Estimating the distribution of fitness effects in aye-ayes (*Daubentonia madagascariensis*), accounting for population history as well as mutation and recombination rate heterogeneity

Vivak Soni[1], Cyril J. Versoza[1], Susanne P. Pfeifer[1,†], and Jeffrey D. Jensen[1,†]

[1] Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ, USA

[†] co-corresponding authors; jointly supervised the project

(susanne@spfeiferlab.org; jeffrey.d.jensen@asu.edu)

Keywords

primate; strepsirrhine; evolutionary genomics; population genetics; selection coefficients; distribution of fitness effects

ABSTRACT

The distribution of fitness effects (DFE) characterizes the range of selection coefficients from which new mutations are sampled, and thus holds a fundamentally important role in evolutionary genomics. To date, DFE inference in primates has been largely restricted to haplorrhines, with limited data availability leaving the other suborder of primates, strepsirrhines, largely under-explored. To advance our understanding of the population genetics of this important taxonomic group, we here map exonic divergence in aye-ayes (*Daubentonia madagascariensis*) – the only extant member of the Daubentoniidae family of the Strepsirrhini suborder. We further infer the DFE in this highly-endangered species, utilizing a recently published high-quality annotated reference genome, a well-supported model of demographic history, as well as both direct and indirect estimates of underlying mutation and recombination rates. The inferred distribution is generally characterized by a greater proportion of deleterious mutations relative to humans, providing evidence of a larger long-term effective population size. In addition however, both immune-related and sensory-related genes were found to be amongst the most rapidly evolving in the aye-aye genome.

MAIN

The distribution of fitness effects (DFE) summarizes the range of selection coefficients from which new mutations are sampled. Consequently, characterizing the DFE holds a fundamentally important role in evolutionary genomics, as it quantifies the fraction of neutrally evolving genomic mutations, provides insights into the expected relative frequencies of purifying relative to positive selection, and informs the expected effects of selection at linked sites, to name but a few implications (see reviews of Eyre-Walker and Keightley 2007; Keightley and Eyre-Walker 2010; Bank et al. 2014a). Moreover, given that the vast majority of fitness-impacting mutations are deleterious, the constant elimination of these variants via purifying selection and the associated background selection (BGS) effects (Charlesworth et al. 1993) represent constantly operating processes shaping levels and patterns of genomic variation in and around functional regions. As such, an accurate characterization of these effects is critical for the construction of any evolutionary baseline model for a given species (Comeron 2014, 2017; Johri et al. 2022a; Howell et al. 2023; Terbot et al. 2023; Soni et al. 2023; Soni and Jensen 2024), and, because these effects may differ strongly depending on the relative proportion of weakly relative to strongly deleterious mutations, the DFE shape again emerges as a fundamental component for any evolutionary modeling or inference (Charlesworth et al. 1993; Hudson and Kaplan 1994; Charlesworth et al. 1995; Ewing and Jensen 2014, 2016; Johri et al. 2020).

Generally speaking, there are two classes of DFE inference, one applicable to lab-tractable organisms that may be experimentally evolved, and one applicable to natural population analysis. The former includes mutation accumulation experiments – in which a

population of organisms can be maintained often in replicate, sampled at regular intervals, and the fitness effects of newly arising mutations characterized with respect to, for example, the wild-type state (e.g., Lenski et al. 1991; Barrick and Lenski 2013; Desai 2013; Böndel et al. 2019; Morales-Arce et al. 2022; Crombie et al. 2024). This class also includes mutagenesis experiments – in which hundreds or thousands of individuals can be maintained that carry one or very few mutations, and their fitness assessed by, for example, relative growth rates (e.g., Hietpas et al. 2011; Jacquier et al. 2013; Bank et al. 2014b; Fowler and Fields 2014; Matuszewski et al. 2015). Both methods represent powerful DFE inference approaches for the organisms in which they can be applied (e.g., *Saccharomyces cerevisiae*, *Caenorhabditis elegans, Chlamydomonas reinhardtii*), with the caveat being that they provide DFE inference only within the context of a lab-grown environment.

With regards to natural population analysis, which will be our focus here, there are generally approaches utilizing divergence data, polymorphism data, or a combination of both. Perhaps the most basic approach utilized to infer aspects of the DFE relies on comparisons between non-synonymous and synonymous divergence. Assuming that synonymous sites are effectively neutral, and thus characterized by a substitution rate equal to their mutation rate (Kimura 1968), one may quantify the fraction of non-synonymous mutations that are deleterious (and thus characterized by reduced fixation probabilities relative to neutrality) by assessing the depression in non-synonymous divergence relative to the synonymous neutral standard (e.g., Eyre-Walker et al. 2002). Similarly, if advantageous mutations are present (characterized by increased fixation probabilities relative to neutrality), one may assess this fraction of the DFE via the acceleration of non-synonymous divergence relative to synonymous

(e.g., Smith and Eyre-Walker 2002). These advances largely owed to the realization that a McDonald-Kreitman-style test (McDonald and Kreitman 1991) could be used to infer proportions of adaptive substitutions (Charlesworth 1994). Synonymous and non-synonymous mutations aside, one may similarly utilize this divergence-based logic to assess selective constraints acting in different genomic regions (e.g., coding relative to intronic relative to intergenic; Andolfatto 2005).

When incorporating polymorphism data into DFE inference, one initial challenge is the need to incorporate the demographic history of the population into the inference procedure, given that this history may also act to shape levels and patterns of variation and thus may potentially result in mis-inference if unaccounted for (see review of Johri et al. 2022b). One of the first advances in this regard utilized the site frequency spectrum (SFS) at putatively neutral synonymous or non-coding sites in order to infer a population history, and then conditioned on that history to infer the DFE at putatively functional non-synonymous sites (Williamson et al. 2005; Keightley and Eyre-Walker 2007). Such step-wise approaches yielded some of the first polymorphism-based DFE estimates for a variety of organisms (Eyre-Walker and Keightley 2007; Boyko et al. 2008; Eyre-Walker and Keightley 2009; Schneider et al. 2011). A related category of methods also arose for utilizing time-sampled polymorphism data in order to infer individual mutational effects based on observed allele frequency changes – as may be applicable to ecological datasets or ancient DNA sampling – with the stochastic effects of genetic drift associated with the given population history being incorporated by estimating an effective population size based on the variance observed in neutral allele frequencies (e.g., Malaspinas et al. 2012; Foll et al. 2015; Ferrer-Admetlla et al. 2016; and see review of Malaspinas 2016).

However, in addition to generally being limited to relatively simple population-size change models (though more complex models have been developed; e.g., Ma et al. 2013; Kim et al. 2017), these initial single- and multi-timepoint approaches also assume independence amongst sites, and thus neglect any role of background selection or other forms of genetic hitchhiking in further shaping levels of polymorphism (see reviews of Charlesworth and Jensen 2021, 2022).

In order to address these polymorphism-based challenges, simultaneous inference approaches have recently been developed. Though accounting for the effects of selection on linked sites within an analytical framework remains challenging, Cvijovic et al. (2018) obtained expressions for the SFS at sites experiencing BGS in a constant size population, and Friedlander and Steinrücken (2022) described a numerical framework to obtain expected SFS and linkage disequilibrium (LD) patterns around a selected region with changing population size. In order to allow for more complex models, progress has also been made using approximate Bayesian computation (ABC) approaches with forward simulations, in order to model both complex population histories and flexible DFE shapes, whilst accounting for the resulting effects of selection on linked sites. For example, Johri et al. (2020) developed a joint ABC approach estimating the DFE densities of neutral, weakly deleterious, moderately deleterious, and strongly deleterious mutations, together with a history of population size change, utilizing aspects of the SFS, LD, and divergence as summary statistics. Notably, the exclusion of BGS effects in previous methods was found to result in an under-estimation of weakly deleterious mutations and an over-estimation of population growth, a bias which is corrected within this ABC framework (Johri et al. 2021). Subsequent work has also demonstrated the potential mis-inference that may arise by neglecting underlying heterogeneity in rates of both mutation and

104 recombination (Soni et al. 2024a). Taken together, this literature thus emphasizes the
105 importance of incorporating population history, the effects of selection at linked sites, and
106 mutation and recombination rate maps / uncertainties when performing DFE inference.

107 In primates specifically, these various divergence- and polymorphism-based approaches
108 have been employed widely, with humans being the best studied in this regard. For example,
109 Keightley and Eyre-Walker (2007) fit a gamma-distributed DFE utilizing a gene set associated
110 with severe disease or inflammatory response, and estimated a large proportion (~40%) of
111 strongly deleterious mutations and a relatively low (~20%) proportion of effectively neutral
112 mutations. Huber et al. (2017) utilized a wider selection of genes, resulting in a DFE skewed
113 towards effectively neutral mutations (~50%; similar to the estimate of Johri et al. 2023 utilizing
114 a different subset of genes), and a smaller proportion of strongly deleterious mutations (~20%).
115 Thus, these differences may well simply and accurately reflect true DFE differences in the
116 underlying gene sets evaluated. Similar inference has also been performed across the great apes
117 (e.g., Castellano et al. 2019; Tataru and Bataillon 2020), and considerations have been extended
118 to general regulatory regions as well (e.g., Simkin et al. 2014; Anderson et al. 2020; Kuderna et
119 al. 2024).

120 Notably however, owing largely to data availability, these estimates have been
121 performed primarily in haplorrhines, with the other suborder of primates, strepsirrhines, being
122 largely unexplored. Yet, a number of recent advances have uniquely enabled investigation in this
123 neglected space of the primate clade. Firstly, Versoza & Pfeifer (2024) have recently provided an
124 annotated chromosome-level genome assembly for aye-ayes (*Daubentonia madagascariensis*),
125 thereby allowing for the essential demarcation of functional and non-function genomic regions

126 needed for performing DFE inference. Secondly, recent work has also generated high-quality
127 direct mutation and recombination rates for aye-ayes from multi-generation pedigree data
128 (Versoza et al. 2024a,b; Versoza, Lloret-Villas, et al. 2024) – as well as indirect fine-scale
129 estimates based on autosomal patterns of LD and neutral divergence (Soni, Versoza, et al. 2024).
130 Finally, utilizing high-coverage whole-genome data from unrelated individuals, Terbot et al.
131 (2024) recently estimated a well-fitting population history for aye-ayes (and see Soni et al.
132 2024b) – which described a severe and ancient population size decline likely associated with the
133 human colonization of Madagascar, as well as a more recent decline likely associated with
134 habitat destruction – thereby providing the needed accounting of the role of population history
135 in shaping observed SFS across the genome.
136
137 Interpreting exonic divergence
138 Building upon these recent results, we replaced the original aye-aye genome in the 447-
139 way mammalian multiple species alignment (Zoonomia Consortium 2020) that includes
140 hundreds of closely related primate species (Kuderna et al. 2023) with the high-quality aye-aye
141 reference genome of Versoza and Pfeifer (2024) to quantify fine-scale exonic divergence in the
142 species. Figure 1 summarizes observed exonic divergence on the aye-aye branch, with the
143 maximum neutral divergence for 1kb and 1Mb windows (Soni, Versoza, et al. 2024) provided for
144 orientation. Although a considerable number of exons were characterized by rates of fixation
145 greater than the maximum neutral divergence observed in 1Mb windows, no exons were found
146 to be in excess of the neutral rate observed in 1kb windows – the more appropriate comparison
147 given that the mean exonic length is <1kb. Thus, exonic divergence was observed to be lower

than the maximum neutral divergence in aye-ayes without exception, as expected from the dominant action of purifying selection in functional regions (Charlesworth et al. 1993).

However, given that even recurrent positive selection is still expected to be rare relative to purifying selection, the absence of entire exons evolving faster than neutrality does not itself eliminate the possibility of positive selection contributing to exonic divergence in subsets of genes. For example, distinct classes of exons were found to occupy the tails of the exonic divergence distribution, and were found to be in excess of the mean neutral fixation rate. Utilizing the genome annotations from the Versoza and Pfeifer (2024) reference genome to calculate the mean divergence per gene, we ran a gene functional analysis using g:Profiler (Kolberg et al. 2023) on all coding regions with a mean divergence greater than the 75th percentile of neutral divergence (0.039694). Figure 2 provides the divergence distribution of all examined exons, compared with the distributions of the two fastest-evolving gene classes – those related to sensory and immune function.

Immune-related genes have long been observed to be amongst the most rapidly evolving across vertebrates, as populations continually respond to challenges of pathogen exposure (e.g., George et al. 2011; Rausell and Telenti 2014), and our results remain consistent with this pattern. With regards to the sensory-related distribution, both the nocturnal activity patterns of aye-ayes (with the suggestion previously being made that dichromacy may enable aye-ayes to perceive color whilst foraging in moonlight conditions; Perry et al. 2007), together with evidence that aye-ayes may discriminate between individuals based on scent (Price and Feistner 1994) and use scent-marking to attract mates (Winn 1994), both suggest potentially significant roles for opsin- and olfactory-related genes throughout the evolutionary history of

the species. Furthermore, Soni et al. (2024b) recently found that a number of sensory functional categories including G-protein coupled receptors and olfactory receptors had strong statistical support for being maintained by long-term balancing selection in aye-ayes – noting that diversity in these genes may increase the number of different odorant-binding sites (Lancet 1994) – further supporting these hypotheses.

Utilizing patterns of exonic divergence to infer the DFE

Divergence is an informative summary statistic when inferring patterns of long-term selection, as the general features of the DFE are likely to remain relatively stable over deep evolutionary time. As such, we ran forward-in-time simulations in SLiM4.0.1 (Haller and Messer 2023) in order to fit observed empirical exonic divergence with a DFE shape consisting of neutral, weakly deleterious, moderately deleterious, and strongly deleterious mutational classes. In brief, we simulated a 54.9 million year divergence time of the aye-aye branch (Horvath et al. 2008), assuming a generation time of 5 years (Ross 2003; Louis et al. 2020) – both of which have additionally been recently supported by whole-genome neutral divergence patterns (Soni, Versoza, et al. 2024) – and utilized the estimated demographic model of Terbot et al. (2024) in order to characterize the recent history of the species. Using our multiple-species alignment to compute the number of divergent sites along the aye-aye branch, we were able to directly compare this empirical observation with the number of fixations accrued in our simulated population during the divergence phase.

As depicted in Figure 3, observed divergence was fit well by a DFE of new mutations characterized by a majority of nearly neutral variants, and a remaining even mix of moderately

and strongly deleterious variants. For comparison, a recent estimate of the DFE from human populations (Johri et al. 2023) has also been included. As shown, humans were characterized by a higher density of neutral variants and a lower density of more strongly deleterious variants relative to aye-ayes, likely consistent with the smaller effective population size of the former. Notably however, this inference in aye-ayes assumes a mutation rate of 0.4e-8/site/generation, as was directly inferred from pedigree data (Versoza et al. 2024a). Yet, this pedigree inference was for the youngest parents in the study (9-11 years of age), and a strong parental age effect was observed. Namely, the oldest parents in the study (24-26 years of age) were characterized by a rate of 2.0e-8/site/generation, with an average rate across the pedigree of 1.1e-8/site/generation. Given that aye-ayes reach sexual maturity by 3 years of age (Winn 1994), reproduction in the wild likely occurs amongst individuals even younger than the youngest in the pedigree, and given support for the 0.4e-8/site/generation rate from recent indirect divergence-based inference (Soni, Versoza, et al. 2024), we believe this to be a reasonable estimate for our conversion here. However, if the true rate were to be even lower owing to parents being generally younger throughout the evolutionary history of the species, the inferred DFE would resultingly become more skewed towards nearly neutral variants and thus potentially more similar to the human estimate. It is also noteworthy that long-term mutation rates on the aye-aye branch higher than 0.4e-8/site/generation become very difficult to reconcile with the fossil-record (see Tavaré et al. 2002; Soni, Versoza, et al. 2024), consistent with the suggestion of generally lower rates in prosimians relative to other primates (see the reviews of Tran and Pfeifer 2018; Chintalapati and Moorjani 2020).

Concluding summary

We have here characterized functional divergence in aye-ayes, finding, as expected, that exonic divergence is generally much reduced relative to neutral divergence. Yet, amongst exons, we also found evidence of an increased rate of divergence in genes related to immune and sensory-related functions, in agreement with previous work across vertebrates for the former, and in primates more specifically for the latter. Employing forward simulations to fit a DFE to observed exonic divergence, we found evidence of an increased proportion of newly arising deleterious variants in aye-ayes relative to humans, likely related to their larger estimated effective population size. These findings also generally support a relatively low mutation rate in aye-ayes compared to other primates, as has been proposed both from indirect neutral divergence as well as from direct pedigree-based inference.

ACKNOWLEDGEMENTS

We would like to thank the Duke Lemur Center for providing the aye-aye samples used in this study, and members of the Jensen Lab and Pfeifer Lab for helpful discussion. Computations were performed on the Sol supercomputer at Arizona State University (Jennewein et al. 2023) and on the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science. This is Duke Lemur Center publication # XXXX.

FUNDING

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM151008 to SPP and the National Science Foundation under Award Number DBI-2012668 to the Duke Lemur Center. VS and JDJ were supported by National Institutes of Health Award Number R35GM139383 to JDJ. CJV was supported by the National Science Foundation CAREER Award DEB-2045343 to SPP. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

REFERENCES

Anderson JA, Vilgalys TP, Tung J. 2020. Broadening primate genomics: new insights into the ecology and evolution of primate gene regulation. Curr Opin Genet Dev. 62:16–22.

Andolfatto P. 2005. Adaptive evolution of non-coding DNA in *Drosophila*. Nature. 437(7062):1149–1152.

Armstrong J, et al. 2020. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature. 587(7833):246–251.

Bank C, Ewing GB, Ferrer-Admetlla A, Foll M, Jensen JD. 2014a. Thinking too positive? Revisiting current methods of population genetic selection inference. Trends Genet. 30(12):540–546.

Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD. 2014b. A Bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: uncovering the potential for adaptive walks in challenging environments. Genetics. 196(3):841–852.

Barrick JE, Lenski RE. 2013. Genome dynamics during experimental evolution. Nat Rev Genet. 14(12):827–839.

Böndel KB, et al. 2019. Inferring the distribution of fitness effects of spontaneous mutations in *Chlamydomonas reinhardtii*. PLoS Biol. 17(6):e300192.

Boyko AR, et al. 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. PLoS Genet. 4(5):e1000083.

Castellano D, Macià MC, Tataru P, Bataillon T, Munch K. 2019. Comparison of the full distribution of fitness effects of new amino acid mutations across the great apes. Genetics. 213(3):953–966.

Charlesworth B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet Res. 63(3):213–227.

Charlesworth D, Charlesworth B, Morgan MT. 1995. The pattern of neutral molecular variation under the background selection model. Genetics. 141(4):1619–1632.

Charlesworth B, Jensen JD. 2021. Effects of selection at linked sites on patterns of genetic variability. Annu Rev Ecol Evol Syst. 52:177–197.

Charlesworth B, Jensen JD. 2022. How can we resolve Lewontin's Paradox? Gen Biol Evol. 14(7):evac096.

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

Chintalapati M, Moorjani P. 2020. Evolution of the mutation rate across primates. Curr Opin Genet Dev. 62:58–64.

Comeron JM. 2014. Background selection as a baseline for nucleotide variation across the *Drosophila* genome. PLoS Genet. 10(6):e1004434.

Comeron JM. 2017. Background selection as null hypothesis in population genomics: insights and challenges from *Drosophila* studies. Philos Trans R Soc Lond B Biol Sci. 372(1736):20160471.

Crombie TA, et al. 2024. Direct inference of the distribution of fitness effects of spontaneous mutations from recombinant inbred *Caenorhabditis elegans* mutation accumulation lines. Genetics. 228(2):iyae136.

Cvijovic I, Good BH, Desai MM. 2018. The effect of strong purifying selection on genetic diversity. Genetics. 209(4):1235-1278.

Desai MM. 2013. Statistical questions in experimental evolution. J Stat Mech. 2013(01):P01003.

Ewing G, Jensen JD. 2014. Distinguishing neutral from deleterious mutations in growing populations. Front Genet. 5:7.

Ewing G, Jensen JD. 2016. The consequences of not accounting for background selection in demographic inference. Mol Ecol. 25(1):135–141.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat Rev Genet. 8(8):610–618.

Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. Mol Biol Evol. 26(9):2097–2108.

Eyre-Walker A, Keightley PD, Smith NGC, Gaffney D. 2002. Quantifying the slightly deleterious model of molecular evolution. Mol Biol Evol. 19(12):2142–2149.

Ferrer-Admetlla A, Leuenberger C, Jensen JD, Wegmann W. 2016. An approximate Markov model for the Wright-Fisher diffusion and its application to time series data. Genetics. 203(2):831–846.

326 Foll M, Shim H, Jensen JD. 2015. A Wright-Fisher ABC-based approach for inferring per-site
327 effective population sizes and selection coefficients from time-sampled data. Mol Ecol
328 Res. 15(1):87–98.
329
330 Fowler DM, Fields S. 2014. Deep mutational scanning: a new style of protein science. Nat
331 Methods. 11(8):801–807.
332
333 Friedlander E, Steinrücken M. 2022. A numerical framework for genetic hitchhiking in
334 populations of variable size. Genetics 220(3):iyac012.
335
336 George RD, et al. 2011. Trans genomic capture and sequencing of primate exomes reveals new
337 targets of positive selection. Genome Res. 21(10):1686–1694.
338
339 Hietpas RT, Jensen JD, Bolon DNA. 2011. Experimental illumination of a fitness landscape. Proc
340 Natl Acad Sci USA. 108(19):7896–7901.
341
342 Haller BC, Messer PW. 2023. SLiM 4: multispecies eco-evolutionary modeling. Am Nat.
343 201(5):E127–E139.
344
345 Hickey G, Paten B, Earl D, Zerbino D, Haussler D. 2013. HAL: a hierarchical format for storing and
346 analyzing multiple genome alignments. Bioinformatics. 29(10):1341–1342.
347
348 Howell AA, et al. 2023. Developing an appropriate evolutionary baseline model for the study of
349 human cytomegalovirus. Gen Biol Evol. 15(4):evad059.
350
351 Horvath JE, et al. 2008. Development and application of a phylogenomic toolkit: resolving the
352 evolutionary history of Madagascar's lemurs. Genome Res. 18(3):489–499.
353
354 Huber CD, Kim BY, Marsden CD, Lohmueller KE. 2017. Determining the factors driving selective
355 effects of new nonsynonymous mutations. Proc Natl Acad Sci USA. 114(17):4465–4470.
356
357 Hudson RR, Kaplan NL. 1994. Gene trees with background selection. In: Non-neutral evolution:
358 theories and molecular data. London, UK: Chapman and Hall. p. 140–153.
359
360 Jacquier H, Birgy A, Le Nagard H, Tenaillon O. 2013. Capturing the mutational landscape of the
361 beta-lactamase TEM-1. Proc Natl Acad Sci USA. 110(32):13067–13072.
362
363 Jennewein DM, et al. 2023. The Sol Supercomputer at Arizona State University. Practice and
364 Experience in Advanced Research Computing, 296–301.
365
366 Johri P, et al. 2022a. Recommendations for improving statistical inference in population
367 genomics. PLoS Biol. 20(5):e3001669.
368

Johri P, Charlesworth B, Jensen JD. 2020. Towards an evolutionarily appropriate null model jointly inferring demography and purifying selection. Genetics. 215(1):173–192.

Johri P, Eyre-Walker A, Gutenkunst RN, Lohmueller KE, Jensen JD. 2022b. On the prospect of achieving accurate joint estimation of selection with population history. Gen Biol Evol. 14(7):evac088.

Johri P, et al. 2021. The impact of purifying and background selection on the inference of population history: problems and prospects. Mol Biol Evol. 38(7):2986–3003.

Johri P, Pfeifer SP, Jensen JD. 2023. Developing an evolutionary baseline model for humans: jointly inferring purifying selection with population history. Mol Biol Evol. 40(5):msad100.

Keightley PD, Eyre-Walker A. 2007. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. Genetics. 177(4):2251–2261.

Keightley PD, Eyre-Walker A. 2010. What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? Philos Trans R Soc Lond B Biol Sci. 365(1544):1187–1193.

Kim BY, Huber CD, Lohmueller KE. 2017. Inference of the distribution of selection coefficients for new nonsynonymous mutations using large samples. Genetics. 206(1):345–361.

Kimura M. 1968. Evolutionary rate at the molecular level. Nature. 217(5129):624–626.

Kolberg L, et al. 2023. g:Profiler – interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). Nucleic Acids Res. 51(W1):W207–W212.

Kuderna LFK, et al. 2023. A global catalog of whole-genome diversity from 233 primate species. Science. 380(6648):906–913.

Kuderna LFK, et al. 2024. Identification of constrained sequence elements across 239 primates. Nature. 625(7996):735–742.

Lancet D. 1994. Olfaction: exclusive receptors. Nature. 372(6504):321–322.

Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. Am Nat. 138:1315–1341.

Louis E, et al. 2020. *Daubentonia madagascariensis.* IUCN Red List Threat. Species.

411 Ma X, et al. 2013. Population genomic analysis of ten genomes reveals a rich speciation and
412 demographic history of orang-utans (*Pongo pygmaeus* and *Pongo abelii*). PLoS One.
413 8(10):e77175.
414
415 Malaspinas A-S. 2016. Methods to characterize selective sweeps using time serial samples: an
416 ancient DNA perspective. Mol Ecol. 25(1):24–41.
417
418 Malaspinas A-S, Malaspinas O, Evans SN, Slatkin M. 2012. Estimating allele age and selection
419 coefficient from time-serial data. Genetics. 192(2):599–607.
420
421 Matuszewski S, Hildebrandt ME, Ghenu AH, Jensen JD, Bank C. 2015. A statistical guide to the
422 design of deep mutational scanning experiments. Genetics. 204(1):77–87.
423
424 McDonald JD, Kreitman M. 1991. Adaptive evolution at the Adh locus in *Drosophila*. Nature.
425 351(6328):652–654.
426
427 Morales-Arce AY, Johri P, Jensen JD. 2022. Inferring the distribution of fitness effects in patient-
428 sampled and experimental virus populations: two case studies. Heredity. 128(2):79–87.
429
430 Perry GH, Martin RD, Verrelli BC. 2007. Signatures of functional constraint at aye-aye opsin
431 genes: the potential of adaptive color vision in a nocturnal primate. Mol Biol Evol.
432 24(9):1963–1970.
433
434 Price EC, Feistner ATC. 1994. Responses of captive aye-ayes (*Daubentonia madagascariensis*) to
435 the scent of conspecifics: a preliminary investigation. Folia Primatologica. 62(1–3):170–
436 174.
437
438 Rausell A, Telenti A. 2014. Genomics of host-pathogen interactions. Curr Opin Immunol. 30:32–
439 38.
440
441 Ross C. 2003. Life history, infant care strategies, and brain size in primates. In: Kappeler PM,
442 Pereira ME, editors. Primate life histories and socioecology. Chicago (IL): Chicago
443 University Press; pp. 266–284.
444
445 Schneider A, Charlesworth B, Eyre-Walker A, Keightley PD. 2011. A method for inferring the rate
446 of occurrence and fitness effects of advantageous mutations. Genetics. 189(4):1427–1437.
447
448 Simkin A, Bailey JA, Gao FB, Jensen JD. 2014. Inferring the evolutionary history of primate
449 microRNA binding sites: overcoming motif counting biases. Mol Biol Evol. 31(7):1894–
450 1901.
451
452 Smith NGC, Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. Nature.
453 415(6875):1022–1024.

454 Soni V, Jensen JD. 2024. Inferring demographic and selective histories from population genomic
455 data using a two-step approach in species with coding-sparse genomes: an application to
456 human data. BioRxiv, preprint.
457 https://www.biorxiv.org/content/10.1101/2024.09.19.613979v2
458
459 Soni V, Johri P, Jensen JD. 2023. Evaluating power to detect recurrent selective sweeps under
460 increasingly realistic evolutionary null models. Evolution. 77(10):2113–2127.
461
462 Soni V, Pfeifer SP, Jensen JD. 2024a. The effect of mutation and recombination rate
463 heterogeneity on the inference of demography and the distribution of fitness effects. Gen
464 Biol Evol. 16(2):evae004.
465
466 Soni V, Terbot JW, Versoza CJ, Pfeifer SP, Jensen JD. 2024b. A whole-genome scan for evidence of
467 recent positive and balancing selection in aye-ayes (*Daubentonia madagascariensis*)
468 utilizing a well-fit evolutionary baseline model. BioRxiv, preprint.
469 https://www.biorxiv.org/content/10.1101/2024.11.08.622667v1
470
471 Soni V [*], Versoza C[*], Terbot J, Jensen JD, Pfeifer SP. 2024. Inferring fine-scale mutation and
472 recombination rate maps in aye-ayes (*Daubentonia madagascariensis*). BioRxiv, preprint.
473 https://www.biorxiv.org/content/10.1101/2024.12.28.630620v1
474
475 Tataru P, Bataillon T. 2020. polyDFE: inferring the distribution of fitness effects and properties of
476 beneficial mutations from polymorphism data. Methods Mol Biol. 2090:125–146.
477
478 Tavaré S, Marshall CR, Will O, Soligo C, Martin RD. 2002. Using the fossil record to estimate the
479 age of the last common ancestor of extant primates. Nature. 416(6882):726–729.
480
481 Terbot JW, Cooper BS, Good JM, Jensen JD. 2023. A simulation framework for modeling the
482 within-patient evolutionary dynamics of SARS-CoV-2. Gen Biol Evol. 15(11):evad204.
483
484 Terbot JW, Soni V, Versoza CJ, Pfeifer SP, Jensen JD. 2024. Inferring the demographic history of
485 aye-ayes (*Daubentonia madagascariensis)* from high-quality, whole-genome, population-
486 level data. Accepted, Gen Biol Evol. BioRxiv, preprint.
487 https://www.biorxiv.org/content/10.1101/2024.11.08.622659v1
488
489 Tran LA, Pfeifer SP. 2018. Germ line mutation rates in old world monkeys. In: John Wiley & Sons,
490 Ltd, editor. eLS. 1st ed. Wiley. p. 1–10.
491
492 Versoza CJ, Ehmke E, Jensen JD, Pfeifer SP. 2024a. Characterizing the rates and patterns of *de*
493 *novo* germline mutations in the aye-aye (*Daubentonia madagascariensis*). Accepted
494 pending revision, Mol Biol Evol. BioRxiv, preprint.
495 https://www.biorxiv.org/content/10.1101/2024.11.08.622690v1
496

497 Versoza CJ, Jensen JD, Pfeifer SP. 2024b. The landscape of structural variation in aye-ayes
498 (*Daubentonia madagascariensis)*. BioRxiv, preprint.
499 https://www.biorxiv.org/content/10.1101/2024.11.08.622672v1
500
501 Versoza CJ [*], Lloret-Villas A [*], Jensen JD, Pfeifer SP. 2024. A pedigree-based map of crossovers
502 and non-crossovers in aye-ayes (*Daubentonia madagascariensis)*. BioRxiv, preprint.
503 https://www.biorxiv.org/content/10.1101/2024.11.08.622675v1
504
505 Versoza CJ, Pfeifer SP. 2024. A hybrid genome assembly of the endangered aye-aye
506 (*Daubentonia madagascariensis*). G3 (Bethesda). 14(10):jkae185.
507
508 Williamson SH, Hernandez R, Fledel-Alon A, Bustamante CD. 2005. Simultaneous inference of
509 selection and population growth from patterns of variation in the human genome. Proc
510 Natl Acad Sci USA. 102(22):7882–7887.
511
512 Winn RM. 1994. Preliminary study of the sexual behaviour of three aye-ayes (*Daubentonia*
513 *madagascariensis*) in captivity. Folia Primatologica. 62(1-3):63–73.
514
515 Zoonomia Consortium. 2020. A comparative genomics multitool for scientific discovery and
516 conservation. Nature. 587(7833):240–245.

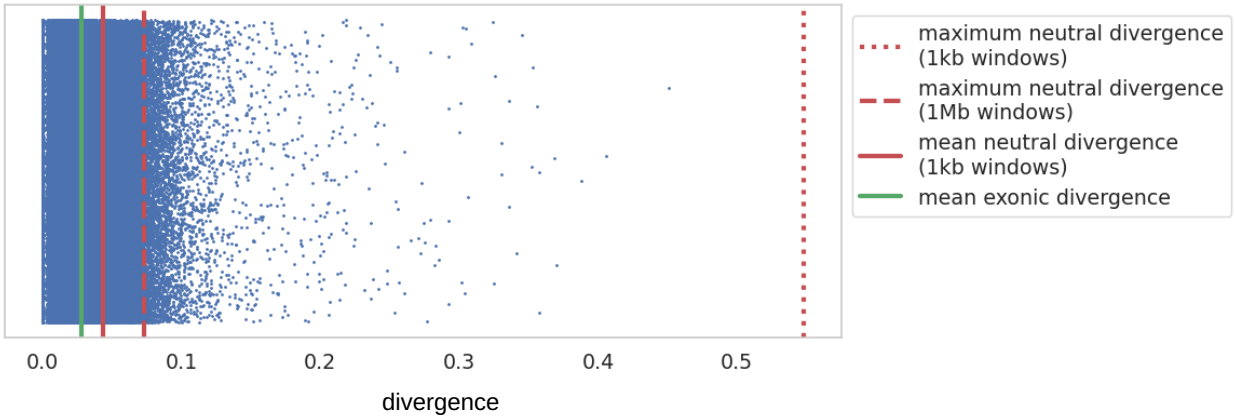

Figure 1: Exonic divergence scatter plot with maximum neutral divergence values marked for windows of size 1Mb (red dashed line) and 1kb (red dotted line), as well as the mean neutral divergence for 1kb windows (red solid line), as calculated from Soni, Versoza, et al. (2024). Exonic divergence was calculated by replacing the aye-aye in the 447-way mammalian multiple species alignment (Zoonomia Consortium 2020; Kuderna et al. 2023) with the high-quality aye-aye reference genome of Versoza and Pfeifer (2024), via the Cactus alignment software (Armstrong et al. 2020), and extracting divergent sites along the aye-aye branch using the HAL software package (Hickey et al. 2013). Exonic divergence was calculated by dividing the number of divergent sites in each exon by the total accessible exonic length, and only exons with a minimum of 100bp of accessible sites were included. Each dot represents an autosomal exon, and the mean exonic divergence is plotted (green solid line).

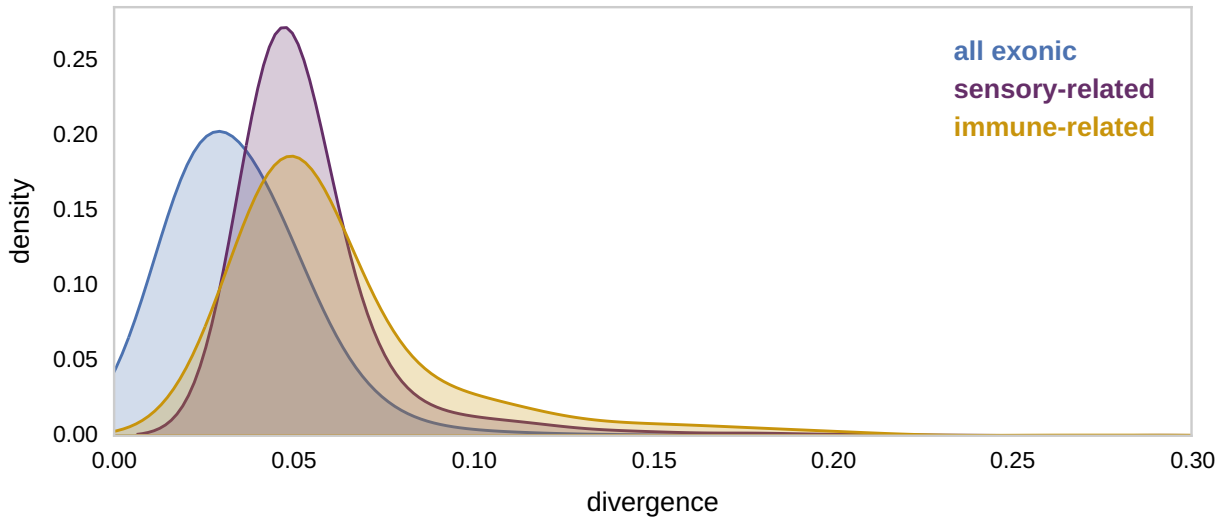

Figure 2: Density plots of exonic divergence in aye-ayes for all exons (blue), exons located in genes implicated in sensory-related functions (purple), and exons located in genes implicated in immune-related functions (gold). Utilizing the Versoza and Pfeifer (2024) genome annotations, the mean divergence per gene was calculated, and gene functional analysis was performed using g:Profiler (Kolberg et al. 2023) on the set of genes with a divergence value greater than the 75th percentile of neutral divergence in aye-ayes (0.0397; Soni, Versoza, et al. 2024).

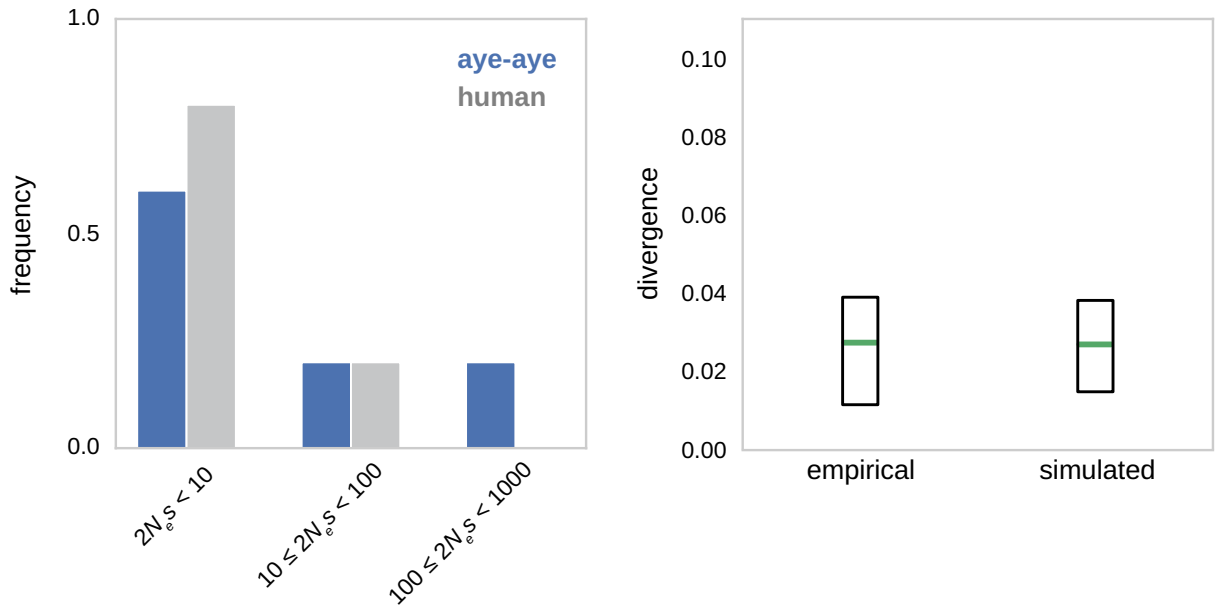

Figure 3: Comparison of the empirical and simulated divergence under the best-fitting DFE. Simulations were run in SLiM4.01 (Haller and Messer 2023) under the Terbot et al. (2024) aye-aye demographic model, assuming 54.9 million years since the branch split (Horvath et al. 2008) and a generation time of 5 years (Ross 2003; Louis et al. 2020). Simulations included a $10N_{ancestral}$ generation burn-in time prior to the demographic model (where $N_{\text{ancestral}}$ is the initial population size and s the reduction in fitness of the mutant homozygote relative to wild type). For this example, the simulated exon length was 2,978bp (i.e., the mean empirical length of exons of size greater than 1kb). Mutation and recombination rate heterogeneity was modelled such that each simulated exon had a rate drawn from a normal distribution, with the mean rate across all 100 simulated exons equal to the mean pedigree estimated rates of 0.4e-8/bp/generation and 0.85cM/Mb for mutation and recombination, respectively (Versoza et al. 2024; Versoza, Lloret-Villas et al. 2024). Exonic mutations were drawn from a DFE comprised of four fixed classes (following Johri et al. 2020), denoted by f_0 with $0 \leq 2N_{\text{ancestral}}\, s < 1$ (i.e., effectively neutral mutations), f_1 with $1 \leq 2N_{\text{ancestral}}\, s < 10$ (i.e., weakly deleterious mutations), f_2 with $10 \leq 2N_{\text{ancestral}}\, s < 100$ (i.e., moderately deleterious mutations), and f_3 with $100 \leq 2N_{\text{ancestral}}\, s$ (i.e., strongly deleterious mutations). **Left panel:** Best-fitting discrete DFE in ayes-ayes (blue), as compared to the Johri et al. (2023) DFE inferred for humans (grey). **Right panel:** Comparison of empirical and simulated divergence values for the best-fitting DFE. Green lines represent the mean value, whilst boxes represent 25 and 75 percentiles.