Article: *Research Article*

Pedigree-based measurement of the de novo mutation rate in the gray mouse lemur reveals a high mutation rate, few mutations in CpG sites, and a weak sex bias

Authors: C. Ryan Campbell[1*], George P. Tiley[1], Jelmer W. Poelstra[1], Kelsie E. Hunnicutt[1,†,], Peter A. Larsen[1,‡], Mario dos Reis[2], & Anne D. Yoder[1]

[1]Department of Biology, Duke University, Durham, NC 27708, USA

[2]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK

[†]Present address: Department of Biological Sciences, University of Denver, Denver, CO 80208

[‡]Present address: Department of Veterinary and Biomedical Sciences, University of Minnesota, St. Paul, MN 55108

*Author for Correspondence: C. Ryan Campbell, Department of Evolutionary Anthropology, Duke University, Durham, USA, c.ryan.campbell@duke.edu

Abstract

Spontaneous germline mutations are the raw material on which evolution acts and knowledge of their frequency and genomic distribution is therefore crucial for understanding how evolution operates at both long and short timescales. At present, the rate and spectrum of de novo mutations have been directly characterized in only a few lineages, and it is therefore critical to examine a wide range of species to determine the generality of patterns that have been identified so far. Our study provides the first direct mutation rate estimate for strepsirrhine primates (i.e., the lemurs and lorises) which comprise nearly half of the primate clade. Using high-coverage linked-read sequencing of a family pedigree (n = 8) of gray mouse lemurs (*Microcebus murinus*), we estimate the mutation rate to be 1.64×10^{-8} (95% CI: 1.41×10^{-8} to 1.98×10^{-8}) mutations per basepair per generation. This estimate is higher than those for most previously characterized mammals, including other primates. Contrary to expectation, we found only a modest overrepresentation of mutations at CpG-sites and of paternal mutations. Comparing mutation rates across taxa, we show that expectations based on the drift barrier hypothesis are met at a broad phylogenetic scale but not within primates. Finally, we compared pedigree-based mutation rates with phylogenetically-based substitution rate estimates for mouse lemurs and six other primate lineages and found considerable differences between the two rate estimates. This finding has implications for divergence-time estimation and calls for further study.

Introduction

Spontaneous germline mutations (SGMs) are single basepair errors that occur in DNA transmission as it is passed from parent to offspring in sexually reproducing organisms. The accrual of these errors provides not only the raw material for evolution but can also serve as a measure of the evolutionary time across phylogeny (Zuckerkandl and Pauling 1965; Kimura and Ohta 1971; Langley and Fitch 1974). The rate at which these mutations are introduced into genomes is thus a crucial metric of evolution at the genomic level as well as a measure of fundamental biological processes (Kondrashov and Kondrashov 2010). And by further parsing this measure by genomic regions, we gain an intimate perspective of intragenomic evolutionary patterns. Finally, by quantifying the variation in mutation rates across the tree of life, we can develop increasingly refined hypotheses regarding the relationship between mutation rates and life history characteristics.

Approaches for estimating mutation rates in vertebrates generally fall into one of two categories: phylogenetically-based (indirect) versus pedigree-based (direct) estimation. While phylogenetically-based methods were the standard before the genomics revolution of the past decade, new sequencing technologies have enabled pedigree-based mutation rate measurements for previously uncharacterized species. By comparing the genomes of individuals with known genealogical relationships – typically, parent to offspring – investigators are putatively able to quantify mutations in real time (Scally and Durbin 2012; Smeds, et al. 2016; Feng, et al. 2017; Pfeifer 2017; Thomas, et al. 2018; Koch, et al. 2019). Phylogenetically-based approaches, on the other hand, are dependent on external calibrations to obtain substitution rates in units of absolute time. In a typical study, divergence between homologous genomic regions across a phylogeny is calibrated with fossils or other external criteria at specific nodes in order to determine substitution rates (Thorne, et al. 1998; Sanderson 2002; Thorne and Kishino 2002; Drummond, et al. 2006; Yang 2006), relying on the assumption that the rate at which substitutions accumulate between species is equal to the mutation rate for neutral sites (Kimura 1983). Thus, in principle, phylogenetically-based and pedigree-based methods should produce equivalent estimates of the mutation rate.

Phylogenetically-based methods are known to suffer from various sources of uncertainty including violation of the molecular clock, inaccuracies in external calibration points, incomplete lineage sorting, and the difficulties of recovering multiple overlapping changes (i.e., "multiple hits") at any given site (dos Reis, et al. 2018). Although a number of solutions to these problems have been proposed (Heath, et al. 2014; Ogilvie, et al. 2017), some limitations are difficult to overcome (Magallóan and Sanderson 2005; Near, et al. 2005; Herrera and Dávalos 2016). Pedigree-based mutation rate estimates are not affected by the same uncertainties, however, and allow for the characterization of mutation rate variation across the genome and the various biases in the mutation spectrum (Ségurel, et al. 2014; Harris, et al. 2017). Previously, these estimates have relied on well-assembled genomes available only to model organisms (Scally and Durbin 2012; Venn, et al. 2014; Uchimura, et al. 2015; Jónsson, et al. 2017), and have therefore been limited in taxonomic scope (for a list of mammals, see *Table 1*). Fortunately, recent genome assembly strategies (Dudchenko, et al. 2017) have increasingly enabled chromosome-level assemblies of non-model organisms (Larsen, et al. 2017) and pedigree-based mutation rate estimation is now feasible for virtually any species, as long as related individuals with known pedigrees are available (Smeds, et al. 2016; Feng, et al. 2017; Pfeifer 2017; Martin, et al. 2018; Koch, et al. 2019). Pedigree-based mutation rate estimates from across the tree of life have the potential to enhance our understanding of mutation rate evolution (Lynch 2010), including the relationship between mutation rate and genome architecture (Kim, et al. 2006).

Even so, pedigree-based studies are not without challenges. Perhaps most challenging for estimating the number of de novo mutations in offspring is that the mutation rate is several orders of magnitude lower than the sequencing error rate, even for the most accurate sequencing methods. Thus, pedigree-based estimates of the SGM rate must employ stringent bioinformatic filters to distinguish between true biological mutations and sequencing error. And herein lies one of the method's most significant challenges: stringency can be so vigorously applied that many true SGMs are not called (i.e., false negatives are common), and somewhat counterintuitively, the mutation rate can be under- rather than overestimated.

97 Here, we utilize two strategies for keeping false negative rates low while ensuring the validity
98 of our called mutations. First, linked short reads from 10x Genomics (Weisenfeld, et al. 2017)
99 provide increased mapping scores and confidence in calling individual variants when coupled
100 with mutation rate estimation methods (Long, et al. 2016; Winter, et al. 2018), especially in
101 repeat-rich mammalian genomes (Chaisson, et al. 2015). Additionally, the phasing information
102 provided by linked reads can determine the parent of origin through just two generations of
103 sequencing. Phased haplotypes with known parental origin then allow us to assign individual
104 mutations as coming from either the maternal or paternal germline. To control for false
105 negatives, we explicitly evaluate the callable proportion of the genome by introducing known
106 false mutations to the sequencing data for one individual and subsequently testing how
107 accurate our bioinformatic pipeline was in recovering these errors (Keightley, et al. 2014; Xie,
108 et al. 2016). We refer to this approach as "allele drop", which can be used to correct observed
109 SGM rates regardless of the variant calling pipeline.
110
111 We applied these advances in sequencing technology and computational methods to produce
112 the first pedigree-based estimated mutation rate for the gray mouse lemur (*Microcebus*
113 *murinus*), the first strepsirrhine primate to be characterized in a pedigree-based SGM study.
114 The gray mouse lemur is a member of a radiation of morphologically cryptic primates
115 distributed throughout Madagascar (Hotaling, et al. 2016), for which an accurate temporal
116 context is needed to, for instance, for comparing the timing of speciation with historical climatic
117 events. Previous divergence time studies have had to rely on either distant external fossil
118 calibrations (Yang and Yoder 2003; dos Reis, et al. 2018) or by using pedigree-based mutation
119 rate estimates from distant relatives (Yoder, et al. 2016).
120
121 By measuring the mutation rate in mouse lemurs with a pedigree-based approach, we aim to
122 simultaneously expand our knowledge of mutation rate variation across lineages and to
123 facilitate the estimation of divergence times within this specific radiation. To do so, we deeply
124 sequenced a family pedigree of gray mouse lemurs (*Microcebus murinus*, n = 8), including a
125 mother, father, and two offspring, to accurately identify and count SGMs and to assign the
126 parent-of-origin of de novo mutations.

Results and Discussion

Mutation Rate

To identify de novo mutations, we used DeNovoGear (Ramu, et al. 2013) and validated the inferred mutations using an independent mutation identification analysis with VarScan2 (Koboldt, et al. 2009). In total, we assessed 4,542,770 potential variants across eight related individuals to discover 134 de novo mutations in two focal offspring, one male and one female (*Fig 1*). Among these 134 mutations, 125 (71 in the female and 63 in the male) were located on autosomes and nine (seven in the female and two in the male) were located on the X chromosome. The average depth of coverage for the 134 mutations was 170 reads (170.2, SD=79.95). We estimate the mutation rate in this family quartet to be 1.64×10^{-8} single nucleotide mutations per basepair per generation per individual. The single-base mutation-rate estimate is a weighted average of the number of mutations on the autosomes (125 mutations) and X chromosome (nine mutations) divided by the number of sites where a mutation was detectable given our methods (Long, et al. 2016; Winter, et al. 2018).

To estimate the number of detectable sites, we artificially generated 1,000 mutations placed randomly across the entire genome within the sequence data of the offspring, similar to previous efforts to account for false negative results (Keightley, et al. 2014; Xie, et al. 2016). On the autosomes, we detected 801 of 952 generated mutations, while 39 of 48 generated mutations were detected on the X chromosome. Therefore, we estimate our detection rate to be 84.1% on autosomes and 81.3% on the X chromosome, which yields a total of 2.088 billion detectable sites (out of a total of 2.487 billion).

To calculate the confidence interval around the estimated mutation rate, we assumed that the total number of mutations inherited by an offspring follows a Poisson distribution (*Supplemental Text 1*). Given that an average of 67 mutations were found in each of the focal offspring, the 95% confidence interval is 56.5 to 79.1 mutations per genome, or 1.41×10^{-8} to 1.98×10^{-8} mutations per site per generation. To calculate the expected number of false positive and false negative results, we sequenced the male offspring twice and compared variant presence and absence across the technical replicates (*Supplemental Text 2*). We

158 calculated 6.46 false positives and 37.51 false negatives from the total of 134 de novo
159 mutations and 2.088 billion potential mutation sites.
160
161 ## Relatively low numbers of mutations at CpG sites
162 CpG sites have consistently been found to have higher mutation rates relative to other site
163 classes, a pattern discovered using sequence comparisons made several decades ago (Bird
164 1980) and ascribed to the frequent deamination of methylated cytosines (Friedberg, et al.
165 2005). We found a total of six mutations at parental CpG sites, constituting 4.5% of all de novo
166 mutations, a more than two-fold overrepresentation given that 1.9% of the genome consists of
167 CpG sites. This pattern is much less pronounced than in other studies, were CpG site
168 mutations have been shown to contribute 12-25% of total mutations, representing an
169 overrepresentation of least ten-fold given the number of CpG sites in the genome (Venn, et al.
170 2014; Smeds, et al. 2016; Thomas, et al. 2018; Besenbacher, et al. 2019; Gao, et al. 2019).
171 Given the limited sample size of our study, and the small fraction of the genome this
172 represents, however, caution is warranted in interpreting this result. Confidence intervals
173 calculated following the genome-wide Poisson distribution example (as in *Supplemental Text*
174 *1*) suggest that the true number of CpG site mutations in the mouse lemur genome could be as
175 high as 10.9 (over 8% of total mutations and a 4.25-fold increase). Additionally, if the true CpG
176 site mutation rate in mouse lemurs was such that 10% of total mutations were found there, the
177 values measured in this study would still fall within the 95% confidence interval (13.4 CpG site
178 mutations, 95% confidence interval 6.0 to 21).
179
180 The elevated mutation rate at CpG sites is linked to methylation, and is therefore not expected
181 in regions of the genome with high GC content (CpG islands), where CpG sites are much less
182 likely be methylated. As expected, none of the 134 de novo mutations were found within CpG
183 islands, which constitute roughly 4% of the *Microcebus murinus* genome.
184 The discrepancy in the relative frequency of mutations at CpG sites between mouse lemurs
185 and other studied species could have arisen from a high proximity of CpG sites to genic
186 regions, where mutations are more likely to be detrimental. A lower substitution rate in
187 proximity to genes has been observed empirically (Narang and Wilson Sayres 2016).
188 Accordingly, we investigated this possibility, but found a normal distance between genes and

189 CG sites (*Fig S1*). Also to be considered, the mouse lemur genomes here compared are of
190 lesser overall quality than human or chimpanzee, and a lower mapping quality in CG-rich
191 regions could be affecting our ability to detect CpG site mutations.
192
193 Finally, the low number of CpG site mutations has the statistical effect of impacting additional
194 mutational metrics: transition-transversion ratio, strong-to-weak and weak-to-strong ratio, and
195 potentially, sex-bias.
196
197 ## Mutation spectrum
198 We characterized two aspects of the mutation spectrum in mouse lemurs. First, the ratio of
199 transitions to transversions (Ti:Tv) was estimated to be 1.03 (68 transitions and 66
200 transversions). Second, the ratio of strong-to-weak mutations (SW; C/G > A/T) to weak-to-
201 strong mutations (WS; A/T > C/G), SW:WS, was also estimated to be 1.03 (31 SW and 30 WS
202 mutations). The most common two categories of de novo mutation type were G>A and C>T
203 (*Fig 2*).
204
205 Our estimates of the Ti:Tv and SW:WS are lower than values found in other animals. For
206 instance, values of Ti:Tv in previous pedigree-based studies varied between 1.97 and 2.67
207 (Agier and Fischer 2011; Kong, et al. 2012; Venn, et al. 2014; Smeds, et al. 2016; Assaf, et al.
208 2017; Thomas, et al. 2018; Besenbacher, et al. 2019). Relatively low values for these ratios
209 are expected in light of the low CpG mutation rate reported above, given that C>T mutations at
210 CpG sites are strong-to-weak transitions. For instance, without an elevation in the mutation
211 rate at CpG sites, the Ti:Tv and SW:WS ratios would drop to 1.55 and 1.33 respectively, in a
212 study of chimpanzees (Venn, et al. 2014). Thus, it is difficult to differentiate the effect of
213 reduced CpG site mutation on the other metrics of the mutation spectrum.
214
215 ## Reduced Male Mutational Bias
216 A male mutational bias has long been hypothesized for diploid sexually reproducing organisms
217 based on the idea that the increased number of cell divisions in sperm versus egg should lead
218 to higher numbers of mutations in the male germline than the female germline (Haldane 1946;
219 Kong, et al. 2012; Lindsay, et al. 2018). Indeed, a strong male mutation rate bias has been

220 observed in the vast majority of pedigree-based mutation rate estimates to date (Venn, et al.
221 2014; Rahbari, et al. 2015; Lindsay, et al. 2018; Thomas, et al. 2018; Gao, et al. 2019) and in
222 many studies of phylogenetically-based rates (Shimmin, et al. 1993; Ellegren and Fridolfsson
223 1997; Axelsson, et al. 2004; Zhang 2004; Goetting-Minesky and Makova 2006).
224
225 Using the long phasing blocks generated by the linked-read method, we were able to
226 determine the parent of origin for 94 out of 134 (70%) de novo mutations. Thus, the number of
227 mutations confidently assigned to a parent are notably higher in our analysis compared to
228 previous studies that used short read sequencing alone, such as 35% (Venn, et al. 2014) or
229 38% (Thomas, et al. 2018). In our study, among the assigned mutations, 54% (n = 51) were
230 found on the offspring's paternal haplotype while the remaining 46% (n = 43) were found on
231 the offsprings' maternal haplotype, a ratio of male-to-female mutations of approximately 1.2.
232
233 This ratio of male-to-female mutations in gray mouse lemur is therefore considerably less
234 pronounced than that observed in other primate species (2.1-5.5) (Venn, et al. 2014; Rahbari,
235 et al. 2015; Thomas, et al. 2018; Besenbacher, et al. 2019) and mouse (2.7) (Lindsay, et al.
236 2018). However, it is identical to the ratio found in collared flycatchers (Smeds, et al. 2016).
237 One of the driving factors of the male mutational bias is likely the time of first reproduction after
238 puberty (Ségurel, et al. 2014), and given that mouse lemurs experience puberty and time of
239 first reproduction nearly simultaneously (Blanco 2011; Zohdy, et al. 2014; Blanco, et al. 2015),
240 this may explain our low ratio of male-to-female mutations. Additionally, there are stark
241 differences in the methylation process within male and female germline cells, with male cells
242 experiencing markedly more methylation (Reik and Dean 2001; Kobayashi, et al. 2013). This
243 discrepancy yields more methylation-related mutations in males than females as mammals
244 age (Gao, et al. 2019). Thus, in the absence of these methylation-focused mutations, and with
245 a relatively short time to puberty, the mouse lemur would have multiple factors pointing to a
246 more balanced mutational sex ratio, as was observed. On the other hand, however, mouse
247 lemurs have exaggerated symptoms of sperm competition (Eberle and Kappeler 2004), large
248 testes size (Wong 2014), which might be expected to lead to elevated male mutation rates,
249 thus further complicating the interpretation the a low male bias evident in our study.
250

Relationship of Mutation Rate Estimates to Life History Traits

The drift barrier hypothesis states that effective population size may explain some variation in mutation rates across species due to a larger efficiency of selection acting on DNA replication fidelity in larger populations, especially across large phylogenetic distances (Lynch 2010; Sung, et al. 2012). Using an MSMC analysis (Schiffels and Durbin 2014), we estimated the effective population size of gray mouse lemur across the last 2 million years to be approximately 41,000 (*Fig S2*). When analyzed in a broad phylogenetic context, we find that the mouse lemur lies comfortably along the regression that shows a negative correlation between effective population size and per-generation mutation rate across animals (*Fig 3*). In stark contrast, this relationship is not upheld in a comparison with other primates with the two species with the largest N_e (orangutan and mouse lemur) also showing the highest mutation rate (*Fig 3*). It is possible that the drift barrier hypothesis does not operate across these relatively small timescales and differences in population size. Moreover, mouse lemurs are the only strepsirrhine primate represented by a pedigree-based analysis of mutation rate and are considerably diverged from the other primates analyzed with a shared common ancestor $\geq$ 60 million years ago (dos Reis, et al. 2018). Further insight into the relationship between mutation rate, effective population size, and other life history traits within primates thus awaits mutation rate estimates of additional species, especially within the strepsirrhine clade.

A negative correlation between per-year mutation rates and generation time has been long-observed in studies of molecular evolution (Wu and Li 1985; Ohta 1993; Li, et al. 1996; Goetting-Minesky and Makova 2006) and underlies the classic hominid slow-down hypothesis (Goodman 1962; Moorjani, et al. 2016). Compared to other primates, mouse lemurs have both a high per-year mutation rate and a short generation time (3.75 y) (Yoder, et al. 2016), consistent with this general expectation. Across primates, the relationship between per-year mutation rates and generation time is present but weak (*Fig 4*).

Discrepancies between pedigree-based and phylogenetically-based estimates

One revelation in the rapidly developing literature on SGM rates has been that the estimated de novo rate in humans is less than half that predicted by phylogenetic studies, prompting speculation that mutation rates may have changed rapidly among primates (Scally and Durbin

2012). This observation has created some alarm, especially with regard to the impacts on divergence age estimates for hominids (Besenbacher, et al. 2019). Therefore, we compared pedigree-based estimates of mutation rates for mouse lemurs and five other primates (humans *Homo sapiens*, chimpanzee *Pan troglodytes*, gorilla *Gorilla gorilla*, orangutan *Pongo abeli*, green monkey *Chlorocebus sabaeus*, and owl monkey *Aotus nancymaae*) with phylogenetically-based estimates from a recent analysis of primate divergence times (dos Reis, et al. 2018) (*Table S1*, *Fig 4*).

Phylogenetic substitution rates are estimated per-year, so we rescaled them by generation time for comparison with pedigree-based per-generation estimates of mutation rates. Although such estimates introduce uncertainty from generation time estimation as well as fossil calibration error, there are two notable observations from the comparison: 1) point estimates for pedigree-based and phylogenetically-based estimates differ considerably for all species, with pedigree-based estimates falling outside of the confidence intervals of phylogenetically-based estimates in 4 out of 7 cases; 2) the direction of disagreement varies across lineages: whereas pedigree-based estimates are higher than phylogenetically-based estimates for mouse lemur, owl monkey, green monkey, and orangutan, the opposite is true for human, chimp, and gorilla. Thus, were it not for the result observed in orangutans, it would be tempting to speculate that phylogenetically-based methods have overestimated mutation rates for hominids.

There are several potential explanations for the widespread disagreement between pedigree-based and phylogenetically-based estimates. Phylogenetically-based estimates may be biased downwards if substitutions are not fully neutral. Even mutations at third-codon positions may be affected by weak negative selection and may therefore be less likely to be fixed at the population level. Indeed, previous studies have found evidence for low phylogenetically-based compared to pedigree-based estimates (Denver, et al. 2000; Howell, et al. 2003; Winter, et al. 2018). For pedigree-based estimates, the degree to which somatic mutations and/or inter-individual variation might impact these estimates is as yet unclear (Ségurel, et al. 2014). Additional analyses will be needed to reconcile the differences between pedigree-based and phylogenetically-based estimates of the mutation rate making this an area ripe for study.

313

314 ### Effect of mutation rate on divergence time mutation rate estimates

315 Using the estimated de novo mutation rate, we have recalculated branch lengths in absolute
316 time for a genus-level phylogeny of mouse lemurs (Yoder, et al. 2016). Applying the newly
317 estimated rate of 1.64×10^{-8}, divergence time estimates are calculated to be considerably
318 more recent compared to previous estimates (*Fig 5*), and the smaller confidence interval
319 reduces uncertainty in age estimates (*Table S2*). Similarly, a recent investigation of great ape
320 divergence times revealed that lineage-specific mutation rate estimates for chimp, gorilla, and
321 orangutan led to more recent estimates of common ancestor ages that more closely agreed
322 with fossil evidence than if the human mutation rate was assumed for all lineages
323 (Besenbacher, et al. 2019). Given the endangered status of mouse lemurs and virtually all
324 other strepsirrhine species, our ability to provide a temporal context to speciation and to
325 estimate demographic parameters such as effective population size may yield critical
326 information for directing conservation policy and efforts.

327

328 ## Conclusions

329 In this study, we provide a pedigree-based estimate of the de novo mutation rate for gray
330 mouse lemurs. The results underscore the importance of measuring mutation rate with linked-
331 read sequencing technology that increases mapping confidence, provides haplotypes for
332 determining parent-of-origin, thereby eliminating the need to sequence across more than two
333 generations. Furthermore, we have used allele-drop simulations to determine the appropriate
334 denominator for mutation rate calculations. Even so, despite our best efforts to develop an
335 accurate method for determining the de novo mutation rate, our results are not immune to
336 problems presented by small sample sizes, in particular those relating to the accurate recovery
337 of the mutation spectrum.

338

339 Our study also emphasizes the importance of measuring mutation rate across the tree of life in
340 order to study the nature and causes of mutation rate evolution. As noted above, this is the first
341 pedigree-based mutation rate estimate for a strepsirrhine primate, and as such, it is not clear
342 whether the unusually high mutation rate, low CpG mutation rate, and weak sex bias may be
343 representative of strepsirrhines. We show that N_e correlates with mutation rate at broad

phylogenetic scales but does not appear to do so within primates. Finally, we demonstrate widespread disagreement between phylogenetically-based and pedigree-based estimates of the mutation rate, suggesting that reconciling these methods should be a focus of future work. It is our hope that the details of our study will provide an encouraging pathway for that work.

Acknowledgments

We thank the Duke Lemur Center staff, especially Erin Ehmke, Bobby Schopler, and Cathy Williams, for providing the *Microcebus murinus* and *Mirza zaza* tissue samples. CRC thanks the members of his dissertation committee, Charlie Nunn, Tom Mitchell Olds, Jeff Thorne, and Marcy Uyenoyama for their insightful comments and suggestions for improvement throughout the course of this project as well as Jonathan Pritchard, Matthew Hahn, Priya Moorjani, and Felix Wu for helpful feedback during a poster presentation of this research. We thank Simon Gregory's lab for preparing the 10X Genomics libraries and are grateful for the support of Duke Research Computing and the Duke Data Commons (NIH 1S10OD018164-01). This study was funded by a National Science Foundation Grant DEB-1354610 and Duke University startup funds to ADY and she gratefully acknowledges support from the John Simon Guggenheim Foundation and the Alexander von Humboldt Foundation during the writing phase of this project. This is Duke Lemur Center publication no. XXX.

Materials and Methods

Samples

Eight individuals were selected from the DLC's mouse lemur colony consisting of a focal family of four, two parents and two offspring from separate litters, an additional half-sibling to the offspring, and three other individuals in the maternal lineage. Four of the eight selected samples were colony founders, which had been transferred from the mouse lemur colony in Brunois, Paris, France in 2003.

Blood and tissue samples were collected from these individuals during annual veterinary check-ups. High molecular weight DNA was extracted with the Qiagen MagAttract kit (Qiagen, Germantown, MD, USA) and 10X Genomics library preparation was performed at the Duke Molecular Genomics Core.

Sequencing

The eight individuals were used to produce nine sequencing libraries; every individual was sequenced once, and the focal paternal sample was prepared twice and sequenced as two separate samples. Libraries were sequenced at the Duke Center for Genomic Computational

379 Biology (GCB) Sequencing and Genomic Technology Shared Resource across two flowcells
380 (nine total lanes) of a HiSeq 4000. The run was 150 basepairs, paired end with an average
381 insert size of 554bp (range: 527-574bp). A single lane was run as a test of the 10x Genomics
382 LongRanger analysis software and was analyzed to confirm successful indexing and
383 preparation of the samples. After confirmation, the additional eight lanes (an entire flowcell)
384 were run. All data was combined for downstream analyses (SRA). 933,337,210,328 bases
385 were generated across nine samples and nine lanes.
386
387 10x Genomics Pipeline
388 Basecall (.bcl) files were demultiplexed and analysed using 10x Genomic's LongRanger v2.2.1
389 pipeline. Average genomic coverage after filtering was 34.5x across the nine samples. The
390 LongRanger pipeline was run on the Duke Scalable Computing Resource for approximately
391 2,000 hours (10 days of compute time per sample). Sequences were aligned to the latest gray
392 mouse lemur genome assembly (version 3.0, GCF_000165445.3) and variant calling was
393 handled by the Genomic Analysis Toolkit (v3.8), implemented within LongRanger v2.2.1 (Van
394 der Auwera, et al. 2013; Weisenfeld, et al. 2017). The mean N50 scaffold length, across
395 samples, generated within the 10x Genomics LongRanger alignment pipeline was 1.18Mb.
396
397 DeNovoGear
398 LongRanger alignments were used to find de novo mutations within the two focal family
399 offspring. Several methods were used to in finding mutations. First, the software package
400 DeNovoGear (Ramu, et al. 2013) was used to analyze the LongRanger variant call files (.vcf)
401 with default settings. As a follow-up, the software package VarScan2 was used with
402 LongRanger binary alignment files (.bam) to confirm that the de novo mutations found with
403 DeNovoGear could be confirmed.
404
405 Finally, de novo mutation hits were confirmed by checking independently with each replicate of
406 the sire. Hits were also checked for absence in all other pedigree individuals as well as
407 additionally sequenced samples from a diversity panel of gray mouse lemurs. The final list of
408 mutations was filtered for quality by offspring de novo quality (min DNQ of 100), offspring map
409 quality (min MQ of 50), and depth of coverage (min coverage of 20) in the parents.

410

411 ## Detectable Mutation Test

412 We conducted a test to determine the true number of sites at which a mutation would be
413 detectable, hereafter referred to as an “allele drop test” (Long, et al. 2016; Winter, et al. 2018).
414 This test increases the accuracy of the number of base pairs used in the denominator of the
415 mutation rate calculation. It estimates the proportion of sites at which a de novo mutation
416 would have been discoverable by our methods. This test consisted of adding 1,000 fake de
417 novo mutations into the pedigree with the software BAMsurgeon (Ewing, et al. 2015). These
418 mutations were added as heterozygotes, by changing half the aligned bases in the bam file at
419 a site to the non-reference allele. After this manipulation, the same methods were used to find
420 de novo mutations, and the results were mined for the 1,000 sites. By conducting this “allele
421 drop test”, we were able to estimate in which fraction of the genome de novo mutations would
422 have been found, were they to have occurred, or detectable sites. Going forward with the
423 number of detectable sites as the denominator in the mutation rate calculation yields a more
424 accurate measure of the rate than using total bases, total autosomal bases, or even
425 calculations based on a minimum depth of coverage because it produces an accurate picture
426 of where mutations can be found specifically via our methods.

427

428 ## CpG estimates

429 CpG islands were identified by two independent methods and compared to measure the
430 number of mutations within them. First, the software EMBOSS:cpgplot (Chojnacki, et al. 2017)
431 was used on the latest gray mouse lemur genome (3.0, GCF_000165445.3) to identify regions
432 that met the threshold of a CpG island (200 bp, over 50% CG content). Then, to confirm these
433 annotations a fasta file of CpG-annotated regions of gray mouse lemur genome 2.0
434 (GCF_000165445.2) was downloaded from the UCSC genome browser. A blast (Altschul, et
435 al. 1990) database of the latest gray mouse lemur genome (3.0, GCF_000165445.3) was
436 created and the annotations were queried to determine the coordinates of the 2.0 CpG
437 annotations in the 3.0 genome release used for mapping and assembly. While the
438 concordance between these methods was over 95% out of an abundance of caution only the
439 CpG islands (a total of 67,673 annotations) confirmed via both methods were used in the
440 analyses.

441

442 With these coordinates in-house bash were used to determine the location of each mutation
443 with respect to CpG islands and coding regions. To measure distribution of mutations relative
444 to the null expectation, a permutation test was conducted of the 1,000 false mutations
445 generated for the allele drop test. A set of de novo mutations were randomly sampled, without
446 replacement, to compare the distance to CpG islands with the true set.

447

448 ### Parent-of-origin

449 Custom in-house bash scripts were used in combination with the phased variant call files
450 produced by LongRanger to assign the mutations to a maternal or paternal chromosome. In
451 brief, these methods took input of the three family individuals and a mutation location. The
452 surrounding haplotype that contained the mutation was directly compared to the parental
453 haplotypes at the same location to determine a match. As these individuals are all related
454 members of a colony, dam and sire often shared similar haplotypes. Caution was used when
455 the mutation-bearing haplotype was found in both parents, resulting in less than 100% parent-
456 of-origin assignment of mutations.

457

458 ### BPP

459 Using BPP (Yang 2015) we reran data from a previous study (Yoder, et al. 2016) to measure
460 the effect of mutation rate estimates on divergence time estimation for the phylogeny of mouse
461 lemurs. We have written an R package, bppr (available at https://github.com/dosreislab/bppr),
462 for calibrating BPP estimated phylogenies to geological time using estimates of the mutation
463 rate. Using bppr, we estimated mouse lemur divergence times twice: (i) using the mutation rate
464 prior of Yoder et al. (2016), which was based on estimates of mouse and human mutation
465 rates, and (ii) using our new estimates of the de novo rate. While in the new version of BPP
466 (v4.0) the theta prior has changed from a gamma distribution to an inverse-gamma distribution,
467 the resulting tree has the same topology and essentially the same branch lengths as prior
468 publications (Yoder, et al. 2016).

469

470 ## References

471 Agier N, Fischer G. 2011. The Mutational Profile of the Yeast Genome Is Shaped by
472 Replication. Molecular Biology and Evolution 29:905-913.

473 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool.
474 Journal of Molecular Biology 215:403-410.

475 Assaf ZJ, Tilk S, Park J, Siegal ML, Petrov DA. 2017. Deep sequencing of natural and
476 experimental populations of Drosophila melanogaster reveals biases in the spectrum of new
477 mutations. Genome research 27:1988-2000.

478 Axelsson E, Smith NGC, Sundström H, Berlin S, Ellegren H. 2004. Male-Biased Mutation Rate
479 and Divergence in Autosomal, Z-Linked and W-Linked Introns of Chicken and Turkey.
480 Molecular Biology and Evolution 21:1538-1547.

481 Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T, Schierup MH. 2019. Author
482 Correction: Direct estimation of mutations in great apes reconciles phylogenetic dating. Nature
483 Ecology & Evolution 3:859-859.

484 Bird AP. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids
485 Research 8:1499-1504.

486 Blanco MB. 2011. Timely Estrus in Wild Brown Mouse Lemur Females at Ranomafana
487 National Park, Southeastern Madagascar. American Journal of Physical Anthropology
488 145:311-317.

489 Blanco MB, Rasoazanabary E, Godfrey LR. 2015. Unpredictable Environments, Opportunistic
490 Responses: Reproduction and Population Turnover in Two Wild Mouse Lemur Species
491 (Microcebus Rufus and M-Griseorufus) From Eastern and Western Madagascar. American
492 Journal of Primatology 77:936-947.

493 Chaisson MJP, Wilson RK, Eichler EE. 2015. Genetic variation and the de novo assembly of
494 human genomes. Nature Reviews Genetics 16:627.

495 Chojnacki S, Cowley A, Lee J, Foix A, Lopez R. 2017. Programmatic access to bioinformatics
496 tools from EMBL-EBI update: 2017. Nucleic Acids Research 45:W550-W553.

497 Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High Direct Estimate of the
498 Mutation Rate in the Mitochondrial Genome of <em>Caenorhabditis elegans</em&gt.
499 Science 289:2342.

500 dos Reis M, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z, Yoder AD. 2018. Using
501 Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on
502 Divergence Time Estimation: Primates as a Test Case. Systematic Biology 67:594-615.

503 Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed Phylogenetics and Dating
504 with Confidence. PLOS Biology 4:e88.

505 Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol
506 I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-
507 C yields chromosome-length scaffolds. Science 356:92-95.

508 Eberle M, Kappeler PM. 2004. Sex in the dark: determinants and consequences of mixed male
509 mating tactics in Microcebus murinus, a small solitary nocturnal primate. Behavioral Ecology
510 and Sociobiology 57:77-90.

511 Ellegren H, Fridolfsson A-K. 1997. Male–driven evolution of DNA sequences in birds. Nature
512 genetics 17:182-184.

513 Ewing AD, Houlahan KE, Hu Y, Ellrott K, Caloian C, Yamaguchi TN, Bare JC, P'ng C, Waggott
514 D, Sabelnykova VY, et al. 2015. Combining tumor genome simulation with crowdsourcing to
515 benchmark somatic single-nucleotide-variant detection. Nature Methods 12:623.

516 Feng C, Pettersson M, Lamichhaney S, Rubin CJ, Rafati N, Casini M, Folkvord A, Andersson
517 L. 2017. Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation
518 rate. Elife 6.

519 Friedberg EC, Walker GC, Siede W, Wood RD. 2005. DNA repair and mutagenesis: American
520 Society for Microbiology Press.

521 Gao Z, Moorjani P, Sasani TA, Pedersen BS, Quinlan AR, Jorde LB, Amster G, Przeworski M.
522 2019. Overlooked roles of DNA damage and maternal age in generating human germline
523 mutations. Proceedings of the National Academy of Sciences 116:9491.

524 Goetting-Minesky MP, Makova KD. 2006. Mammalian Male Mutation Bias: Impacts of
525 Generation Time and Regional Variation in Substitution Rates. Journal of molecular evolution
526 63:537-544.

527 Goodman M. 1962. Immunochemistry of the Primates and Primate Evolution. Annals of the
528 New York Academy of Sciences 102:219-234.

529 Haldane JBS. 1946. The Mutation Rate of the Gene for Haemophilia, and its Segregation
530 Ratios In Males and Females. Annals of Eugenics 13:262-271.

531 Harris K, Pritchard JK, Kimura M. 2017. Rapid evolution of the human mutation spectrum
532 On the evolutionary adjustment of spontaneous mutation rates. Elife 6:e24284.

533 Heath TA, Huelsenbeck JP, Stadler T. 2014. The fossilized birth–death process for coherent
534 calibration of divergence-time estimates. Proceedings of the National Academy of Sciences
535 111:E2957.

536 Herrera JP, Dávalos LM. 2016. Phylogeny and Divergence Times of Lemurs Inferred with
537 Recent and Ancient Fossils in the Tree. Systematic Biology 65:772-791.

538 Hotaling S, Foley ME, Lawrence NM, Bocanegra J, Blanco MB, Rasoloarison R, Kappeler PM,
539 Barrett MA, Yoder AD, Weisrock DW. 2016. Species discovery and validation in a cryptic
540 radiation of endangered primates: coalescent-based species delimitation in Madagascar's
541 mouse lemurs. Molecular Ecology DOI: 10.1111/mec.13604.

542 Howell N, Smejkal CB, Mackey DA, Chinnery PF, Turnbull DM, Herrnstadt C. 2003. The
543 Pedigree Rate of Sequence Divergence in the Human Mitochondrial Genome: There Is a
544 Difference Between Phylogenetic and Pedigree Rates. The American Journal of Human
545 Genetics 72:659-670.

546 Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT,
547 Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al. 2017. Parental influence on human
548 germline de novo mutations in 1,548 trios from Iceland. Nature 549:519.

549 Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey JW,
550 Jiggins CD. 2014. Estimation of the Spontaneous Mutation Rate in Heliconius melpomene.
551 Molecular Biology and Evolution 32:239-243.

552 Kim S-H, Elango N, Warden C, Vigoda E, Yi SV. 2006. Heterogeneous Genomic Molecular
553 Clocks in Primates. PLOS Genetics 2:e163.

554 Kimura M. 1983. The neutral theory of molecular evolution: Cambridge University Press.

555 Kimura M, Ohta T. 1971. On the rate of molecular evolution. Journal of molecular evolution
556 1:1-17.

557 Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T,
558 Suzuki Y, Ito T, et al. 2013. High-resolution DNA methylome analysis of primordial germ cells
559 identifies gender-specific reprogramming in mice. Genome research 23:616-627.

560 Koboldt DC, Larson DE, Mardis ER, Weinstock GM, Chen K, Ding L, McLellan MD, Wilson RK,
561 Wylie T. 2009. VarScan: variant detection in massively parallel sequencing of individual and
562 pooled samples. Bioinformatics (Oxford, England) 25:2283-2285.

563 Koch EM, Schweizer RM, Schweizer TM, Stahler DR, Smith DW, Wayne RK, Novembre J.
564 2019. De Novo Mutation Rate Estimation in Wolves of Known Pedigree. Molecular Biology and
565 Evolution.

566 Kondrashov FA, Kondrashov AS. 2010. Measurements of spontaneous rates of mutations in
567 the recent past and the near future. Philosophical Transactions of the Royal Society B:
568 Biological Sciences 365:1169-1176.

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

572 Langley CH, Fitch WM. 1974. An examination of the constancy of the rate of molecular
573 evolution. Journal of molecular evolution 3:161-177.

574 Larsen PA, Harris RA, Liu Y, Murali SC, Campbell CR, Brown AD, Sullivan BA, Shelton J,
575 Brown SJ, Raveendran M, et al. 2017. Hybrid de novo genome assembly and centromere
576 characterization of the gray mouse lemur (Microcebus murinus). BMC Biology 15:110.

577 Li W-H, Ellsworth DL, Krushkal J, Chang BHJ, Hewett-Emmett D. 1996. Rates of Nucleotide
578 Substitution in Primates and Rodents and the Generation–Time Effect Hypothesis. Molecular
579 Phylogenetics and Evolution 5:182-187.

580 Lindsay SJ, Rahbari R, Kaplanis J, Keane T, Hurles ME. 2018. Striking differences in patterns
581 of germline mutation between mice and humans. bioRxiv:082297.

582 Long H, Winter DJ, Chang AYC, Sung W, Wu SH, Balboa M, Azevedo RBR, Cartwright RA,
583 Lynch M, Zufall RA. 2016. Low Base-Substitution Mutation Rate in the Germline Genome of
584 the Ciliate Tetrahymena thermophila. Genome Biology and Evolution 8:3629-3639.

585 Lynch M. 2010. Evolution of the mutation rate. Trends Genet 26:345-352.

586 Magallóan SA, Sanderson MJ. 2005. Angiosperm Divergence Times: The Effect of Genes,
587 Codon Positions, and Time Constraints. Evolution 59:1653-1670.

588 Martin HC, Batty EM, Hussin J, Westall P, Daish T, Kolomyjec S, Piazza P, Bowden R,
589 Hawkins M, Grant T, et al. 2018. Insights into Platypus Population Structure and History from
590 Whole-Genome Sequencing. Molecular Biology and Evolution 35:1238-1252.

591 Moorjani P, Amorim CEG, Arndt PF, Przeworski M. 2016. Variation in the molecular clock of
592 primates. Proceedings of the National Academy of Sciences 113:10607.

593 Narang P, Wilson Sayres MA. 2016. Variable Autosomal and X Divergence Near and Far from
594 Genes Affects Estimates of Male Mutation Bias in Great Apes. Genome Biology and Evolution
595 8:3393-3405.

596 Near TJ, Bolnick DI, Wainwright PC. 2005. Fossil Calibrations and Molecular Divergence Time
597 Estimates in Centrarchid Fishes (Teleostei: Centrarchidae). Evolution 59:1768-1782.

598 Ogilvie HA, Bouckaert RR, Drummond AJ. 2017. StarBEAST2 Brings Faster Species Tree
599 Inference and Accurate Estimates of Substitution Rates. Molecular Biology and Evolution
600 34:2101-2114.

601 Ohta T. 1993. Amino acid substitution at the Adh locus of Drosophila is facilitated by small
602 population size. Proceedings of the National Academy of Sciences 90:4548.

603 Pfeifer SP. 2017. Direct estimate of the spontaneous germ line mutation rate in African green
604 monkeys. Evolution 71:2858-2870.

605 Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, Dominiczak A,
606 Morris A, Porteous D, Smith B, et al. 2015. Timing, rates and spectra of human germline
607 mutation. Nature genetics 48:126.

608 Ramu A, Noordam MJ, Schwartz RS, Wuster A, Hurles ME, Cartwright RA, Conrad DF. 2013.
609 DeNovoGear: de novo indel and point mutation discovery and phasing. Nature Methods
610 10:985.

611 Reik W, Dean W. 2001. DNA methylation and mammalian epigenetics. ELECTROPHORESIS
612 22:2838-2843.

613 Sanderson MJ. 2002. Estimating Absolute Rates of Molecular Evolution and Divergence
614 Times: A Penalized Likelihood Approach. Molecular Biology and Evolution 19:101-109.

615 Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding
616 human evolution. Nature Reviews Genetics 13:745-753.

617 Schiffels S, Durbin R. 2014. Inferring human population size and separation history from
618 multiple genome sequences. Nature genetics 46:919-925.

619 Ségurel L, Wyman MJ, Przeworski M. 2014. Determinants of Mutation Rate Variation in the
620 Human Germline. Annual Review of Genomics and Human Genetics 15:47-70.

621 Shimmin LC, Chang BH-J, Li W-H. 1993. Male-driven evolution of DNA sequences. Nature
622 362:745-747.

623 Smeds L, Qvarnström A, Ellegren H. 2016. Direct estimate of the rate of germline mutation in a
624 bird. Genome research 26:1211-1218.

625 Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and
626 mutation-rate evolution. Proc Natl Acad Sci U S A 109:18488-18492.

627 Thomas GWC, Wang RJ, Puri A, Harris RA, Raveendran M, Hughes DST, Murali SC, Williams
628 LE, Doddapaneni H, Muzny DM, et al. 2018. Reproductive Longevity Predicts Mutation Rates
629 in Primates. Current Biology 28:3193-3197.e3195.

630 Thorne JL, Kishino H. 2002. Divergence Time and Evolutionary Rate Estimation with
631 Multilocus Data. Systematic Biology 51:689-702.

632 Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular
633 evolution. Molecular Biology and Evolution 15:1647-1657.

634 Uchimura A, Higuchi M, Minakuchi Y, Ohno M, Toyoda A, Fujiyama A, Miura I, Wakana S,
635 Nishino J, Yagi T. 2015. Germline mutation rates and the long-term phenotypic effects of
636 mutation accumulation in wild-type laboratory mice and mutator mice. Genome research
637 25:1125-1134.

638 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan
639 T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ Data to High-Confidence Variant
640 Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in
641 Bioinformatics 43:11.10.11-11.10.33.

642 Venn O, Turner I, Mathieson I, de Groot N, Bontrop R, McVean G. 2014. Strong male bias
643 drives germline mutation in chimpanzees. Science 344:1272-1275.

644 Weisenfeld NI, Kumar V, Shah P, Church DM, Jaffe DB. 2017. Direct determination of diploid
645 genome sequences. Genome research 27:757-767.

646 Winter DJ, Wu SH, Howell AA, Cartwright RA, Zufall RA, Azevedo RBR. 2018. accuMUlate: a
647 mutation caller designed for mutation accumulation experiments. Bioinformatics (Oxford,
648 England) 34:2659-2660.

649 Wong A. 2014. Covariance between Testes Size and Substitution Rates in Primates.
650 Molecular Biology and Evolution 31:1432-1436.

651 Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in
652 man. Proceedings of the National Academy of Sciences 82:1741.

653 Xie Z, Wang L, Wang L, Wang Z, Lu Z, Tian D, Yang S, Hurst Laurence D. 2016. Mutation rate
654 analysis via parent–progeny sequencing of the perennial peach. I. A low rate in woody
655 perennials and a higher mutagenicity in hybrids. Proceedings of the Royal Society B: Biological
656 Sciences 283:20161016.

657 Yang Z. 2006. Computational molecular evolution: Oxford University Press.

658 Yang ZH. 2015. The BPP program for species tree estimation and species delimitation.
659 Current Zoology 61:854-865.

660 Yang ZH, Yoder AD. 2003. Comparison of likelihood and Bayesian methods for estimating
661 divergence times using multiple gene loci and calibration points, with application to a radiation
662 of cute-looking mouse lemur species. Systematic Biology 52:705-716.

663 Yoder AD, Campbell CR, Blanco MB, dos Reis M, Ganzhorn JU, Goodman SM, Hunnicutt KE,
664 Larsen PA, Kappeler PM, Rasoloarison RM, et al. 2016. Geogenetic patterns in mouse lemurs
665 (genus Microcebus) reveal the ghosts of Madagascar's forests
666 past. Proceedings of the National Academy of Sciences 113:8049.

667 Zhang J. 2004. Evolution of DMY, a Newly Emergent Male Sex-
668 Determination Gene of Medaka Fish. Genetics 166:1887.

669 Zohdy S, Gerber BD, Tecot S, Blanco MB, Winchester JM, Wright PC, Jernvall J. 2014. Teeth,
670 sex, and testosterone: aging in the world's smallest primate. PLoS One 9:e109528.

671 Zuckerkandl E, Pauling L. 1965. Molecules as documents of evolutionary history. Journal of
672 Theoretical Biology 8:357-366.
673

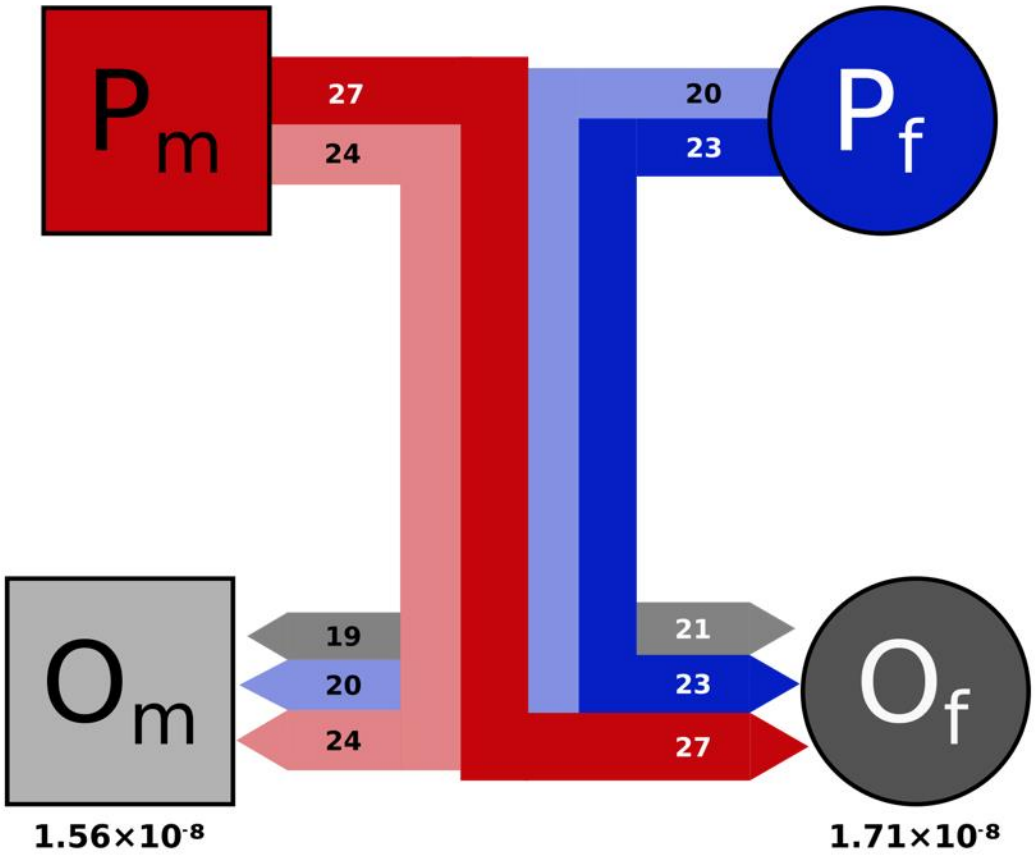
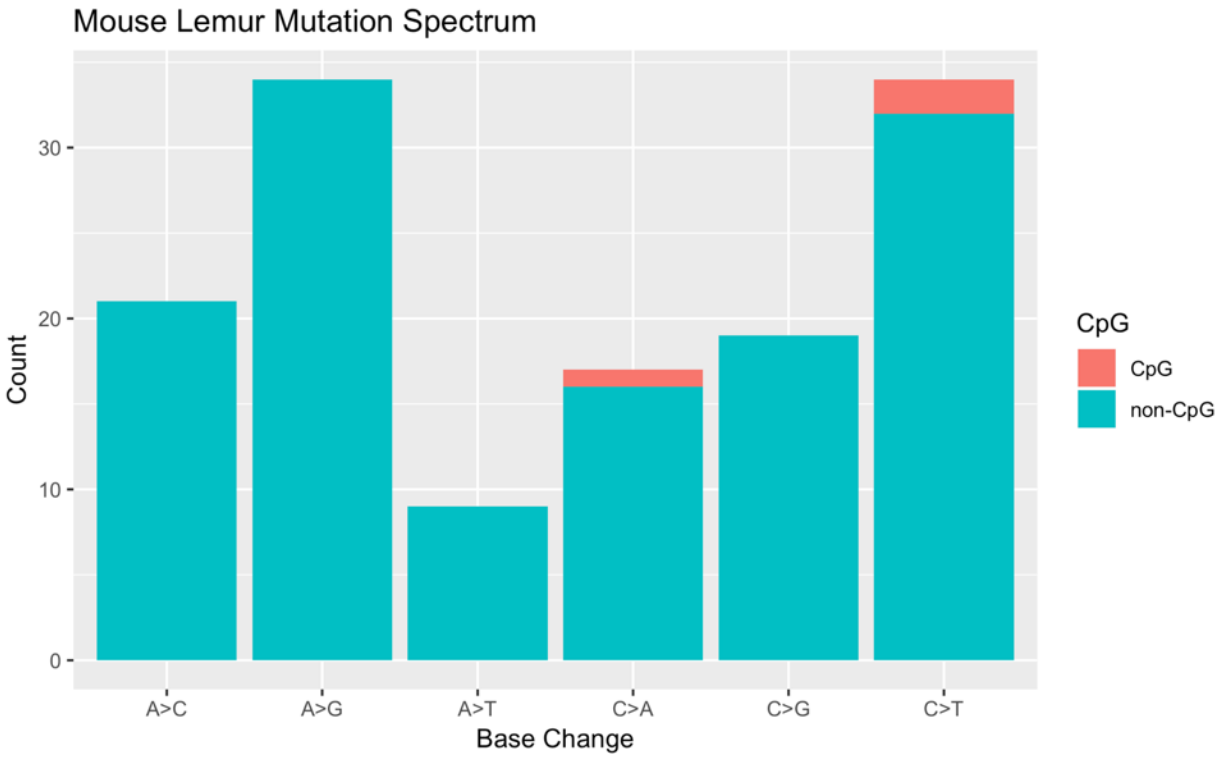
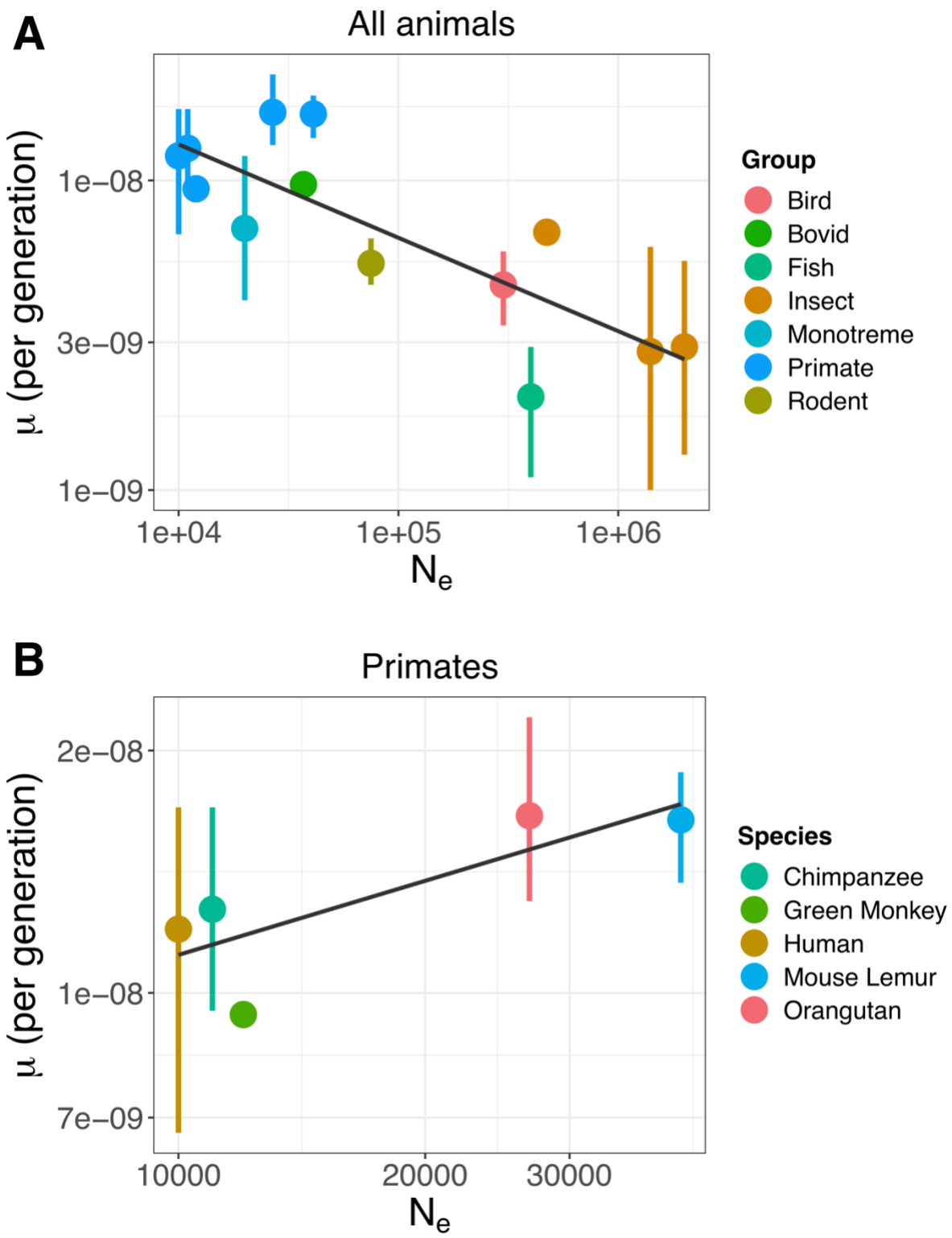


Figure 1. Focal family quartet with parents (P), offspring (labeled O) and sex (subscript m – male, or f - female). Lines represent familial relationships as in a traditional pedigree, with thickness and color reflecting the number and source of *de novo* mutations passed down (red is from male parent, blue from female parent, gray is undetermined origin). Color of line represents source and shading represents destination (lighter shading to O_m, darker to O_f). Numbers within bars show mutation counts and the rate of each individual offspring is listed below.



683
684 **Figure 2.** *de novo* mutation spectrum of the Gray Mouse Lemur, broken down basepair
685 change (x-axis) and CpG site (color).



686
687 **Figure 3.** The linear relationship between mutation rate and effective population size for
688 animals (A) and primates (B) with directly estimated mutation rates (for data see Table S3).
689 Groups and species are identified by color, plots are log-log.
690

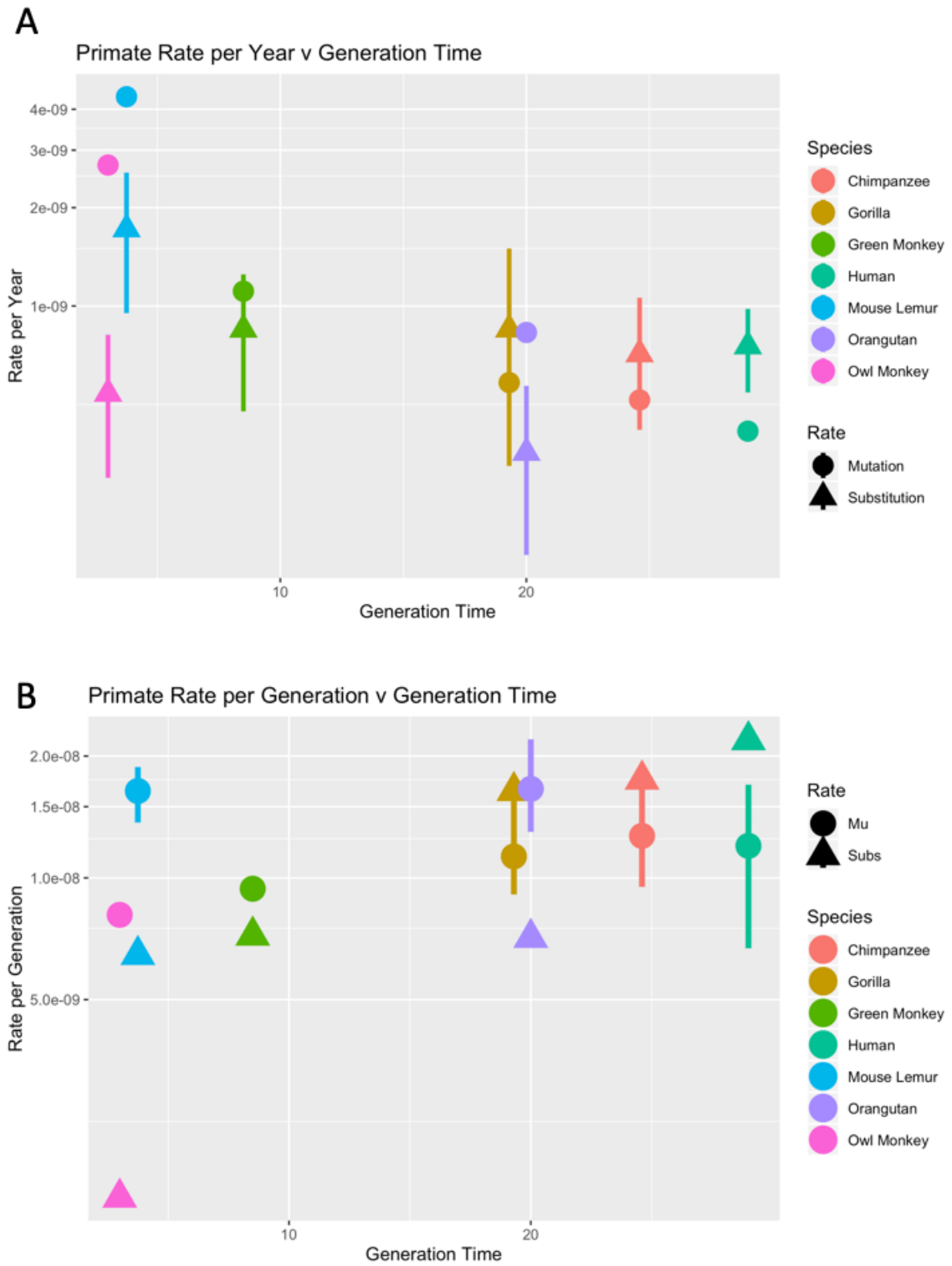
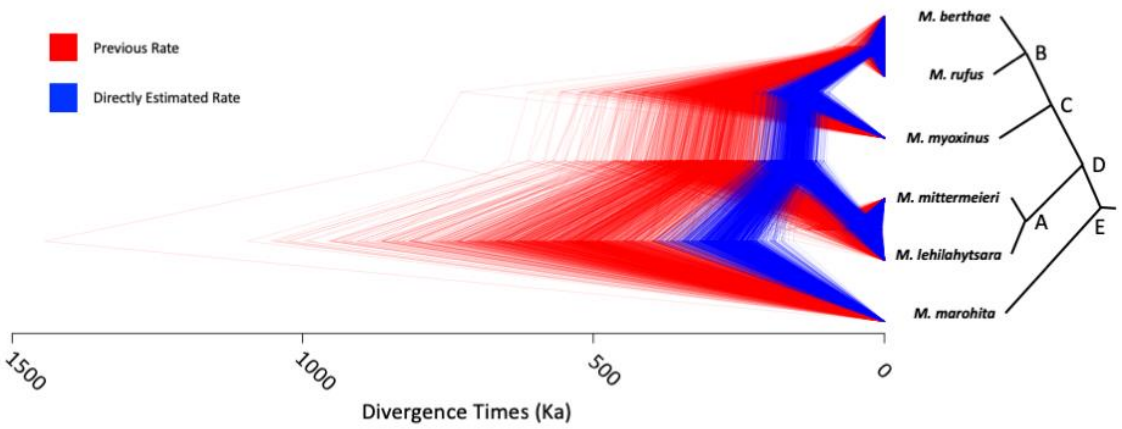


Figure 4. The relationship between mutation rate and substitution rate, (A) per year and (B) per generation, among primates. For the primate species with directly measured mutation rate, plotted against the yearly substitution rate with 95% credible intervals for each measure.



697
698 **Figure 5.** Densitree plot of trees from BPP analysis and bppr plot of previously published
699 mouse lemur RADseq data. Divergence times estimated with either the previous rate estimate
700 (red) or the directly measured rate (blue). Data is also displayed in Table S2, matched by node
701 labels (A-E).
702
703 **Table 1.** Directly estimated mammalian mutation rates.
704

Species	Method	Year	Author	Rate
P. abelii	Pedigree	2019	Besenbacher	1.66×10^{-8}
P. troglodytes	Pedigree	2017	Tatsumoto	1.48×10^{-8}
H. sapiens	Pedigree/Large Scale	2017	Jonsson	1.29×10^{-8}
C. sabaeus	Pedigree	2017	Pfieffer	9.4×10^{-8}
A. nancymaae	Population	2018	Thomas et al.	8.1×10^{-8}
M. musculus	Pedigree/Experimentation	2015	Uchimura	5.4×10^{-9}

705

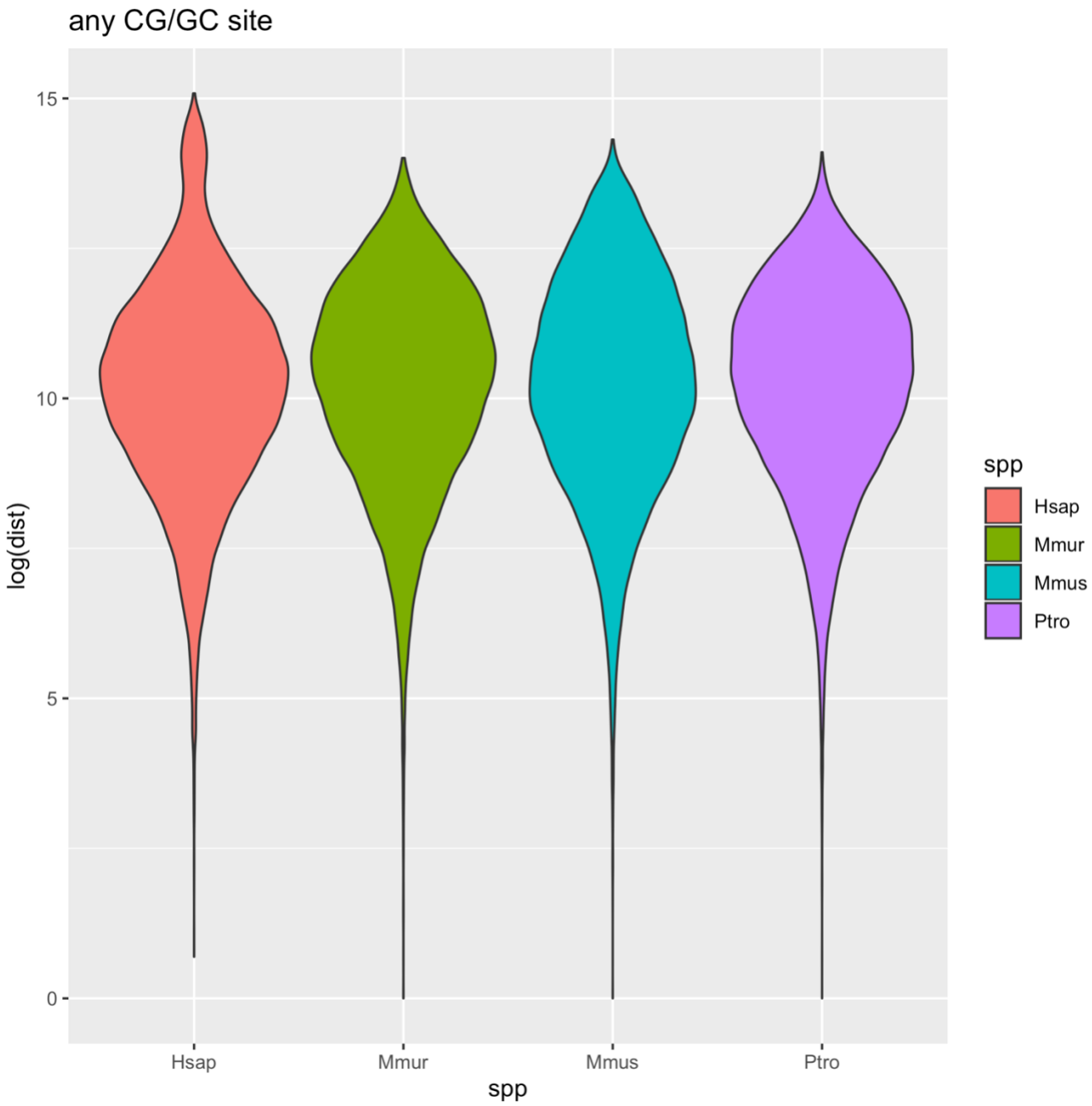


Figure S1. Proximity of CG sites to genic regions in four species of mammals (Hsap – *Homo sapiens*, Mmur – *Microcebus murinus*, Mmus – *Mus musculus*, Ptro – *Pan troglodytes*). No differences were observed in proximity of CG sites to genic regions.

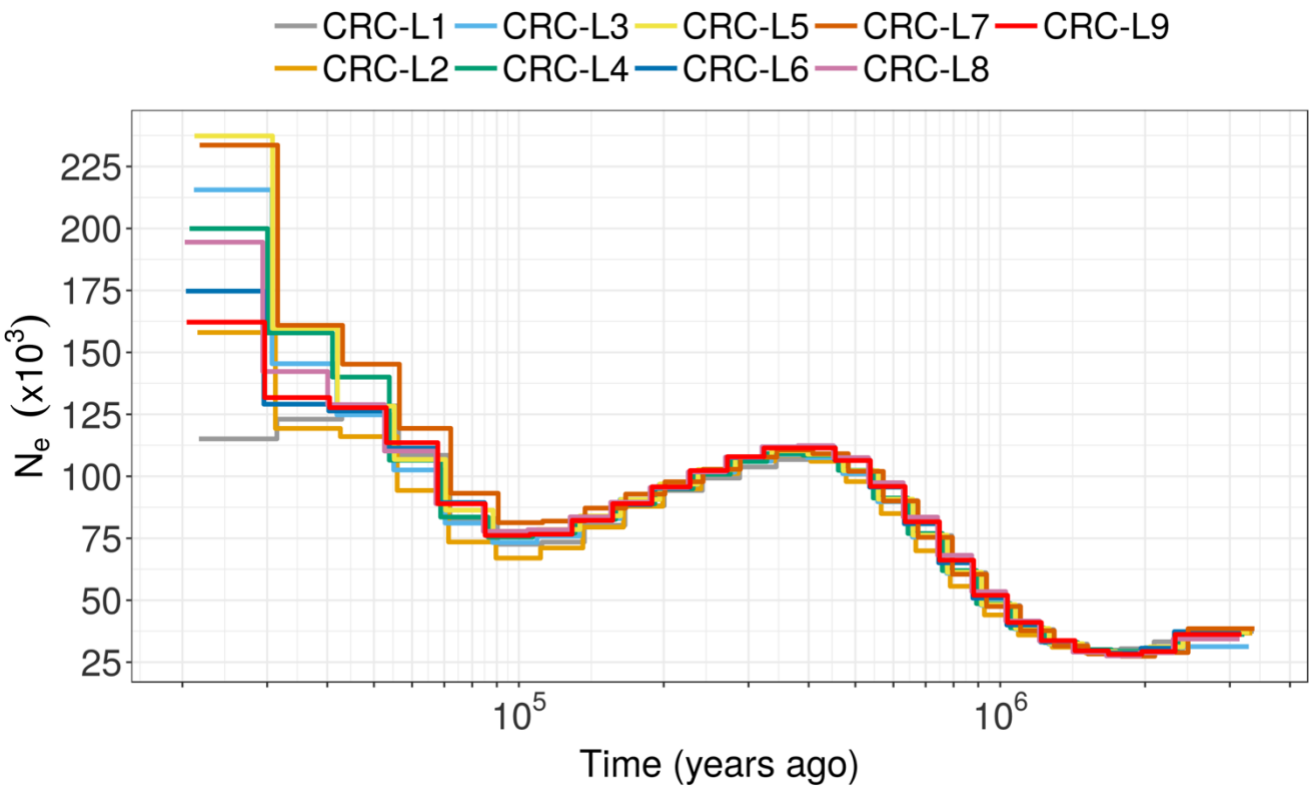


Figure S2. MSMC plot of gray mouse lemur samples in this study, used to estimate the effective population size of gray mouse lemurs. The harmonic mean of the estimated population size of all samples was taken across the last 2 million years, which yielded an N_e of 41,000.

Table S1. Mutation (various pubs) and substitution rates (dos Reis et al. 2018) calculated per year and per generation.

Species	Per Year		Per Generation		Gen Time
	Mutation	Substitution	Mutation	Substitution	
Chlorocebus sabaeus	1.1E-09	8.5E-10	9.4E-09	7.2E-09	8.5
Pan troglodytes	5.2E-10	7.1E-10	1.3E-08	1.8E-08	24.6
Homo sapiens	4.1E-10	7.5E-10	1.2E-08	2.2E-08	29
Gorilla gorilla	5.8E-10	8.5E-10	1.1E-08	1.6E-08	19.3
Pongo abelii	8.3E-10	3.6E-10	1.7E-08	7.1E-09	20
Aotus nancymaae	2.7E-09	5.4E-10	8.1E-09	1.6E-09	3
Microcebus murinus	4.4E-09	1.7E-09	1.6E-08	6.5E-09	3.75

Table S2. Ancestral effective population size (N_e) and estimated divergence times as determined by BPP and bppr (see Materials and Methods). Table includes results for the previous estimate (Yoder, et al. 2016) and the directly measure rate, with node labels corresponding to Figure 5.

Directly Estimated Rate				
Node	**N_e ($\times 10^3$)**	**95% CI**	**Divergence Time (ka)**	**95% CI**
E	155	129, 182	275	198, 357
D	10.8	5.29, 14.5	157	117, 201
A	39.8	25.9, 53.1	26.4	4.79, 54.8
C	2.01	1.19, 2.67	146	107, 187
B	44.5	30.8, 57.7	8.84	0.30, 26.0
Previous Rate Estimate				
Node	**N_e ($\times 10^3$)**	**95% CI**	**Divergence Time (ka)**	**95% CI**
E	307	200, 427	545	321, 796
D	21.5	8.56, 34.5	313	188, 452
A	79.0	42.3, 118	52.4	7.42, 112
C	3.98	1.92, 6.14	290	174, 421
B	88.4	51.5, 130	17.5	0.51, 52.1

739 **Table S3.** Mutation rates (various pubs) and effective population sizes plotted in Figure 3.
740

Directly Estimated Animal Mutation Rates					
Species	**Mutation Rate**	**Ne**	**Author**	**Year**	**Category**
Human	1.20E-08	1.00E+04	Many	2017	Primate
Chimp	1.27E-08	1.10E+04	Besenbacher	2019	Primate
Orangutan	1.66E-08	2.68E+04	Besenbacher	2019	Primate
Green Monkey	9.40E-09	1.20E+04	Pfeifer	2017	Primate
Owl Monkey	8.10E-09	0.00E+00	Thomas	2018	Primate
Mouse Lemur	1.64E-08	4.10E+04	This Study	2019	Primate
Mouse	5.40E-09	7.50E+04	Uchimara	2015	Rodent
Platypus	7.00E-09	2.00E+04	Martin	2018	Monotreme
Cow	9.70E-09	3.70E+04	Harland	2017	Bovid
Flycatcher	4.60E-09	3.00E+05	Smeds	2016	Bird
Herring	2.00E-09	4.00E+05	Feng	2017	Fish
Bee	6.80E-09	4.73E+05	Yang	2015	Insect
Butterfly	2.90E-09	2.00E+06	Keightley	2015	Insect
Fly	2.80E-09	1.40E+06	Keightley	2014	Insect

741

Supplemental Text 1

The uncertainty of the mutation rate estimate from pedigree-sequencing studies

Let λ_m and λ_p be the maternal and paternal mutation rates per genome. The total mutation rate is $\lambda = \lambda_m + \lambda_p$. For example, if we expect an offspring to inherit 30 and 40 mutations form their mother and father respectively, then $\lambda_m = 30$, $\lambda_p = 40$, and $\lambda = 70$. If we assume that the mutations inhereted form the mother and father have a Poisson distribution, then the total number of mutations inherited also follows a Poisson distribution.

Let x_i be the total number of mutations observed in an individual, and suppose we observe mutations in n individuals. We can use the Bayesian method to estimate the mutation rate, l . Note that the gamma distribution is the conjugate prior of the Poisson, and thus we use a gamma prior on l with parameters α (shape) and β (rate). The posterior distribution is

$$
\begin{aligned}
f_\lambda(\lambda \mid x_1, \ldots, x_n) &\propto \lambda^{\alpha-1} e^{-\beta\lambda} \times \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}, \\
&\propto \lambda^{\alpha+\sum x_i - 1} e^{-(\beta+n)\lambda}.
\end{aligned}
$$

Thus, the posterior of λ is a gamma distribution with parameters $\alpha + \sum x_i$ (shape) and $\beta + n$ (rate).

In the mouse lemur pedigree-sequencing results, we see 71 and 63 mutations in two individuals respectively. Let the prior on λ be gamma with α = 2 and β = 0.02. This is a diffuse prior with mean α/β = 100 mutations. Thus the estimate of λ is gamma distributed with shape 2 + 71 + 63 = 136 and rate 0.02 + 2 = 2.02. Thus the posterior mean is $\tilde{\lambda} = 67.3$, posterior standard deviation $\tilde{s}$ $\tilde{s} = \sqrt{136}/2.02 = 5.77$. Note this estimate of λ is given as mutations per genome. To get the rate per nucleotide site, we simply divide by the diploid genome size, g. For example, assuming this is $g = 4\text{x}10^9$ in mouse lemurs, we get $\tilde{\lambda} = 1.68\text{x}10^{-8}$ and $\tilde{s} = 1.44\text{x}10^{-9}$. The 95% credibility interval for the estimate between 56.5 and 79.1 mutations per genome, or between $1.41\text{x}10^{-8}$ and $1.98\text{x}10^{-8}$ mutations per nucleotide site.

Supplemental Text 2

Variant Calling's Effect on *de novo* Mutation Rate Estimation

An attempt to, given a duplicate sample providing a known variant call error rate, quantify the effects of incorrect variant calls on the estimation of de novo *mutations.*

Our knowns are as follows, the number of sites in the genome with sufficient coverage to make *de novo* calls (sites, s), the number of variants (variants, v) in the entire pedigree (v_p), the mother's sample (v_m), the father's sample (v_f), and the offspring sample (v_o). Also the number of unique variants (unique variants, w) in the entire pedigree (w_p), the mother's sample (w_m), the father's sample (w_f), and the offspring sample (w_o). We also know the error rate of the replicated sample, presented here as a fraction of variants called (called error rate, c). This is an average rate of variants that are present in one replicate and absent in the other. We also assume the transmission probability of any given site is 0.5 (transmission probability, t).

Finally we'll add a single unknown variable (e), which is the fraction of erroneous variant calls that are false positives. This variable ranges from 0 to 1, with 1 representing the scenario where all erroneous calls are false positives. We have no reason to expect that this fraction differs across the replicates. This variable will allow us to explore a range of possible error effects. It will also allow us to be conservative (taking the value of e that yields the most error) when reaching a final conclusion regarding error rates.

The probability that a *de novo* mutation call is a false positive is simply the probability of a false positive variant within the offspring that is not already a variant in the pedigree or a parental false negative at a site that is unique to the individual and transmitted.

$$P(\mu_{f+}) = P(unique\, offspring\, false\, positive) + P(shared\, paternal\, false\, negative) + P(shared\, maternal\, false\, negative)$$

$$P(unique\ offspring\ false\ positive) = c * e * \frac{w_o}{v_o}$$

$$P(shared\ paternal\ false\ negative) = \frac{c * v_p}{s} * (1 - e) * \frac{w_o}{v_o}$$

$$P(shared\ maternal\ false\ negative) = \frac{c * v_m}{s} * (1 - e) * \frac{w_o}{v_o}$$

So, substituting the these into the original equation:

$$P(\mu_{f+}) = c * e * \frac{w_o}{v_o} + \frac{c * v_m}{s} * (1 - e) * \frac{w_o}{v_o} + \frac{c * v_f}{s} * (1 - e) * \frac{w_o}{v_o}$$

$$P(\mu_{f+}) = \frac{c * w_o}{v_o} * (e + \frac{v_p * v_m * (1 - e)}{s})$$

The probability that a genomic site with sufficient coverage is an uncalled false negative *de novo* mutation call is essentially the converse of the false positive. It is the probability of a false negative within the offspring that is not already a variant in the pedigree or a parental false positive at a site that is unique to the individual and transmitted.

$$P(\mu_{f-}) = P(unique\ offspring\ false\ negative) + 2 * P(shared\ parental\ false\ positive)$$

$$P(unique\ offspring\ false\ negative) = (1 - \frac{w_o}{s}) * \frac{c}{s} * (1 - e)$$

$$P(shared\ parental\ false\ positive) = c * e * \frac{w_o}{s}$$

So, substituting the these into the original equation:

$$P(\mu_{f-}) = (1 - \frac{w_o}{s}) * \frac{c}{s} * (1 - e) + c * e * \frac{w_o}{s} + c * e * \frac{w_o}{s}$$

$$P(\mu_{f-}) = \frac{c}{s} * ((1 - \frac{w_o}{s}) * (1 - e) + 2 * w_o * e)$$

$$P(\mu_{f-}) = c * v_o * ((1 - e) * (s - v_p) + 2 * e * w_o)$$

Using as known values the following rough estimates:

$$c = 0.025$$
$$e = 0.5$$
$$s = 1.968 * 10^9$$
$$\mu_{dn} = 134$$
$$v_{all} = 30,000$$
$$w_{all} = 1,500$$

The probabilites of false negative and false positives are as follows:

$$P(\mu_{f+}) = 4.83 * 10^{-2}$$

$$P(\mu_{f-}) = 1.91 * 10^{-8}$$

Given 134 *de novo* mutations from 1,968,000,000 sites of sufficient coverage, we expect:

$$E(\mu_{f+}) = \mu_{dn} * P(\mu_{f+})$$
$$E(\mu_{f+}) = 6.46$$

$$E(\mu_{f-}) = s * P(\mu_{f-})$$
$$E(\mu_{f-}) = 37.51$$