Current Biology
Report

Reproductive Longevity Predicts Mutation Rates in Primates

Gregg W.C. Thomas,[1,2,8,*] Richard J. Wang,[1] Arthi Puri,[2] R. Alan Harris,[3,4] Muthuswamy Raveendran,[3,4] Daniel S.T. Hughes,[3,4] Shwetha C. Murali,[3,4] Lawrence E. Williams,[5] Harsha Doddapaneni,[3,4] Donna M. Muzny,[3,4] Richard A. Gibbs,[3,4] Christian R. Abee,[5] Mary R. Galinski,[6,7] Kim C. Worley,[3,4] Jeffrey Rogers,[3,4] Predrag Radivojac,[2] and Matthew W. Hahn[1,2,*]

[1]Department of Biology, Indiana University, 107 S. Indiana Avenue, Bloomington, IN 47405, USA
[2]Department of Computer Science, Indiana University, 107 S. Indiana Avenue, Bloomington, IN 47405, USA
[3]Human Genome Sequencing Center, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA
[4]Department of Molecular and Human Genetics, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 77030, USA
[5]Keeling Center for Comparative Medicine and Research, University of Texas, MD Anderson Cancer Center, 650 Cool Water Drive, Bastrop, TX 78602, USA
[6]Emory Vaccine Center, Yerkes National Primate Research Center, Emory University, 201 Dowman Drive, Atlanta, GA, USA
[7]Division of Infectious Diseases, Department of Medicine, Emory University, 201 Dowman Drive, Atlanta, GA, USA
[8]Lead Contact
*Correspondence: grthomas@indiana.edu (G.W.C.T.), mwh@indiana.edu (M.W.H.)
https://doi.org/10.1016/j.cub.2018.08.050

SUMMARY

Mutation rates vary between species across several orders of magnitude, with larger organisms having the highest per-generation mutation rates. Hypotheses for this pattern typically invoke physiological or population-genetic constraints imposed on the molecular machinery preventing mutations [1]. However, continuing germline cell division in multicellular eukaryotes means that organisms with longer generation times and of larger size will leave more mutations to their offspring simply as a byproduct of their increased lifespan [2, 3]. Here, we deeply sequence the genomes of 30 owl monkeys (*Aotus nancymaae*) from six multigeneration pedigrees to demonstrate that paternal age is the major factor determining the number of *de novo* mutations in this species. We find that owl monkeys have an average mutation rate of 0.81×10^{-8} per site per generation, roughly 32% lower than the estimate in humans. Based on a simple model of reproductive longevity that does not require any changes to the mutational machinery, we show that this is the expected mutation rate in owl monkeys. We further demonstrate that our model predicts species-specific mutation rates in other primates, including study-specific mutation rates in humans based on the average paternal age. Our results suggest that variation in life history traits alone can explain variation in the per-generation mutation rate among primates, and perhaps among a wide range of multicellular organisms.

RESULTS AND DISCUSSION

The rate at which new mutations arise is a key parameter of life on Earth, contributing to both individual disease risk and the evolution of novel traits. The mutation rate per generation varies among taxa, from as low as 1×10^{-10} per base in Archaea to more than 1×10^{-8} in mammals [1]. Two classes of models have been proposed to explain this variation. In one, the physiological and biochemical costs of increased fidelity during DNA replication limit the minimum mutation rate achievable [4, 5]. Selection for faster replication in smaller organisms constrains the accuracy with which the cellular machinery can copy DNA, resulting in an inverse relationship between body size and mutation rate. Alternatively, a population-genetic model invokes the limits to natural selection in organisms with smaller population sizes [6–8]. This model posits a higher rate of mutation in larger organisms because of their generally smaller population size [9].

One difficulty in teasing apart the forces driving the evolution of the mutation rate among multicellular organisms is the fact that lifespan varies as much as the per-generation mutation rate. In multicellular organisms, the number of mutations passed on to offspring in a single generation is a combination of the errors made in each round of germline replication and the accumulation of unrepaired DNA damage. One hundred years after the first observation of increased disease incidence in the children of older parents [2, 10], whole-genome sequencing in humans revealed the precise contribution of parental age to the number of *de novo* mutations in their offspring [3, 11–17]. In particular, the number of mutations passed on to the next generation is largely dependent on the age of the father [3], though there is a non-negligible contribution from the age of the mother [12–15, 17]. This is a consequence of the fact that after a set number of germline mitoses during development in both males and females, the male germline resumes cell division at puberty [18, 19]. A similar effect of paternal age has been found in chimpanzees [20], suggesting that the age of reproduction may generally be an important determinant of the per-generation mutation rate.

Studying closely related primates offers a unique opportunity to examine the role that life history traits—such as age of puberty and average generation time—may play in determining mutation rates. We sequenced the genomes of 30 owl monkeys (*Aotus*

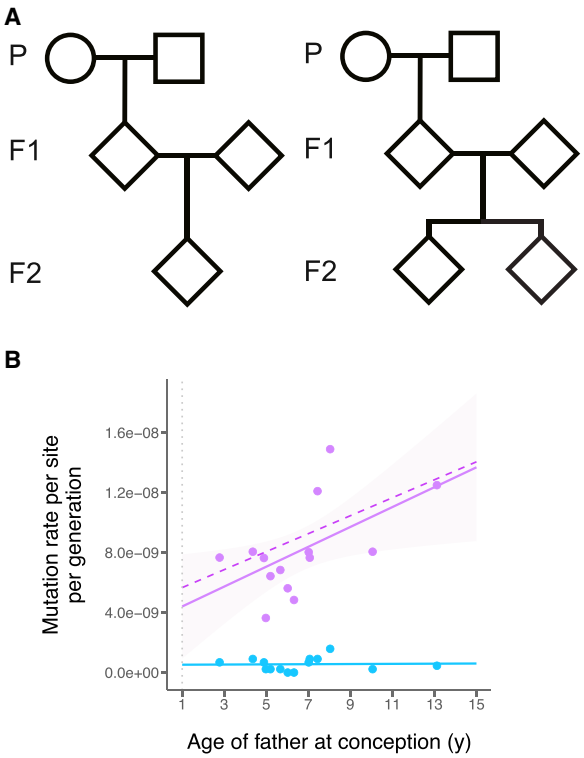

Figure 1. Pedigree Structures and Mutation Rates in Owl Monkeys

(A) We used six multi-generation pedigrees in these two formats. Four families have a single F2 offspring (left), and two families have two F2 offspring (right). In total, 14 independent trios can be constructed from these pedigrees.

(B) Mutation rate estimates from the 14 owl monkey trios (purple points). A simple linear regression has been fit to these points (solid purple line; shaded area indicates 95% confidence interval) to show that the number of mutations increases with the father's age. Our model of reproductive longevity (dashed purple line) is not significantly different from the fit of the linear regression. The rate of non-replicative mutations, such as those that occur at CpG sites (blue dots), are not correlated with reproductive longevity (blue line). The dotted vertical gray line indicates expected age of puberty.

See also Data S1 and Figure S1.

nancymaae) within six multi-generation pedigrees (Figure 1A; Data S1A) in order to estimate the effect of parental age on the mutation rate. Owl monkeys reach sexual maturity at ~1 year of age [21] and can live up to 20 years in captivity [22]. Our sample includes individuals conceived by sires ranging from 3–13 years old and dams ranging from 3–12 years old, with an average age of 6.64 and 6.53 for sires and dams, respectively (Data S1A). These ages are comparable to those observed in the wild, as owl monkeys are solitary for some time before joining a mating group at around age 4 [23]. The genomes of all parents and offspring were sequenced to an average of 37× coverage (range: 35×–38×) using paired-end Illumina reads. Sequencing multi-generation pedigrees allows us to determine whether *de novo* mutations arose in either sires or dams, as well as to validate mutations transmitted to the next generation.

We observe 283 *de novo* mutations across 14 trios (Data S1B) and estimate an average mutation rate for owl monkeys of 0.81×10^{-8} per site per generation (Data S1C). In addition to stringent quality filters (see STAR Methods), the average transmission frequency of *de novo* mutations passed from F1 individuals to F2 individuals across families was 0.502, giving us high confidence in our final set of mutations. As in humans, we find a strong association between paternal age and the number of *de novo* mutations (Figure 1B), with 2.92 additional mutations accumulating per year (R^2 = 0.25, d.f. = 12, p = 0.040). Also as expected, we find no effect of age on CpG mutations (Figure 1B, blue points and line), as these are not associated with replication errors. We were able to assign phase to 105 of the 283 *de novo* mutations via transmission to the third generation in our pedigrees (Data S1B). We find that 71 of these 105 phased mutations are paternal, with the number of mutations passed on increasing with the age of the father (R^2 = 0.58, d.f. = 4, p = 0.048). We did not find an increasing number of mutations with maternal age (R^2 = 0.07, d.f. = 4, p = 0.307) or age of the offspring (R^2 = −0.02, d.f. = 12, p = 0.388). This is the first direct observation of the paternal age effect outside of apes.

Inspection of the types of mutations found in the genomes of owl monkeys shows a transition:transversion (Ts:Tv) ratio of 1.97. This is in close agreement with the observed human Ts:Tv ratio of 2.10 [3]. In fact, the overall mutational spectrum between humans, chimpanzees, and owl monkeys appears almost identical, with the only difference being a slightly higher proportion of A→T mutations in owl monkeys (Figure 2). We also observe that 12.0% of mutations in owl monkeys occur at CpG sites, with CpG sites having a much higher Ts:Tv ratio (4.67), similar to observations in humans [3, 14]. Multinucleotide mutations (MNMs) are mutations that occur in close proximity to one another (<20 bp apart), most likely caused by a single mutational event [24]. Here, we find six MNMs consisting of two mutations each, indicating that 2.1% of *de novo* mutations in owl monkeys are the result of MNMs (Data S1B). This fraction is also in agreement with that observed within humans [14, 24].

The mutation rate we observe in owl monkeys is 32.5% lower than the average human estimate of 1.2×10^{-8} mutations per site per generation [3, 25]. Although traditional models of mutation rate evolution invoke changes to the underlying replication machinery as the main cause of such differences, we asked whether a shift in reproductive timing could explain the lower rate in owl monkeys. The effects of paternal age on per-generation mutation rates have previously been modeled by combining estimates of the rate of mutation from different life stages [19, 25, 26]. The germline in males and females undergo a fixed number of divisions before birth, but the male germline continues dividing upon reaching sexual maturity. This phenomenon suggests that the length of time between puberty and the conception of offspring in an individual—which we define here as the "reproductive longevity" of males—plays a key role in determining the number of mutations passed on to the next generation. Although paternal age is sufficient for predicting mutation rates within a species [3, 11–17], the concept of reproductive longevity makes it possible to predict mutation rates between species with varying ages of puberty. We modeled the owl monkey mutation rate as a linear combination of the mutations accumulated as a result of a constant number of germline divisions *in utero* and those accumulated during continued germline divisions post-puberty. The rate of mutation in these two stages was estimated from human studies, and sexual maturity was set at 1 year of age (STAR Methods).

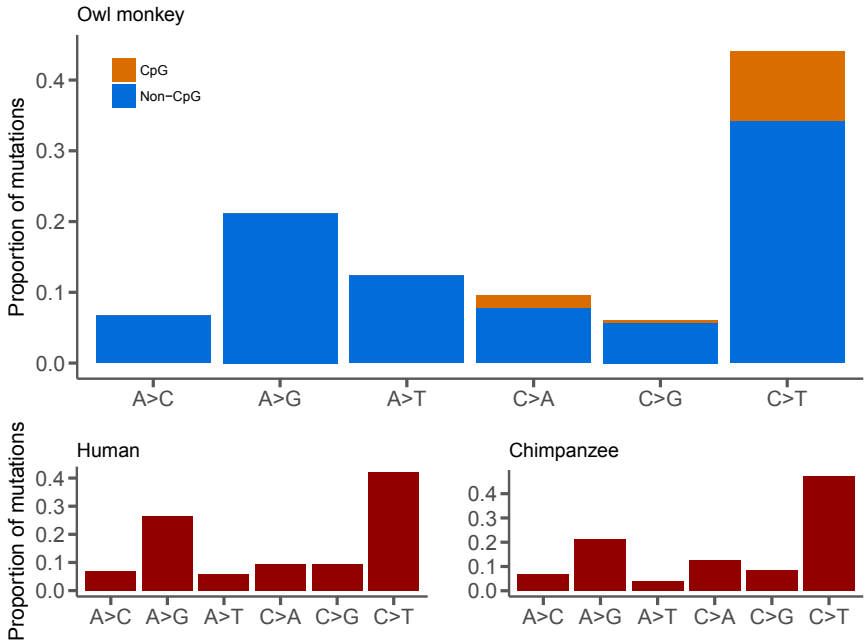

Figure 2. Comparison of Mutational Spectra from Owl Monkeys, Humans, and Chimpanzees

There is a slight but significant difference in the frequency of A→T mutations between owl monkeys and humans ($\chi^2 = 25.7$, d.f. = 4, $p < 0.05$), but otherwise no difference between mutational spectra for these three species. Human data were averaged across four studies (see Data S1D for references), and chimpanzee data were extrapolated from Figure 3A in Venn et al. [20]. Mutation categories include their reverse complement. See also Data S1B.

Our minimal model provides an excellent fit to the observed owl monkey data (Figure 1B, dashed line). In fact, a linear regression of the observed number of mutations with paternal age at conception is not significantly better than the predictions provided by our model ($F = 0.996$, d.f. = 13, $p = 0.994$). The main determinant of the mutation rate is reproductive longevity in sires, which determines the number of mitotic germline divisions before spermatogenesis. For instance, a 13-year-old owl monkey male (who reached sexual maturity at 1) will have the same reproductive longevity as a 25-year-old human male (who reached sexual maturity at 13). Our model therefore predicts the same estimated mutation rate if *de novo* mutations are sampled from offspring of these individuals, and this is what was observed (Figure 1B). Because reproductive longevity reflects replicative mutations, we observed no effect of father's age on non-replicative mutations, such as those found at CpG sites (Figure 1B, blue).

Given the fit of our model to owl monkey data, we calculated the expected mutation rates as a function of age for other primates, accounting for changes in the time to sexual maturity in each species. A model of reproductive longevity provides a good fit to the data from primate species for which direct mutation rate estimates are available (Figure 3; Data S1D). Our model explains why chimpanzees and humans have very similar per-generation mutation rates despite differences in average generation time: the earlier time to sexual maturity in chimpanzees causes reproductive longevity to be the same in both species. The model also accurately predicts estimated mutation rates reported from various studies in humans where sampled parents were of different average age (Figure 3). Much of the variation in reported mutation rates in human studies is due to differences in the average reproductive longevity of sampled individuals ($R^2 = 0.54$, d.f. = 7, $p = 0.01$). Variation in the age of reproduction across pedigrees will affect inferences regarding genetic variation in the mutation rate, as consistent differences in these ages may incorrectly be interpreted as heritable differences in this trait.

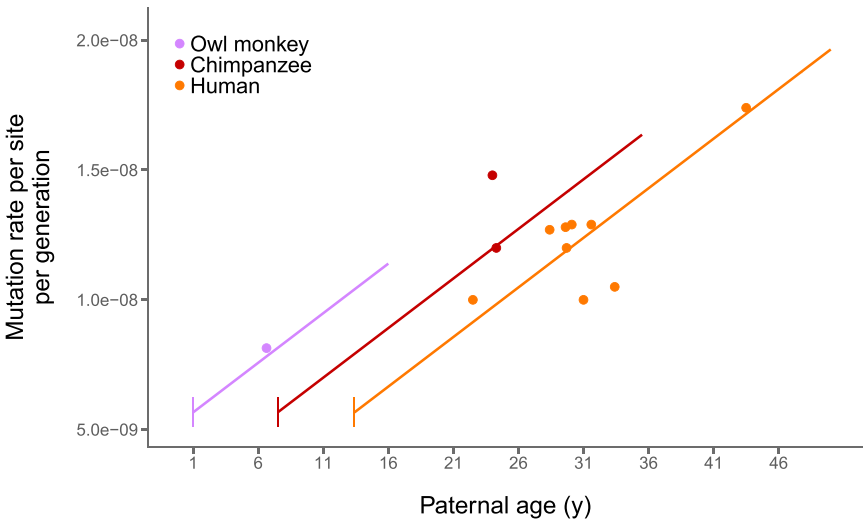

Figure 3. A Model of Reproductive Longevity Fits Estimated Primate Mutation Rates

Humans, chimpanzees, and owl monkeys are the only primates that currently have high-quality estimates of mutation rates via pedigree sequencing. Here we plot the average rate from each published study (points; see Data S1D for references). Predictions from our model of reproductive longevity (Equations 3, 4, 5, 6, 7, and 8 in Method Details) using human mutational parameters—varying only life history traits—are also shown (lines). Vertical line segments represent the age of puberty for each species. See also Data S1C, Data S1D, and Figure S2.

The association between mutation rates and reproductive longevity implies that changes in life history traits rather than changes to the mutational machinery are responsible for the evolution of these rates. Species that have evolved greater reproductive longevity will have a higher mutation rate per generation without any underlying change to the replication, repair, or proofreading proteins. The similarities between the mutational spectra of humans, chimpanzees, and owl monkeys (Figure 2) are further evidence that the molecular mechanisms responsible for mutation have not changed between these species. Many differences in the details of germline cell division may exist between these primates, but these differences do not appear to affect either pre-birth or post-puberty mutation accumulation. For instance, varying levels of sexual selection between species in the form of sperm competition leads to variation testis and ejaculate size [27]. This sort of variation most likely also affects mutation rates through changing the germline replication rate [28], which can be accommodated in our model (see STAR Methods). The underlying consistency of mutation rates must also be reconciled with variation in the long-term substitution rate among primates [25, 26, 29, 30], as mutation rates are mechanistically tied to substitution rates (see STAR Methods and Figure S3). Nevertheless, the close fit between the observed and expected mutation rates suggests that reproductive longevity is the major determinant of variation in mutation rates.

Studies of mutation rate evolution will continue to accumulate across the tree of life as sequencing costs continue to plummet. In order to understand the forces affecting this important evolutionary parameter, future studies must recognize that the mutation rate is a function-valued trait: it is a function of reproductive longevity and other life history traits. Evidence from other species—for instance, arthropods [31] and long-lived plants [32]—suggests that reproductive longevity affects the mutation rate in many taxa, though the details of germline cell division will differ among lineages. If such a pattern holds widely in multicellular organisms, the effect of variation in life history traits should provoke a reexamination of the causes underlying the correlation between body size and the per-generation mutation rate. At the very least, the null model for changes in the per-generation mutation rate must include reproductive longevity.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Sequencing
  - Mapping and variant calling
  - Filtering of putative mutations
  - Phasing mutations
  - Estimating mutation rates
  - Modeling mutation rates
  - Estimating mutational parameters from humans
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and one data file and can be found with this article online at https://doi.org/10.1016/j.cub.2018.08.050.

ACKNOWLEDGMENTS

This research was supported by the Precision Health Initiative at Indiana University. This research was also supported in part by federal funds from the US National Institute of Allergy and Infectious Diseases, NIH, Department of Health and Human Services under contract no. HHSN272201200031C, which supported the Malaria Host-Pathogen Interaction Center (MaHPIC).

AUTHOR CONTRIBUTIONS

G.W.C.T., M.W.H., and P.R. conceived analyses; G.W.C.T., R.J.W., and A.P. performed analyses; M.W.H. supervised analyses; C.R.A. directs and supervises the owl monkey colony; L.E.W. manages the owl monkey colony and assisted with sample acquisition and metadata; M.R. managed the sequencing process; M.R. and R.A.H called SNPs; R.A.H., D.H., and S.M., assembled and analyzed the reference genome; H.D. supervised library production and tech development; D.M.M. directs operations of sequencing and developed pipeline operations; R.G. directs the sequencing center and supervises sequencing operations; M.R.G. provided project rationale; K.C.W. and J.R. planned and supervised the genome assembly; and M.W.H., P.R., and J.R. designed the project.

DECLARATION OF INTEREST

The authors declare no competing interests.

Received: May 22, 2018
Revised: July 26, 2018
Accepted: August 22, 2018
Published: September 27, 2018

REFERENCES

1. Lynch, M. (2010). Evolution of the mutation rate. Trends Genet. *26*, 345–352.
2. Crow, J.F. (2000). The origins, patterns and implications of human spontaneous mutation. Nat. Rev. Genet. *1*, 40–47.
3. Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., et al. (2012). Rate of *de novo* mutations and the importance of father's age to disease risk. Nature *488*, 471–475.
4. Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA *88*, 7160–7164.
5. Kondrashov, A.S. (1995). Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. Genet. Res. *66*, 53–69.
6. Lynch, M. (2006). The origins of eukaryotic gene structure. Mol. Biol. Evol. *23*, 450–468.
7. Lynch, M. (2007). The Origins of Genome Architecture (Sinauer Associates).
8. Lynch, M. (2008). The cellular, developmental and population-genetic determinants of mutation-rate evolution. Genetics *180*, 933–943.
9. Damuth, J. (1981). Population-density and body size in mammals. Nature *290*, 699–700.
10. Weinberg, W. (1912). Zur vererbung des zwergwuchses. Arch. Rassen-u Gesell. Biol. *9*, 710–718.

11. Michaelson, J.J., Shi, Y., Gujral, M., Zheng, H., Malhotra, D., Jin, X., Jian, M., Liu, G., Greer, D., Bhandari, A., et al. (2012). Whole-genome sequencing in autism identifies hot spots for *de novo* germline mutation. Cell *151*, 1431–1442.
12. Wong, W.S., Solomon, B.D., Bodian, D.L., Kothiyal, P., Eley, G., Huddleston, K.C., Baker, R., Thach, D.C., Iyer, R.K., Vockley, J.G., and Niederhuber, J.E. (2016). New observations on maternal age effect on germline de novo mutations. Nat. Commun. *7*, 10486.
13. Goldmann, J.M., Wong, W.S., Pinelli, M., Farrah, T., Bodian, D., Stittrich, A.B., Glusman, G., Vissers, L.E., Hoischen, A., Roach, J.C., et al. (2016). Parent-of-origin-specific signatures of *de novo* mutations. Nat. Genet. *48*, 935–939.
14. Besenbacher, S., Sulem, P., Helgason, A., Helgason, H., Kristjansson, H., Jonasdottir, A., Jonasdottir, A., Magnusson, O.T., Thorsteinsdottir, U., Masson, G., et al. (2016). Multi-nucleotide *de novo* mutations in humans. PLoS Genet. *12*, e1006315.
15. Girard, S.L., Bourassa, C.V., Lemieux Perreault, L.P., Legault, M.A., Barhdadi, A., Ambalavanan, A., Brendgen, M., Vitaro, F., Noreau, A., Dionne, G., et al. (2016). Paternal age explains a major portion of *de novo* germline mutation rate variability in healthy individuals. PLoS ONE *11*, e0164212.
16. Rahbari, R., Wuster, A., Lindsay, S.J., Hardwick, R.J., Alexandrov, L.B., Turki, S.A., Dominiczak, A., Morris, A., Porteous, D., Smith, B., et al.; UK10K Consortium (2016). Timing, rates and spectra of human germline mutation. Nat. Genet. *48*, 126–133.
17. Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M.T., Hjorleifsson, K.E., Eggertsson, H.P., Gudjonsson, S.A., et al. (2017). Parental influence on human germline *de novo* mutations in 1,548 trios from Iceland. Nature *549*, 519–522.
18. Goriely, A. (2016). Decoding germline *de novo* point mutations. Nat. Genet. *48*, 823–824.
19. Ségurel, L., Wyman, M.J., and Przeworski, M. (2014). Determinants of mutation rate variation in the human germline. Annu. Rev. Genomics Hum. Genet. *15*, 47–70.
20. Venn, O., Turner, I., Mathieson, I., de Groot, N., Bontrop, R., and McVean, G. (2014). Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. Science *344*, 1272–1275.
21. Dixson, A.F., Gardner, J.S., and Bonney, R.C. (1980). Puberty in the male owl monkey (*Aotus trivirgatus griseimembra*): A study of physical and hormonal development. Int. J. Primatol. *1*, 129–139.
22. Rowe, N. (1996). The Pictorial Guide to the Living Primates (Pogonias Press).
23. Huck, M., Rotundo, M., and Fernandez-Duque, E. (2011). Growth and development in wild owl monkeys (*Aotus azarai*) of Argentina. Int. J. Primatol. *32*, 1133–1152.
24. Schrider, D.R., Hourmozdi, J.N., and Hahn, M.W. (2011). Pervasive multinucleotide mutational events in eukaryotes. Curr. Biol. *21*, 1051–1054.
25. Scally, A. (2016). The mutation rate in human evolution and demographic inference. Curr. Opin. Genet. Dev. *41*, 36–43.
26. Thomas, G.W.C., and Hahn, M.W. (2014). The human mutation rate is increasing, even as it slows. Mol. Biol. Evol. *31*, 253–257.
27. Dixson, A., and Anderson, M. (2001). Sexual selection and the comparative anatomy of reproduction in monkeys, apes, and human beings. Annu. Rev. Sex Res. *12*, 121–144.
28. Presgraves, D.C., and Yi, S.V. (2009). Doubts about complex speciation between humans and chimpanzees. Trends Ecol. Evol. *24*, 533–540.
29. Yi, S.V. (2013). Morris Goodman's hominoid rate slowdown: the importance of being neutral. Mol. Phylogenet. Evol. *66*, 569–574.
30. Moorjani, P., Amorim, C.E.G., Arndt, P.F., and Przeworski, M. (2016). Variation in the molecular clock of primates. Proc. Natl. Acad. Sci. USA *113*, 10607–10612.
31. Latta, L.C., 4th, Morgan, K.K., Weaver, C.S., Allen, D., Schaack, S., and Lynch, M. (2013). Genomic background and generation time influence deleterious mutation rates in *Daphnia*. Genetics *193*, 539–544.
32. Schmid-Siegert, E., Sarkar, N., Iseli, C., Calderon, S., Gouhier-Darimont, C., Chrast, J., Cattaneo, P., Schütz, F., Farinelli, L., Pagni, M., et al. (2017). Low number of fixed somatic mutations in a long-lived oak tree. Nat. Plants *3*, 926–929.
33. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997, http://arxiv.org/abs/1303.3997.
34. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
35. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. *43*, 491–498.
36. Li, H. (2014). Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics *30*, 2843–2851.
37. Besenbacher, S., Liu, S., Izarzugaza, J.M., Grove, J., Belling, K., Bork-Jensen, J., Huang, S., Als, T.D., Li, S., Yadav, R., et al. (2015). Novel variation and *de novo* mutation rates in population-wide de novo assembled Danish trios. Nat. Commun. *6*, 5969.
38. Smeds, L., Mugal, C.F., Qvarnström, A., and Ellegren, H. (2016). High-resolution mapping of crossover and non-crossover recombination events by whole-genome re-sequencing of an avian pedigree. PLoS Genet. *12*, e1006044.
39. Francioli, L.C., Polak, P.P., Koren, A., Menelaou, A., Chun, S., Renkens, I., van Duijn, C.M., Swertz, M., Wijmenga, C., van Ommen, G., et al.; Genome of the Netherlands Consortium (2015). Genome-wide patterns and properties of *de novo* mutations in humans. Nat. Genet. *47*, 822–826.
40. Amster, G., and Sella, G. (2016). Life history effects on the molecular clock of autosomes and sex chromosomes. Proc. Natl. Acad. Sci. USA *113*, 1588–1593.
41. Nielsen, C.T., Skakkebaek, N.E., Richardson, D.W., Darling, J.A.B., Hunter, W.M., Jørgensen, M., Nielsen, A., Ingerslev, O., Keiding, N., and Müller, J. (1986). Onset of the release of spermatozoa (spermarche) in boys in relation to age, testicular growth, pubic hair, and height. J. Clin. Endocrinol. Metab. *62*, 532–535.
42. Marson, J., Meuris, S., Cooper, R.W., and Jouannet, P. (1991). Puberty in the male chimpanzee: progressive maturation of semen characteristics. Biol. Reprod. *44*, 448–455.
43. Tatsumoto, S., Go, Y., Fukuta, K., Noguchi, H., Hayakawa, T., Tomonaga, M., Hirai, H., Matsuzawa, T., Agata, K., and Fujiyama, A. (2017). Direct estimation of *de novo* mutation rates in a chimpanzee parent-offspring trio by ultra-deep whole genome sequencing. Sci. Rep. *7*, 13561.
44. Drost, J.B., and Lee, W.R. (1995). Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among Drosophila, mouse, and human. Environ. Mol. Mutagen. *25* (*Suppl 26*), 48–64.
45. Heller, C.G., and Clermont, Y. (1963). Spermatogenesis in man: an estimate of its duration. Science *140*, 184–186.
46. Scally, A. (2016). Mutation rates and the evolution of germline structure. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150137.
47. Plant, T.M. (2010). Undifferentiated primate spermatogonia and their endocrine control. Trends Endocrinol. Metab. *21*, 488–495.
48. de Rooij, D.G., van Alphen, M.M.A., and van de Kant, H.J.G. (1986). Duration of the cycle of the seminiferous epithelium and its stages in the rhesus monkey (*Macaca mulatta*). Biol. Reprod. *35*, 587–591.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Blood samples from 30 owl monkeys (*Aotus nancymaae*)	Owl Monkey Breeding and Research Resource at the Keeling Center	https://www.ncbi.nlm.nih.gov/biosample?LinkName=bioproject_biosample_all&from_uid=451475
Chemicals, Peptides, and Recombinant Proteins		
KAPA Hyper PCR-free library reagents	KAPA Biosystems	KK8505
Deposited Data		
Sequence reads for 30 owl monkeys (*Aotus nancymaae*)	This paper	NCBI BioProject Accession PRJNA451475; https://www.ncbi.nlm.nih.gov/bioproject/451475
Software and Algorithms		
BWA-MEM version 0.7.12-r1039	[33]	https://github.com/lh3/bwa
Picard MarkDuplicates version 1.105	Broad Institute	http://broadinstitute.github.io/picard/
GATK version 3.3-0	[34]	https://software.broadinstitute.org/gatk/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gregg Thomas (grthomas@indiana.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Thirty owl monkeys (*Aotus nancymaae*) were selected for genome sequencing from the Owl Monkey Breeding and Research Resource at the Keeling Center based on available pedigrees, aiming for a spread of parental ages (Data S1A). Blood samples were taken from the femoral vein of unanesthetized animals. The animals were manually restrained in a supine position with one care staff holding the animal while another takes the sample, under approved IACUC protocols.

METHOD DETAILS

Sequencing

Genomic DNA isolated from the blood samples was used to perform whole genome sequencing. We generated standard PCR-free Illumina paired-end sequencing libraries. Libraries were prepared using KAPA Hyper PCR-free library reagents (KK8505, KAPA Biosystems) in Beckman robotic workstations (Biomek FX and FXp models). We sheared total genomic DNA (500 ng) into fragments of approximately 200-600 bp in a Covaris E220 system (96-well format) followed by purification of the fragmented DNA using AMPure XP beads. A double size selection step was employed, with different ratios of AMPure XP beads, to select a narrow size band of sheared DNA molecules for library preparation. DNA end-repair and 3′-adenylation were then performed in the same reaction followed by ligation of the barcoded adaptors to create PCR-Free libraries, and the library run on the Fragment Analyzer (Advanced Analytical Technologies, Ames, Iowa) to assess library size and presence of remaining adaptor dimers. This was followed by qPCR assay using KAPA Library Quantification Kit using their SYBR FAST qPCR Master Mix to estimate the size and quantification. These WGS libraries were sequenced on the Illumina HiSeq-X instrument to generate 150 bp paired-end reads. All flow cell data (BCL files) are converted to barcoded FASTQ files.

Mapping and variant calling

BWA-MEM version 0.7.12-r1039 [33] was used to align Illumina reads to the owl monkey reference assembly Anan_2.0 (GenBank assembly accession GCA_000952055.2) and to generate BAM files for each of the 30 individuals. Picard MarkDuplicates version 1.105 (http://broadinstitute.github.io/picard/) was used to identify and mark duplicate reads. Single nucleotide variants (SNVs) and small indels (up to 60bp) were called using GATK version 3.3-0 following best practices [34, 35]. HaplotypeCaller was used to generate gVCFs for each sample. Joint genotype calling was performed on all samples using GenotypeGVCFs to generate

a VCF file. GATK hard filters (SNPs: "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0"; Indels: "QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0") (https://software.broadinstitute.org/gatk/documentation/article?id=2806) were applied and calls that failed the filters were removed.

GATK's PhaseByTransmission was used to identify Mendelian violations that represent possible *de novo* variants. After removing Mendelian violations (MVs) that resulted from missing genotypes or had other anomalies (i.e., 5 MVs with read depth of 0 and 1,984 MVs with allelic depth of 0,0), we obtained 45,432 putative Mendelian violations. We also identified 62 scaffolds as deriving from the X chromosome. These scaffolds had significantly higher homozygosity and lower mean read depth among males (one-tailed t test, $q < 0.05$ for both mean homozygosity and read depth). MVs on these scaffolds and scaffolds shorter than 10 kb were removed. This resulted in an initial set of 34,189 putative MVs.

Filtering of putative mutations

Stringent filters are necessary to avoid potential false positive calls of *de novo* mutations [3, 36, 37]. To address this issue we applied the following filters to our initial set of MVs:

1. Removed 32,638 MVs with allelic balance less than 0.4 or greater than 0.6 in the child.
2. Removed 112 MVs that are not homozygous reference in both parents.
3. Removed 636 MVs with read depth below 20 or above 60 in any individual in the trio.
4. Removed 520 MVs where the alternate allele is present in an unrelated individual in the sample.

We define allelic balance as the fraction of reads that are a non-reference allele at a given site, meaning that a true heterozygous site should have allelic balance of roughly 0.5. Importantly, we observed that 95% of all initial MVs have allelic balance less than 0.4 (Figure S1A). This indicates that many of these initial calls are false positives. After these four filtering steps we find a total of 283 *de novo* mutations across our 14 trios (Data S1B).

Phasing mutations

Genotypes from three generations allow us to trace the parent of origin for *de novo* mutations transmitted to the third generation. We accomplished this by phasing chromosomal segments with respect to the grandparents (P generation in Figure 1A). Phase informative sites were identified in each family and assembled into haplotype blocks. We selected bi-allelic informative sites where: the grandparents had different genotypes, their offspring was heterozygous, and this individual's partner and offspring were not both heterozygous. The transmission of alleles at these sites can be unambiguously traced to one of the grandparents. We assembled these sites into blocks under the assumption that no more than one recombination occurred per 0.5 Mb interval [20, 38]. The phases of haplotype blocks supported by fewer than 100 informative sites were left unassigned, as were the phases of short scaffolds (less than 0.5 Mb). The parent of origin for *de novo* mutations transmitted to the third generation can then be established from the phase of their corresponding haplotype block.

Estimating mutation rates

To estimate mutation rates per generation per site (μ_g) we must consider rates of error. Our stringent filters ensure that we have few to no false positives; however, we expect that these filters removed a number of true *de novo* variants, leading to a substantial false negative rate (α). To estimate α resulting from the allelic balance filter, we used the distribution of allelic balance from the total set of 471,532,403 heterozygous autosomal sites in our sample. Unlike the initial set of MVs, the distribution of allelic balance for these sites conforms to the expected distribution for true heterozygous sites, with a single peak at about 0.5 (Figure S1B). We find that the number of heterozygous sites with allelic balance below 0.4 or above 0.6 is 206,358,774 resulting in an estimate of $\alpha = 0.44$. With a less stringent allelic balance filter of 0.3-0.7 the false negative rate falls to 0.29, but changing this filter does not greatly impact the number of mutations called (Figure S1C). These numbers represent false negative estimates from the allelic balance filter alone and in that sense only represent the upper-bound from that filter. False negatives may occur during other filtering steps, or due to mis-calls from the variant identification process, however these numbers are difficult to estimate. Therefore, we correct the observed number of mutations (m_g) in each trio using $\alpha = 0.44$ and an assumed false positive rate of 0. After correction we estimate that there are about 36 *de novo* mutations passed on in a single owl monkey generation.

To calculate the mutation rate per site, we counted the number of callable sites in each trio (C). A site was determined to be callable if it passed filters (1) and (4) in the child, filter (2) in the parents, and filter (3) in all individuals in the trio. We find an average number of callable sites of 2,207,614,768 in our 14 trios (range: 2,198,415,883-2,214,425,687). Mutation rates were then calculated by dividing the number of observed mutations (corrected for α) in a trio by 2 times the number of callable sites:

$$\mu_g = \frac{m_g}{((1-\alpha) * (2 * C))} \quad \text{(Equation 1)}$$

This results in mutation rates ranging from 0.63×10^{-8} to 1.5×10^{-8} with an average mutation rate of 0.81×10^{-8} among the 14 trios (Data S1C). Mutation rate was then regressed on father's age (A_M) (Figure 1B, solid line) with the resulting formula for a best fit line:

$$\mu_g = 3.74\text{x}10^{-9} + (A_M * 6.62\text{x}10^{-10}) \quad \text{(Equation 2)}$$

With an average haploid genome size of 2.21 billion base pairs, this means that 16.53 mutations accumulate in males and females before puberty at age 1 and that there are 2.92 additional mutations for every year of the father's life after puberty in owl monkeys.

Modeling mutation rates

Large-scale pedigree sequencing projects in humans have shown the importance of different life-stages in the determination of mutation rates [3, 11–17, 37, 39]. Models for predicting mutation rates generally account for the three important life stages in the mammalian germline [19, 26]. These life stages are (1) female (F), (2) male before puberty $(M0)$, (3) and male after puberty $(M1)$. The relative contribution of each of these stages must be accounted for when estimating mutation rates per generation [26] or per year [19, 26, 40]. Here, we re-frame this model in terms of reproductive longevity. Reproductive longevity depends on both the age of puberty in males (P_M) and the age of the father at conception of his offspring (A_M) and we find that it is the main determinant of mutation rate variation in primates. We define the value of reproductive longevity (RL) as:

$$RL = (A_M - P_M) \quad \text{(Equation 3)}$$

RL therefore measures the amount of time mutations have accumulated post-puberty in a male, which only occurs during stage *M*1.

To see how reproductive longevity affects the per-generation mutation rate, μ_g, we must model the combined contribution from all life stages. In any given period of time *t*, the mutation rate due to errors in DNA replication, μ_t, is simply a product of the mutation rate per cell division, μ_c, and the number of cell divisions that occur, d_t:

$$\mu_t = \mu_c * d_t \quad \text{(Equation 4)}$$

Since females (stage *F*) and pre-puberty males (stage *M*0) have a fixed number of cell divisions, their contribution to the mutation rate per-generation is constant and requires only the substitution of appropriate terms into Equation 4:

$$\text{Female contribution to } \mu_g : \mu_{gF} = \mu_c * d_F \quad \text{(Equation 5)}$$

$$\text{Pre-puberty male contribution to } \mu_g : \mu_{gM0} = \mu_c * d_{M0} \quad \text{(Equation 6)}$$

However, in males after puberty (stage *M*1) the number of cell divisions is a linear function of time, and the mutation rate per-generation in this life stage therefore depends on the yearly rate of cell division (d_{yM1}) and reproductive longevity (*RL*):

$$\text{Post-puberty male contribution to } \mu_g : \mu_{gM1} = \mu_c * d_{yM1} * RL \quad \text{(Equation 7)}$$

Finally, since an autosome will spend roughly half of its time in females and half in males, the mutation rate per generation (μ_g) for a given species is the average of the male and female contributions:

$$\mu_g = \frac{\mu_{gF} + (\mu_{gM0} + \mu_{gM1})}{2} \quad \text{(Equation 8)}$$

Given estimates of the underlying mutational parameters, this model allows us to predict the mutation rate as a function of reproductive longevity. In order to assess reproductive longevity in species that reach puberty at different times, we used published values for P_M. For owl monkeys, we set P_M at 1 year [21] (purple line in Figure 3), for humans, we used a value of P_M of 13.4 years [41] (orange line in Figure 3), for chimpanzees we used 7.5 years [42] (red line in Figure 3). The ages at conception for all parents in all studies of the mutation rate (points in Figure 3; Data S1D) were taken from the original papers [3, 11, 12, 14–17, 20, 37, 43].

This mutational model can easily be extended to calculate mutation rates per year (μ_y) [26, 40] by averaging the mutational contribution from each life stage per generation and weighting by the amount of time that passes:

$$\mu_y = \frac{(\mu_{gF} + (\mu_{gM0} + \mu_{gM1}))}{(A_F + P_M + RL)} \quad \text{(Equation 9)}$$

Considering yearly rates is useful when comparing long term evolutionary rates (*k*) between species since the neutral mutation rate (μ) is inextricably linked to the neutral substitution rate $(\mu_y = k_y)$.

Unlike μ_g, which is only dependent on the age of puberty and age at conception in males, μ_y is also dependent on the age of conception in the female (A_F; Figure S3). This means that increasing A_F will most likely decrease the yearly mutation rate because it increases the absolute amount of time without increasing the number of germline cell divisions. However, variation in either P_M or *RL* will have more complicated effects as they appear in both the numerator (as *RL* in $\mu_{g_{M1}}$) and the denominator. Increasing *RL* at some points in parameter space will increase μ_y, while decreasing it at others. Increasing P_M tends to increase μ_y(Figure S3).

Estimating mutational parameters from humans

Empirical observations from developmental studies and large-scale pedigree data from humans inform us about some of the underlying mutational parameters of our model (Equations 5, 6, and 7). For example, we use 31 and 34 as estimates for the number of cell divisions in human females (d_F) and males before puberty (d_{M0}) [44]. We use 16 days as the length of a single spermatogenic cycle (t_{sc}) [45], which means we expect $d_{yM1} = 23$ spermatogenic cycles to occur in a year if all spermatagonial cells are constantly dividing (but see next paragraph).

The remaining parameter of the model, μ_c, can be estimated from human pedigrees. We confirm the estimate of μ_c made by Amster and Sella [40] by using the μ_{gF} observed in Kong et al. [3] of 14.2, the number of female germline divisions, and rearranging Equation 5:

$$\mu_c = \frac{14.2}{31} = 0.458 \quad \text{(Equation 10)}$$

or 1.74×10^{-10} given a haploid genome size of 2.63 billion base pairs [3]. We assume this rate is the same between females and males before puberty. However, the observation that 2.01 mutations are passed on per year from the father after puberty [3] (the mutation rate per year in this lifestage, μ_{yM1}) could imply two things about μ_c in this life-stage: either the mutation rate per cell division has been reduced by an order of magnitude in males after puberty to 0.33×10^{-11} [40] or there are fewer than the expected 23 cell divisions per year [46]. There is no evidence to support such a dramatic reduction in the mutation rate per cell division, especially since there does not appear to be a large shift in mutational mechanisms between life stages [17]. The hypothesis that fewer cell divisions have taken place is also more likely based on observations that, of the two types of spermatagonial cells observed in humans, pale and dark, only pale cells actively divide [46, 47]. If dividing pale cells transition into non-dividing dark cells and vice versa, then not all spermatagonial cells necessarily undergo 23 spermatogenic cycles in a year and we must re-estimate d_{yM1}. If we assume the mutation rate per cell division in humans is constant before and after puberty, we can estimate the expected number of spermatogenic cycles per year (d_{yM1}):

$$dy_{M1}^{Human} = \frac{\mu_{yM1}}{\mu_c} = \frac{2.01}{0.458} = 4.39 \quad \text{(Equation 11)}$$

This implies that roughly only 19% of spermatagonial cells are in the pale dividing state at any given time.

Though either a decreased μ_c in males after puberty or a decreased proportion of dividing spermatagonial cells can be fit equally well to the model, we make predictions with the latter assumption. When predicting a mutation rate function for owl monkeys (Figure 1B) we also decrease the length of the spermatogenic cycle to $t_{sc}^{Owl\ monkey} = 10.2$ days [48] and adjust the expected d_{yM1} assuming 19% of spermatagonial cells are undergoing spermatogenesis at one time:

$$dy_{M1}^{Owl\ monkey} = \left(\frac{365}{t_{sc}^{Owl\ monkey}}\right) * 0.19 = 6.88 \quad \text{(Equation 12)}$$

However, when comparing mutation rate functions between species (Figure 3) all underlying mutational parameters are those estimated above from observations in humans, in order to demonstrate that minimal changes to the model can still make accurate predictions of mutation rate functions. Using species-specific parameters of spermatogenesis does not change our results (Figure S2).

Using Equation 9, we are also able to predict μ_y for an assumed age of puberty and average age of conception for humans and owl monkeys. For humans, with an age of puberty of roughly 13.4 years and average age of conception for both males and females of 30 years, we estimate a yearly mutation rate of 0.4×10^{-9} mutations per site per year (orange point in Figure S3). This is remarkably close to the calculated average yearly rate from several studies of human mutation rates: from 0.43×10^{-9} mutations per site per year [17] to 0.5×10^{-9} mutations per site per year [25]. With an average age of puberty of 7.5 years and average age at reproduction of 24.3, we predict the yearly chimp mutation rate is to be 0.48×10^{-9} per site per year (red point in Figure S3), on par with the previous estimate of 0.46×10^{-9} [20]. For owl monkeys we assumed a puberty age of 1 year and average ages of conception of 6.64 and 6.53 years for males and females, respectively. Using these values we estimate a yearly mutation rate of 1.2×10^{-9} mutations per site per year, three times higher than the yearly human rate (purple point in Figure S3).

QUANTIFICATION AND STATISTICAL ANALYSIS

All data were analyzed using Python v2.7 and R v3.4.1. Linear regression was performed on the observed mutation rate per trio and paternal age to obtain the solid lines in Figures 1B and S2. To assess how well our model predicts this relationship, we performed an *F*-test on the residuals of the observed relationship (solid lines in Figures 1B and S2) and the predicted relationship (dashed lines in Figures 1B and S2). Comparing variance in the residuals between the two lines captures variation in both the slope and intercept of the predicted and observed lines. A similar *F*-test was performed on the human study points in Figures 3 and S3.

DATA AND SOFTWARE AVAILABILITY

Raw sequence reads for the 30 owl monkey individuals have been deposited as NCBI BioProject: PRJNA451475 (https://www.ncbi.nlm.nih.gov/bioproject/451475).

ADDITIONAL RESOURCES

All code used to analyze data and generate figures is available as an R Markdown document at the following link: https://github.com/gwct/owl-monkey.

Current Biology, Volume *28*

Supplemental Information

Reproductive Longevity Predicts Mutation Rates in Primates

Gregg W.C. Thomas, Richard J. Wang, Arthi Puri, R. Alan Harris, Muthuswamy Raveendran, Daniel S.T. Hughes, Shwetha C. Murali, Lawrence E. Williams, Harsha Doddapaneni, Donna M. Muzny, Richard A. Gibbs, Christian R. Abee, Mary R. Galinski, Kim C. Worley, Jeffrey Rogers, Predrag Radivojac, and Matthew W. Hahn

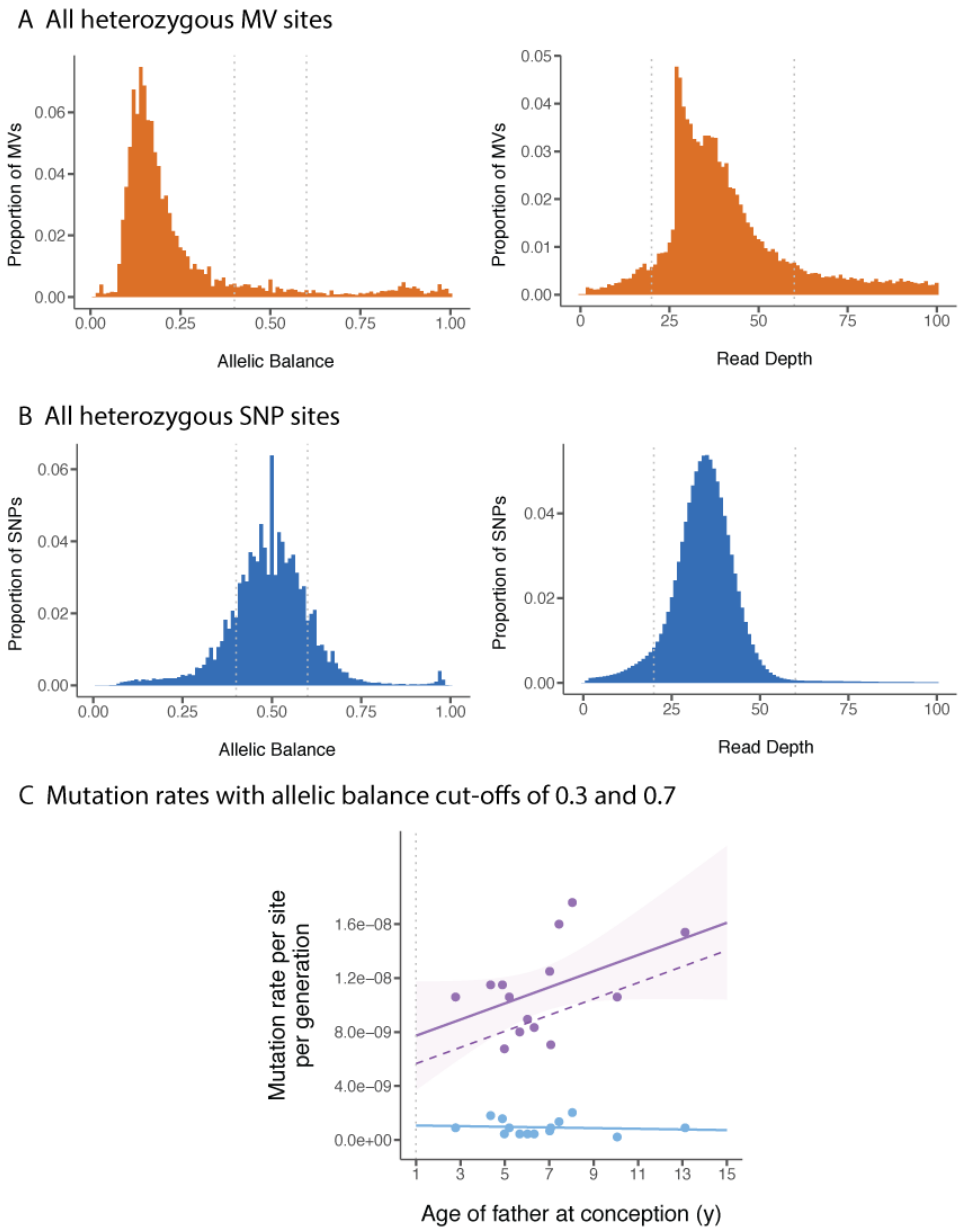

Figure S1: **Read depth and allelic balance distributions and the effect of varying the allelic balance cut-off on rate estimates. Related to Figure 1. A,** read depth and allelic balance distributions for all unfiltered Mendelian violations (MVs) in the 30 owl monkey individuals. The cut-offs used to filter MVs are indicated by the vertical dotted grey lines. **B**, read depth and allelic balance distributions for all SNP sites in the 30 owl monkey individuals. Filtering cut-offs are again indicated by the vertical dashed grey lines for comparison. **C,** mutation rate estimates for the 14 owl monkey trios when using a less stringent allelic balance cut-off to 0.3 and 0.7 (purple dots). A linear regression still shows a correlation with father's age (solid purple line; (R^2=0.15, d.f.=12 P=0.10); shaded area indicates 95% confidence interval) that is not significantly different from our model's prediction (dashed purple line; F=1.0, d.f.=13, P=1.0). Mutations at CpG sites (blue dots) are not correlated with father's age (blue line). The grey dotted line indicates age of puberty for owl monkeys.

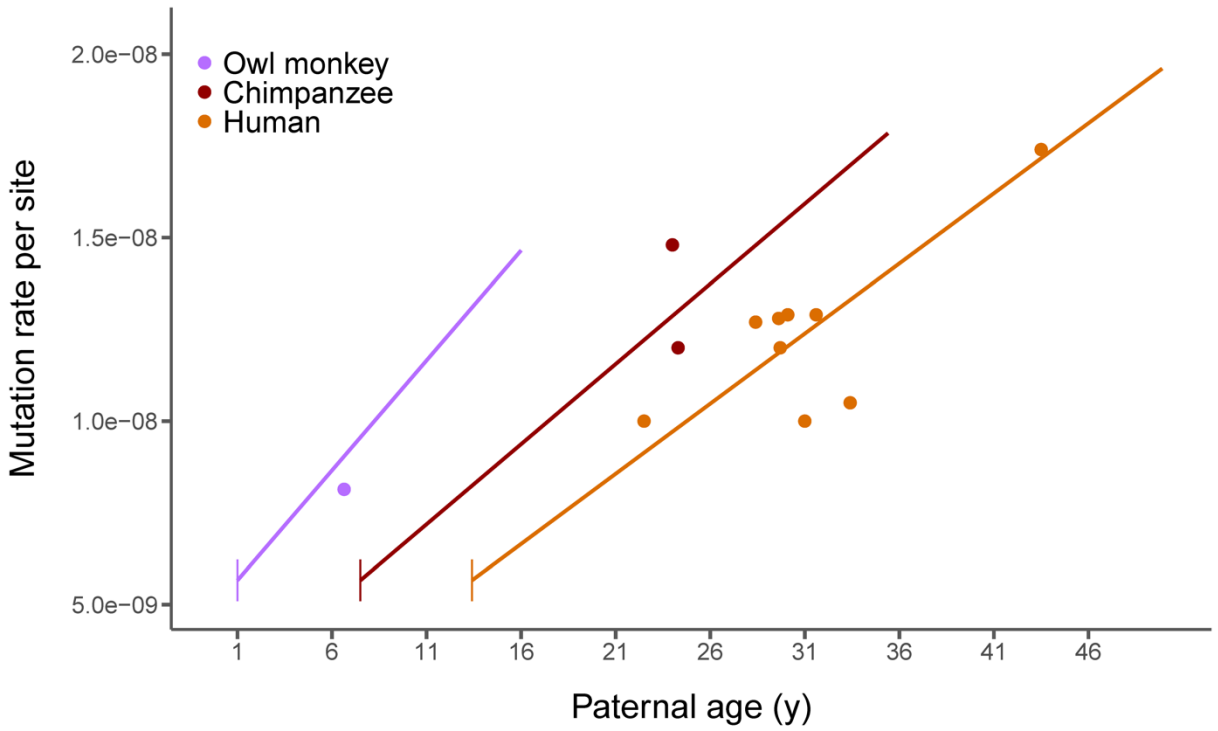

Figure S2: Functions of mutational accumulation predicted using species specific rates of spermatogenesis. **Related to Figure 3.** Using species specific rates for the three species with high-quality mutation estimates when making predictions from our model of reproductive longevity (lines; equations 3-8 in Methods) does not greatly affect the fit of the model to published estimates of mutation rates (points; see Data S1D for references). Vertical line segments represent the age of puberty for each species. In this figure we used $t_{sc}^{Owl\ monkey} = 10.2$ [S1], $t_{sc}^{Chimp} = 14$ [S2], $t_{sc}^{Human} = 16$ [S3].

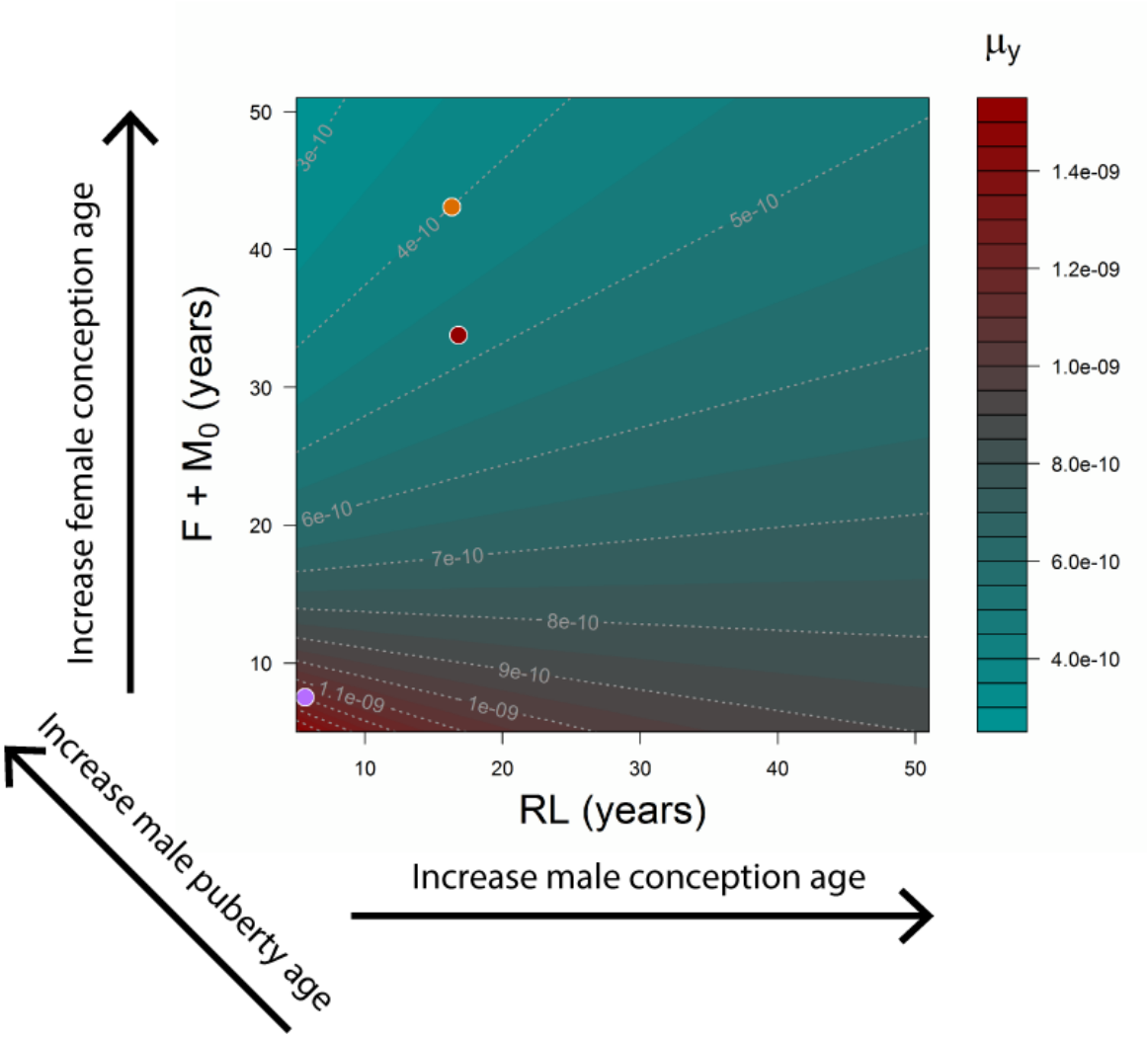

Figure S3: **Modeling mutation rate per year over various parental ages at reproduction and puberty. Related to STAR Methods.** Mutation rates per year (μ_y) are a function of the length of all three life stages of the mammalian germline: Female (F), males before puberty (M_0), and males after puberty (RL). Increasing F consistently decreases μ_y, while increasing M_0 by increasing the age of puberty in males tends to increase μ_y (note that the y-axis plots the sum of these two parameters). Increasing the length of RL either by decreasing the age of puberty or increasing the average age of conception in males has varying effects based on the length of the other two stages. Points are the predicted mutation rates per year for humans (orange), chimpanzees (red), and owl monkeys (purple).

Supplementary References

S1. Derooij, D.G., Vanalphen, M.M.A., and Vandekant, H.J.G. (1986). Duration of the cycle of the seminiferous epithelium and its stages in the rhesus-monkey (*Macaca mulatta*). Biology of Reproduction *35*, 587-591.

S2. Smithwick, E.B., Young, L.G., and Gould, K.G. (1996). Duration of spermatogenesis and relative frequency of each stage in the seminiferous epithelial cycle of the chimpanzee. Tissue Cell *28*, 357-366.

S3. Heller, C.G., and Clermont, Y. (1963). Spermatogenesis in man: an estimate of its duration. Science *140*, 184-186.