nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-019-0890-6

Exploiting selection at linked sites to infer the rate and strength of adaptation

Lawrence H. Uricchio[1,3]*, Dmitri A. Petrov[1] and David Enard[2]*

Genomic data encode past evolutionary events and have the potential to reveal the strength, rate and biological drivers of adaptation. However, joint estimation of adaptation rate (α) and adaptation strength remains challenging because evolutionary processes such as demography, linkage and non-neutral polymorphism can confound inference. Here, we exploit the influence of background selection to reduce the fixation rate of weakly beneficial alleles to jointly infer the strength and rate of adaptation. We develop a McDonald–Kreitman-based method to infer adaptation rate and strength, and estimate $\alpha = 0.135$ in human protein-coding sequences, 72% of which is contributed by weakly adaptive variants. We show that, in this adaptation regime, α is reduced ~25% by linkage genome-wide. Moreover, we show that virus-interacting proteins undergo adaptation that is both stronger and nearly twice as frequent as the genome average ($\alpha = 0.224$, 56% due to strongly beneficial alleles). Our results suggest that, while most adaptation in human proteins is weakly beneficial, adaptation to viruses is often strongly beneficial. Our method provides a robust framework for estimation of adaptation rate and strength across species.

The relative importance of selection and drift in driving species diversification has been a matter of debate since the origins of evolutionary biology. In the earliest formulations of evolutionary theory, natural selection was proposed to be the predominant driver of differences among species[1,2]. Subsequent theorists argued that random genetic drift could be a more important contributor to species differences[3–6], with random changes accumulating over evolutionary time due to reproductive isolation between populations. Although it is now clear that natural selection plays a substantial role in both diversification and constraint in many species[7–10], considerable uncertainty remains regarding the relative importance of stochastic drift, mutation, selection and linkage, with no clear consensus among evolutionary geneticists[11–14]. A better mechanistic understanding of these processes and how they jointly shape genetic diversity could help to resolve old evolutionary puzzles, such as the narrow range of observed genetic diversity across species[15] and the apparently low rate of adaptation in primates[16].

With the exception of rapidly evolving microbial species, most adaptation events occur too slowly for direct observation over the time-scale of a scientific study. Therefore, detailed study of the molecular basis of adaptation has required the development of computational methods to infer adaptation rates (denoted by α, defined as the proportion of fixed differences between species that confer fitness benefits) directly from genetic sequence data. Most existing approaches derive from the McDonald–Kreitman (MK) test[7,17] and related Poisson random field framework[18], both of which use divergence and polymorphism data to infer adaptation rates. Note that a recent approach uses polymorphism data alone to infer the distribution of fitness effects (DFE) of fixing mutations[19]. The critical idea behind each of these methods is to compare evidence for differentiation at alleles that are likely to have fitness effects (for example, non-synonymous alleles that change protein function by altering the amino acid sequence) to alleles that are less likely to have fitness effects (for example, synonymous alleles that do not change amino acid sequences).

In the classic MK framework, the rate of divergence at putatively functional sites (D_N, often defined as non-synonymous differences within proteins) is compared to putatively neutral diverged sites (D_S, often defined as synonymous differences). Polymorphic sites within both the functional and non-functional classes (P_N and P_S, respectively) are used as a background to calibrate the expected rate of divergence under a neutral model. If mutations at functional sites are assumed to be either virtually lethal or neutral, then the ratio $\frac{D_N}{D_S}$ has the same expected value as $\frac{P_N}{P_S}$ given that virtually lethal mutations contribute to neither P_N nor D_N. When $\frac{D_N}{D_S}$ exceeds $\frac{P_N}{P_S}$, this is interpreted as evidence of adaptation because sites with functional effects on proteins are over-represented among the fixed differences relative to the neutral expectation. A simple equation was developed that uses the same logic as the MK test to estimate adaptation rate α:

$$\alpha \approx 1 - \frac{D_S}{D_N}\frac{P_N}{P_S} \tag{1}$$

and this approach was used to provide evidence for a high rate of adaptation in *Drosophila*[17]. In principle, non-adaptive processes (that is, processes that do not increase fitness) such as guanine and cytosine (GC)-biased gene conversion[20] could also lead to an excess of non-synonymous fixed differences between species, but only if these processes differentially affect synonymous and non-synoymous mutations.

Unfortunately, this elegant framework is susceptible to many biases, most notably driven by the presence of weakly deleterious polymorphism in the class P_N. Deleterious polymorphism effectively makes the test overly conservative, because deleterious alleles are unlikely ever to reach fixation and therefore lead to overestimation of the expected background rate of substitutions in the functional class. The idea was proposed to include only common polymorphic alleles (for example, alleles at frequency 15% or greater), which should remove many deleterious alleles[21]; however, this approach has been shown to provide conservative adaptation rate estimates in many contexts[22]. More recently, it was shown that even removal of all polymorphism <50% is insufficient to correct this bias,

[1]Department of Biology, Stanford University, Stanford, CA, USA. [2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. [3]Present address: Department of Integrative Biology, University of California, Berkeley, CA, USA. *e-mail: uricchio@berkeley.edu; denard@email.arizona.edu

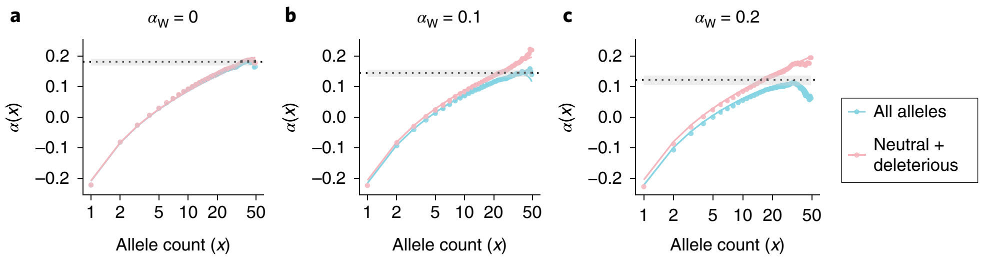

Fig. 1 | aMK estimates as a function of adaptation strength. a–c, We plot $\alpha(x)$ as a function of derived allele count (x) in a sample of 50 chromosomes. The true value of $\alpha = 0.2$ in each panel, with varying contributions from weakly ($2Ns = 10$) and strongly adaptive alleles ($2Ns = 500$), where N is the population size and s is the selection coefficient. The solid lines show the results of our analytical approximation (equation (11) in Supplementary Information), while the points show the value of $\alpha(x)$ from forward simulations. The blue points and curves show the calculation as applied to all polymorphic loci, while in the pink points and curves we have removed positively selected alleles from the calculation. The dotted line shows the estimated value of α from the simulated data using existing asymptotic-MK methods[24,47], while the grey bars show the 95% confidence interval around the estimate.

especially when slightly deleterious mutations are common and the rate of adaptive evolution is high[23]. To mitigate this effect, an asymptotic implementation of the MK test, called aMK, was introduced. In this implementation, $\frac{P_N}{P_S}$ in equation (1) is replaced by $\frac{P_N(x)}{P_S(x)}$, where $P_N(x)$ and $P_S(x)$ are the number of segregating non-synonymous and synonymous alleles at frequency x, respectively[23]. An exponential curve is fit to the resulting $\alpha(x)$ function, which can be calculated for all values of x in the interval (0,1) for a sample of sequenced chromosomes. The intercept of the best-fit exponential curve at $x = 1$ is a good approximation for α, as this effectively removes all slightly deleterious polymorphism at all frequencies. This approach was shown to be robust to both the underlying distribution of deleterious effects and recent demographic events[23]. The aMK test has inspired new approaches to inferring adaptation in mitochondrial genes[24] and revealed a high rate of adaptation in proteins interacting with pathogens[25].

While aMK extends the elegant MK framework for estimation of adaptation rate, it does not explicitly account for the possibility that beneficial alleles contribute to segregating polymorphism. It is unknown whether aMK is robust to the presence of weakly beneficial alleles, but there is reason to believe that beneficial alleles would be problematic because these are preferentially found at very high frequencies[19], and thus their effect would not be eliminated by the asymptotic procedure. The recent emphasis on adaptation from standing variation[26–30], and the reported evidence for weakly beneficial polymorphism in *Drosophila*[31], suggest that robust methods for inferring adaptation strength over longer evolutionary time-scales are needed.

A key limitation of existing MK-based approaches is that they provide estimates of adaptation rate but not adaptation strength, and therefore it is not clear whether weakly beneficial mutations contribute substantially to the fixation process. The underlying processes driving weak and strong adaptation might differ, and the ability to separately estimate rates of weak and strong adaptation could provide insight into the biological drivers of adaptation. We hypothesized that such a method could be developed by exploiting the impact of background selection (BGS) on the fixation rate of weakly beneficial alleles. BGS removes neutral and weakly beneficial variation via linkage to deleterious loci[32], while the fixation rate of strongly adaptive alleles is not substantially affected[33]. Given that the strength of BGS varies widely and predictably across the human genome[34], a method that interrogates the rate of adaptation as a function of BGS might be able jointly to infer the rate and strength of adaptation.

Here, we probe the performance of aMK when weakly beneficial alleles substantially contribute to segregating polymorphism, and we show that aMK underestimates α in this adaptation regime. We additionally show that when adaptation is weak, true α is predicted to vary substantially across the genome as a function of the strength of BGS. We exploit this signal of co-variation between α and BGS in the weak-adaptation regime to develop an approximate Bayesian computation (ABC) method, which we call ABC-MK, that separately infers the rate of adaptation for both weakly and strongly beneficial alleles. Both our approach and aMK rely on similar input data, but we use a model-based fitting procedure that directly accounts for BGS and weakly beneficial alleles. We apply our method to human genetic data to provide evidence that adaptation in humans is primarily weakly beneficial and varies as a function of BGS strength. Interestingly, adaptation rate estimates on virus-interacting proteins (VIPs) support a much higher rate of strong adaptation, suggesting that adaptation to viruses is both frequent and strongly fitness-increasing. We address seven potential sources of confounding, and discuss our results in light of recent research on adaptation in humans.

Results

Estimates of α are conservative for weakly beneficial selection. The aMK approach is known to converge to the true α at high frequency under the assumption that positively selected mutations make negligible contributions to the frequency spectrum[23]. This assumption is likely to be met when beneficial alleles confer large fitness benefits, because selective sweeps occur rapidly and beneficial alleles are rarely observed as polymorphic. However, when selection is predominantly weak, attaining a substantial α requires much higher mutation rates for beneficial alleles and longer average transit time to fixation, introducing the possibility that weakly beneficial alleles will contribute non-negligibly to the frequency spectrum.

To test whether aMK is sensitive to polymorphic weakly adaptive alleles, we used simulated polymorphism and divergence data to estimate the rate of adaptation using published aMK software[35]. In our simulations, we set the true value of α to 0.2 and varied the contribution of weakly and strongly beneficial alleles to the adaptation process (see Methods and Supplementary Information). When adaptation was due entirely to strongly adaptive alleles, the estimated value of α ($\hat{\alpha}$) was close to the true value but slightly conservative ($\hat{\alpha} = 0.181 \pm 0.01$; Fig. 1a). As we increased the contribution of weakly beneficial alleles (α_W) to α, estimates of α became increasingly conservative ($\hat{\alpha} = 0.144 \pm 0.01$ when $\alpha_W = 0.1$, and $\hat{\alpha} = 0.122 \pm 0.015$ when $\alpha_W = 0.2$; Fig. 1b,c). Removing polymorphism of frequency >0.5 has been suggested as an approach to account for potential biases induced by high-frequency-derived alleles, which could be mis-polarized in real datasets[25]. Restriction to alleles of frequency <0.5 produced similar (but conservative) estimates for all three models ($\hat{\alpha} = 0.14271, 0.14529$ and 0.14264 for $\alpha_W = 0.0$, 0.1 and 0.2, respectively), probably because the frequency spectrum is not strongly dependent on the rate of weakly beneficial

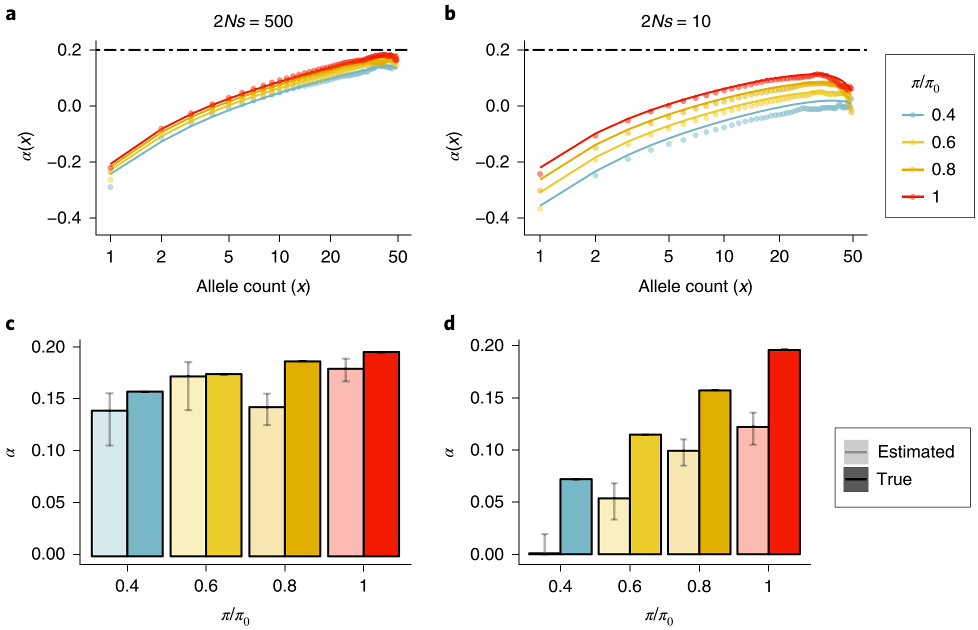

Fig. 2 | The effect of BGS on α. a,**b**, $\alpha(x)$ is plotted as a function of derived allele count (x) for various background selection (π/π_0) values. Adaptive alleles are strongly beneficial ($2Ns = 500$) (**a**) or are weakly beneficial ($2Ns = 10$) (**b**). The lines represent analytical approximations, while the points represent the results of stochastic simulations. The dashed lines at $\alpha = 0.2$ represent the true rate of adaptation in the absence of BGS. **c**,**d**, True (dark colours) and estimated (light colours) α for each of the corresponding models in **a**,**b**, which corresponds to strong adaptation ($2Ns = 500$) (**c**) or weak adaptation ($2Ns = 10$) (**d**). Estimates of α were made using existing asymptotic-MK software[24], and the error bars correspond to 95% confidence intervals reported by the software. For each parameter combination, we used 2×10^5 independent simulations of 10^3 coding base pairs each.

mutation for low-frequency alleles. Lastly, we performed a much larger parameter sweep across α values and selection coefficients. We found that α estimates became increasingly conservative as the proportion of weakly deleterious alleles increased, and as the strength of selection at beneficial alleles decreased (Supplementary Fig. 12a and Supplementary Information). Asymptotic-MK estimates of α are only weakly dependent on the distribution of deleterious selection coefficients (Supplementary Fig. 12).

To better understand why parameter estimates decreased as the proportion of weakly adaptive alleles increased, we performed analytical calculations of $\alpha(x)$ using diffusion theory[36,37]. Since we use large sample sizes in our analysis herein, we replace the terms $P_N(x)$ and $P_S(x)$ in $\alpha(x)$ with $\sum_x P_N(x)$ and $\sum_x P_S(x)$ in our calculations, which trivially asymptote to the same values as the original formulation but are not strongly affected by sample size (see Supplementary Information). We find that the downward bias in estimates of α is due to segregation of weakly adaptive alleles, and removal of these alleles from the simulated and calculated $\alpha(x)$ curves restored the convergence of $\alpha(x)$ to the true α at high frequency (Fig. 1a–c, red curves). In real data, it is not possible to perfectly partition positively selected and deleterious polymorphic alleles. Hence, in later sections we focus on using the shape of the $\alpha(x)$ curve to infer the strength and rate of adaptation under models that include linkage and complex demography.

Background selection reduces true α when adaptation is weak. We have shown that weakly beneficial alleles may impact aMK analyses by contributing to segregation of polymorphism. This presents an opportunity to study whether aMK estimates vary as a function of BGS strength. BGS, the action of linkage between deleterious alleles and neutral alleles, reduces genetic diversity in the human genome[34] and affects neutral divergence rates[38], and is predicted to decrease the fixation probability of weakly adaptive alleles[33]. Hence, we hypothesized that if adaptation is partially driven by weakly beneficial alleles in some species, BGS could play a role in modulation of adaptation rate across the genome.

To better understand how BGS might affect aMK inference in the presence of weakly beneficial alleles, we performed analytical calculations and simulations of $\alpha(x)$ with various levels of BGS. We set $\alpha = 0.2$ in the absence of BGS, and then performed simulations while fixing the rate of adaptive mutations and changing the level of BGS (ranging from $\frac{\pi}{\pi_0} = 0.4$ to 1.0, where π is neutral nucleotide diversity as compared to the neutral diversity in the absence of linked selection, π_0). We find that when adaptation is strong, BGS has a modest effect on $\alpha(x)$ and the true value of α (Fig. 2a,c), mostly driven by an increase in the rate of fixation of deleterious alleles (Supplementary Fig. 2e). When adaptation is weak, BGS removes a substantial portion of weakly adaptive alleles and precludes these from fixing, resulting in much stronger dependence of $\alpha(x)$ on BGS and a substantial reduction in the true value of α (Fig. 2b,d and Supplementary Fig. 2c). Similar to the previous section, estimates of α were conservative across all models but the underestimation was much more pronounced for weak adaptation (Fig. 2c,d).

Human adaptation rate is shaped by linked selection. Our modelling results show that α is likely to be underestimated when weakly beneficial alleles contribute substantially to the frequency spectrum, and that background selection may reduce adaptation rate when fitness benefits of adaptive alleles are small. Since BGS is thought to drive broad-scale patterns of diversity across the human genome[34], we hypothesized that directly accounting for the action of BGS on adaptation rate could provide new insights into the evolutionary mechanisms driving adaptation. Moreover, the fact that weak adaptation is strongly affected by BGS, while strong adaptation is not, suggests that strong and weak adaptation could be differentiated in genomic data by comparing regions of differing BGS strength (from $\frac{\pi}{\pi_0} = 0.2$ to $\frac{\pi}{\pi_0} = 1$). We therefore designed an ABC-based method to infer α while accounting for both BGS and weakly beneficial alleles.

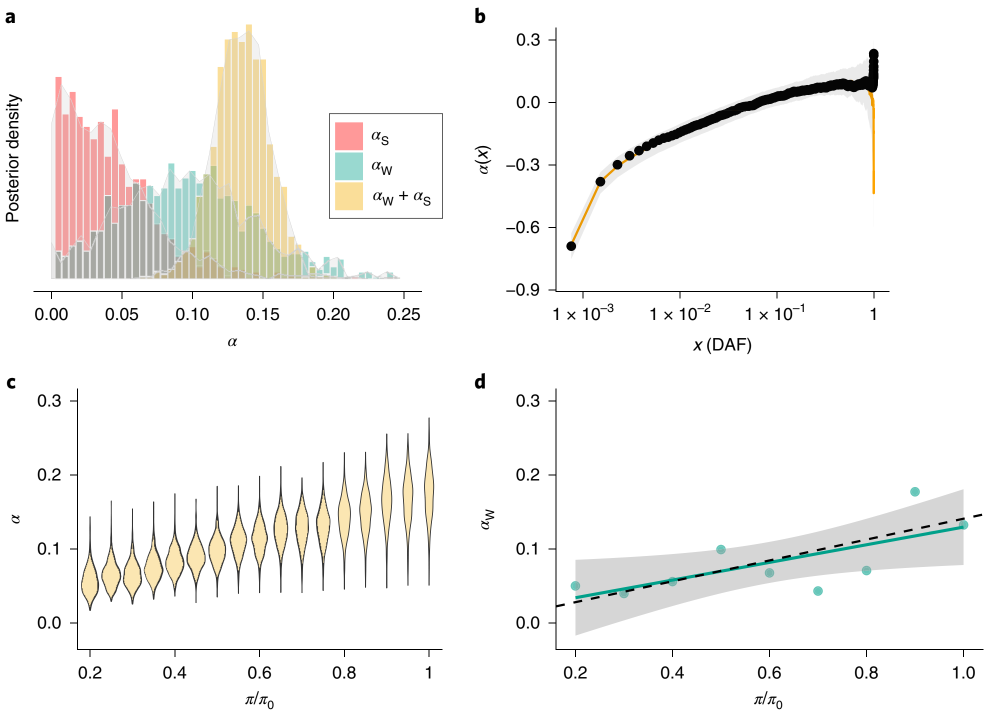

Fig. 3 | Adaptation rate and strength estimates for human genomic data. a, Posterior distribution of α_W, α_S and $\alpha = \alpha_S + \alpha_W$ as inferred by application of our ABC approach to 661 samples of African ancestry from TGP phase3. **b**, $\alpha(x)$ as a function of derived allele frequency (DAF) for genomic data (black points) plotted along with the mean posterior estimate from our model (orange line) and 99% confidence interval (grey envelope), as obtained by an independent set of simulations using posterior parameter estimates. **c**, Inferred posterior distribution of α as a function of BGS strength in the human genome. **d**, Mean posterior estimates of α_W, as determined by separate fitting of the model to alleles from each independent background selection strength bin. A linear model fit to the data (green line) supported statistically significant co-variation between π/π_0 and α_W ($P = 0.0343$). The black dashed line shows the predicted change in α_W as a function of π/π_0 given the mean estimate of α_W.

We applied our inference procedure (ABC-MK) to empirical $\alpha(x)$ data computed from human genomes obtained from the Thousand Genomes Project (TGP) for all 661 samples with African ancestry[39]. We find strong posterior support for a substantial component of α driven by weakly beneficial alleles ($\hat{\alpha}_W = 0.097$; see Fig. 3a and see Table 1 for area of 95% highest posterior density), as well as posterior support for a smaller component of α from strongly beneficial alleles (α_S) ($\hat{\alpha}_S = 0.041$). We estimate that total $\hat{\alpha} = 0.135$, nearly twice the estimate obtained with the same dataset using the original aMK approach ($\hat{\alpha} = 0.076$, see Supplementary Information; we note that while our estimate is similar to previous estimates[23,40], we used a much larger set of genes in our inference and hence the estimates are not directly comparable). In addition to rates of positive selection, our approach provides estimates of negative selection strength. We find support for mean strength of negative selection of $2Ns \approx -220$ (Supplementary Fig. 9C), which is consistent with recent studies using large sample sizes[41] but weaker than earlier estimates using small samples[40,42].

In addition to estimation of evolutionary parameters, we sought to better understand how BGS might impact adaptation rates across the genome. We resampled parameter values from our posterior estimates of each parameter, and ran a new set of forward simulations using these parameter values. We then calculated α as a function of BGS in our simulations. We find that α co-varies strongly with BGS, with α in the lowest BGS bins being 33% of α in the highest bins (Fig. 3c). Integrating across the whole genome, our results suggest that human adaptation rate in coding regions is reduced by approximately 25% by BGS (Supplementary Fig. 9d). To confirm that these model projections are supported by the underlying data, we split the genome into BGS bins and separately estimated adaptation rate in each bin. Although these estimates are substantially noisier than our inference on the full dataset, we find that the rate of adaptation due to weakly beneficial alleles decreases as a function of BGS strength in accordance with the model predictions (Fig. 3d). In contrast, estimates of the mean strength of negative selection against non-synoymous mutations did not co-vary with BGS strength (Supplementary Fig. 20). Lastly, to validate that our model recapitulates $\alpha(x)$ values that we observe in real data, we also used our independent forward simulations to re-compute $\alpha(x)$. We find that our model is in close agreement with observed data across the majority of the frequency spectrum. The model and data deviate at high frequency, but both are within sampling uncertainty (Fig. 3b, grey envelope).

Previous research has shown that VIPs have undergone faster rates of adaptation than the genome background[25]. However, the strength of selection acting on these genes is unknown and, given our BGS results, it is plausible that the higher rate of adaptation in VIPs is driven by lower overall background selection at VIPs rather than increased selection pressure for adaptation. In contrast, if pathogens have imposed large fitness costs on humans it is possible that VIPs would support both higher adaptation rates and greater adaptation strength. We ran our method while restricting it to an expanded set of 4,066 VIPs for which the divergence and polymorphism data were available. We found evidence for strikingly higher adaptation rates in VIPs than the genome background ($\alpha = 0.224$) and a much larger contribution from strongly adaptive alleles ($\alpha_S = 0.126$; Fig. 4). The higher value of α for VIPs cannot be explained by BGS, because VIPs undergo slightly stronger BGS than average genes; the mean BGS strength at VIPs is 0.574 as compared to 0.629 for all genes (in units of π/π_0). Taking $\alpha_S = 0.126$ as a point estimate for the rate of strongly beneficial substitutions in VIPs and $\alpha_S = 0.041$ genome-wide, we estimate that 61% of all strongly

Table 1 | Datasets and corresponding adaptation rates

Dataset	NS	SYN	$\hat{\alpha}$	$\hat{\alpha}_W$	$\hat{\alpha}_S$
Whole-genome	29,925	38,135	0.135 (0.096,0.17)	0.097 (0.0,0.21)	0.041 (0.0,0.13)
VIPs	6,249	10,309	0.224 (0.17,0.28)	0.098 (0.0,0.24)	0.126 (0.018,0.26)
Non-VIPs	23,676	27,826	0.12 (0.09,0.15)	0.077 (0.01,0.13)	0.042 (0.0,0.09)

Estimated α values represent the mean of posterior distribution. NS and SYN represent the number of non-synonymous and synonymous fixations, respectively. Values in parentheses represent the area of 95% highest posterior density.

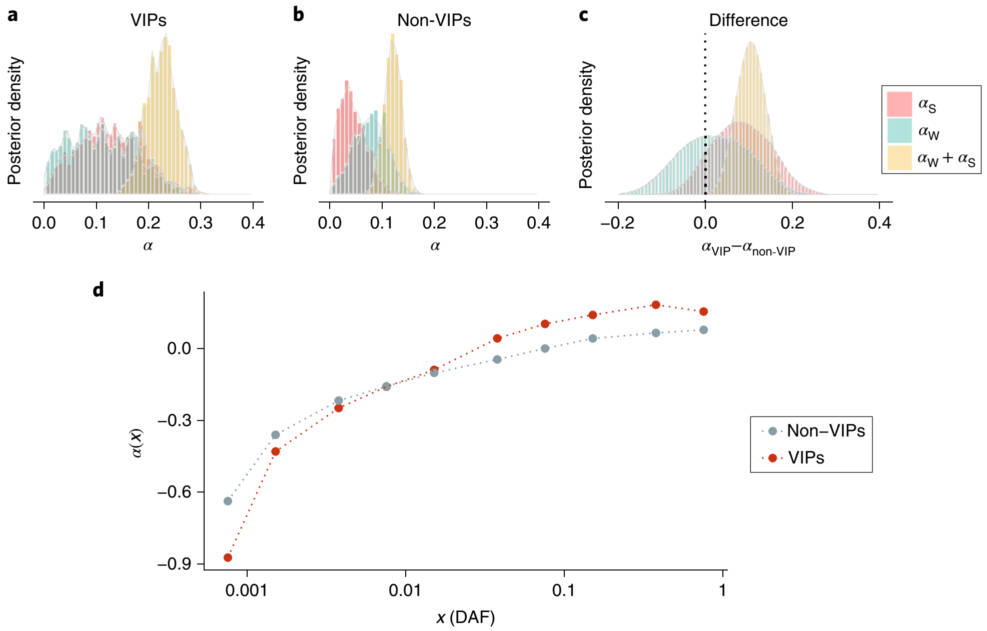

Fig. 4 | Virally interacting genes support a high rate and strength of adaptation. **a**, Posterior distributions for α, α_W and α_S for VIPs (4,066 genes). **b**, The same quantities for non-VIPs (12,962 genes). **c**, Posterior distribution of the difference in α between VIPs and non-VIPs. **d**, $\alpha(x)$ as a function of derived allele frequency x for VIPs and non-VIPs, specifically at the values of x that we use for statistical inference.

beneficial substitutions occurred in VIPs (Table 1). Moreover, we estimate that the posterior probability that α is greater in VIPs than non-VIPs is 99.97%, while the posterior probability that α_S is greater in VIPs is 88.9% (Fig. 4c). Bootstrap samples of non-VIPs (1,000 replicates) resulted in no α_S estimates as high as those obtained from VIPs (Supplementary Fig. 19). These results are concordant with the $\alpha(x)$ summary statistics for VIPs, which had larger values at high-frequency alleles than non-VIPs (Fig. 4d). Interestingly, $\alpha(x)$ is lower for VIPs than non-VIPs at low frequency, suggesting increased overall levels of conservation among VIPs (see also Supplementary Fig. 9, where we find support for stronger negative selection against non-synonymous mutations in VIPs).

Discussion

A long-running debate in evolutionary biology has concerned the relative importance of drift and selection in determining the rate of diversification among species[3,4,6,13]. While previous studies have shown that there is a substantial signal of adaptation in *Drosophila*[17], estimates of adaptation rate in humans are much lower[13]. Here, we extended the classic MK framework to account for weakly beneficial alleles, and we provide evidence for a high rate of weakly adaptive mutation in humans. We show that a state-of-the-art approach to adaptation rate estimation that does not account for beneficial polymorphism provides conservative estimates of α ($\hat{\alpha}=0.076$ for these data)[23], while our method nearly doubles the estimated human adaptation rate (to $\hat{\alpha}=0.135$). Most of the adaptation signal that we detected was due to weakly beneficial alleles. Interestingly, VIPs supported a much higher rate of adaptation than the genome background ($\hat{\alpha}=0.226$), especially for strongly beneficial substitutions ($\hat{\alpha}_S=0.126$ compared to $\hat{\alpha}_S=0.041$ genome-wide). Our results provide an evolutionary mechanism that partially explains the apparently low observed rate of human adaptation in previous studies, and extends support for viruses as a major driver of adaptation in humans[25].

It has long been known that recombination could, in principle, affect the evolutionary trajectories of both beneficial and deleterious alleles[33,43,44], and studies in *Drosophila*[45,46] and dogs[47] have provided evidence for the effect of recombination on divergence and load. Despite the expectation that recombination could have a strong effect on adaptation in humans, studies have differed on how recombination affects human divergence and polymorphism. One human genomic study explored the ratio $\frac{D_N}{D_S}$ as a function of recombination rate, and found no evidence for an effect of recombination on divergence rate[48]. Our results may partially explain why $\frac{D_N}{D_S}$ does not fully capture the effect of recombination on divergence in humans. As BGS increases in strength, the rate of accumulation of deleterious alleles increases while the rate of fixation of weakly adaptive alleles decreases. These two effects partially offset each other, which should reduce the sensitivity of $\frac{D_N}{D_S}$ as a tool in detecting the effect of recombination on divergence. A more recent study provided evidence that recombination affects the accumulation of deleterious polymorphic alleles[49], but did not provide detailed information about the effect of recombination on adaptation. Our results are consistent with the idea that weakly deleterious alleles are predicted to segregate at higher frequencies in regions under strong BGS, and we additionally show that BGS affects the accumulation of weakly beneficial alleles in humans.

While classic MK approaches estimate only the rate of adaptation, our method extends the MK framework to provide information on both the rate and strength of selection. While previous approaches used to estimate the strength of adaptation have either focused on the dip in diversity near sweeping alleles[31,45,50–52] or directly inferred the DFE from the frequency spectrum[19], our approach capitalizes on an orthogonal signal of the reduction in fixation rate of weakly beneficial alleles induced by selection at linked sites. We developed an ABC method to capture this signal, but less computationally intensive methods could also be used—for example, the original aMK approach could be applied in bins of BGS strength. If a substantial proportion of adaptation is due to weakly beneficial alleles, such an analysis should result in a strong correlation between BGS strength and (potentially conservative) α estimates. However, it should be noted that cryptic co-variation between gene functions (such as VIPs) and BGS strength could confound such inferences.

We supposed that the main effects of linked selection in humans are due to background selection, but in principle genetic draft could drive similar patterns. Draft is expected to substantially reduce genetic diversity when sweeps occur frequently, and can impede the fixation of linked beneficial alleles[53,54]. Previous work has also shown that strong draft can alter the fixation rate and frequency spectra of neutral and deleterious alleles[23]. We performed simulations of strong draft in 1-MB flanking sequences surrounding a gene evolving under natural selection, and tested the magnitude of the deviation from theoretical predictions under a model of background selection alone. Consistent with previous work, we observed that draft increases the fixation rate of deleterious alleles and thereby decreases α (ref. [23]). However, the effect on $\alpha(x)$ is only modest at the frequencies that we used in our inference procedure (that is, <75%), even when the strength and rate of positive selection were much higher than we and others have inferred in humans (although there is a modest deviation around 75% frequency, the highest frequency we used in our inference; Supplementary Fig. 4c,d). This implies that draft due to selected sites outside genes would have to be much stronger than that due to positive selection inside exons to drive the effects that we infer in the human genome. We note that it is likely that in species undergoing both strong, frequent sweeps and BGS (for example, *Drosophila*—see ref. [31]), draft will contribute to the removal of weakly beneficial polymorphism.

Selection has left many imprints on the human genome, with studies reporting signatures of selective sweeps[52], soft sweeps[29], background selection[34], negative selection[40,42] and polygenic adaptation[28]. Nevertheless, considerable uncertainty remains about the relative importance of these evolutionary mechanisms, especially in regard to the rate and strength of positive selection. Recent work has suggested that the contrasting adaptation rate estimates of previous studies[51,52] can be reconciled by arguing that most adaptation signals in humans are consistent with adaptation from standing variation[29]. Our results show that the frequency spectra and patterns of divergence are also consistent with the idea that many adaptive alleles segregate for much longer than is expected for a classic sweep, and hence also help to reconcile the results of previous studies.

In addition to determining the rate, strength and mechanisms of adaptation, there is an ongoing effort to find the biological processes most important for driving adaptation. Previous work has shown that viruses are a critical driver of adaptation in mammals[25], but the strength of the fitness advantages associated with resistance to (or tolerance of) infection remains unclear. Our approach clarifies that strongly adaptive fixed differences are also enriched, approximately threefold, in VIPs relative to non-VIPs. In contrast, weak adaptation rate was not substantially different between VIPs and non-VIPs, suggesting that weak adaptation may proceed through mechanisms that are shared across proteins regardless of function (for example, optimization of stability). While we have focused on VIPs here due to the expected fitness burdens associated with infection, in future research our approach could be used to investigate adaptation in any group of genes, or extended to partitioning of genes into strong and weak adaptation classes.

The model that we fit to human data does an excellent job of recapitulating the observed patterns in the TGP data, but we were concerned that several possible confounding factors could have influenced our results. We showed that seven confounding factors (ancestral mis-polarization[55], demographic model mis-specification[56,57], BGS model mis-specification, co-variation of BGS and sequence conservation, GC-biased gene conversion[20], selection on synonymous alleles[58] and mis-specification of strongly and/or weakly beneficial selection coefficients) are unlikely to have substantially influenced the results (see Supplementary Information), but it should be noted that the adaptive process in our model is exceedingly simple and it is very likely that the evolutionary processes driving diversification are much more complex. We supposed that adaptation proceeds in two categories, weak and strong selection, each of which is described by a single selection coefficient. In reality, adaptive alleles are likely to have selection coefficients drawn from a broad distribution, and adaptation is likely to proceed by a variety of mechanisms, including sweeps[52], polygenic adaptation[28] and selection from standing variation[29]. While our results show that BGS shapes adaptation rate across the genome, our method does not differentiate among adaptation mechanisms. We expect that future research will further clarify the relative importance of various selection mechanisms in shaping genomic patterns of diversity in the genomes of humans and other organisms[10,59].

Our method is flexible in that it could be applied to any species for which both divergence/polymorphism data and estimates of background selection strength are available. As with the original aMK approach, we showed that the α estimates we obtained are not highly sensitive to recent demographic uncertainty. Our approach may therefore be effective in providing more accurate estimates of adaptation rate in non-model species. Despite recent advances, the evolutionary mechanisms that shape genetic diversity across species (which could include linked selection, population size and/or population demography) remain the subject of debate[11,12,15]. Future work using and extending our method, which accounts for the effect of weakly beneficial alleles on adaptation rate estimates, could help to resolve this open question.

Methods

Divergence and polymorphism data. We retrieved the number of polymorphic sites and their allele frequencies in human coding sequences, as well as the number of human-specific fixed substitutions in coding sequences since divergence with chimpanzees. Fixed substitutions were identified by parsimony based on alignments of human (hg19 assembly), chimpanzee (panTro4 assembly) and orangutan (ponAbe2 assembly) coding sequences. Human coding sequences from Ensembl v.73 (ref. [60]) were blatted[61] on the panTro4 and ponAbe2 assemblies and the best corresponding hits were blatted back on the hg19 human assembly to finally identify human–chimp–orangutan best reciprocal orthologous hits. We used the Blat-fine option to ensure that even short exons at the edge of coding sequences would be included in the hits. We further used a Blat protein -minIdentity threshold of 60%. The corresponding human, chimpanzee and orangutan coding sequences were then aligned with the PRANKs coding sequence evolution model[62] after removal of codons containing undefined positions.

For each human coding gene in Ensembl we considered all possible protein-coding isoforms and aligned each isoform individually among human, chimpanzee and orangutan. The numbers of polymorphic or divergent sites are therefore the numbers over all possible isoforms of a human gene (however, the same polymorphic or divergent site present in multiple isoforms was counted only once). If a polymorphic or divergent site was synonymous in an isoform but non-synonymous in another isoform, we counted that as a single non-synonymous polymorphic or divergent site. Only fixed divergent sites were included, meaning that substitutions still polymorphic in humans were not counted as divergent.

The derived allele frequency of polymorphic sites herein corresponds to the frequency across all African populations from TGP phase 3, which comprises 661 individuals spread across seven different subpopulations[39]. Allele frequencies were extracted from vcf files provided by the TGP for the phase 3 data. In total, 17,740 human–chimpanzee–orangutan orthologues were included in the analysis. Supplementary Data Table 1 provides the number of synonymous and non-synonymous polymorphic or divergent sites for each of these 17,740 orthologues, as well as the allelic frequencies of the polymorphic sites. Polymorphic sites were counted only if they overlapped those parts of human coding sequences that were aligned with chimpanzee and orangutan coding sequences. The ancestral and derived allele frequencies were based on the ancestral alleles inferred by TGP phase 3 and available in the previously mentioned vcf files[39].

Model-based simulations and calculations. We tested the robustness of the aMK approach to the presence of weakly beneficial alleles using simulation and theory. We simulated simultaneous negative and positive selection in coding sequences using model-based forward simulations under a range of scenarios[63,64]. We supposed that non-synonymous alleles are under selection while synonymous alleles are neutral. In each simulation, we set $\alpha = \alpha_W + \alpha_S = 0.2$, where α_W is the component of α due to weakly beneficial mutations ($2Ns = 10$) and α_S represents strongly beneficial alleles ($2Ns = 500$). Note that α is not treated as a parameter in the analyses herein; we use analytical theory to calculate the mutation rates for deleterious alleles and advantageous alleles that result in the desired α, meaning that α is a model output and not a model input. We drew deleterious selection coefficients from a gamma distribution inferred from human sequence data[40], and we varied α_W from 0 to 0.2. We used the simulated allele frequency spectra and fixed differences to calculate the $\alpha(x)$ summary statistics. The results of these simulations are provided in Fig. 1, and additional simulation details are included in the Supplementary Information.

We also performed analytical calculations under the same evolutionary model using results from diffusion theory. These calculations are described in the Supplementary Information (see the sections entitled 'Analytical approximation to $\alpha(x)$' and 'Background selection and adaptive divergence'). Software to perform these calculations is available at https://github.com/uricchio/mktest.

Using ABC-MK to infer adaptation rate and strength. We developed an ABC approach for estimation of α_W and α_S in the presence of BGS and complex human demography[65]. We sampled parameters from previous distributions corresponding to the shape and scale of deleterious selection coefficients (assumed to be gamma distributed) and the rate of mutation of weakly and strongly beneficial mutations. We performed forward simulations[63,64] of simultaneous negative and positive selection at a coding locus under a demographic model inferred from NHLBI Exome project African American samples[66], with varying levels of background selection from $\pi/\pi_0 = 0.2$ to $\pi/\pi_0 = 1.0$ and the sampled parameter values. We then calculated $\alpha(x)$ using these simulated data, sampling alleles from the simulations such that the distribution of BGS values in the simulation matches that in the empirical data as calculated by a previous study[34]. We used $\alpha(x)$ values at a subset of frequencies x as summary statistics in ABC (specifically, at derived allele counts 1, 2, 5, 10, 20, 50, 100, 200, 500 and 1,000 in a sample of 1,322 chromosomes). To improve efficiency, we employed a resampling-based approach that allows us to query many parameter values using the same set of forward simulations (see Supplementary Information).

We tested our approach by estimating parameter values (population-scaled mutation rates θ_S, θ_W and the parameters of a gamma distribution controlling negative selection strength) and quantities of interest (α_W, α_S, α) from simulated data. We found that the method produces highly accurate estimates for most inferred parameters and α values (including α_W, α_S and total α; Supplementary Fig. 6). Some parameter values (particularly those corresponding to the DFE) over deleterious alleles and mutation rates of beneficial alleles) were somewhat noisily inferred. We found that estimations of α were not very sensitive to various types of model mis-specification (See Supplementary Information, 'Robustness analyses'), but α_W and α_S were modestly affected by mis-specification of the demographic model or the DFE of alleles driving BGS. We term our approach ABC-MK.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Supplemental Data Table 1 is provided on the publisher's website. The data that we used to parameterize our model are also available online at https://github.com/uricchio/mktest. Columns in Supplementary Data Table 1 are as follows: 1, Ensembl coding gene identification; 2, number of non-synonymous polymorphic sites; 3, respective derived allele frequencies of these sites separated by commas; 4, number of synonymous polymorphic sites; 5, respective frequency-derived allele frequencies of these sites; 6, number of fixed non-synonymous substitutions on the human branch; and 7, number of fixed synonymous substitutions on the human branch.

Code availability

The code that we used to parameterize our model is freely available online at https://github.com/uricchio/mktest.

Received: 25 September 2018; Accepted: 28 March 2019;
Published online: 6 May 2019

References

1. Darwin, C. *On the Origin of Species* (Murray, 1859).
2. Wallace, A. R. *Darwinism: an exposition of the theory of natural selection with some of its applications* (MacMillan & Co., 1889).
3. Wright, S. On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution* **2**, 279–294 (1948).
4. Kimura, M. et al. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
5. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96 (1973).
6. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275 (1977).
7. McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the ADH locus in *Drosophila*. *Nature* **351**, 652 (1991).
8. Fay, J. C., Wyckoff, G. J. & Wu, C.-I. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**, 1024 (2002).
9. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
10. Charlesworth, B. & Charlesworth, D. Neutral variation in the context of selection. *Mol. Biol. Evol.* **35**, 1359–1361 (2018).
11. Corbett-Detig, R. B., Hartl, D. L. & Sackton, T. B. Natural selection constrains neutral diversity across a wide range of species. *PLoS Biol.* **13**, e1002112 (2015).
12. Coop, G. Does linked selection explain the narrow range of genetic diversity across species? Preprint at *bioRxiv* https://doi.org/10.1101/042598 (2016).
13. Kern, A. D. & Hahn, M. W. The neutral theory in light of natural selection. *Mol. Biol. Evol.* **35**, 1366–1371 (2018).
14. Jensen, J. D. et al. The importance of the neutral theory in 1968 and 50 years on: a response to Kern and Hahn 2018. *Evolution* **73**, 111–114 (2018).
15. Leffler, E. M. et al. Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biol.* **10**, e1001388 (2012).
16. Galtier, N. Adaptive protein evolution in animals and the effective population size hypothesis. *PLoS Genet.* **12**, e1005774 (2016).
17. Smith, N. G. C. & Eyre-Walker, A. Adaptive protein evolution in *Drosophila*. *Nature* **415**, 1022 (2002).
18. Sawyer, S. A. & Hartl, D. L. Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176 (1992).
19. Tataru, P., Mollion, M., Glémin, S. & Bataillon, T. Inference of distribution of fitness effects and proportion of adaptive substitutions from polymorphism data. *Genetics* **207**, 1103–1119 (2017).
20. Ratnakumar, A. et al. Detecting positive selection within genomes: the problem of biased gene conversion. *Philos. Trans. R. Soc. London B* **365**, 2571–2580 (2010).
21. Fay, J. C., Wyckoff, G. J. & Wu, C.-I. Positive and negative selection on the human genome. *Genetics* **158**, 1227–1234 (2001).
22. Eyre-Walker, A. & Keightley, P. D. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Mol. Biol. Evol.* **26**, 2097–2108 (2009).
23. Messer, P. W. & Petrov, D. A. Frequent adaptation and the McDonald–Kreitman test. *Proc. Natl Acad. Sci. USA* **110**, 8615–8620 (2013).
24. James, J. E., Piganeau, G. & Eyre-Walker, A. The rate of adaptive evolution in animal mitochondria. *Mol. Ecol.* **25**, 67–78 (2016).
25. Enard, D., Cai, L., Gwennap, C. & Petrov, D. A. Viruses are a dominant driver of protein adaptation in mammals. *eLife* **5**, e12469 (2016).
26. Pritchard, J. K., Pickrell, J. K. & Coop, G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr. Biol.* **20**, R208–R215 (2010).
27. Messer, P. W. & Petrov, D. A. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol. Evol.* **28**, 659–669 (2013).
28. Berg, J. J. & Coop, G. A population genetic signal of polygenic adaptation. *PLoS Genet.* **10**, e1004412 (2014).
29. Schrider, D. R. & Kern, A. D. Soft sweeps are the dominant mode of adaptation in the human genome. *Mol. Biol. Evol.* **34**, 1863–1877 (2017).
30. Uricchio, L. H., Kitano, H. C., Gusev, A. & Zaitlen, N. A. An evolutionary compass for detecting signals of polygenic selection and mutational bias. *Evol. Lett.* **3**, 69–79 (2019).
31. Elyashiv, E. et al. A genomic map of the effects of linked selection in *Drosophila*. *PLoS Genet.* **12**, e1006130 (2016).
32. Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).

33. Barton, N. H. Linkage and the limits to natural selection. *Genetics* **140**, 821–841 (1995).
34. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
35. Haller, B. C. & Messer, P. W. asymptoticMK: a web-based tool for the asymptotic McDonald–Kreitman test. *G3 (Bethesda)* **7**, 1569–1575 (2017).
36. Evans, S. N., Shvets, Y. & Slatkin, M. Non-equilibrium theory of the allele frequency spectrum. *Theor. Popul. Biol.* **71**, 109–119 (2007).
37. Kimura, M. Diffusion models in population genetics. *J. Appl. Probab.* **1**, 177–232 (1964).
38. Phung, T. N., Huber, C. D. & Lohmueller, K. E. Determining the effect of natural selection on linked neutral divergence across species. *PLoS Genet.* **12**, e1006199 (2016).
39. Consortium TGP et al. A global reference for human genetic variation. *Nature* **526**, 68 (2015).
40. Boyko, A. R. et al. Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet.* **4**, e1000083 (2008).
41. Kim, B. Y., Huber, C. D. & Lohmueller, K. E. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. *Genetics* **206**, 345–361 (2017).
42. Eyre-Walker, A., Woolfit, M. & Phelps, T. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* **173**, 891–900 (2006).
43. Hill, W. G. & Robertson, A. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
44. Smith, J. M. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
45. Macpherson, J. M., Sella, G., Davis, J. C. & Petrov, D. A. Genomewide spatial correspondence between nonsynonymous divergence and neutral polymorphism reveals extensive adaptation in *Drosophil*a. *Genetics* **177**, 2083–2099 (2007).
46. Castellano, D., Coronado-Zamora, M., Campos, J. L., Barbadilla, A. & Eyre-Walker, A. Adaptive evolution is substantially impeded by Hill–Robertson interference in *Drosophila*. *Mol. Biol. Evol.* **33**, 442–455 (2015).
47. Marsden, C. D. et al. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl Acad. Sci. USA* **113**, 152–157 (2016).
48. Bullaughey, K. L., Przeworski, M. & Coop, G. No effect of recombination on the efficacy of natural selection in primates. *Genome Res.* **18**, 544–554 (2008).
49. Hussin, J. G. et al. Recombination affects accumulation of damaging and disease-associated mutations in human populations. *Nature Genet.* **47**, 400 (2015).
50. Jensen, J. D., Thornton, K. R. & Andolfatto, P. An approximate Bayesian estimator suggests strong, recurrent selective sweeps in *Drosophila*. *PLoS Genet.* **4**, e1000198 (2008).
51. Hernandez, R. D. et al. Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924 (2011).
52. Enard, D., Messer, P. W. & Petrov, D. A. Genome-wide signals of positive selection in human evolution. *Genome Res.* **24**, 885–895 (2014).
53. Comeron, J. M. & Kreitman, M. Population, evolutionary and genomic consequences of interference selection. *Genetics* **161**, 389–410 (2002).
54. Uricchio, L. H. & Hernandez, R. D. Robust forward simulations of recurrent hitchhiking. *Genetics* **197**, 221–236 (2014).
55. Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Context dependence, ancestral misidentification, and spurious signatures of natural selection. *Mol. Biol. Evol.* **24**, 1792–1800 (2007).
56. Ewing, G. B. & Jensen, J. D. The consequences of not accounting for background selection in demographic inference. *Mol. Ecol.* **25**, 135–141 (2016).
57. Torres, R., Szpiech, Z. A. & Hernandez, R. D. Human demographic history has amplified the effects of background selection across the genome. *PLoS Genet.* **14**, e1007387 (2018).
58. Huang, Y.-F. & Siepel, A. Estimation of allele-specific fitness effects across human protein-coding sequences and implications for disease. Preprint at *bioRxiv* https://doi.org/10.1101/441337 (2018).
59. Huber, C. D., Kim, B. Y., Marsden, C. D. & Lohmueller, K. E. Determining the factors driving selective effects of new nonsynonymous mutations. *Proc. Natl Acad. Sci. USA* **114**, 4465–4470 (2017).
60. Yates, A. et al. Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2015).
61. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
62. Löytynoja, A. & Goldman, N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**, 579 (2010).
63. Hernandez, R. D. & Uricchio, L. H. SFS_CODE: more efficient and flexible forward simulations. Preprint at *bioRxiv* https://doi.org/10.1101/025064 (2015).
64. Uricchio, L. H., Torres, R., Witte, J. S. & Hernandez, R. D. Population genetic simulations of complex phenotypes with implications for rare variant association tests. *Genet. Epidemiol.* **39**, 35–44 (2015).
65. Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035 (2002).
66. Tennessen, J. A. et al. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).

Acknowledgements

We thank A. Aw, N. Rosenberg and members of the Rosenberg and Petrov laboratories for helpful discussions. L.H.U. was partially supported by an IRACDA fellowship through NIGMS grant No. K12GM088033. L.H.U was supported by National Institutes of Health grant R01 HG005855 and National Science Foundation grant DBI-1458059 (to N. Rosenberg). We also thank the Stanford/SJSU IRACDA Program for support.

Author contributions

Designed the research: L.H.U., D.A.P., D.E. Performed the modeling and simulations: L.H.U. Analyzed the data: L.H.U., D.A.P.. Designed inference procedure: L.H.U. Wrote the paper: L.H.U. Edited and approved paper: L.H.U., D.A.P., D.E.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-019-0890-6.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.H.U. or D.E.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Uricchio LH & Enard D

Last updated by author(s): Mar 18, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement

☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐ ☒ A description of all covariates tested

☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
Give P values as exact values whenever suitable.

☐ ☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☒ ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection: We used open access human genetic data from the Thousand Genome Project (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/) and B-values (i.e., estimates of Background selection strength in humans) from previous research (http://www.phrap.org/othersoftware.html) as described in our methods section. Our final dataset, which summarizes data from these datasets, is available at https://github.com/uricchio/mktest in the data folder.

Data analysis: We used the Blat tool (https://genome.ucsc.edu/FAQ/FAQblat.html) to identify orthologs as described in the methods section. We used custom software for the remaining analyses. Our software is available at https://github.com/uricchio/mktest -- note that the software is freely available but is currently set to compile on the Stanford cluster and run on the Stanford Sherlock cluster.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The supplemental table, along with the data that we used to parameterize our model, is available online at https://github.com/uricchio/mktest

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We use models of weak adaptation to show that previous methods to estimate adaptation rate provide inaccurate estimates when adaptation is weak. We then extend existing methods to correct this issue, and show that variation in the strength of linked selection can be used to additionally estimate both adaptation rate and strength. We apply our method to human genomes to show that human polymorphism and divergence data is consistent with weak adaptation genome-wide, while virus interacting proteins support both faster and stronger adaptation than the genome background.
Research sample	We used genetic data from the 661 samples of African origin from the Thousand Genomes Project in our study. We chose these samples because 1) African samples have been understudied in previous human genetics research and 2) the best-fitting demographic models for this continental group are simpler than European demographic models, simplifying our analysis.
Sampling strategy	Since we used previously sampled data we were restricted to the samples available and did not make any sampling design choices herein.
Data collection	We used publicly available data.
Timing and spatial scale	The data were collected by the 1000 genomes consortium. Sample collection is described here: https://media.nature.com/original/nature-assets/nature/journal/v526/n7571/extref/nature15393-s1.pdf. According to this document, the samples were sequenced between October 2012 and March 2013.
Data exclusions	We included genetic data from all human coding regions for which we could map an ortholog and a B-value, as described in the methods section. The total data set is described in Table 1.
Reproducibility	We validated our estimation procedures by simulating a large number of datasets and applying our approach to the simulated data.
Randomization	All samples were combined into a single group (i.e., we perform estimation on genetic data from individuals from a single continental group, and do not compare across groups).
Blinding	Blinding was not necessary because we do not tune our analysis to the data in any way -- we prepared our estimation procedure by applying it to simulated data and simply report the estimates obtained from the real data.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging