Supplemental Materials

Greater strength of selection and higher proportion of beneficial amino acid changing mutations in humans compared to mice and *Drosophila melanogaster*

Ying Zhen[1,2,3*], Christian D. Huber[1], Robert W. Davies[4], Kirk E. Lohmueller[1,5*]

Table of Contents

Supplemental Text

Demographic and DFE inferences for mice

We used methods established in Huber et al. (2017) to infer demography and the DFE of neutral and deleterious mutations from mouse polymorphism data. In short, we first used the synonymous SFS to infer demographic parameters for the Simple model using ∂a∂i (Sawyer and Hartl 1992; Gutenkunst et al. 2009). We infer that the ancestral population size is approximately 206,500 which expanded 2.4-fold 293,000 generations ago (Supplemental Table S3). Conditional on this demographic model, we estimated the DFE for new nonsynonymous mutations in mice. We assume that the DFE follows a gamma-distribution and estimate its shape parameter α to be 0.21 and scale parameter to be 0.083 (Supplemental Table S3). These estimates are within the same magnitude of previous estimates from Huber *et al.* (2017), which used a much smaller dataset (<0.1% of the total sites used in our study).

Estimating α on the human lineage

Human-chimpanzee differences were polarized by the macaque sequence. Differences between human and chimpanzee were assigned to human lineage if the bases differ between human and macaque, but were the same between chimpanzee and macaque in the pairwise genome alignments. Differences that are in regions where the human and macaque sequences were un-alignable and sites that differ among human, chimp, and macaque cannot be polarized (3.8% total) and were filtered out. The total length of coding regions was scaled by this filter accordingly. A total of 30,530 synonymous substitutions and 20,013 nonsynonymous differences were observed on human lineage. We assume the human lineage and the outgroup lineage contribute

equally to neutral D_S, thus the total D_S between human and chimpanzee would be 61,060. We then adjusted divergence time to match predicted D_S to 61,060, and use this adjusted divergence time, DFE of deleterious mutations and demography of human to predict D_N as described in *Model-based estimates of α*. The human lineage α is then calculated using equation (2) in the main text.

Details on the analysis of SSWW sites

To test whether BGC and hypermutable CpG sites drive the observed pattern of positive selection across species, we filtered human and mouse data to keep only strong to strong or weak to weak mutations (herein called SSWW mutations), which are not affected by BGC and are not CpG changes. These changes include only A to T, T to A, C to G and G to C changes. These changes are only a small subset of all variable sites. The nonsynonymous and synonymous sequence lengths (L_{NS}, L_S) depend on the transition/transversion ratio and the CpG mutational bias. SSWW mutations are all transversions and do not included any CpG mutations, leading to a multiplier of L_{NS} = 5.21 x L_S in both primates and rodents. To determine this multiplier: 1) we used the numbers of 0-, 2-, 3- and 4- fold sites in the human exome from Veeramah et al. (2014); 2) we consider all 2-fold sites to be nonsynonymous (because SSWW mutations are all transversions); and 3) we do not consider a mutational bias of CpG sites (because CpG sites are not included in the SSWW set). In addition, because SSWW mutations are only a small subset of all mutations, mutation rates need to be scaled down to the SSWW specific mutation rate. We used the observed number of synonymous SSWW polymorphisms to estimate the mutation rate of human SSWW mutations to be 3.14×10^{-9},

which is comparable to previous estimates (Kong et al. 2012; Lachance and Tishkoff 2014). For humans, the filtered SSWW polymorphisms have a similar SFS as the full dataset (Supplemental Fig. S5A). Thus we use the demographic and DFE parameters estimated from the full data. Following the method used for the full data, we re-estimated the human-chimpanzee divergence time that fits best to the observed SSWW D_S. We then predict the SSWW D_N using this newly estimated divergence time, DFE, and demographic models (Supplemental Table S1).

We estimated that under the Simple demographic model (*Nout=Nanc=Nanc.in*) and the Complex demographic model (*Nout≠Nanc≠Nanc.in*), approximately 13.1% or approximately 26.3% of the observed SSWW D_N in humans using chimpanzee as outgroup was driven by positive selection, respectively (Table 1). These estimates of α from SSWW sites are slightly elevated but are comparable to the estimates from the full dataset (Table 1).

For mice, however, the SFS of SSWW polymorphism has a very different shape compared to the SFS from the full dataset (Supplemental Fig. S5B). Thus, we re-estimated the demographic and DFE parameters for the mouse SSWW mutations (Supplemental Table S3). First, we used the observed number of synonymous SSWW polymorphisms to estimate the mutation rate for mice SSWW mutations to be 5.99×10^{-10}. Then, using the SFS for SSWW synonymous polymorphisms, we inferred that the ancestral population size in mice is approximately 246,256 which expanded 1.7-fold approximately 262,000 generations ago (Supplemental Table S3). Conditional on this demographic model, we estimated the DFE for new nonsynonymous SSWW mutations in mice. We assume that the DFE follows a gamma distribution and estimate its shape

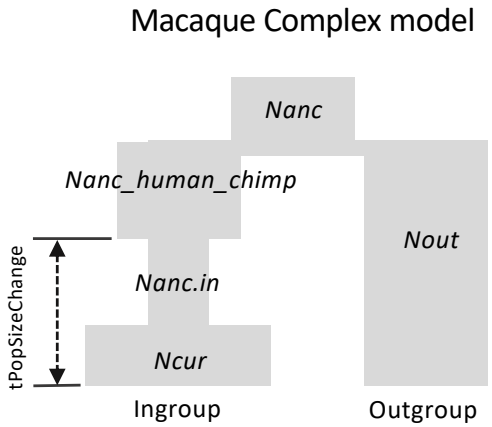
parameter α to be 0.21 and scale parameter beta to be 0.050 (Supplemental Table S3). These estimates are within the same magnitude of the estimates from the full dataset and a previous study (Huber et al. 2017). We re-estimated the mouse-rat divergence time that fits best with the observed SSWW D_S.

We then estimated that under the Simple demographic model and the Complex demographic model, approximately 19.5% and 9.7% of the observed SSWW D_N in mice was driven by positive selection, respectively (Table 1). These estimates of proportion of nonsynonymous substitutions fixed by positive selection from SSWW mutations are much lower than those estimated from the full dataset (Table 1). This suggests that biased gene conversion and CpG mutational processes may account for some of the nonsynonymous substitutions between mouse and rat.

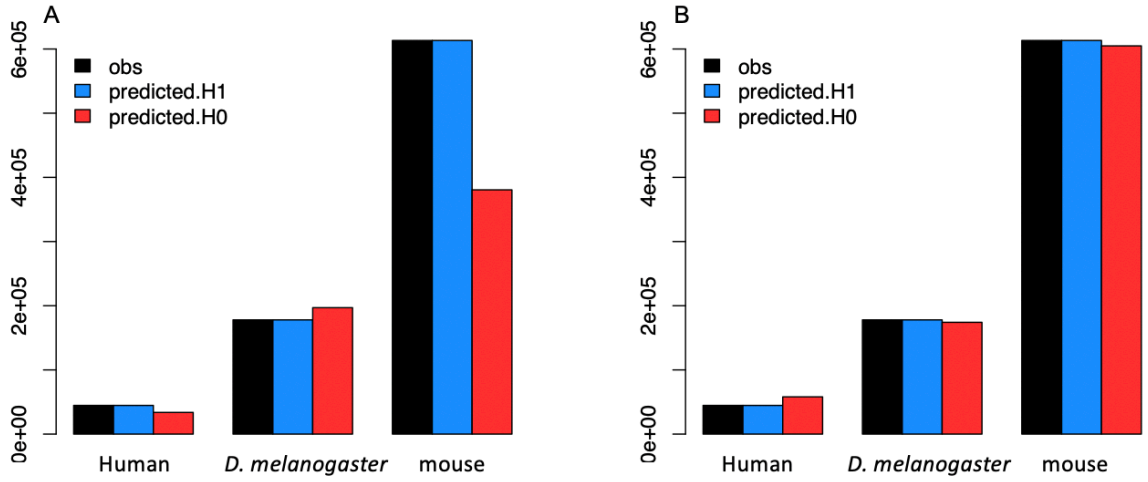
Coalescent simulations to compare human polymorphism under Simple and Complex demographic models

To evaluate whether patterns of neutral polymorphism would be predicted to be different under the human Simple and Complex demographic models, we conducted coalescent simulations under these models using *ms* (Hudson 2002). Specifically, we simulated 1000 replicates for each scenario and calculated the mean number of synonymous segregating sites across replicates. Both models showed similar numbers of neutral segregating sites (34075 for the Simple Model and 33810 for the Complex model), suggesting that using population sizes more appropriate for the outgroup population will not affect polymorphism data in the ingroup sample.

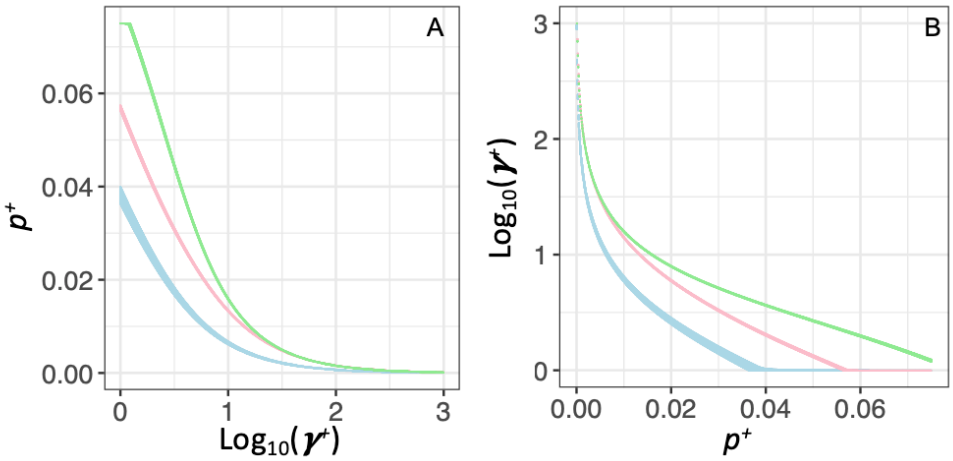
Supplemental Fig. S1
Complex demographic model used for divergence between human (ingroup) and macaque (outgroup). In this demographic model, at the timing of the human-chimp split approximately 105,000 generations ago (2.6 Myr assuming 25 years/generation), the human population changes size to that of the human-chimp ancestral size of 60,000 (see Results). *Nout* is set to 73,000 (Hernandez et al. 2011).



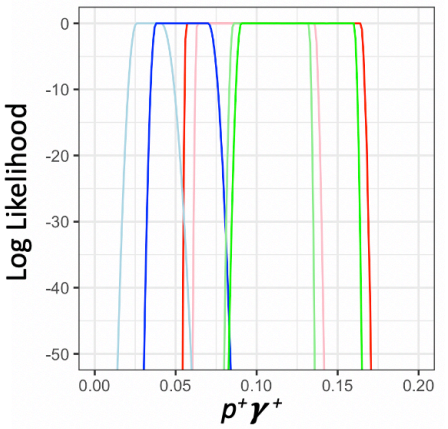
Supplemental Fig. S2
Observed and predicted number of nonsynonymous divergent sites between each of our three focal species and their outgroups. (A) Predictions using MLEs of p^+ and s^+. (B) Predictions using MLEs of p^+ and $\boldsymbol{\gamma}^+$. Predicted numbers are from two models: the full model, H1, where each species has its own positive selection parameters and the constrained model, H0, where the positive selection parameters are constrained to be the same across all three taxa.



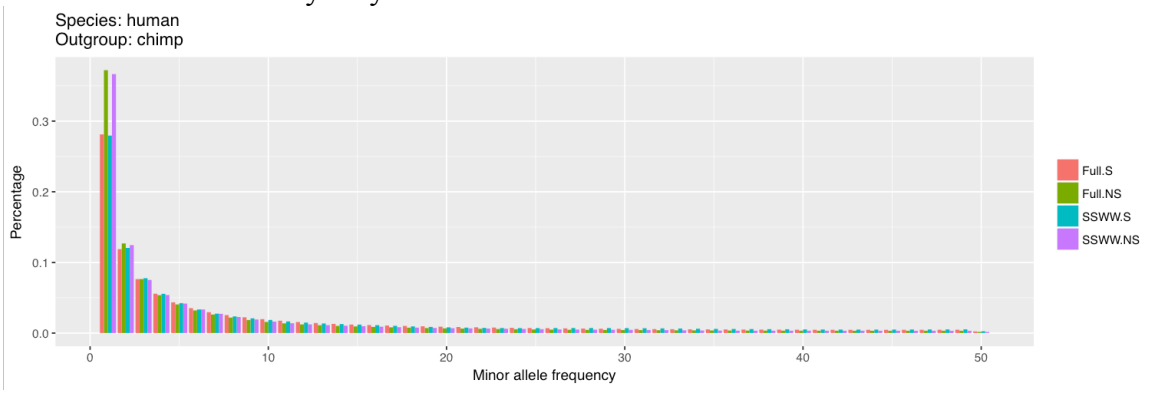
Supplemental Fig. S3
Conditional log-likelihood (LL) surfaces. (A) Maximizing p^+ given particular values of γ^+ and (B) maximizing γ^+ given particular values of p^+. Only grid points within 3 LL units of the MLEs of for each parameter for each species are shown. Light blue denotes human, pink denotes *D. melanogaster*, and light green denotes mouse.

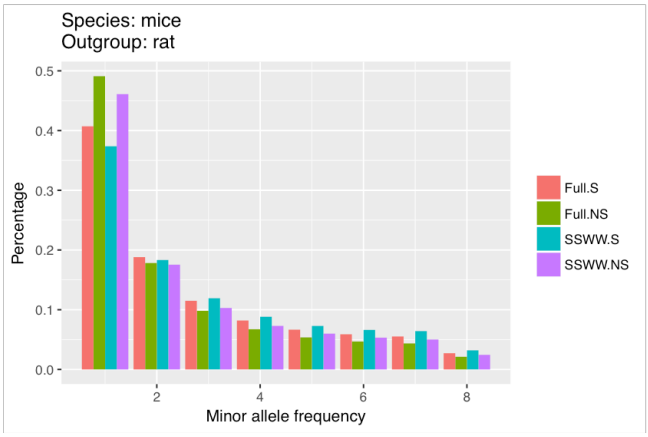


Supplemental Fig. S4
The composite parameter $p^+\gamma^+$ across species. Log-likelihood curves for $p^+\gamma^+$ in the three different taxa. Red denotes the inference for *D. melanogaster*, green denotes the inference for mouse, blue denotes the inference for humans using the chimpanzee as the outgroup. Lighter colors denote the Simple model. Darker colors denote the Complex model that better models the ancestral demography and population size of the outgroup.

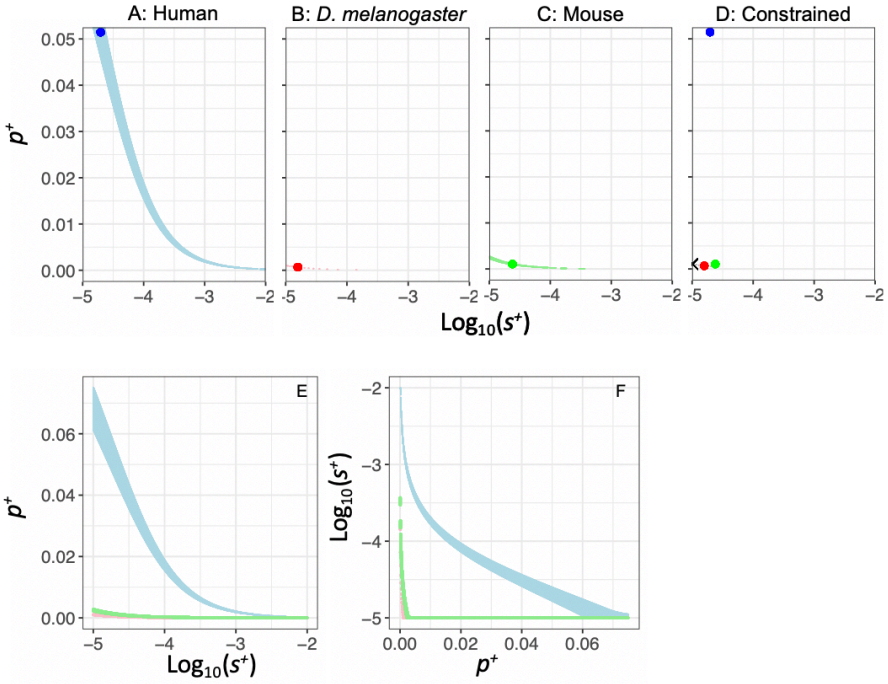


Supplemental Fig. S5
Effects of biased gene conversion on the folded SFS for (A) humans and (B) mice. Full denotes the data without any filtering for biased gene conversion. SSWW denotes the SFS for strong to strong or weak to weak substitutions only. S denotes synonymous SNPs. NS denotes nonsynonymous SNPs.

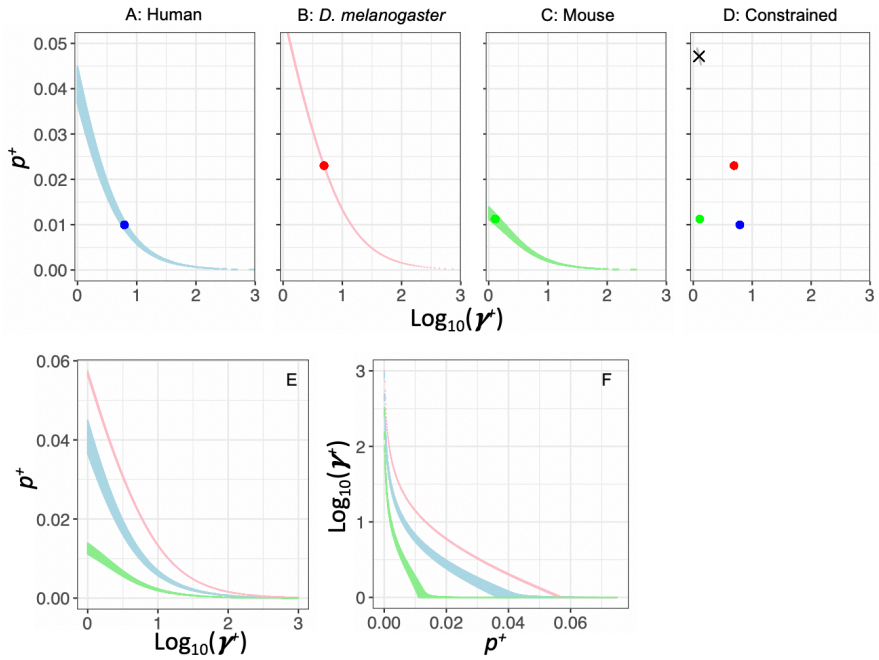




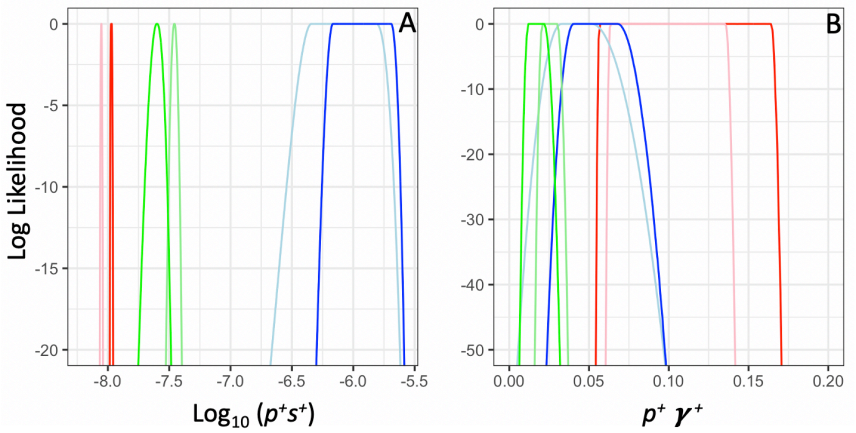
Supplemental Fig. S6
Log-likelihood surfaces for sites unaffected by biased gene conversion (SSWW sites in mammals). (A-C) show the log-likelihood surfaces for p^+ and s^+ for different taxa. (D) shows the constrained model, H0, where p^+ and s^+ are constrained to be the same across all three taxa. Log-Likelihoods are calculated using grid search method of $\log_{10}(s^+)$ in the range of -5 to -2 and p^+ in the range of 0-7.5%. Blue denotes human, red denotes *D. melanogaster*, and green denotes mouse. The large points represent the MLE for each species. The black cross in panel D represents the MLE of the constrained model, and the lighter colors show grid points within 3 LL units of each MLE. The Complex model is used for each species and we use chimpanzee as the outgroup for humans. (E) shows the conditional log-likelihood surface maximizing p^+ given particular values of s^+ and (F) shows the conditional log-likelihood surface maximizing s^+ given particular values of p^+. In panels E-F, only grid points within 3 LL units of the MLEs of for each parameter for each species are shown.



Supplemental Fig. S7
Log-likelihood (LL) surfaces for sites unaffected by biased gene conversion (SSWW sites in mammals) for p^+ and γ^+. (A-C) show the log-likelihood surfaces for p^+ and γ^+ for different taxa. (D) shows the constrained model, H0, where p^+ and γ^+ are constrained to be the same across all three taxa. Log-Likelihoods are calculated using grid search method of $\log_{10}(\gamma^+)$ in the range of 0 to 3 and p^+ in the range of 0-7.5%. Blue denotes human, red denotes *D. melanogaster*, and green denotes mouse. The large points represent the MLE for each species. The black cross in panel D represents the MLE of the constrained model, and the lighter colors show grid points within 3 LL units of each MLE. The Complex model is used for each species and we use chimpanzee as the outgroup for humans. (E) shows the conditional log-likelihood surface maximizing p^+ given particular values of γ^+ and (F) shows the conditional log-likelihood surface maximizing γ^+ given particular values of p^+. In panels E-F, only grid points within 3 LL units of the MLEs of for each parameter for each species are shown.



Supplemental Fig. S8
The composite parameters p^+s^+ and $p^+\gamma^+$ across taxa using SSWW sites that are not influenced by BGC or hypermutable CpG sites. (A) Log-likelihood surfaces for p^+s^+ in the three different taxa. (B) Log-likelihood surfaces for $p^+\gamma^+$ in the three different taxa. Red denotes the inference for *D. melanogaster*, green denotes the inference for mouse, blue denotes the inference for humans using the chimpanzee as the outgroup. Lighter colors denote the Simple model. Darker colors denote the Complex model that better models the ancestral demography and population size of the outgroup.



Supplemental Table S1. Observed and predicted counts of polymorphism and divergence.

species	outgroup	demographic model	dataset	D_S	D_N	P_S	P_N	α 95% CI
human	chimpanzee	observed	full	65,021	44,665	33,684	32,520	
		predicted:Simple Model		65,021	39,925			0.099, 0.113
		predicted:Complex Model		65,021	33,681			0.240, 0.252
	macaque	observed		323,581	182,532	34,113	32,798	
		predicted:Simple Model		323,581	196,401			-0.080, -0.072
		predicted:Complex Model		323,581	135,101			0.257, 0.263
	chimpanzee	observed	SSWW	5,077	8,084	2,344	5,116	
		predicted:Simple Model		5,077	7,026			0.115, 0.147
		predicted:Complex Model		5,077	5,961			0.249, 0.276
D. melanogaster	*D. simulans*	observed	full	404,537	177,936	466,188	231,267	
		predicted:Simple Model		404,537	83,193			0.531, 0.534
		predicted:Complex Model		404,537	70,660			0.601, 0.604
mice	rat	observed	full	1,133,269	613,281	181,039	73,591	
		predicted:Simple Model		1,133,269	334,398			0.454, 0.456
		predicted:Complex Model		1,133,269	360,664			0.411, 0.413
		observed	SSWW	58,882	51,462	10,713	10,662	
		predicted:Simple Model		58,882	41,422			0.189, 0.201
		predicted:Complex Model		58,882	46,471			0.090, 0.103

Predicted counts of D_N come from the gamma-DFE with only neutral and deleterious mutations. The differences between observed and predicted D_N are attributed to positive selection and are used to estimate α. 95% CIs on α were calculated from a parametric bootstrap by resampling D_N from a Poisson distribution with mean equal to the observed D_N. As such, they do not include other forms of uncertainty due to demography or the DFE.

Supplemental Table S2. Effect of slightly deleterious mutations on α.

			MAF*				
Species	**Outgroup**	**Sample Size**	**0**	**>5%**	**>10%**	**>20%**	**>30%**
Human	Chimpanzee	100	-0.41	-0.09	-0.04	-0.01	0.00
Human	Macaque	100	-0.70	-0.33	-0.26	-0.22	-0.21
D. melanogaster	*D. simulans*	100	-0.13	0.39	0.44	0.49	0.51
Mouse	Rat	16	0.25	na	0.36	0.40	0.40

*MAF=0 includes SNPs at all frequencies when computing α. MAF>5% includes only SNPs at frequencies larger than 5%, etc.

Supplemental Table S3. Demographic and DFE parameters estimated from polymorphism data.

parameter (unit)	human	Drosophila	mice	human SSWW	mice SSWW
sample size	100	100	16	100	16
total sites	19089129	15819843	26642307	19089129	26642307
Nanc	7067	2790000	206519	7067	246256
Ncurr/Nanc	2.34	2.73	2.37	2.34	1.71
tau (2Nanc)	0.43	0.09	0.71	0.43	0.53
DFE: alpha	0.19	0.35	0.21	0.19	0.21
DFE: beta (s)	0.074	0.00038	0.083	0.074	0.050
mu	2.5E-08	1.5E-09	5.4E-09	3.14E-09	5.99E-10
NS/S length ratio	2.31	2.85	2.31	5.21	5.21

Nanc: ancestral population size. Ncurr: current population size. Tau: is the time at which the population changed in size. NS: nonsynonymous. S: synonymous. mu: mutation rate per site per generation.

Supplemental Table S4. Effective population sizes of chimpanzee, macaque, and ancestral primate populations from the literature.

	Ne	Reference
chimpanzee	30900–61800	Prado-Martinez et al 2013
	25000-35000	Fischer et al 2004
	18300	Hvilsom et al 2014
	~30000	Hvilsom et al 2012
human-chimpanzee ancestor	52,000–96,000	Chen and Li 2001
	12,000–21,000	Yang et al 2002
	65000	Hobolth et al 2007
	47,000	Hobolth et al 2011
	35,000-65,000	Ruvolo et al 1997
	99 (95–102) x 1000	Burgess et al 2008
	50000, 63000	Prado-Martinez et al 2013
	33000	Hara et al 2012
	27716-41263	Schrago 2014a
	~50000	Wall 2003
	47500	Schrago 2014b
macaque	73000 (ancestral)	Hernandez et al 2007
	52350, 61800 (Indian rhesus)	Xue et al 2016
	71200, 82080 (Chinese rhesus)	Xue et al 2016
human-macaque ancestor	48000	McVicker et al 2009

Supplemental Table S5. Parameters of the Simple demographic model and the Complex demographic model for each species

species	human						*D. melanogaster*		mice			
demographic model	Simple Model	Complex Model	Simple Model	Complex Model	Simple Model	Complex Model	Simple Model	Complex Model	Simple Model	Complex Model	Simple Model	Complex Model
outgroup	chimpanzee	chimpanzee	macaque	macaque	chimpanzee	chimpanzee	*D. simulans*	*D. simulans*	rat	rat	rat	rat
dataset	full	full	full	full	SSWW	SSWW	full	full	full	full	SSWW	SSWW
sample size	100	100	100	100	100	100	100	100	16	16	16	16
syn theta	9538.37	9538.37	9620.51	9620.51	656.78	656.78	187784.40	187784.40	85008.11	85008.11	4329.45	4329.45
tdiv	6.39	3.19	33.21	30.73	7.30	4.10	1.79	0.65	12.91	13.13	13.02	13.40
tau	0.18	0.18	0.18	0.18	0.18	0.18	0.03	0.03	0.30	0.30	0.31	0.31
omega	0.43	0.43	0.43	0.43	0.43	0.43	0.37	0.37	0.42	0.42	0.58	0.58
omega_outgroup	0.43	1.81	0.43	4.41	0.43	1.81	0.37	1.50	0.42	0.20	0.58	0.20
omega_ancestral	0.43	3.63	0.43	2.90	0.43	3.63	0.37	1.50	0.42	0.20	0.58	0.20

All scaled to current ingroup population size (Ncurr in Supplemental Table S3).
tdiv: divergence time of ingroup and outgroup species;
tau: time from present to the first change in ingroup Ne;
omega: ancestral ingroup population size;
omega_outgroup: outgroup long-term effective population size;
omega_ancetral: ancestral population size before ingroup and outgroup divergence

Only for the Complex model for human using the macaque as the outgroup, we have two additional parameters: omega2 and tPopSizeChange to model the change of human ingroup population size to human-chimpanzee ancestral population size (omega2=3.63) at human-chimpanzee divergence time (tPopSizeChange=3.19).

Supplemental Table S6. Comparison of the beneficial selection coefficients and proportion of new beneficial mutations across all three taxa.

A. Test whether s^+ and p^+ differ across taxa using the full datasets

1

Hypothesis	species	outgroup	demorgraphic model	p+	log10(s+)	LL	abbr.
Full model (H1)	**Human**	**chimpanzee**	**Complex**	**1.55E-02**	**-3.949**	**-6.27**	**human3**
	Human	Macaque	Complex	7.20E-03	-4.422	-6.98	human3mac
	Human	chimpanzee	Simple	2.39E-02	-4.429	-6.27	human2
	D.melanogaster	***D.simulans***	**Complex**	**6.75E-04**	**-4.801**	**-6.96**	**fly3**
	D.melanogaster	*D.simulans*	Simple	6.00E-04	-4.831	-6.96	fly2
	Mouse	**Rat**	**Complex**	**1.02E-02**	**-4.797**	**-7.58**	**mice3**
	Mouse	Rat	Simple	1.21E-02	-4.954	-7.58	mice2

2

	models comparison	sum of LL	p+	log10(s+)	constrained LL	likelihood ratio	p-value
Constrained (H0): same s+, p+	**human3=fly3=mice3**	**-20.82**	**7.50E-05**	**-3.772**	**-62507.88**	**124974.12**	<1E-16
	human3=fly3	**-13.24**	**1.05E-03**	**-4.995**	**-1576.27**	**3126.07**	<1E-16
	human3=fly2	-13.24	8.25E-04	-4.969	-1588.05	3149.63	<1E-16
	human2=fly3	-13.24	1.05E-03	-4.995	-268.00	509.52	<1E-16
	human2=fly2	-13.24	8.25E-04	-4.970	-271.30	516.12	<1E-16
	human3mac=fly3	-13.94	1.05E-03	-4.992	-6835.68	13643.47	<1E-16
	human3mac=fly2	-13.94	9.00E-04	-5.000	-6943.80	13859.73	<1E-16
	human3=mice3	**-13.85**	**1.64E-02**	-5.000	**-852.13**	**1676.56**	<1E-16
	human3=mice2	-13.85	1.35E-02	-5.000	-968.96	1910.21	<1E-16
	human2=mice3	-13.85	1.64E-02	-5.000	-92.47	157.22	<1E-16
	human2=mice2	-13.85	1.34E-02	-5.000	-117.56	207.41	<1E-16
	human3mac=mice3	-14.56	1.68E-02	-5.000	-840.342	1651.57	<1E-16
	human3mac=mice2	-14.56	1.38E-02	-5.000	-1487.82	2946.52	<1E-16
	fly3=mice3	**-14.55**	**7.50E-05**	**-3.772**	**-60903.15**	**121777.21**	<1E-16
	fly3=mice2	-14.55	7.50E-05	-3.735	-73462.87	146896.65	<1E-16
	fly2=mice3	-14.55	7.50E-05	-3.852	-62865.88	125702.67	<1E-16
	fly2=mice2	-14.55	7.50E-05	-3.812	-76531.50	153033.91	<1E-16

Under the full model, MLEs are approximately 1.55% of new nonsynonymous mutations are beneficial with s^+ of approximately 1.12×10^{-4} in humans (outgroup: chimpanzee), approximately 0.0675% of new nonsynonymous mutations are beneficial with s^+ of 1.58×10^{-5} in *D. melanogaster*, and approximately 1.02% of new nonsynonymous mutations are beneficial with s^+ of 1.60×10^{-5} in mice. However, as there is a ridge in the likelihood surface, it is not possible to estimate s^+ and p^+ separately. As such, other models with a larger s^+ and a smaller p^+ fit the observed date equally well.

B. Test whether γ^+ and p^+ differ across taxa using the full datasets

1

model	species	outgroup	demorgraphic model	p^+	γ^+	log-likelihood	abbr.
Full model (H1)	**Human**	**chimpanzee**	**Complex**	**0.01**	**6.05**	**-6.27**	**human3**
	Human	Macaque	Complex	0.01	1.17	-6.98	human3mac
	Human	chimpanzee	Simple	0.02	1.65	-6.27	human2
	D.melanogaster	***D.simulans***	**Complex**	**0.02**	**4.92**	**-6.96**	**fly3**
	D.melanogaster	*D.simulans*	Simple	0.04	2.79	-6.96	fly2
	Mouse	**Rat**	**Complex**	**0.04**	**3.76**	**-7.58**	**mice3**
	Mouse	Rat	Simple	0.05	2.15	-7.58	mice2

2

	models comparison	sum of LL	p^+	γ^+	constrained LL	likelihood ratio	p-value
Constrained (H0): same gamma, p+	**human3=fly3=mice3**	**-20.82**	**0.00**	**61.66**	**-1791.51**	**3541.39**	<1E-16
	human3=fly3	**-13.24**	**0.06**	**1.00**	**-292.10**	**557.73**	<1E-16
	human3=fly2	-13.24	0.06	1.00	-498.49	970.50	<1E-16
	human2=fly3	-13.24	0.06	1.00	-319.50	612.54	<1E-16
	human2=fly2	-13.24	0.06	1.00	-455.30	884.14	<1E-16
	human3mac=fly3	-13.94	0.02	1.00	-26283.69	52539.50	<1E-16
	human3mac=fly2	-13.94	0.02	1.00	-22821.27	45614.66	<1E-16
	human3=mice3	**-13.85**	**0.08**	**1.00**	**-1309.62**	**2591.53**	<1E-16
	human3=mice2	-13.85	0.00	845.28	-859.97	1692.23	<1E-16
	human2=mice3	-13.85	0.08	1.10	-936.87	1846.03	<1E-16
	human2=mice2	-13.85	0.08	1.09	-895.77	1763.84	<1E-16
	human3mac=mice3	-14.56	0.02	1.00	-44692.64	89356.16	<1E-16
	human3mac=mice2	-13.85	0.02	1.00	-54279.71	108531.71	<1E-16
	fly3=mice3	**-14.55**	**0.00**	**58.48**	**-14.55**	**0.00**	**1**
	fly3=mice2	-14.55	0.01	9.08	-14.56	0.02	0.99
	fly2=mice3	-14.55	0.00	993.12	-424.24	819.39	<1E-16
	fly2=mice2	-14.55	0.01	12.76	-14.56	0.02	0.99

Under the full model, MLEs are approximately 1% of new nonsynonymous mutations are beneficial with γ^+ of 6.05 in humans, approximately 2% of new nonsynonymous mutations are beneficial with γ^+ of 4.92 in *D. melanogaster*, and approximately 4% of new nonsynonymous mutations are beneficial with γ^+ of 3.76 in mice. However, as there is a ridge in the likelihood surface, it is not possible to estimate γ^+ and p^+ separately. As such, other models with a larger γ^+ and a smaller p^+ fit the observed date equally well.

C. Test whether s^+ and p^+ differ across taxa using only SSWW changes for human and mouse

[1]

Hypothesis	species	outgroup	demorgraphic model	p+	log10(s+)	LL	abbr.
Full model (H1)	**Human**	**chimpanzee**	**Complex**	**5.15E-02**	**-4.706**	**-5.42**	**human3bgc**
	Human	chimpanzee	Simple	3.26E-02	-4.500	-5.42	human2bgc
	D.melanogaster	***D.simulans***	**Complex**	**6.75E-04**	**-4.801**	**-6.96**	**fly3**
	D.melanogaster	*D.simulans*	Simple	6.00E-04	-4.831	-6.96	fly2
	Mouse	**Rat**	**Complex**	**1.05E-03**	**-4.620**	**-6.34**	**mice3bgc**
	Mouse	Rat	Simple	2.10E-03	-4.780	-6.34	mice2bgc

[2]

	models comparison	sum of LL	p+	log10(s+)	constrained LL	likelihood ratio	p-value
Constrained (H0): same s+, p+	**human3bgc=fly3=mice3bgc**	**-18.72**	**1.05E-03**	**-4.994**	**-431.44**	**825.43**	<1E-16
	human3bgc=fly3	**-12.38**	**1.05E-03**	**-4.995**	**-340.67**	**656.59**	<1E-16
	human3bgc=fly2	-12.38	8.25E-04	-4.970	-342.95	661.15	<1E-16
	human2bgc=fly3	-12.38	1.05E-03	-4.995	-84.76	144.76	<1E-16
	human2bgc=fly2	-12.38	8.25E-04	-4.970	-85.51	146.26	<1E-16
	human3bgc=mice3bgc	**-11.76**	**2.70E-03**	**-5.000**	**-323.63**	**623.74**	<1E-16
	human3bgc=mice2bgc	-11.76	3.53E-03	-5.000	-313.98	604.44	<1E-16
	human2bgc=mice3bgc	-11.76	2.55E-03	-5.000	-79.16	134.79	<1E-16
	human2bgc=mice2bgc	-11.76	3.53E-03	-5.000	-76.06	128.60	<1E-16
	fly3=mice3bgc	**-13.31**	**7.50E-05**	**-3.842**	**-94.75**	**162.89**	<1E-16
	fly3=mice2bgc	-13.31	7.50E-05	-3.840	-524.63	1022.64	<1E-16
	fly2=mice3bgc	-13.31	7.50E-05	-3.926	-118.44	210.26	<1E-16
	fly2=mice2bgc	-13.31	7.50E-05	-3.923	-613.76	1200.90	<1E-16

D. Test whether $\boldsymbol{\gamma}^+$ and p^+ differ across taxa using only SSWW changes for human and mouse

[1]

model	species	outgroup	demorgraphic model	p^+	$\boldsymbol{\gamma}^+$	log-likelihood	abbr.
Full model (H1)	**Human**	**chimpanzee**	**Complex**	**0.01**	**6.18**	**-5.42**	**human3bgc**
	Human	chimpanzee	Simple	0.01	4.80	-5.42	human2bgc
	D.melanogaster	***D.simulans***	**Complex**	**0.02**	**4.92**	**-6.96**	**fly3**
	D.melanogaster	*D.simulans*	Simple	0.04	2.79	-6.96	fly2
	Mouse	**Rat**	**Complex**	**0.01**	**1.29**	**-6.34**	**mice3bgc**
	Mouse	Rat	Simple	0.02	1.11	-6.34	mice2bgc

[2]

	models comparison	sum of LL	p^+	$\boldsymbol{\gamma}^+$	constrained LL	likelihood ratio	p-value
Constrained (H0): same gamma, p+	**human3bgc=fly3=mice3bgc**	**-18.72**	**0.05**	**1.26**	**-2266.67**	**4495.88**	**<1E-16**
	human3bgc=fly3	**-12.38**	**0.06**	**1.00**	**-52.69**	**80.62**	**<1E-16**
	human3bgc=fly2	-12.38	0.06	1.00	-90.94	157.12	<1E-16
	human2bgc=fly3	-12.38	0.06	1.00	-45.22	65.67	5.5E-15
	human2bgc=fly2	-12.38	0.06	1.00	-65.69	106.62	<1E-16
	human3bgc=mice3bgc	**-11.76**	**0.01**	**3.31**	**-124.80**	**226.08**	**<1E-16**
	human3bgc=mice2bgc	-11.76	0.02	1.66	-72.08	120.63	<1E-16
	human2bgc=mice3bgc	-11.76	0.00	301.30	-34.60	45.68	1.2E-10
	human2bgc=mice2bgc	-11.76	0.02	1.00	-20.80	18.07	1.2E-04
	fly3=mice3bgc	**-13.31**	**0.05**	**1.31**	**-2231.12**	**4435.63**	<1E-16
	fly3=mice2bbc	-13.31	0.05	1.00	-1992.65	3958.68	<1E-16
	fly2=mice3bgc	-13.31	0.06	1.00	-2784.95	5543.28	<1E-16
	fly2=mice2bgc	-13.31	0.06	1.00	-2559.72	5092.83	<1E-16

[1]The top panel denotes the unconstrained model where each species is allowed to have its own s^+ (or $\boldsymbol{\gamma}^+$) and p^+. The s^+ (or $\boldsymbol{\gamma}^+$) and p^+ columns denote the maximum likelihood estimates (MLEs) of these parameters. The "abbr." gives the abbreviation for this demographic model used in the lower portion of the table. For example, "human3mac" means human Complex model using macaque as outgroup species.

[2]The bottom panel denotes the constrained model where s^+ (or $\boldsymbol{\gamma}^+$) and p^+ were constrained to be the same across taxa. For example, "human3=fly3=mice3" means we constrained Complex models of human, *D. melanogaster* and mouse to have the same s^+ (or $\boldsymbol{\gamma}^+$) and p^+. "Sum of LL" denotes the sum of the log-likelihoods across species for the full models listed in the second column. "Constrained LL" denotes the log-likelihood of the constrained model listed across the relevant species. "Likelihood ratio" denotes the difference in log-likelihood between the full and constrained models. *P*-values assume that twice the likelihood ratio is asymptotically distributed following a chi-square distribution.

Additional References

Burgess R, Yang Z. 2008. Estimation of hominoid ancestral population sizes under bayesian coalescent models incorporating mutation rate variation and sequencing errors. *Mol Biol Evol* **25**: 1979–1994.

Chen F-C, Li W-H. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am J Hum Genet* **68**: 444–456.

Fischer A, Wiebe V, Pääbo S, Przeworski M. 2004. Evidence for a complex demographic history of chimpanzees. *Mol Biol Evol* **21**: 799–808.

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS Genet* **5**: e1000695.

Hara Y, Imanishi T, Satta Y. 2012. Reconstructing the demographic history of the human lineage using whole-genome sequences from human and three great apes. *Genome Biol Evol* **4**: 1133–1145.

Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J, et al. 2007. Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science* **316**: 240–243.

Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, 1000 Genomes Project Consortium, Sella G, Przeworski M. 2011. Classic selective sweeps were rare in recent human evolution. *Science* **331**: 920–924.

Hobolth A, Christensen OF, Mailund T, Schierup MH. 2007. Genomic relationships and speciation times of human, chimpanzee, and gorilla inferred from a coalescent hidden Markov model. *PLoS Genet* **3**: e7.

Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. *Genome Res* **21**: 349–356.

Huber CD, Kim BY, Marsden CD, Lohmueller KE. 2017. Determining the factors driving selective effects of new nonsynonymous mutations. *Proc Natl Acad Sci* **114**: 4465–4470.

Hudson RR. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* **18**: 337–338.
Hvilsom C, Carlsen F, Heller R, Jaffré N, Siegismund HR. 2014. Contrasting demographic histories of the neighboring bonobo and chimpanzee. *Primates* **55**: 101–112.

Hvilsom C, Qian Y, Bataillon T, Li Y, Mailund T, Sallé B, Carlsen F, Li R, Zheng H, Jiang T, et al. 2012. Extensive X-linked adaptive evolution in central chimpanzees. *Proc Natl Acad Sci* **109**: 2054–2059.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* **488**: 471.

Lachance J, Tishkoff SA. 2014. Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. *Am J Hum Genet* **95**: 408–420.

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

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* **499**: 471–475.

Ruvolo M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol Biol Evol* **14**: 248–265.

Sawyer SA, Hartl DL. 1992. Population genetics of polymorphism and divergence. *Genetics* **132**: 1161–1176.

Schrago CG. 2014a. The effective population sizes of the anthropoid ancestors of the human–chimpanzee lineage provide insights on the historical biogeography of the great apes. *Mol Biol Evol* **31**: 37–47.

Schrago CG. 2014b. The limiting distribution of the effective population size of the ancestor of humans and chimpanzees. *J Theor Biol* **357**: 55–61.

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

Wall JD. 2003. Estimating ancestral population sizes and divergence times. *Genetics* **163**: 395–404.

Xue C, Raveendran M, Harris RA, Fawcett GL, Liu X, White S, Dahdouli M, Rio Deiros D, Below JE, Salerno W, et al. 2016. The population genomics of rhesus macaques (*Macaca mulatta*) based on whole-genome sequences. *Genome Res* **26**: 1651–1662.

Yang Z. 2002. Likelihood and bayes estimation of ancestral population sizes in hominoids using data from multiple loci. *Genetics* **162**: 1811–1823.