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]

[1]Department of Ecology and Evolutionary Biology, UCLA; [2]School of Life Sciences, Westlake University, China; [3]Institute of Biology, Westlake Institute for Advanced Study, China; [4]Genetics and Genomic Biology and The Centre for Applied Genomics, Hospital for Sick Children, Canada; [5]Department of Human Genetics, UCLA

Does the amount of adaptive evolution vary across species?

Previous studies have shown differences in adaptive evolution across species; however, specific reasons why they differ remain elusive. Here, we use improved modeling of weakly deleterious mutations and the demographic history of the outgroup species and ancestral population and estimate that at least 20% of nonsynonymous substitutions between humans and an outgroup species were fixed by positive selection, higher than previously estimated. Next, we directly estimate the proportion and selection coefficients (p^+ and s^+, respectively) of newly arising beneficial nonsynonymous mutations in humans, mice, and *Drosophila melanogaster*. We develop a novel composite likelihood framework to test whether these parameters differ across species. Overall, we reject a model with the same p^+ and s^+ of beneficial mutations across species, and estimate that humans have a higher p^+s^+ compared to *D. melanogaster* and mice.

α is commonly used to report the amount of adaptive evolution across species

- α : % of nonsynonymous substitutions that are adaptive

Under neutrality: $\frac{D_N}{D_S} = \frac{p_N}{p_S}$

when α of fixed divergence is due to positive selection then $\alpha = 1 - \frac{p_N D_S}{p_S D_N}$

Here divergence (*D*), polymorphism (*p*), nonsynonymous (*N*), synonymous (*S*)

- **Estimates of α vary tremendously across species** – tending to be higher in insects, but much lower in primates and plants. It's not clear why.
- α statistic is influenced by other factors, such as fixation of weakly deleterious mutations, thus **sizes of outgroup and ancestral populations could potentially influence the estimate of α.**

Approaches:

- **Correct for the sizes of outgroup and ancestral populations to infer α**
- **Directly estimate p^+ and s^+ or γ^+ of strongly beneficial nonsynonymous mutations**
- **Test whether p^+ and s^+ or γ^+ differ across taxa**

α in primates is higher than previously estimated

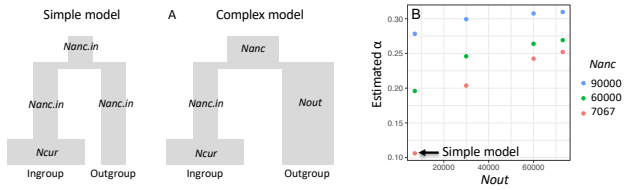

Figure 1. The ancestral and outgroup population sizes greatly influence α.

(A). In the Simple model, the size of the ancestral population as well as the size of the outgroup are assumed to be the same as the ancestral size of the ingroup (*Nanc.in*). This assumption is relaxed in the Complex model. (B). Effect of *Nanc* and *Nout* on estimates of α for humans using the chimpanzee as an outgroup. Arrow points to the estimate of α from the Simple model where *Nanc* = *Nout* = *Nanc.in* = 7067.

			Method of Inference					
	Species	Outgroup	MK-method: all	MK-method: MAF>20%	Model-based: Simple Model	Model-based: Complex Model	DFEα: α	DFEα: omega_a
Full data	Human	Chimpanzee	-0.41	-0.01	0.11	0.25	0.24	0.07
	Human	Macaque	-0.70	-0.22	-0.08	0.26	0.02	0.01
	Human lineage	-	-	-	0.06	0.16	-	-
	D. melanogaster	*D. simulans*	-0.13	0.49	0.53	0.60	0.71	0.14
	Mice	Rat	0.25	0.40	0.45	0.41	0.51	0.12
SSWW only	Human	Chimpanzee	-0.37	0.08	0.13	0.26	-	-
	Mice	Rat	-0.14	0.08	0.20	0.10	-	-

Table 1. Estimates of α using different methods.

For the human-chimpanzee comparison, using the more realistic Complex demographic model, when *Nout* = 30,000 (Prado-Martinez et al. 2013; Hvilsom et al. 2012; Fischer et al. 2004) and *Nanc* = 60,000 (Chen and Li 2001; Hobolth et al. 2007; Prado-Martinez et al. 2013), α is approximately 25% - much higher than previous estimates, implying that there is a greater contribution of positive selection to nonsynonymous divergence than previously appreciated.

Directly estimate p^+ and s^+ of strongly beneficial nonsynonymous mutations

For humans, mice, and *D. melanogaster*, we estimate the proportion of strongly beneficial mutations (p^+) and their selection coefficient (s^+) for each species.

The number of nonsynonymous differences between a pair of species (D_N) is assumed to be Poisson-distributed (Sawyer and Hartl 1992), with rate parameter equal to:

$$E[D_N] = 2N\mu\left[\int G(s)u(s)\,(1-p^+) + u(s^+)p^+\right]$$

DFE of deleterious mutations; Prob fixation of delet. mutation; Proportion of delet. mutations; Prob fixation of beneficial mutations; Proportion of beneficial mutations

p^+ and s^+ differ across species

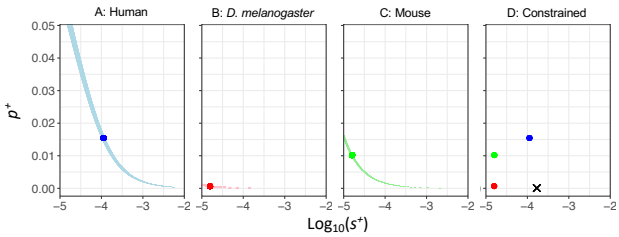

Figure 2. Log-likelihood surfaces for p^+ and s^+ for different species. The Complex model is used for each species and we use the chimpanzee as the outgroup for humans.

We find that the full model H1, where each taxon is allowed to have its own p^+ and s^+, fits D_N significantly better than the constrained null model, where p^+ and s^+ are constrained to be the same across all three taxa (LRT statistic Λ=124,974, df=4, $P<10^{-16}$)

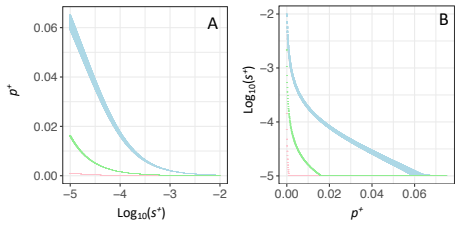

Figure 3. Conditional log-likelihood surfaces. (A) Maximizing p^+ given particular values of s^+, and (B) maximizing s^+ given particular values of p^+. Only grid points within 3 LL units of the MLEs for each parameter for each species are shown. Light blue - human, pink - *D. melanogaster*, and light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

The composite parameter p^+s^+ differs across species

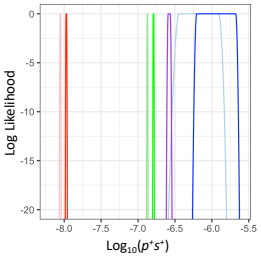

Figure 4. Log-likelihood surface of p^+s^+ for different species.

Regardless of which demographic model is used, the log-likelihood curves from the different species do not overlap within the top 500 log-likelihood units, suggesting p^+s^+ is significantly different among taxa.

Red - *D. melanogaster*, green - mouse, blue -human using the chimpanzee as the outgroup, and purple -human using the macaque 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.

Conclusions

- **At least 20% of nonsynonymous divergence between human and an outgroup is fixed by positive selection, much higher than previous estimates.**
- **A realistic demographic model, particularly correcting for sizes of outgroup and ancestral populations, is critical for this inference.**
- **Proportion and selection coefficient (p^+ and s^+, respectively) of newly arising beneficial nonsynonymous mutations differ across primates, rodents, and *Drosophila*.**
- **Humans have a significantly higher p^+s^+ than do *D. melanogaster* and mice.**
- **Selection coefficient scaled by current population size ($\gamma^+=2Ns^+$) and p^+ differ across primates, rodents, and *Drosophila* (data not shown).**
- **These results are robust to the choice of outgroups, biased gene conversion, and hypermutable CpG sites.**

Acknowledgements

We thank Lawrence Uricchio and David Enard for advice, discussions, and Tanya Phung, Jazlyn Mooney, Clare Marsden and Jesse Garcia for helpful comments on our manuscript. This work was supported by a Searle Scholars Fellowship and NIH Grant R35GM119856 (to K.E.L.) and foundation of Westlake University and NSFC 31900315 (to Y.Z.). We acknowledge support from a QCB Collaboratory Postdoctoral Fellowship to Y.Z. and the QCB Collaboratory Community directed by Matteo Pellegrini.

Email: zhenying@westlake.edu.cn

BioRxiv: https://doi.org/10.1101/427583