a total $\rho > 10$ (where ρ denotes the population-scaled recombination rate, $4N_e r$) the variance in $\bar{t}$ approaches zero [64]. For an average 100kb window in the human genome ($r = 10^{-8}/$ bp, $N_e = 10,000$), ρ is 40 and thus this assumption holds true. Any difference in $\bar{t}$ between loci is then only attributable to differences in background selection: $\text{E}[\bar{t}|B] = 2N_a B$. Further, variation in ancestral (d_a) and total (d_t) divergence results from a Poisson distributed number of mutations added to the genealogy, such that $Var[d_a|B] = E[d_a|B] = 4N_a B\mu L$ and $Var[d_t|B] = E[d_t|B] = E[d_a|B]+2t_{split}\mu L$ where L is the sequence length of a locus. The law of total variance can be used to compute the variance in total divergence across loci with varying levels of background selection:

$$Var[d_t] = Var_B[E[d_t|B]] + E_B[Var[d_t|B]]$$

Thus, variance in total divergence can be decomposed into variance due to background selection and variance due to the mutational process. For simplicity, the first locus experiences no background selection ($B_1 = 1$), and the second locus experiences some fixed amount of background selection ($0 \leq B_2 \leq 1$). Under this model, we computed the variance due to background selection as:

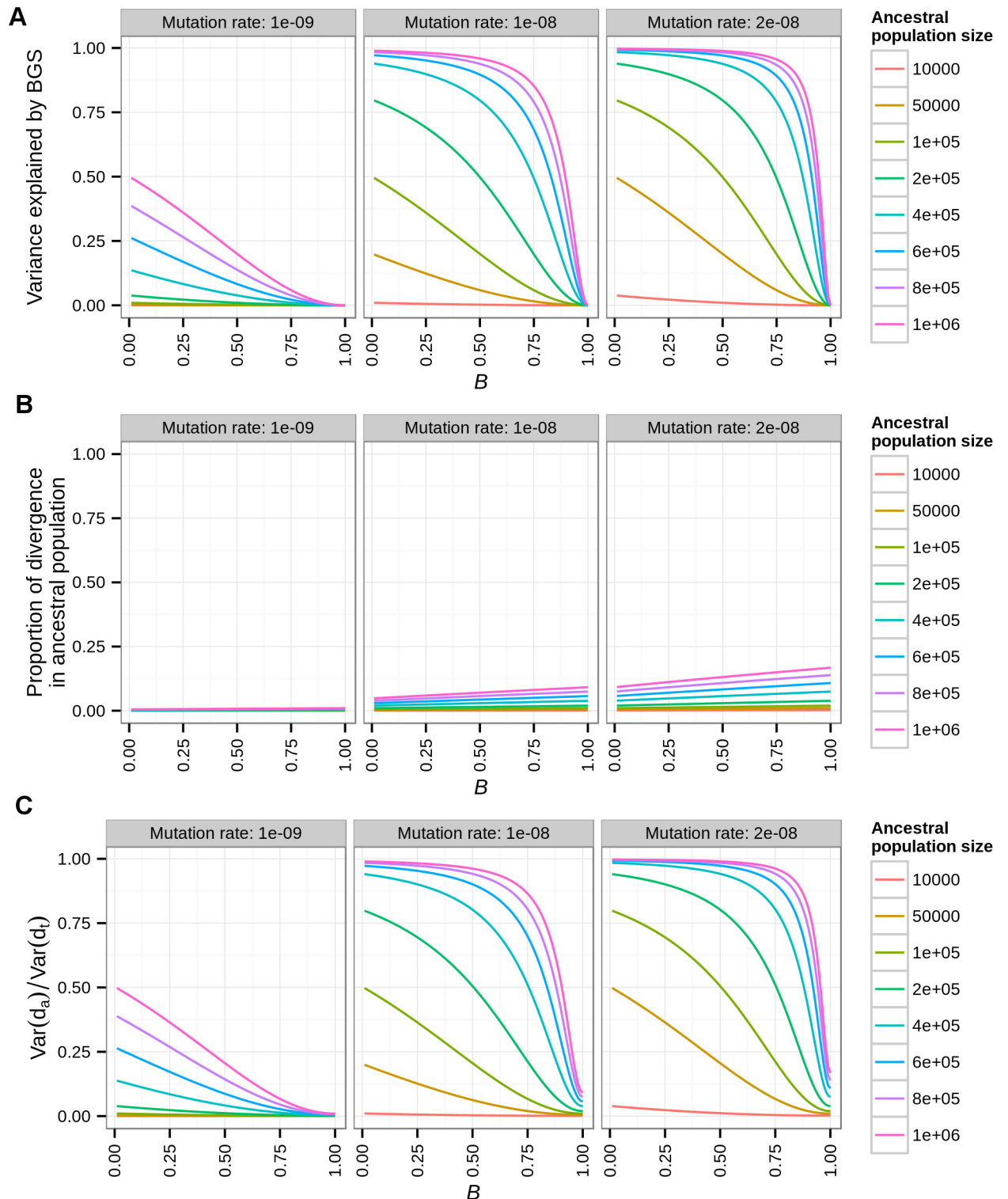$$Var_B[E[d_t|B]] = ((E[d_t|B = 1] \; - \; E[d_t|B = B_2])/2)^2.$$

Fig 5. A two-locus model for the effect of background selection on divergence. (A) The variance in divergence between two loci explained by background selection (BGS) as a function of the strength of background selection at the second locus (B_2). (B) The expected proportion of divergence due to polymorphism in the ancestral population as a function of B_2. (C) The variance in divergence between the two loci explained by polymorphism in the ancestral population as a function of B_2. Different columns denote different mutation rates. Colored lines denote different ancestral population sizes (N_a). Note that the variance in divergence attributable to background selection is greater than the expected proportion of divergence contributed by ancestral polymorphism.

doi:10.1371/journal.pgen.1006199.g005

We then computed the variance due to the mutational process as:

$$E_B[Var[d_t|B]] = (Var[d_t|B = 1] + Var[d_t|B = B_2])/2.$$

We assumed an old split time, such that the divergence that accumulated from present time to population split is similar to the human-mouse divergence (40%). Both loci have a sequence length (L) of 100kb. Our theoretical analysis of variance approach shows that with this old split time and assuming a low mutation rate of 1×10^{-9}/bp, more than 20% of the variation in the divergence can be explained by background selection in the ancestral population with the following conditions: ancestral population size $> 600,000$ and $B < 0.2$ (Fig 5A, **panel 1, blue, purple, and pink lines**). Note that under these conditions, the proportion of divergence that accumulated in the ancestral population can be as low as 0.3% (Fig 5B, **panel 1**). However, the proportion of the variance in divergence that is attributable to the ancestral population is larger than 20% (Fig 5C, **panel 1**), mainly due to background selection leading to differences in $\bar{t}$ between loci. With a larger mutation rate (2×10^{-8}/bp), background selection results in a stronger effect on variation in divergence even when ancestral population size is relatively small ($>50,000$; Fig 5A, **yellow line**). When assuming a moderately large population size of 200,000, and a moderate strength of background selection ($B = 0.75$), then as much as 50% of variance in divergence can be explained by background selection (Fig 5C, **light green line**). Nonetheless, the proportion of divergence that accumulated in the ancestral population in this case is still only 3.4%. Collectively, even for old split times, where the vast majority of divergence accumulated after the population split, with certain assumptions about the ancestral population size, mutation rate, and strength of background selection, the variance in the divergence could be explained by background selection.

Coalescent simulations predict background selection can reduce neutral divergence between species with long split times

Because the theoretical model described above ignores regions of low recombination and only considers one pair of loci at a time, we used coalescent simulations (similar to what we outlined above) to examine whether background selection could generate the positive correlation between estimates of background selection in primates and divergence between distantly related species using more realistic models. Since we were not particularly concerned with any specific species, we simplified these simulations by setting the mutation rate to 2.5×10^{-8}/bp.

We found that across all population sizes and split times examined, background selection generated a positive correlation between recombination and divergence as well as a positive correlation between divergence and B-values, even for pairs of species that split up to $100N$ generations ago (S4 Fig, **black lines and dashed lines**). This correlation remained strong even when the proportion of the divergence due to ancestral polymorphism was small. For example, for a pair of populations with $t_{split} = 100N$ generations and an ancestral population of size 50,000, only 1.53% of the divergent sites are due to ancestral polymorphism (S4 Fig, **red lines**). However, this model predicts a correlation of 0.211 between recombination and divergence and a correlation of 0.377 between recombination and B-values. Although ancestral polymorphism only contributes in a small way to the total divergence, the variance in the amount of ancestral polymorphism across the windows accounts for nearly 60% of the variance in divergence across different windows (S5 Fig, **black lines**). In general, the correlations decreased as both the split time increased and the size of the ancestral population decreased (S5 Fig). This behavior is expected as the contribution of the variance in levels of ancestral polymorphism to the variance in divergence decreases with increasing split time and decreasing ancestral population size (S5 Fig).

Examining the direct effects of natural selection on observed correlations

While we have shown under a variety of models that natural selection can affect putatively neutral divergence and generate the correlations that we observe empirically, other selective scenarios could explain these patterns. An alternative explanation for the empirical correlations reported in Figs 2 and 3 is that the filtering criteria we used to obtain neutral sites did not effectively remove all non-neutral sites. Therefore, the observed correlations could be due to the direct effects of purifying selection reducing genetic divergence. As sites under purifying selection may be located close to conserved functional elements and could conceivably result in low B-values, this is a potentially plausible explanation for our findings. As our current filters removed the 10% of the genome that was most likely under the direct effect of selection based upon the top 10% of GERP scores, we reasoned that additional sites under purifying selection would have elevated GERP scores relative to neutrality.

To test this hypothesis, we repeated our correlation analyses by first obtaining the neutral human-primate divergence and neutral human-rodent divergence using different GERP score cutoffs (i.e. 5% to 25%). When examining human and primate pairs, the correlation between neutral human-primate divergence and functional content decreased as a function of increasing GERP cutoff score (Fig 6A). Nevertheless, the negative correlation between neutral human-primate divergence and functional content remained significant even after removing any site whose GERP score fell within the top 25% of the distribution (Spearman's $\rho_{human-chimp}$ = -0.189, $P < 10^{-16}$, Spearman's $\rho_{human-orang}$ = -0.122, $P < 10^{-16}$, S6A and S6B Fig). On the other hand, the relationship between neutral human-primate divergence and human recombination rate were not affected by varying GERP score cutoffs (Fig 6B, S6C and S6D Fig).

When examining human and rodent pairs, we found that the negative correlation between human-rodent divergence and functional content decreased as a function of increasing GERP score cutoff. Further, the relationship became nonsignificant when filtering any site whose GERP score fell within the top 15th percentile (Fig 6C, S7A and S7B Fig). The positive correlation between neutral human-rodent divergence and McVicker's B values decreased as a function of increasing GERP score cutoff, but remained significantly positive even after removing any sites whose GERP score fell within the top 25th percentile (Fig 6D, S7C and S7D Fig). Still, this latter pattern indicates that the direct effects of natural selection are unlikely to explain our findings, unless the selected sites are not in the upper 25% of the GERP score distribution.

To test whether background selection could explain these correlations when removing the 25% of the genome with the most conserved GERP scores, we used our coalescent simulation framework. These simulations match the empirical distribution of divergence across the genome (S8 Fig, S7 Table) and use the parameters given in S6 Table. For human-chimp divergence, none of the 500 neutral coalescent simulations resulted in a Spearman's ρ between divergence and human recombination rate as large as observed empirically after filtering sites affected by biased gene conversion (S9A Fig, **white histogram**). On the other hand, simulations including background selection in the ancestral population generated a Spearman's ρ between divergence and human recombination rate similar to what was observed empirically after filtering sites affected by biased gene conversion (S9A Fig, **gray histogram**). Similarly, for human-mouse divergence, while none of the 500 coalescent simulations using the neutral model could generate a Spearman's ρ between divergence and McVicker's B-values as large as the empirical correlation, models including background selection in the ancestral population could generate this correlation (S9B Fig).
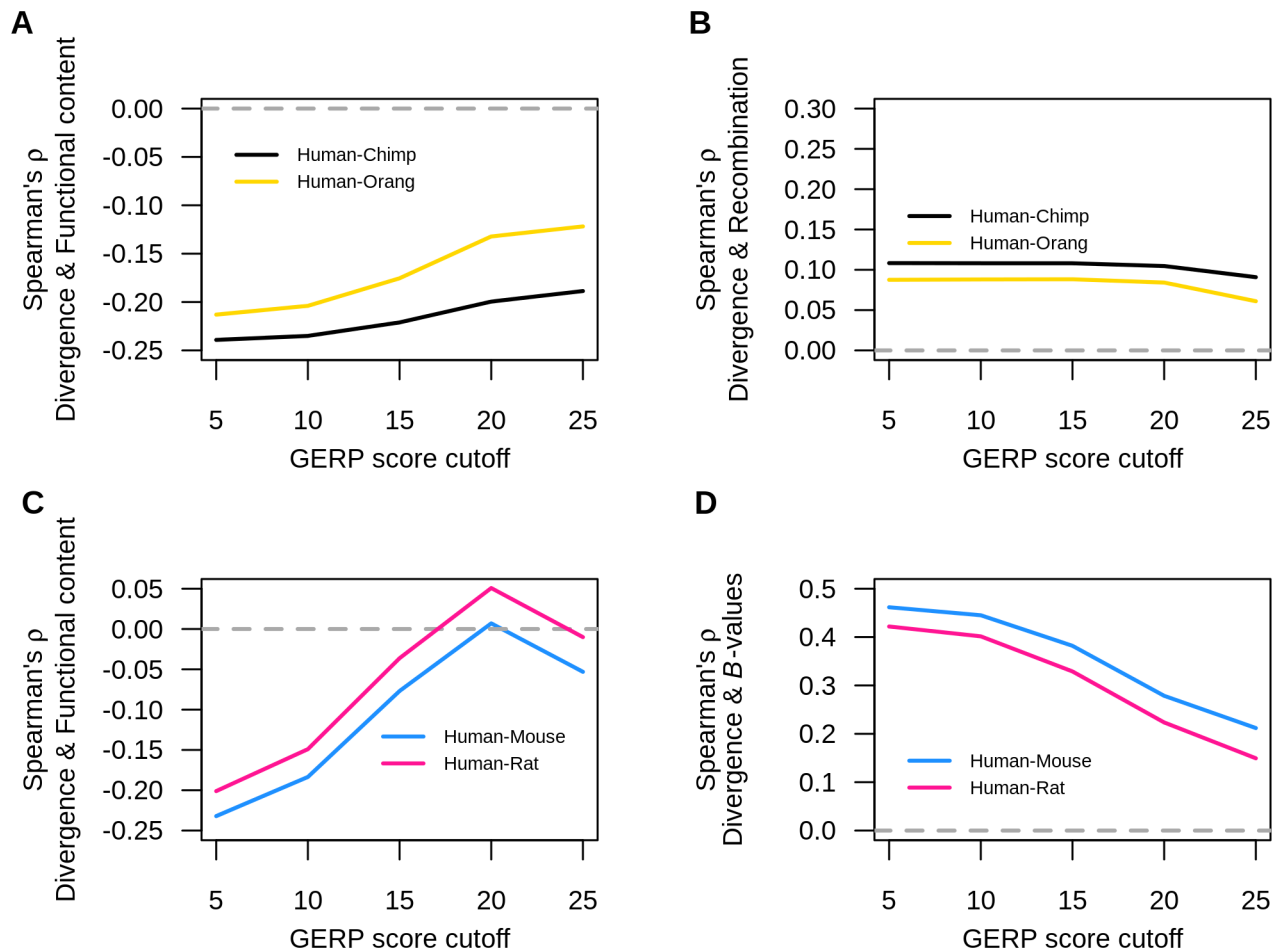
Fig 6. Relationship between divergence and functional content, human recombination, and McVicker's *B*-values as a function of GERP score cutoff. (A) Human-primate divergence versus functional content. (B) Human-primate divergence versus human recombination rate. (C) Human-rodent divergence versus functional content. (D) Human-rodent divergence versus McVicker's *B*-values.

doi:10.1371/journal.pgen.1006199.g006

Discussion

Here we have examined patterns of divergence between pairs of species with various degrees of divergence. We document several signatures that are consistent with the action of natural selection reducing divergence at linked neutral sites. First, for all pairs of species considered, we find that neutral divergence is lowest in regions of the genome with the greatest functional content (Figs 2 and 3). This pattern may be expected if more selection occurs in regions of the genome with greater functional content. Second, human-primate neutral divergence strongly correlates with human recombination rate and the correlation persists after accounting for hypermutable CpG sites, GC content, and biased gene conversion. Regions of low recombination show lower levels of divergence, which is consistent with selection having a greater effect on linked neutral sites in regions of low recombination. The correlation between human-primate divergence and human recombination is higher in regions with greater overlap with RefSeq transcripts, indicative of a greater reduction in neutral divergence in regions near genes as opposed to far from genes (S2 Fig). Third, human-rodent neutral divergence strongly correlates with the strength of background selection estimated for primates. These correlations persist after accounting for CpG sites, GC content, and biased gene conversion. Importantly, coalescent simulations

including background selection can generate several of these correlations. However, neutral coalescent models without background selection do not.

One interesting observation made was that while most of our correlation analyses were robust to the confounding effect of biased gene conversion, the correlation between human-primate neutral divergence and recombination rate was affected significantly by biased gene conversion. This suggests that while some of the correlation between recombination and divergence can be driven by biased gene conversion, it cannot explain the entire correlation. This result also argues that when testing for a correlation between divergence and recombination, the effect of biased gene conversion should be taken into account.

While we found that models incorporating background selection predict correlations comparable to the empirical data, in principle, several other evolutionary processes may be able to generate these patterns. First, selective sweeps in the ancestral population could reduce divergence just like background selection. Given that we are unlikely to be able to survey patterns of polymorphism in the human-mouse ancestor in more than two lineages, it will be difficult or nearly impossible to distinguish between these two types of selection at linked neutral sites. Thus, one should interpret our use of B-values as reflecting a reduction in divergence due to the combined effects of both background selection and selective sweeps, as suggested in McVicker et al. [15].

A second possibility is that the negative correlation between divergence and functional content as well as the positive correlation between divergence and B-values could be driven by variation in mutation rate across the genome. Indeed, McVicker et al. [15] attributed a positive correlation between B-values and human-dog divergence to the effects of variable mutation rates. However, for this mechanism to explain our results, it would require that mutation rates would have to be lower closer to genes and in regions of the genome thought to experience more background selection (i.e. in regions with lower B-values). There is some limited evidence of this effect in Arabidopsis where mutation rates are higher in regions of the genome with greater heterozygosity [65]. However, the extent to which these results apply to mammalian genomes remains unclear. Further, other studies in humans do not support the view that mutation rates are systematically lower in regions of the genome more subjected to selection. Recent estimates of the *de novo* mutation rate have not found any evidence of a reduction close to genes [32]. Further, Palamara et al. [66] found that their estimates of the mutation rate do not differ as a function of B-values. Variation in mutation rate across the genome, while inflating the variance in divergence across the genome, would not be predicted to generate correlations between B-values and divergence as well as the correlation between functional content and divergence. Thus, we can rule it out as the sole explanation for the empirical patterns seen in our study.

Further, mutagenic recombination is unlikely to explain the empirical patterns in our study because the correlation between divergence and functional content does not depend on recombination rate. The negative correlation between divergence and functional content remained strong when controlling for variation in recombination rates (S2 Table), suggesting our results are unlikely to be driven by mutagenic recombination. Nevertheless, our results do not rule out the possibility of mutagenic recombination and this topic certainly warrants further investigation.

Another possibility is that the reduction in neutral divergence near genes and in regions with lower B-values could be due to the direct effects of purifying selection removing variation from the population. Current evidence from a variety of comparative genomic studies suggests <10% of the genome is under purifying selection [39,41–49]. We attempted to mitigate the direct effects of purifying selection by employing a conservative set of filters in order to obtain putatively neutral sites. When removing the 10% of the genome that is most conserved, using a

variety of conservation metrics, the correlations persisted, suggesting they were not driven by the direct effects of selection. However, when we removed the top 15% of sites with the most conserved GERP score, the correlation between human-rodent divergence and functional content disappeared. This finding suggests that either the GERP scores themselves are affected by background selection, or, instead, that this correlation is driven, in part, by the direct effects of purifying selection. However, in order for direct purifying selection to explain the correlation, either more of the genome (at least 15%) would have to be under selection than suggested by current estimates [39,41–49] or many of the sites in the top 15% most conserved GERP scores would have to be neutrally evolving. Additionally, the negative correlation between human-chimp divergence and functional content, the positive correlation between human-chimp divergence and recombination rate, and the positive correlation between human-mouse divergence and *B*-values, remained even after removing the 25% of the genome that is most conserved (Fig 6). This implies that even such a large amount of functional sites under selection cannot explain all of our results. Finally, an additional line of evidence suggesting that the putatively neutral sites close to genes are not subjected to the direct effects of purifying selection stems from the fact that they show similar levels of neutral divergence to four-fold degenerate sties (S1 Fig, S1 Table). Thus, our putatively neutral noncoding sites have levels of divergence comparable to those seen for sites solely subjected to background selection.

Additionally, our filters rely on functional annotations and conservation to remove functionally important sites directly under the effects of selection. It is formally possible that the direct effects of selection could generate the correlations seen in our study if there are sites under selection that were invisible to the conservation-based filters used in our study. This could occur if there are recently derived, lineage-specific functional elements under selection that cannot be picked up by conservation metrics, or if there are sequences subject to purifying selection in the ancestral population but subsequently became neutral and therefore were not conserved. While we cannot exclude such a scenario, current population genetic evidence provides, at most, limited support for such an explanation [44,46,47,67].

One limitation in this study is that we made many assumptions regarding the parameters used in the simulations such as the ancestral population size, generation times, and mutation rates over the last 5–7 million years between human and chimp and 75 million years between human and mouse. There is much uncertainty surrounding all of these parameters [41,68–72]. Overall, we used a set of parameters in which the simulated divergence dataset from the coalescent simulations matched closely with the mean and standard deviation of the empirical divergence dataset. This allowed us to assess whether a simple neutral model could result in the correlations as large as observed empirically or whether a model with background selection needed to be invoked. We utilized the coalescent simulations as a proof of concept and therefore, the parameters we used in these sets of simulations should not be taken as estimates of the true values. Estimation of these parameters (ancestral population size, mutation rate, split time, etc.) is beyond the scope of this study and certainly warrants further in-depth investigation.

Other studies have argued that selection will not affect linked neutral divergence between distantly related species because the genealogy in the ancestral population only comprises a small proportion of the total genealogy between one chromosome from each of the two species [9,18,27]. This means that ancestral polymorphism will only account for a small proportion of the total divergence between distantly related species. It was thought that the signature of selection reducing the genealogy in the ancestral population would be diluted by the mutations that occurred since the split. As such, there would be no detectable signature of selection. Our theoretical results and simulations show the proportion of ancestral polymorphism actually is a poor predictor of the correlation between divergence and recombination as well as between

divergence and *B*-values. For example, consider a pair of species that split *N* generations ago with an ancestral population size of 25,000. In this model, 40% of the divergence is attributable to ancestral polymorphism (S4A Fig). Now consider a second pair of species that split 100*N* generations ago where N_a = 200,000. Here <5% of the divergence is due to ancestral polymorphism (S4D Fig). Previous intuition suggests the effect of background selection would be stronger in the first pair of species because they split more recently and ancestral polymorphism makes a greater contribution to divergence. However, our simulations show the exact opposite pattern (S4A and S4D Fig). The correlation between *B*-values and divergence is higher in the model with the more ancient split (Spearman's ρ = 0.610) than the one with the more recent split (Spearman's ρ = 0.452). Similar results are seen for the correlation between recombination rate and divergence. The reason for this discrepancy is that the main driver of these correlations is not the average amount of ancestral polymorphism, but rather the contribution to the variance in divergence due to the variance in ancestral polymorphism. Even when ancestral polymorphism makes only a small contribution to the overall average divergence, a substantial amount of the variance in total divergence across the genome can still be explained by variance in ancestral polymorphism, particularly if the ancestral population size is large. Our theoretical results suggest that the variance in the amount of background selection in different regions of the genome can account for a lot of the variance in total divergence, even for species that split long ago. In sum, our theoretical results and simulations suggest that previous intuition has understated the importance of even small amounts of ancestral polymorphism on the variability of genome-wide patterns of divergence between species.

Our results have important implications for understanding patterns of genetic variation and divergence across genomes. First, our findings add to the growing literature suggesting the importance of background selection at shaping genome-wide patterns of variability across species [1,7,13,15,16,60,62,73–80]. Our new contribution to this literature is demonstrating that natural selection can affect neutral divergence, even between distantly related species. Second, our work suggests that estimators of mutational properties that rely on contrasting patterns of divergence across different parts of the genome that may be differentially affected by background selection may yield biased results. This effect has been studied within primates in greater detail in recent work [81]. Third, the fact that we detect evidence of background selection between distantly related species suggests that there is still some information about the distribution of coalescent genealogies across the genome. This distribution of coalescent genealogies can be exploited to obtain more reliable estimates regarding the human-mouse ancestral population size. While several methods exist to estimate ancestral demographic parameters from divergence [82–86], we suggest that these methods may be applicable for very distantly related species. Our finding that background selection can increase the variance in coalescent times across the genome suggests these methods as well as other statistical methods which seek to infer demographic history from the distribution of coalescent times across the genome, such as the PSMC approach [87], should account for the increased variance in coalescent times across the genome due to background selection. Not accounting for background selection could result in inferring spurious demographic events to account for the additional variance in coalescent times across the genome as has recently been suggested for positive selection [88]. Lastly, our results suggest a need for caution when using patterns of divergence to calibrate neutral mutation rates. Some of the variation in divergence across the genome may be due to varying coalescent times, further accentuated by selection, rather than differing mutation rates [20,89]. Future work could explore the extent to which selection at linked neutral sites can explain the discrepancies between different types of estimates of mutation rates [33,34].

Methods

Data sets

We obtained the pairwise (.axt) alignments between human/chimpanzee (hg18/panTro2), human/orang (hg18/ponAbe2), human/mouse (hg18/mm9), human/rat (hg18/rn4), and human/zebrafish (hg18/danRer15) from the UCSC genome browser [90]. These alignments are the net of the best human chained alignments for each region of the genome [91]. For quality control, we excluded sites that (1) were missing in either of the species in the alignment, (2) were located within 10Mbp from the starting or ending position of a centromere, (3) were located within 10Mbp from the ending position of a telomere, (4) were located in repetitive elements.

We obtained the coordinate positions for the exons, RefSeq transcripts, and different phastCons measures calculated from different phylogenetic scopes [82] from the UCSC table browser [92] with the following specifications:

1. Exons: clade: Mammal, genome: Human, assembly: Mar. 2006 (NCBI36/hg18), group: Genes and Gene predictions, track: UCSC Genes, table: knownGene.

2. RefSeq transcripts: clade: Mammal, genome: Human, assembly: Mar. 2006 (NCBI36/hg18), group: Genes and Gene predictions, track: RefSeq Genes, table: refGene.

3. phastCons Vertebrates: clade: Mammal, genome: Human, assembly: Mar. 2006 (NCBI36/hg18), group: Comparative Genomics, track: Conservation, table: Vertebrate El (phastConsElements44way).

4. phastCons Primates: clade: Mammal, genome: Human, assembly: Mar. 2006 (NCBI36/hg18), group: Comparative Genomics, track: Conservation, table: Primate El (phastConsElements44wayPrimates).

5. phastCons Mammals: clade: Mammal, genome: Human, assembly: Mar. 2006 (NCBI36/hg18), group: Comparative Genomics, track: Conservation, table: Mammal El (phastConsElements44wayPlacental).

GERP scores were downloaded for hg18 from http://mendel.stanford.edu/SidowLab/downloads/gerp/. We used RS scores (range from -11.6 to 5.82) to obtain the conserved sites to remove. S8 Table summarizes the cutoffs we used.

For each window, we computed the recombination rate using the high resolution pedigree-based genetic map assembled by deCODE [50]. The B-value for each window was obtained from McVicker et al. [15]. Four-fold divergence was calculated by counting the number of between species differences that overlapped four-fold sites, divided by the total number of four-fold sites within each window. Functional annotation was done following Lohmueller et al. [13]. Briefly, we translated the Consensus Coding Sequence (CCDS) genes from the UCSC Genome Browser into proteins and determined which nucleotide changes did not alter the encoded amino acid. If transcripts overlapped, we retained the longest one.

Correlation analyses

To calculate the divergence between each pair of species, we divided the human genome into 100kb non-overlapping windows. For each window, we computed the total number of sites that passed the filtering criteria which resulted in the total number of neutral sites in each 100kb window. To reduce variation, we only considered windows in which the total number of eligible sites was greater than 10,000 (for analyses using 50kb as window size, we only considered windows in which the total number of eligible sites was greater than 5,000). Then we computed the divergence by tabulating the number of sites that are different between the two

species being compared. To account for multiple mutational hits for the distantly related species pairs (human-mouse and human-rat), we applied the Kimura two-parameter model [93].

To compute Spearman's ρ, we used the *cor* function in R. We used the *pcor* function to calculate partial correlation [94].

Controlling for confounding factors

To filter out possible hypermuteable CpG sites, we excluded sites that were preceded by a C or were followed by a G in hg18 [15]. To control for the effects of biased gene conversion, we removed all AT→GC substitutions across the genome.

Coalescent simulations

We modeled background selection as a simple reduction in effective population size in the ancestral population [4,5,7,15,58–62]. This was done by scaling the ancestral population size N_a, by the *B*-values. We used the *B*-values from McVicker et al. [15]. Each simulation replicate consisted of two parts. The first part modeled genetic variation in the ancestral population, and included the effects of background selection. For each window i, we simulated an ancestral recombination graph (ARG) with a population-scaled recombination rate $4N_aB_ir_i$, where N_a is the ancestral population size, B_i is the strength of background selection affecting window i, and r_i is the recombination rate for window i. Mutations were added to the genealogy assuming a population-scaled mutation rate $\theta = 4N_aB_i\mu_{a,i}L_i$, where $\mu_{a,i}$ is the ancestral per-base pair mutation rate for window i and L_i is the number of successfully aligned neutral bases in window i. Simulations were done using the program *ms* [63]. Note, we included recombination in the ancestral population because it affects the variance in coalescent times across windows and this variance in coalescent times will in turn affect the variance in levels of divergence, which will ultimately affect the strength of the correlation between divergence and recombination. Thus, we aimed to capture this variance as accurately as possible. This part of the simulation generated the amount of divergence due to ancestral polymorphism, which we call d_a.

We then added the mutations that arose since (i.e. more recently than) the split. The divergence from the present time to the split time follows a Poisson distribution, where the rate parameter equals the expected divergence between two populations. For each window of the genome, d_s was simulated using the *rpois* function in R. Finally, the total divergence within a window is the sum of divergence generated in the ancestral population (d_a) and the divergence generated since the two species split (d_s).

For human chimp divergence (Fig 4A, S9A Fig), $d_s = 2t_{split}\mu L$ where d_s is the expected divergence from the present time to the split time in the divergence model, t_{split} is the split time, μ is the mutation rate, and L is the length of each sequence. When computing both d_a and d_s for human-chimp divergence in Fig 4A, we drew μ from a gamma distribution with shape = 16.82 and scale 1.7×10^{-10} (S6 Table). In S9A Fig, we drew μ from a gamma distribution with shape = 15.68 and scale 1.8×10^{-10} (S6 Table). These parameters were chosen to match the observed mean and standard deviation of the distribution of human-chimp divergence (after removing all AT to GC differences as such changes could be due to biased gene conversion) as well as the observed correlation coefficient between divergence and recombination rate (S3A, S3B, S8A and S8B Figs). The split times and ancestral population sizes are roughly comparable to previous estimates from genetic data [84,85,95,96].

Due to the differences in generation times and mutation rates between the human and mouse lineages, we modified our approach for these simulations (Fig 4B, S9B and S10 Figs). First, here $d_s = (t_{mouse}\mu_{mouse} + t_{human}\mu_{human})L$, where t_{mouse} is the number of generations on the lineage leading to the mouse from t_{split} till the present day, t_{human} is the number of generations

on the lineage leading to human experienced from t_{split} till the present day, μ_{mouse} is the mutation rate along the mouse lineage, and μ_{human} is the mutation rate along the human lineage. There is much uncertainty surrounding these parameters. However, the following values are broadly consistent with what has been reported previously and match the observed mean and standard deviation of human-mouse divergence (S3C, S3D, S8C and S8D Figs, S7 Table). First, we assumed t_{split} = 75 million years ago. We then assumed mice have 1 generation per year, giving t_{mouse} = 75 x 10^6 generations. We assumed humans have 25 years per generation, making t_{human} = 3 x 10^6. We then set μ_{mouse} = 3.8 x 10^{-9} per generation and μ_{human} = 3.75 x 10^{-8} per generation (S10 Fig). These estimates are broadly consistent with previous reports and allow for approximately twice as much divergence on the mouse lineage as compared to the human lineage [41].

For the simulations in Fig 4B, we assumed that μ_a was equal to 2 x 10^{-8} per generation, which is the average of μ_{human} and μ_{mouse}. We accounted for variation in mutation rates across different regions of the genome by drawing μ_a from a gamma distribution [97]. We kept the ratio of μ_a to μ_{mouse} constant across all windows of the genome. For example, μ_a / μ_{mouse} = 5.26. Then if $\mu_{a,i}$ is the rate for the ith region drawn from the gamma distribution, we set $\mu_{mouse,i}$ equal to $\mu_{a,i}$ / 5.26. A similar procedure was used to find $\mu_{human,i}$. Note that for the simulations in S9B Fig, we used the average mutation rate of 2.7 X 10^{-8}, but we kept the ratio of μ_a to μ_{mouse} and the ratio of μ_a to μ_{mouse} to be the same as the simulations in Fig 4B. Increasing the variance in the mutation rate across regions increased the variance in divergence across windows of the genome and decreased the correlation between divergence and the B-values. We then examined different values of N_a and parameters of the gamma distribution that matched the observed mean and standard deviation of the distribution of human-mouse divergence as well as the observed correlation coefficient between divergence and B-values. The ancestral population size, shape, and scale parameters of the gamma distribution used for the simulations in Fig 4B and S9B Fig are reported in S6 Table. The simulated human-mouse divergence using these parameters matched closely with the empirical human-mouse divergence (S3C, S3D, S8C and S8D Figs, S7 Table).

Supporting Information

S1 Fig. Four-fold degenerate sites show similar levels of divergence as our putatively neutral noncoding sites. Each point represents the divergence within a 100kb window. (A) Human-chimpanzee, (B) Human-orangutan, (C) Human-mouse, and (D) Human-rat. (PDF)

S2 Fig. The correlation between human recombination rate and neutral divergence is stronger near genes. Correlation (Spearman's ρ) between neutral divergence and human recombination as a function of the amount of overlap with a RefSeq transcript. Black line denotes the correlations between human-chimpanzee neutral divergence and human recombination rate. Yellow line denotes the correlations between human-orangutan neutral divergence and human recombination rate. (PDF)

S3 Fig. Observed and modeled genome-wide distributions of human-chimp divergence and human-mouse divergence in 100 kb windows. Gray lines denote 500 simulated genome-wide distributions of divergence. Red line denotes the observed distribution of neutral divergence. Note, the distribution of simulated divergence is comparable to that from empirical data. (A) Simulated human-chimp divergence without the effects of background selection (BGS). (B) Simulated human-chimp divergence with the effects of background selection. (C) Simulated

human-mouse divergence without the effects of background selection. (D) Simulated human-mouse divergence with the effects of background selection. We filtered all AT→GC changes between the human and chimp sequences as they could be affected by biased gene conversion. Thus, the distribution of human-chimp divergence shown here is lower than the overall divergence.
(PDF)

S4 Fig. Background selection is predicted to affect neutral divergence across a range of split times and ancestral population sizes. Solid line shows the expected correlation coefficients (Spearman's ρ) between neutral divergence and recombination rate as a function of split time. Dashed line shows the expected Spearman's ρ between neutral divergence and McVicker's B-values as a function of split time. Red lines denote the proportion of the divergence due to polymorphism that arose in the ancestral population. Error bars denote ± one standard error of the mean. Panels A-D denote different ancestral population sizes (N_a). Note that the correlations are greater than 0 for a range of split times and ancestral population sizes, even when the proportion of divergence due to ancestral polymorphism is low.
(PDF)

S5 Fig. Variance of the total divergence attributable to the variance in levels of ancestral polymorphism. Black lines show the ratio of the variance of divergence in the ancestral population to the variance of the total divergence as a function of split time. Red lines denote the proportion of the divergence due to polymorphism that arose in the ancestral population. Panels A-D denote different ancestral population sizes (N_a).
(PDF)

S6 Fig. Correlations between human-primate divergence and genomic features persist when filtering the 25% of the genome with the highest GERP scores. (A) Neutral human-chimp divergence shows a negative correlation with functional content. (B) Neutral human-orang divergence shows a negative correlation with functional content. (C) Neutral human-chimp divergence shows a positive correlation with human recombination rate. (D) Neutral human-orang divergence shows a positive correlation with human recombination rate. Each point represents the mean divergence and functional content (A and B) or recombination rate (C and D) in 1% of the 100kb windows binned by functional content or recombination rate. Red lines indicate the loess curves fit to divergence and functional content (A and B) and divergence and recombination rate (C and D). Note that the last bin containing less than 1% of the windows was omitted from the plot. While the graph presents binned data, the correlations reported in the text are from the unbinned data.
(PDF)

S7 Fig. Correlations between human-rodent divergence and genomic features change when filtering the 25% of the genome with the highest GERP scores. (A) Neutral human-mouse divergence no longer correlates with functional content. (B) Neutral human-rat divergence does not correlate with functional content. (C) Neutral human-mouse divergence shows a positive correlation with McVicker's B-values. (D) Neutral human-rat divergence shows a positive correlation with McVicker's B-values. Each point represents the mean divergence and functional content (A and B) or B-values (C and D) in 1% of the 100kb windows binned by functional content or B-values. Red lines indicate the loess curves fit to divergence and functional content (A and B) and divergence and B-values (C and D). Note that the last bin containing less than 1% of the windows was omitted from the plot. While the graph presents binned data, the correlations reported in the text are from the unbinned data.
(PDF)

S8 Fig. Observed and modeled genome-wide distributions of human-chimp divergence and human-mouse divergence in 100kb windows when filtering sites whose GERP scores fall into the top 25% of the genome-wide distribution. Gray lines denote 500 simulated genome-wide distributions of divergence. Red line denotes the observed distribution of neutral divergence. Note, the distribution of simulated divergence is comparable to that from empirical data. (A) Simulated human-chimp divergence without the effects of background selection (BGS). (B) Simulated human-chimp divergence with the effects of background selection. (C) Simulated human-mouse divergence without the effects of background selection. (D) Simulated human-mouse divergence with the effects of background selection. We filtered all AT→GC changes between the human and chimp sequences as they could be affected by biased gene conversion. Thus, the distribution of human-chimp divergence shown here is lower than the overall divergence.
(PDF)

S9 Fig. Models incorporating background selection can recapitulate the empirical correlations after removing sites with GERP scores falling in the top 25% of the genome-wide distribution. (A) Models of background selection predict a positive correlation between neutral human-chimp divergence and human recombination rate. Because our model does not include biased gene conversion, the empirical correlation was calculated omitting AT to GC sequence differences. (B) Models of background selection predict a positive correlation between neutral human-mouse divergence and McVicker's B-values. The white histogram denotes 500 simulations without including background selection. The gray histogram denotes 500 simulations incorporating background selection. Red lines represent the correlations computed from the empirical data. Thus, plausible levels of background selection can match the observed correlations when using the most stringent filtering criteria while neutral simulations cannot.
(PDF)

S10 Fig. Human-mouse mutational parameters used for the simulations.
(PDF)

S1 Table. Summary statistics of divergence at four-fold sites and at putatively neutral regions.
(PDF)

S2 Table. Correlation coefficients of human-primate divergence and functional content.
(PDF)

S3 Table. Correlation coefficients of human-primate divergence and recombination rate.
(PDF)

S4 Table. Correlation coefficients of human-rodent divergence and functional content.
(PDF)

S5 Table. Correlation coefficients of human-rodent divergence and McVicker's B-values.
(PDF)

S6 Table. Summary of parameters used for the coalescent simulations.
(PDF)

S7 Table. Comparison of the mean and standard deviation of the empirical and simulated divergence.
(PDF)

S8 Table. GERP RS score cutoff.
(PDF)

Acknowledgments

Author Contributions

Conceived and designed the experiments: TNP CDH KEL.

Analyzed the data: TNP CDH.

Contributed reagents/materials/analysis tools: TNP CDH KEL.

Wrote the paper: TNP CDH KEL.

References

1. Cutter AD, Payseur BA. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet. 2013; 14: 262–274. doi: 10.1038/nrg3425 PMID: 23478346

2. Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res. 1974; 23: 23–35. doi: 10.1017/S0016672300014634 PMID: 4407212

3. Kaplan NL, Hudson RR, Langley CH. The "hitchhiking effect" revisited. Genetics. 1989; 123: 887–899. PMID: 2612899

4. Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993; 134: 1289–1303. PMID: 8375663

5. Hudson RR, Kaplan NL. Deleterious background selection with recombination. Genetics. 1995; 141: 1605–1617. PMID: 8601498

6. Nordborg M, Charlesworth B, Charlesworth D. The effect of recombination on background selection. Genet Res. 1996; 67: 159–174. PMID: 8801188

7. Charlesworth B. The effects of deleterious mutations on evolution at linked sites. Genetics. 2012; 190: 5–22. doi: 10.1534/genetics.111.134288 PMID: 22219506

8. Nachman MW. Single nucleotide polymorphisms and recombination rate in humans. Trends Genet. 2001; 17: 481–485. PMID: 11525814

9. Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M. A neutral explanation for the correlation of diversity with recombination rates in humans. Am J Hum Genet. 2003; 72: 1527–1535. doi: 10.1086/375657 PMID: 12740762

10. Hellmann I, Prufer K, Ji H, Zody MC, Paabo S, Ptak SE. Why do human diversity levels vary at a mega-base scale? Genome Res. 2005; 15: 1222–1231. doi: 10.1101/gr.3461105 PMID: 16140990

11. Hellmann I, Mang Y, Gu Z, Li P, de la Vega FM, Clark AG, et al. Population genetic analysis of shotgun assemblies of genomic sequences from multiple individuals. Genome Res. 2008; 18: 1020–1029. doi: 10.1101/gr.074187.107 PMID: 18411405

12. Cai JJ, Macpherson JM, Sella G, Petrov DA. Pervasive hitchhiking at coding and regulatory sites in humans. PLoS Genet. 2009; 5: e1000336. doi: 10.1371/journal.pgen.1000336 PMID: 19148272

13. Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, Vinckenbosch N, et al. Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. PLoS Genet. 2011; 7: e1002326. doi: 10.1371/journal.pgen.1002326 PMID: 22022285

14. Payseur BA, Nachman MW. Gene density and human nucleotide polymorphism. Mol Biol Evol. 2002; 19: 336–340. PMID: 11861892

15. McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of natural selection in homi-nid evolution. PLoS Genet. 2009; 5: e1000471. doi: 10.1371/journal.pgen.1000471 PMID: 19424416

16. Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, et al. Classic selective sweeps were rare in recent human evolution. Science. 2011; 331: 920–924. doi: 10.1126/science.1198878 PMID: 21330547

17. Enard D, Messer PW, Petrov DA. Genome-wide signals of positive selection in human evolution. Genome Res. 2014; 24: 885–895. doi: 10.1101/gr.164822.113 PMID: 24619126

18. Birky CW, Walsh JB. Effects of linkage on rates of molecular evolution. Proc Natl Acad Sci USA. 1988; 85: 6414–6418. PMID: 3413105

19. Cutler DJ. Clustered mutations have no effect on the overdispersed molecular clock: a response to Huai and Woodruff. Genetics. 1998; 149: 463–464. PMID: 9660672

20. Edwards SV, Beerli P. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evol Int J Org Evol. 2000; 54: 1839–1854.

21. Cruickshank TE, Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol. 2014; 23: 3133–3157. doi: 10.1111/mec.12796 PMID: 24845075

22. Reed FA, Akey JM, Aquadro CF. Fitting background-selection predictions to levels of nucleotide variation and divergence along the human autosomes. Genome Res. 2005; 15: 1211–1221. doi: 10.1101/gr.3413205 PMID: 16140989

23. Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. Nature. 1992; 356: 519–520. doi: 10.1038/356519a0 PMID: 1560824

24. McGaugh SE, Heil CS, Manzano-Winkler B, Loewe L, Goldstein S, Himmel TL, et al. Recombination modulates how selection affects linked sites in Drosophila. PLoS Biol. 2012; 10: e1001422. doi: 10.1371/journal.pbio.1001422; PMID: 23152720

25. Noor MA. Mutagenesis from meiotic recombination is not a primary driver of sequence divergence between Saccharomyces species. Mol Biol Evol. 2008; 25: 2439–2444. doi: 10.1093/molbev/msn186 PMID: 18723832

26. Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MA. Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. Proc Natl Acad Sci USA. 2008; 105: 10051–10056. doi: 10.1073/pnas.0801848105 PMID: 18621713

27. Begun DJ, Holloway AK, Stevens K, Hillier LW, Poh Y-P, Hahn MW, et al. Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. PLoS Biol. 2007; 5: e310. doi: 10.1371/journal.pbio.0050310 PMID: 17988176

28. Tyekucheva S, Makova KD, Karro JE, Hardison RC, Miller W, Chiaromonte F. Human-macaque comparisons illuminate variation in neutral substitution rates. Genome Biol. 2008; 9: 1–13. doi: 10.1186/gb-2008-9-4-r76

29. Lercher MJ, Hurst LD. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. 2002; 18: 337–340. PMID: 12127766

30. Pratto F, Brick K, Khil P, Smagulova F, Petukhova GV, Camerini-Otero RD. Recombination initiation maps of individual human genomes. Science. 2014; 346: 1256442. doi: 10.1126/science.1256442 PMID: 25395542

31. Arbeithuber B, Betancourt AJ, Ebner T, Tiemann-Boege I. Crossovers are associated with mutation and biased gene conversion at recombination hotspots. Proc Natl Acad Sci USA. 2015; 112: 2109–2114. doi: 10.1073/pnas.1416622112 PMID: 25646453

32. Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, et al. Genome-wide patterns and properties of de novo mutations in humans. Nat Genet. 2015; 47: 822–826. doi: 10.1038/ng.3292 PMID: 25985141

33. Scally A, Durbin R. Revising the human mutation rate: implications for understanding human evolution. Nat Rev Genet. 2012; 13: 745–753. doi: 10.1038/nrg3295 PMID: 22965354

34. Ségurel L, Wyman MJ, Przeworski M. Determinants of mutation rate variation in the human germline. Annu Rev Genomics Hum Genet. 2014; 15: 47–70. doi: 10.1146/annurev-genom-031714-125740 PMID: 25000986

35. Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the Drosophila genome? PLoS Genet. 2009; 5: e1000495. doi: 10.1371/journal.pgen.1000495 PMID: 19503600

36. Kimura M. The Neutral Theory of Molecular Evolution. Cambridge: Cambridge University Press; 1983.

37. Ohta T. Slightly deleterious mutant substitutions in evolution. Nature. 1973; 246: 96–98. PMID: 4585855

38. Akashi H, Osada N, Ohta T. Weak selection and protein evolution. Genetics. 2012; 192: 15–31. doi: 10.1534/genetics.112.140178; PMID: 22964835

39. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol. 2010; 6: e1001025. doi: 10.1371/journal.pcbi.1001025 PMID: 21152010

40. Moorjani P, Amorim CEG, Arndt PF, Przeworski M. Variation in the molecular clock of primates. bioRxiv. 2016; 036434. doi: 10.1101/036434

41. Mouse Genome Sequencing Consortium, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002; 420: 520–562. doi: 10.1038/nature01262 PMID: 12466850

42. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005; 15. doi: 10.1101/gr. 3715005

43. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 2010; 20: 110–121. doi: 10.1101/gr.097857.109 PMID: 19858363

44. Gulko B, Hubisz MJ, Gronau I, Siepel A. A method for calculating probabilities of fitness consequences for point mutations across the human genome. Nat Genet. 2015; 47: 276–283. doi: 10.1038/ng.3196 PMID: 25599402

45. Meader S, Ponting CP, Lunter G. Massive turnover of functional sequence in human and other mammalian genomes. Genome Res. 2010; 20: 1335–1343. doi: 10.1101/gr.108795.110 PMID: 20693480

46. Rands CM, Meader S, Ponting CP, 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. 2014; 10: e1004525. doi: 10.1371/journal.pgen.1004525 PMID: 25057982

47. Ward LD, Kellis M. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. Science. 2012; 337: 1675–1678. doi: 10.1126/science.1225057 PMID: 22956687

48. Cooper GM, Brudno M, Stone EA, Dubchak I, Batzoglou S, Sidow A. Characterization of evolutionary rates and constraints in three mammalian genomes. Genome Res. 2004; 14: 539–548. doi: 10.1101/gr. 2034704 PMID: 15059994

49. Schrider DR, Kern AD. Inferring selective constraint from population genomic data suggests recent regulatory turnover in the human brain. Genome Biol Evol. 2015; 7: 3511–3528. doi: 10.1093/gbe/evv228 PMID: 26590212

50. Kong A, Thorleifsson G, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, et al. Fine-scale recombination rate differences between sexes, populations and individuals. Nature. 2010; 467: 1099–1103. doi: 10.1038/nature09525 PMID: 20981099

51. Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, et al. A fine-scale chimpanzee genetic map from population sequencing. Science. 2012; 336: 193–198. doi: 10.1126/science.1216872 PMID: 22422862

52. Stevison LS, Woerner AE, Kidd JM, Kelley JL, Veeramah KR, McManus KF, et al. The time scale of recombination rate evolution in great apes. Mol Biol Evol. 2015; msv331. doi: 10.1093/molbev/msv331

53. Galtier N, Duret L. Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. Trends Genet. 2007; 23: 273–277. doi: 10.1016/j.tig.2007.03.011 PMID: 17418442

54. Duret L, Arndt PF. The impact of recombination on nucleotide substitutions in the human genome. PLoS Genet. 2008; 4: e1000071. doi: 10.1371/journal.pgen.1000071 PMID: 18464896

55. Duret L, Galtier N. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 2009; 10: 285–311. doi: 10.1146/annurev-genom-082908-150001 PMID: 19630562

56. Berglund J, Pollard KS, Webster MT. Hotspots of biased nucleotide substitutions in human genes. PLoS Biol. 2009; 7: e26. doi: 10.1371/journal.pbio.1000026 PMID: 19175294

57. Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, et al. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 2004; 14. doi: 10.1101/gr.1970304

58. Coop G, Ralph P. Patterns of neutral diversity under general models of selective sweeps. Genetics. 2012; 192: 205–224. doi: 10.1534/genetics.112.141861 PMID: 22714413

59. Corbett-Detig RB, Hartl DL, Sackton TB. Natural selection constrains neutral diversity across a wide range of species. PLoS Biol; 13(4): e1002112. doi: 10.1371/journal.pbio.1002112 PMID: 25859758

60. Charlesworth B. The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the Drosophila X chromosome. Genetics. 2012; 191: 233–246. doi: 10.1534/genetics.111.138073 PMID: 22377629

61. Charlesworth D, Charlesworth B, Morgan MT. The pattern of neutral molecular variation under the background selection model. Genetics. 1995; 141: 1619–1632. PMID: 8601499

62. Comeron JM. Background selection as baseline for nucleotide variation across the Drosophila genome. PLoS Genet. 2014; 10: e1004434. doi: 10.1371/journal.pgen.1004434 PMID: 24968283

63. Hudson RR. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics. 2002; 18: 337–338. doi: 10.1093/bioinformatics/18.2.337 PMID: 11847089

64. Wakely J. Coalescent Theory: An Introduction. W. H. Freeman; 2008.

65. Yang S, Wang L, Huang J, Zhang X, Yuan Y, Chen J-Q, et al. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. Nature. 2015; 523: 463–467. doi: 10.1038/nature14649 PMID: 26176923

66. Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, et al. Leveraging distant relatedness to quantify human mutation and gene-conversion rates. Am J Hum Genet. 2015; 97: 775–789. doi: 10.1016/j.ajhg.2015.10.006 PMID: 26581902

67. Palazzo AF, Gregory TR. The case for junk DNA. PLoS Genet. 2014; 10: e1004351. doi: 10.1371/journal.pgen.1004351 PMID: 24809441

68. Kumar S, Subramanian S. Mutation rates in mammalian genomes. Proc Natl Acad Sci USA. 2002; 99: 803–808. doi: 10.1073/pnas.022629899 PMID: 11792858

69. Smith NGC, Webster MT, Ellegren H. Deterministic mutation rate variation in the human genome. Genome Res. 2002; 12: 1350–1356. doi: 10.1101/gr.220502 PMID: 12213772

70. Hardison RC, Roskin KM, Yang S, Diekhans M, Kent WJ, Weber R, et al. Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. Genome Res. 2003; 13: 13–26. doi: 10.1101/gr.844103 PMID: 12529302

71. Hodgkinson A, Eyre-Walker A. Variation in the mutation rate across mammalian genomes. Nat Rev Genet. 2011; 12: 756–766. doi: 10.1038/nrg3098 PMID: 21969038

72. Geraldes A, Basset P, Smith KL, Nachman MW. Higher differentiation among subspecies of the house mouse (Mus musculus) in genomic regions with low recombination. Mol Ecol. 2011; 20: 4722–4736. doi: 10.1111/j.1365-294X.2011.05285.x PMID: 22004102

73. Wright SI, Andolfatto P. The impact of natural selection on the genome: emerging patterns in Drosophila and Arabidopsis. Annu Rev Ecol Evol Syst. 2008; 39: 193–213. doi: 10.1146/annurev.ecolsys.39.110707.173342

74. Cutter AD, Choi JY. Natural selection shapes nucleotide polymorphism across the genome of the nematode Caenorhabditis briggsae. Genome Res. 2010; 20: 1103–1111. doi: 10.1101/gr.104331.109 PMID: 20508143

75. Flowers JM, Molina J, Rubinstein S, Huang P, Schaal BA, Purugganan MD. Natural selection in gene-dense regions shapes the genomic pattern of polymorphism in wild and domesticated rice. Mol Biol Evol. 2012; 29: 675–687. doi: 10.1093/molbev/msr225 PMID: 21917724

76. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, et al. Comparative population genomics of maize domestication and improvement. Nat Genet. 2012; 44: 808–811. doi: 10.1038/ng.2309 PMID: 22660546

77. Halligan DL, Kousathanas A, Ness RW, Harr B, Eory L, Keane TM, et al. Contributions of protein-coding and regulatory change to adaptive molecular evolution in murid rodents. PLoS Genet. 2013; 9: e1003995. doi: 10.1371/journal.pgen.1003995 PMID: 24339797

78. Campos JL, Halligan DL, Haddrill PR, Charlesworth B. The relation between recombination rate and patterns of molecular evolution and variation in Drosophila melanogaster. Mol Biol Evol. 2014; 31: 1010–1028. doi: 10.1093/molbev/msu056 PMID: 24489114

79. Slotte T. The impact of linked selection on plant genomic variation. Brief Funct Genomics. 2014; 13: 268–275. doi: 10.1093/bfgp/elu009 PMID: 24759704

80. Wilson Sayres MA, Lohmueller KE, Nielsen R. Natural selection reduced diversity on human Y chromosomes. PLoS Genet. 2014; 10: e1004064. doi: 10.1371/journal.pgen.1004064 PMID: 24415951

81. Narang P, Wilson Sayres MA. Long-term natural selection affects patterns of neutral divergence on the X chromosome more than the autosomes. bioRxiv. 2015; 023234. doi: 10.1101/023234

82. Takahata N. An attempt to estimate the effective size of the ancestral species common to two extant species from which homologous genes are sequenced. Genet Res. 1986; 48: 187–190. doi: 10.1017/S001667230002499X PMID: 3569902

83. Rannala B, Yang Z. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics. 2003; 164: 1645–1656. PMID: 12930768

84. Wall JD. Estimating ancestral population sizes and divergence times. Genetics. 2003; 163: 395–404. PMID: 12586724

85. Siepel A. Phylogenomics of primates and their ancestral populations. Genome Res. 2009; 19: 1929–1941. doi: 10.1101/gr.084228.108 PMID: 19801602

86. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. Bayesian inference of ancient human demography from individual genome sequences. Nat Genet. 2011; 43: 1031–4. doi: 10.1038/ng.937 PMID: 21926973

87. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011; 475: 493–496. doi: 10.1038/nature10231 PMID: 21753753

88. Schrider D, Shanku AG, Kern AD. Effects of linked selective sweeps on demographic inference and model selection. bioRxiv. 2016; 047019. doi: 10.1101/047019

89. Gillespie JH, Langley CH. Are evolutionary rates really variable? J Mol Evol. 1979; 13: 27–34. PMID: 458870

90. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002; 12: 996–1006. doi: 10.1101/gr.229102 PMID: 12045153

91. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci USA. 2003; 100: 11484–11489. doi: 10.1073/pnas.1932072100 PMID: 14500911

92. Karolchik D, Baertsch R, Diekhans M, Furey TS, Hinrichs A, Lu YT, et al. The UCSC genome browser database. Nucleic Acids Res. 2003; 31. doi: 10.1093/nar/gkg129

93. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol. 1980; 16: 111–120. PMID: 7463489

94. Kim S. ppcor: An R package for a fast calculation to semi-partial correlation coefficients. Commun Stat Appl Methods. 2015; 22: 665–674. doi: 10.5351/CSAM.2015.22.6.665 PMID: 26688802

95. Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, et al. Great ape genetic diversity and population history. Nature. 2013; 499: 471–475. doi: 10.1038/nature12228 PMID: 23823723

96. Langergraber KE, Prüfer K, Rowney C, Boesch C, Crockford C, Fawcett K, et al. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. Proc Natl Acad Sci USA. 2012; 109: 15716–15721. doi: 10.1073/pnas.1211740109 PMID: 22891323

97. Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. Proc Natl Acad Sci U S A. 2005; 102: 18508–18513. doi: 10.1073/pnas.0507325102 PMID: 16352722