doi:10.1111/evo.13383

Direct estimate of the spontaneous germ line mutation rate in African green monkeys

Susanne P. Pfeifer[1,2,3,4]

[1]*School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland*

[2]*Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland*

[3]*School of Life Sciences, Arizona State University (ASU), Tempe, Arizona 85281*

[4]*E-mail: susanne.pfeifer@asu.edu*

Received July 6, 2017
Accepted October 9, 2017

Here, I provide the first direct estimate of the spontaneous mutation rate in an Old World monkey, using a seven individual, three-generation pedigree of African green monkeys. Eight de novo mutations were identified within ~1.5 Gbp of accessible genome, corresponding to an estimated point mutation rate of 0.94×10^{-8} per site per generation, suggesting an effective population size of ~12000 for the species. This estimation represents a significant improvement in our knowledge of the population genetics of the African green monkey, one of the most important nonhuman primate models in biomedical research. Furthermore, by comparing mutation rates in Old World monkeys with the only other direct estimates in primates to date–humans and chimpanzees–it is possible to uniquely address how mutation rates have evolved over longer time scales. While the estimated spontaneous mutation rate for African green monkeys is slightly lower than the rate of 1.2×10^{-8} per base pair per generation reported in chimpanzees, it is similar to the lower range of rates of 0.96×10^{-8}–1.28×10^{-8} per base pair per generation recently estimated from whole genome pedigrees in humans. This result suggests a long-term constraint on mutation rate that is quite different from similar evidence pertaining to recombination rate evolution in primates.

KEY WORDS: **African green monkey, germ line, mutation rate, pedigree, vervet monkey, whole genome sequencing.**

Regarded as the raw material of evolution, germ line mutations continuously give rise to new heritable variation in the genome, making them the primary source of genetic diversity within populations as well as divergence between species. Thus, characterizing the mechanisms by which mutations arise and accurately estimating their spontaneous rate of occurrence is of great importance both mechanistically, but also for improved evolutionary inference related to the genetic basis of disease, characterizing relationships among populations and species, and dating demographic and selective events. In fact, much of our current understanding of the chronology of human evolution assumes a constant mutation rate over time (a "molecular clock") to date events from genetic data for which there is no fossil record available, despite known variation both in mutation rates as well as in the mutation spectra within primates (Elango et al. 2006; Kim et al. 2006; Harris 2015; Amster and Sella 2016; Gao et al. 2016; Moorjani et al. 2016a; Harris and Pritchard 2017; Matheieson and Reich 2017), as expected from differing life-history traits ranging from generation times to mating system to rates of spermatogenesis (Yi et al. 2002; Wilson Sayres et al. 2011; Xu et al. 2012).

Despite this significance as a parameter in evolutionary genetics, little is known about the rate at which mutations are introduced in to the genomes of different species, including primates. Accurately estimating germ line mutations is often challenging owing to their slow rate of occurrence (Drake et al. 1998; Lynch 2010a), making experiments studying mutation rates both complex and imprecise. Because mutations are comparatively rare in mammals, estimates of mutation rates have historically been obtained from two sources: (1) classical genetic approaches, using equilibrium frequencies of fully penetrant dominant monogenic Mendelian disorders with major phenotypic effects in pedigrees (e.g., Haldane 1935; Kondrashov 2003; Lynch 2010b); or (2) phylogenetic approaches based on the expectation that under

neutrality the mutation rate is equal to the rate of divergence (Kimura 1968), thus focusing on putatively neutral sites (e.g., pseudogenes; Kondrashov and Crow 1993; Drake et al. 1998; Nachman and Crowell 2000; Chimpanzee Seq. Anal. Consort. 2005).

However, there is considerable uncertainty in the estimates obtained from these indirect methods as they rely on often inaccurate parameter estimates. Thereby, the major drawback of calculating mutation rates from incidents of genetic disease is that the mutational target size (i.e., the number of sites producing the penetrant phenotype) together with the strength of selection must be known a priori to obtain a per-site mutation rate (Haldane 1935; Kondrashov 2003; Lynch 2010b). Moreover, disease incident-based methods are, by their nature, limited both to specific types of diseases as well as to specific regions in the genome, thus they do not provide a comprehensive picture of genome-wide mutation rates. In contrast, uncertainty in mutation rate estimates calculated from sequence divergence between species is mainly caused by uncertainty in the estimates of divergence times (often drawn from fossil evidence), ancestral population sizes, and generation times. Contributing to the uncertainty in estimation is the assumption of neutrality—with the incorporation of nonneutral sites generally expected to result in an underestimation of mutation rates owing to the action of purifying selection. Further, other evolutionary processes that act analogously to selection (e.g., GC-biased gene conversion; Duret and Galtier 2009) will influence substitution rates even at sites that do not affect fitness. In addition, as information is averaged across multiple generations, no information about individual mutation rates or differences between sexes can be obtained.

In recent years, advances in high-throughput sequencing technologies have permitted the large-scale analysis of genomes of different individuals of a species, making it now possible to directly estimate mutation rates on a genome-wide level from parent-offspring trios or larger multigeneration pedigrees, even in nonmodel organisms. In animals, pedigree studies have thus far been carried out in humans (1000 Genomes Project Consortium 2010; Awadalla et al. 2010; Roach et al. 2010; Conrad et al. 2011; Campbell et al. 2012; Kong et al. 2012; Michaelson et al. 2012; Jiang et al. 2013; Besenbacher et al. 2015; Francioli et al. 2015; Yuen et al. 2015; Goldmann et al. 2016; Rahbari et al. 2016; Wong et al. 2016), chimpanzees (Venn et al. 2014), *Drosophila melanogaster* (Keightley et al. 2014), the butterfly *Heliconius melpomene* (Keightley et al. 2015), the European honey bee *Apis mellifera* (Yang et al. 2015), the bird *Ficedula albicollis* (Smeds et al. 2016), and the bumblebee *Bombus terrestris* (Liu et al. 2017) (Fig. 1). Pedigree-based analyses provide an opportunity to survey mutation rates in a comprehensive and relatively unbiased way, substantially improving our understanding of how much variation in mutation rates exist in natural populations, and over which time scales these rates evolve.

For anthropocentric reasons, humans have been a focal point of many studies addressing the process of mutation, given the importance of this knowledge for better understanding both genetic disease and population history. Direct investigation of germ line mutations from subsequent generations in whole genome pedigrees have yielded rates of $0.96 \times 10^{-8} - 1.28 \times 10^{-8}$ per base pair per generation, but 95% CI of most studies were wide: $0.67 \times 10^{-8} - 1.7 \times 10^{-8}$ (1000 Genomes Project Consortium 2010; Awadalla et al. 2010; Roach et al. 2010; Conrad et al. 2011; Campbell et al. 2012; Kong et al. 2012; Michaelson et al. 2012; Jiang et al. 2013; Besenbacher et al. 2015; Francioli et al. 2015; Yuen et al. 2015; Goldmann et al. 2016; Rahbari et al. 2016; Wong et al. 2016). A similar rate of 1.2×10^{-8} per base pair per generation has been reported in chimpanzees but with differences in the distribution of germ line mutations, both among different types of mutations as well as between sexes in the two species (Venn et al. 2014).

In humans, there is a well-known discrepancy between mutation rates obtained from whole genome pedigree data and those obtained from phylogenetic approaches—the former being roughly twofold lower than the indirectly inferred sex-averaged mutation rate based on the genetic divergence between humans and chimpanzees. Specifically, the human mutation rate inferred from phylogeny is estimated to be 2.5×10^{-8} in pseudogenes (where selection is not believed to be a confounding factor), based on a human-chimp nucleotide divergence per site of 0.013, a species divergence time of five million years, and an ancestral effective population size of 10,000 individuals (Nachman and Crowell 2000). Taken together, this methodological difference in estimation raises the possibility that mutation rates may have evolved relatively rapidly across primates (Scally and Durbin 2012). Thereby, a gradual "hominoid slowdown" of the germ line mutation rate along the human ancestral lineage relative to other primates, perhaps due to an increase in generation time, might contribute to the observed decrease in yearly mutation rate (Goodman 1963; Yi et al. 2002; Elango et al. 2006; Scally and Durbin 2012; Yi 2013; Scally 2016). On the other hand, it has been suggested that this effect may be mitigated to an extent by the paternal age effect (i.e., an increased number of mutations with parental age), as well as evolutionary changes in other life-history traits, behavior, and cycle times of cellular processes in gametogenesis (Ségurel et al. 2014; Amster and Sella 2016; Scally 2016; and see Moorjani et al. 2016a for an in-depth discussion on how to reconcile human mutation rate estimates from pedigrees with estimates from rates of substitution).

Here, uniquely using whole-genome sequence data of a three-generation pedigree (Fig. 2), I investigate the rate and patterns of germ line mutations in the most populous African nonhuman primate, the African green monkey. This Old World monkey, also

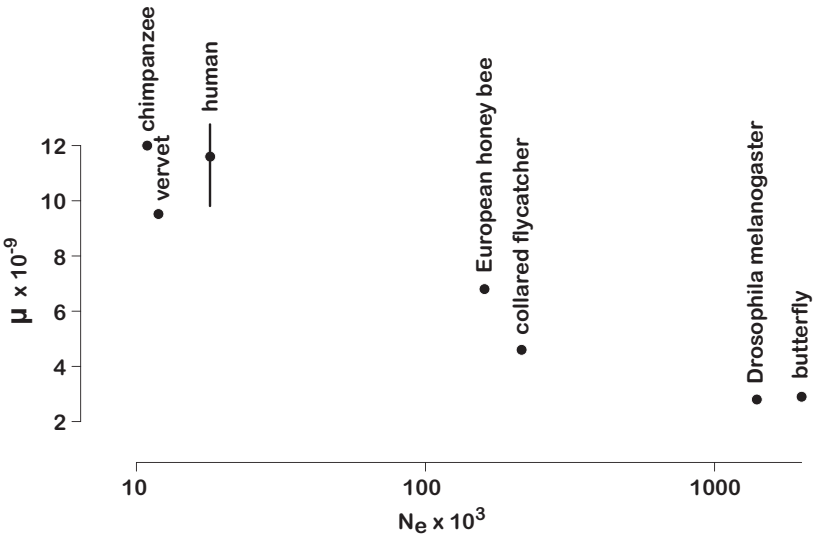

Figure 1. Relationship between mutation rates per site per generation (μ) and effective population sizes (N_e) directly estimated from parent-offspring trios or larger multi-generation pedigrees in humans (average rate reported by recent large-scale pedigree studies (Kong et al. 2012; Michaelson et al. 2012; Yuen et al. 2015; Rahbari et al. 2016; Wong et al. 2016); inter-individual variance around the average mutation rate estimated is indicated as a black line), chimpanzees (Venn et al. 2014), *Drosophila melanogaster* (Keightley et al. 2014), the butterfly *Heliconius melpomene* (Keightley et al. 2015), the European honey bee *Apis mellifera* (Yang et al. 2015; Wallberg et al. 2014), the collared flycatcher *Ficedula albicollis* (Smeds et al. 2016; Burri et al. 2015), and vervet monkeys (this study).

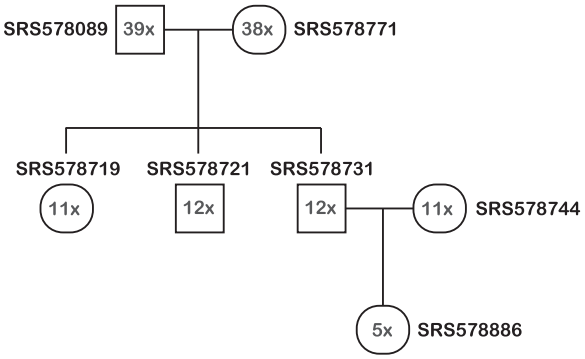

Figure 2. Structure of the seven individual, three generation African green monkey pedigree (square: male; circle: female), indicating average genome-wide sequencing depth per individual.

known as the vervet monkey, shared a most recent common ancestor with humans roughly 25 million years ago (Kumar and Hedges 1998). Germ line mutations were identified using the parents and children, and confirmed by stable inheritance to the grandchild. This strategy avoids the need of genotyping or resequencing for validation, as demonstrated by Wang and Zhu (2014), who found that the de novo germ line mutation rate in humans estimated from a three-generation pedigree is consistent with estimates obtained from studies using genotyping and resequencing for validation. In addition, utilizing data from a third generation greatly helps in discerning genuine germ line mutations from somatic mutations—a potential problem when using blood and other nongerm line samples for sequencing (Ségurel et al. 2014).

Accurate calling of de novo mutations (DNM) from high-throughput sequencing data is challenging and complicated by the inevitable presence of sequencing errors, often leading to spurious variant calls (see review of Pfeifer 2017a). Thus, several highly stringent computational filters were developed to mitigate false positives. This included the removal of artifacts caused by misalignment in highly repetitive genomic regions, the incorporation of uncertainty via genotype likelihoods, as well as the implementation of test statistics to ensure that de novo mutation sites were not known to be polymorphic in previously sampled populations (Huang et al. 2015; Pfeifer 2017b). In addition, candidate mutations were manually curated using the Integrated Genomics Viewer (Thorvaldsdóttir et al. 2012) to identify false positives caused by misaligned reads. Apart from increasing specificity, the application of a stringent set of filter criteria also decreases the number of loci at which genuine de novo mutations can be identified, potentially leading to an underestimate of the mutation rate. Thus, to obtain a more accurate estimate of the spontaneous mutation rate per site in the species, an unbiased estimate of the number of callable genomic sites was obtained and false-negative rates in both DNM discovery and validation were calculated.

Germ line mutations depend on several biological processes that are well known to vary between different primate species. For example, the lifespan of African green monkeys in captivity ranges between 11 and 13 years (likely an upper limit for their age in the wild as they are heavily preyed upon; Fairbanks and McGuire 1985) with a generation time of ~8.5 years

(Warren et al. 2015). In contrast, the average human generation time ranges from 26 to 30 years (Moorjani et al. 2016b), with a life expectancy of ~71.5 years (according to the United Nations). Thus, under the generation-time effect hypothesis, it is expected that vervets will have a higher mutation rate per unit time than humans (Parker 1970; Li et al. 1996). In addition, both strength of sperm competition as well as the average duration of spermatogenesis, which vary between species (e.g., ranging from 36 days in rhesus macaque (Abee et al. 2012) to 74 days in humans (Heller and Clermont 1964; Amann 2008)), will likely affect mutation rates in different primates. As a consequence, a comparison of de novo mutation rates between an Old World monkey and great apes can help to better illuminate variation in mutation rates across primates, presenting a significant step forward in our understanding of the molecular clock. Furthermore, this estimate provides a more correct scaling factor for characterizing the population history of the African green monkey itself–a species, which stands as one of the most important nonhuman primate models in biomedical research (e.g., Broussard et al. 2001; Lemere et al. 2004; Emborg 2007; Chapman et al. 2016).

Materials and Methods

SAMPLES AND ALIGNMENT

Publicly available whole-genome sequence data for seven African green monkey individuals from a three-generation pedigree (Fig. 2), housed at the Wake Forest University Primate Center Vervet Research Colony, was downloaded from SRA (i.e., blood samples were previously sequenced using Illumina HiSeq2000 for seven *C. a. sabaeus* individuals with the identifiers SRS578089, SRS578771, SRS578719, SRS578721, SRS578731, SRS578744, and SRS578886; deposited under BioProject accession number PRJNA240242). Reads from each read group were aligned to the repeat-masked *Chlorocebus sabaeus* reference genome v.1.1, publicly available from NCBI (GenBank accession number GCA_000409795.2; Warren et al. 2015) using BWA-MEM v.0.7.13 (Li and Durbin 2009). A reference assembly for Epstein-Barr virus (NCBI Reference Sequence NC_007605.1) was included in the alignment step, functioning as a decoy to absorb reads that did not originate from African green monkey DNA, thus helping to decrease the number false-positive variant calls.

Aligned reads were deduplicated using Picard v.2.1.1 (http://picard.sourceforge.net) and multiple sequence realignments were calculated around insertions and deletions (indels) using the Genome Analysis Toolkit (GATK) IndelRealigner v.3.5 (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013), simultaneously adjusting Base Alignment Qualities (Li 2011)—a procedure that has been shown to effectively remove the majority of false variant calls due to misaligned indels (Ness et al. 2012). Next, base quality scores were recalibrated using GATK's BaseRecalibrator v.3.5 using ~500 k variants from the genome-wide SNP panel of the Vervet Genetic Mapping Project (downloaded from the European Variant Archive: study number PRJEB7923; Huang et al. 2015). After preprocessing each read group individually, reads originating from the same sample were merged and per-sample duplicates were marked using Picard v.2.1.1 to eliminate PCR duplicates in the analyses.

GENOTYPE CALLING

For each individual, variant sites were called using GATK's HaplotypeCaller v.3.5 using the default heterozygosity rate of 0.001 (McKenna et al. 2010; DePristo et al. 2011; Van der Auwera et al. 2013). Next, a single-sample genotype calling strategy was employed, calling genotypes separately for both parents (SRS578089 and SRS578771) and the child (SRS578731) using GATK's GenotypeGVCFs v.3.5. In contrast to commonly applied multi-sample genotype calling strategies, often leading to an undercalling of rare variants (Han et al. 2014), this single-sample genotype calling strategy enables the detection of rare de novo mutations in the trio. To obtain counts for reads supporting the alternative allele in each individual of the pedigree, all individuals were in addition jointly genotyped using GATK's GenotypeGVCFs v.3.5. Thereby, nonvariant sites were included in the genotype call to generate a genome mask across individuals to avoid false positives in misaligned regions or regions with sparse coverage. Nonvariant sites were subject to the same filter criteria than variant sites (see "Identification of DNM candidates") and, as such the genome mask provides valuable information on the number of nucleotides for which there was complete power to identify de novo mutations, thus enabling the calculation of a mutation rate per site.

IDENTIFICATION OF DNM CANDIDATES

Due to the fact that de novo mutations should only occur in a single individual, the variant quality recalibration (VQSR) suggested by GATK's Best Practices can not be applied to this dataset as it would likely filter out genuine de novo mutations. Instead, a set of highly stringent hard filter criteria was applied to weed out potential false positives. Following Keightley et al. (2014, 2015), sites with sparse coverage in either of the parents (i.e., read depth smaller than 10 in either parent) were excluded from the dataset (note that this also excludes all bases called as "N"). In addition, sites with a low quality (i.e., either marked by GATK with the "LowQual" flag or exhibiting a mapping quality less than 60, corresponding to an error rate of 1 in 10^6) were removed to avoid false variant calls due to spurious alignments. Due to the fact that repetitive genomic regions present a serious challenge in read alignment, often leading to an excess of heterozygous genotype calls in misaligned regions, SNPs within repeats were excluded from further analyses. To enable validation using stable

Table 1. Mutations segregating in the pedigree.

chr	# SNPs	# candidates	# validated
1	248,175	12	0
2	200,896	4	0
3	198,591	7	0
4	201,216	10	0
5	168,589	4	0
6	106,687	18	0
7	265,224	6	0
8	314,029	3	0
9	271,392	12	1
10	244,420	19	0
11	247,680	17	2
12	197,695	13	2
13	207,313	7	0
14	221,750	9	2
15	193,128	8	0
16	133,397	9	0
17	173,220	6	0
18	162,808	9	0
19	70,662	4	0
20	259,136	18	0
21	249,749	6	0
22	186,341	9	0
23	163,915	5	0
24	174,233	8	0
25	207,136	4	0
26	119,265	7	1
27	115,546	7	0
28	45,821	0	0
29	55,067	3	0
Total	5,403,081	244	8

Candidate mutations were required to have a high confidence in both parents (i.e., both parents were homozygous for the reference allele and no reads were present that supported the alternative allele), to be heterozygous in only one of the three children (SRS578731) with at least 20% of the reads supporting the alternative allele, while the other two children (SRS578719 and SRS578721) were purely homozygous for the reference allele (thus no other individuals in the same generation were heterozygous or homozygous for the alternative allele). On the autosomes, 244 sites with Mendel inconsistent genotypes (i.e., de novo mutation candidates) were detected. Following Wang and Zhu (2014), candidates were validated using their stable inheritance to the next generation (i.e., at each candidate site, the partner (SRS578744) was purely homozygous for the reference allele and at least one read supporting the alternative allele was present in the grandchild (SRS578886)) —limiting the data set to eight putative de novo mutations.

inheritance to the next generation, sites with incomplete genotype information across individuals were removed, leading to a dataset of 5,403,081 variants (Table 1) and 1,470,115,269 monomorphic sites on the autosomes (corresponding to 57% of the autosomal genome being accessible; Table 2).

To guard against false positives through mis-genotyping, candidate mutations were required to have a high confidence in both parents (i.e., following Keightley et al. (2014, 2015), the parents were required to be homozygous for the reference allele with no reads present that supported the alternative allele, indicative of potential parental mosaicism (i.e., two or more genotypes are present in a sample of cells)). Note that the exclusion of segregating sites prohibits the detection of de novo mutation at these loci, potentially down-biasing the mutation rate estimate. In addition, candidate mutations were required to be heterozygous in only one of the three children (SRS578731) with at least 20% of the reads supporting the alternative allele, while the other two children (SRS578719 and SRS578721) were purely homozygous for the reference allele (i.e., no alternative reads were present in the children where the variant was not called, thus no other individuals in the same generation were heterozygous or homozygous for the alternative allele).

On the autosomes, 244 sites with Mendel inconsistent genotypes (i.e., de novo mutation candidates) were detected in the focal child (SRS578731) (Table 1). Mutation rate estimates obtained from parent-offspring data or larger multi-generational pedigrees in humans have shown considerable variation (e.g., Wong et al. (2016) observed a nearly fourfold inter-individual difference in human mutation rates). Potentially consistent with this observation, variation in the number of sites with Mendelian inconsistencies was detected in the three children of this study (i.e., SRS578719: 274 candidates; SRS578721: 427 candidates; SRS578731: 244 candidates)—however, given the lack of a third generation for two of the three siblings, numbers of de novo mutation could not be directly compared.

Due to the fact that de novo mutations are rare events, candidates also detected as variation segregating in unrelated individuals are likely false positives. Therefore, the candidate mutations were screened against segregating variation in the species, previously sampled from over 700 African green monkey individuals (Huang et al. 2015: 497,163 SNPs; Pfeifer 2017b: 1,795,643 SNPs). All 244 putative de novo mutations were monomorphic in the previously studied large population samples.

MANUAL CURATION OF DNM CANDIDATES

Misalignment in paralogous regions is a known problem and indeed one of the most frequent causes of false-positive variants in resequencing studies (see recent review by Pfeifer 2017a). The available African green monkey reference assembly was (like those of many other nonmodel organisms) based on genomic data obtained from a single individual (an adult male monkey, animal ID 1994-021), thus it was difficult during assembly to distinguish variation present as heterozygosity in this single individual from truly paralogous regions (e.g., Hahn et al. 2014). In addition, several regions might not (or at least not accurately) be represented in this draft reference assembly. These gaps and errors in the assembly might lead to alignment errors (especially of short reads

Table 2. Callable sites.

chr	Length	Sufficient coverage	High quality	Nonrepetitive	100% GT information
1	126,035,930	120,275,254	115,411,714	78,853,092	71,966,181
2	90,373,283	87,566,436	84,948,823	58,398,689	53,541,668
3	92,142,175	89,700,001	87,123,493	58,635,341	52,890,395
4	91,010,382	88,632,790	85,640,969	56,554,215	51,148,587
5	75,399,963	72,421,153	70,137,203	46,243,571	42,513,477
6	50,890,351	47,267,014	44,893,201	26,811,123	24,592,658
7	135,778,131	132,760,149	128,455,288	83,516,621	74,982,045
8	139,301,422	136,190,030	132,172,547	88,846,358	80,777,906
9	125,710,982	122,461,790	118,350,287	80,003,545	73,101,754
10	128,595,539	123,610,612	118,940,579	80,428,786	72,891,148
11	128,539,186	124,467,290	120,073,910	78,626,508	71,517,695
12	108,555,830	105,515,789	102,306,437	68,804,549	62,776,659
13	98,384,682	92,963,694	88,387,552	58,766,597	53,073,392
14	107,702,431	103,181,035	99,127,204	67,231,055	61,275,527
15	91,754,291	89,225,215	86,338,615	57,153,823	51,775,743
16	75,148,670	72,055,543	69,362,452	45,297,441	41,669,479
17	71,996,105	69,406,376	66,789,491	44,766,462	40,530,208
18	72,318,688	70,362,993	68,405,526	47,059,773	42,757,570
19	33,263,144	30,960,562	29,619,462	20,077,273	18,595,431
20	130,588,469	126,863,391	122,556,523	81,720,132	74,802,389
21	127,223,203	124,095,424	119,972,244	80,371,457	72,843,440
22	101,219,884	97,167,218	93,725,922	63,494,722	57,788,494
23	82,825,804	80,672,286	78,285,473	52,539,904	47,825,896
24	84,932,903	80,139,066	76,297,180	50,763,095	46,083,604
25	85,787,240	83,493,261	80,904,427	54,134,475	49,200,553
26	58,131,712	56,065,124	54,188,382	36,265,275	33,218,955
27	48,547,382	46,644,345	45,143,604	30,399,915	27,598,353
28	21,531,802	19,704,930	18,802,977	11,179,119	10,277,958
29	24,206,276	22,572,809	21,692,110	14,748,179	13,501,185
Total	2,607,895,860	2,516,441,580	2,428,053,595	1,621,691,095	1,475,518,350

Number of nucleotides for which there was complete power to identify de novo mutations after filtering (i.e., callable sites). Excluded from any analyses were sites (i) with sparse coverage in either of the parents (i.e., read depth smaller than 10 in either parent), (ii) with a low quality (i.e., either marked by GATK with the "LowQual" flag or exhibiting a mapping quality less than 60, corresponding to an error rate of 1 in 10^6), (iii) within repetitive regions, or (iv) with incomplete genotype information (GT).

like the ones used in this study), and worse, mappers often give reads originating from paralogous regions high mapping qualities due to a lack of alternative placements—making it hard to filter them out. Unfortunately, these issues can computationally often only be mitigated by using a better reference assembly and thus, a manual curation step was necessary to weed out false positives.

Candidate mutations were manually curated using the Integrated Genomics Viewer (Thorvaldsdóttir et al. 2012), following the guidelines outlined in Keightley et al. 2014, to visually identify false positives caused by misaligned reads or sequencing errors. Candidate mutations were assumed to be false positives if (i) variants in complete association with the alternative base calls at the candidate sites co-occurred on the same reads, or (ii) they occurred in regions with an unusually high clustering of (often low quality) nonreference alleles, frequently only supported by two reads of the same read pair (i.e., forward and reverse strand) in close proximity to a genuine segregating variant—indicative of spurious alignments.

Two hundred twenty-two out of 244 candidate mutations failed the visual inspection (i.e., failed variants per above categories: (i) 126 and (ii) 96)—a ∼90% reduction of the dataset. Earlier work by Keightley et al. (2014) identified 88 candidate mutations from a 14-individual pedigree in the model organism Drosophila of which 78 candidate mutations (88.6%) failed the manual inspection. As such, a large reduction of the dataset during the manual curation can be expected (especially taking into account that the African green monkey draft reference assembly is of lesser quality than that of the widely studied model organism Drosophila). Consistent with the assumption that the manual

Table 3. De novo candidate mutations.

chr	Position	Annotation	Base call	Depth father	Depth mother	Depth focal child	Depth sister	Depth brother	Depth partner	Depth grand-child	FNR
			WT/M	WT/M	WT/M	WT/M	WT/M	WT/M	WT/M	WT/M	
5	9,987,594	Intergenic	C/T	32/0	40/0	5/5	13/0	7/0	10/0	1/0	0.970
9	**81,525,708**	**Intergenic**	**C/T**	**39/0**	**45/0**	**4/8**	**14/0**	**15/0**	**6/0**	**3/1**	**-**
10	28,304,386	Intronic	G/A	29/0	50/0	14/5	9/0	9/0	8/0	2/0	0.425
11	23,292,192	Intergenic	G/T	42/0	11/0	4/8	12/0	6/0	10/0	1/0	0.970
11	**42,745,992**	**Intergenic**	**C/G**	**27/0**	**38/0**	**3/4**	**6/0**	**10/0**	**7/0**	**3/3**	**-**
11	**95,728,494**	**Intronic**	**T/A**	**32/0**	**42/0**	**4/5**	**7/0**	**18/0**	**13/0**	**7/3**	**-**
12	**77,257,469**	**Intergenic**	**G/T**	**38/0**	**39/0**	**12/4**	**6/0**	**15/0**	**15/0**	**1/3**	**-**
12	**80,638,728**	**Intronic**	**C/T**	**33/0**	**42/0**	**3/5**	**7/0**	**17/0**	**12/0**	**3/9**	**-**
13	38,128,814	Intronic	G/A	20/0	38/0	5/7	10/0	9/0	23;0	3/0	0.203
13	66,221,313	UTR3	G/A	46/0	26/0	4/11	15/0	15/0	10/0	5/0	0.089
14	**96,796,726**	**Intergenic**	**G/A**	**29/0**	**36/0**	**9/4**	**17/0**	**10/0**	**14/0**	**1/2**	**-**
14	**105,189,500**	**Intergenic**	**G/T**	**39/0**	**43/0**	**5/8**	**11/0**	**10/0**	**10/0**	**2/1**	**-**
16	31,414,248	Intronic	C/T	44/0	26/0	8/6	18/0	15/0	14/0	2/0	0.425
16	42,187,036	Intronic	G/C	38/0	23/0	6/4	12/0	9/0	16/0	4/0	0.130
16	59,836,232	Intronic	G/A	35/0	37/0	8/5	7/0	8/0	9/0	2/0	0.425
17	40,493,231	Intergenic	G/T	34/0	35/0	6/7	10/0	12/0	16/0	5/0	0.089
20	41,242,056	Intronic	C/T	42/0	43/0	10/12	22/0	13/0	10/0	7/0	0.051
20	49,016,544	Intronic	C/G	37/0	26/0	9/4	7/0	8/0	6/0	3/0	0.203
21	63,204,706	Intergenic	A/G	40/0	32/0	7/5	18/0	13/0	18/0	11/0	0.031
26	**28,715,974**	**Intergenic**	**C/T**	**29/0**	**38/0**	**4/3**	**12/0**	**6/0**	**8/0**	**2/1**	**-**
27	28,774,595	Intronic	G/A	41/0	36/0	9/5	8/0	18/0	7/0	10/0	0.030
29	20,847,272	Intronic	C/T	38/0	36/0	11/4	20/0	7/0	11/0	2/0	0.425

Annotated de novo candidate mutations and depth of sequencing coverage for the wild type (WT) and the mutant (M) nucleotide in all individuals of the study as well as false-negative rates (FNR) associated with the validation by stable inheritance to the grandchild. Stably inherited germ line mutations are highlighted in bold.

filtering can indeed correctly identify false positives due to sequencing/alignment errors, none of the 222 candidate mutations that failed the visual inspection were inherited to the grandchild.

The remaining 22 mutations passed the manual curation filters, exhibiting read patterns consistent with a genuine de novo mutation, and were taken forward for further analyses.

ANNOTATION

Sites were annotated using ANNOVAR v2016Feb01 (Wang et al. 2010) with the annotation of the vervet genome build (NCBI *Chlorocebus sabaeus* Annotation Release 100) consisting of 29,648 genes. Among the identified de novo mutation candidates 11 were located in intronic regions, ten in intergenic regions, and one within an UTR3 (Table 3).

ESTIMATION OF THE FALSE NEGATIVE RATE IN DNM DISCOVERY

Synthetic mutations were simulated in the African green monkey data to estimate the rate of false negatives (FNR; i.e., genuine de novo mutations that were not identified) in de novo mutation discovery. Specifically, BAMSurgeon v1.0 (Ewing et al. 2015a) was used to modify autosomal sequencing reads overlapping 1000 randomly selected sites in the focal offspring (SRS578731) to introduce synthetic mutations. Following Keightley et al. (2014, 2015), mutation frequencies were drawn from the empirical distribution of allele-balance observed at autosomal heterozygous sites in the offspring at which the parents were confidently called as homozygous for different alleles (i.e., sites where one parent was called purely homozygous for the alternative allele and the other parent was called purely homozygous for the reference allele; both covered by at least 20 reads). Thereby, empirical distributions were generated for read depths of 1–20 (read depths larger than 20 were counted toward the read depth 20 bin; Supplementary Table 1). The modified reads were reanalyzed using the identical protocol than for the real data (see "Genotype calling"), including manual curation for a random subsample of 50 mutations (none of the visually inspected synthetic mutations occurred in a region subject to mis-mapping). The rate of false negatives in de novo mutation discovery was then calculated as FNR = 1 – (# called synthetic mutations/# callable synthetic mutations). Note that the number of callable synthetic mutations might be smaller

than the 1000 randomly selected sites due to the fact that certain sites might have been excluded by the application of the filtering criteria outlined in "Identification of DNM candidates."

VALIDATION OF DNMs AND ESTIMATION OF THE FALSE NEGATIVE RATE IN DNM VALIDATION

Following Wang and Zhu (2014), manually curated candidates were validated using their stable inheritance to the next generation (i.e., at each candidate site, the partner (SRS578744) was purely homozygous for the reference allele and support for the alternative allele was present in the grandchild (SRS578886))—limiting the dataset to eight putative de novo mutations (Table 1). Thereby, the low depth of coverage in the grandchild makes it challenging to determine allele transmission with absolute certainty. As a result, false-negative rates in de novo mutation validation have been estimated for read depths of 1–20 (read depths larger than 20 were counted toward the read depth 20 bin) by computing the number of missed heterozygous calls in the grandchild for sites at which one parent was purely homozygous for the reference allele and the other one was purely homozygous for the alternative allele (Table 3).

Results and Discussion

IDENTIFICATION AND VALIDATION OF DNMs

A seven individual, three generation pedigree (parents, children, and one grandchild; Fig. 2) was used to estimate the spontaneous mutation rate in African green monkeys. Overall, 5.4 million autosomal variants were segregating in the pedigree (Table 1), concordant with expectations based on previously reported nucleotide levels in this species (Huang et al. 2015; Pfeifer 2017b). Of these, 244 putative de novo mutation candidates were detected. These are defined as heterozygous in the child (SRS578731) with at least 20% of the reads supporting the alternative allele but completely homozygous for the reference allele in both parents (SRS578089 and SRS578771) as well as the other two children (SRS578719 and SRS578721) (i.e., no reads supporting the alternative allele in either of these individuals were present; Table 1). None of the de novo mutation candidates were known to segregate in the species (based on polymorphism data of more than 700 individuals (Huang et al. 2015; Pfeifer 2017b)).

All putative de novo mutation candidates were visually screened using the Integrated Genomics Viewer (Thorvaldsdóttir et al. 2012) to detect sequencing and alignment errors present in the reads surrounding the focal site. Twenty-two candidate mutations passed the manual curation step, exhibiting read patterns consistent with a genuine de novo mutation.

Eight out of the 22 mutations were successfully transmitted from the child to the grandchild (SRS578886), while the partner (SRS578744) was purely homozygous for the reference allele, confirming stable Mendelian inheritance of these de novo germ line mutations (Table 1). Out of these eight detected de novo mutations, six were located in intergenic regions and two in intronic regions (Table 3; the distribution of these observed de novo mutations among intergenic, intronic, and exonic categories did not deviate from expectations ($\chi^2 = 2.24$, *d.f.* $= 2$, $P = 0.33$)). Although multiple nucleotide mutation (MNM) events, resulting from repaired structural breaks (Behringer and Hall 2015), have been observed in humans (Amos 2010; Michaelson et al. 2012; Francioli et al. 2015) as well as in other eukaryotes (Schrider et al. 2011), no MNMs were observed in this study, potentially due to the small sample size.

SPONTANEOUS MUTATION RATE ESTIMATE

The mutation rate per base pair per generation can be estimated as $N/(2L)$, where N is the number of observed de novo mutations and L is the number of accessible sites in the genome (i.e., the number of sites for which there was power to detect mutations if they occurred). From the two- and three-generation pedigree, the mutation rate for African green monkeys was estimated to be 0.75×10^{-8} and 0.54×10^{-8} per base pair per generation (uncorrected for false negatives), respectively, assuming an accessible genome size of ~1.5 Gbp (corresponding to 57% of the genome; Table 2). The false-negative rate in de novo mutation discovery was estimated by simulating 1000 synthetic point mutations from the autosomal sequencing reads in the focal offspring. Among these 1000 synthetic mutations, 596 occurred at callable sites. Five-hundred three of these callable synthetic mutations were detected by the utilized pipeline, corresponding to a false-negative rate in de novo mutation discovery of 15.6%. The false negative rate in de novo mutation validation was estimated by computing the number of missed heterozygous calls in the grandchild for sites at which one parent was purely homozygous for the reference allele and the other one was purely homozygous for the alternative allele (Table 3). The corrected mutation rate estimate reported here of 0.88×10^{-8} and 0.94×10^{-8} per base pair per generation in the two- and three-generation pedigree, respectively, represents the first direct estimate for any Old World monkey. However, it is important to note that this rate is likely an underestimate for two reasons. First, due to the low depth of coverage in the grandchild, false-negative rates for the validation by stable inheritance are in some of the cases relatively high—thus, more than the identified eight de novo mutations might have been transmitted to the grandchild. Second, the manual curation makes it impossible to accurately determine the denominator (L)—a well-appreciated (but difficult to quantify) problem when inferring spontaneous mutation rates from pedigrees (Ségurel et al. 2014).

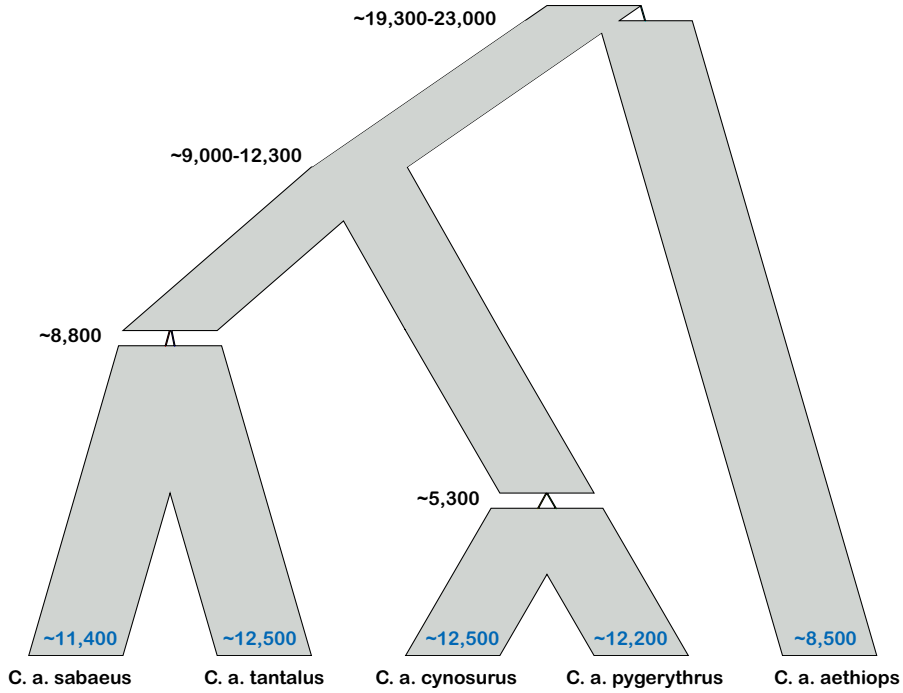

Figure 3. **Demographic history of the African green monkey. Divergence times (provided in generations) have been estimated using a molecular clock based on putatively neutral, fixed differences between the genomes of the populations (Pfeifer 2017b), assuming a mutations rate of 0.94 x 10^{-8} per base pair per generation. Effective population sizes are reported at the tip of the branches (in blue). This figure was generated using PopPlanner (Ewing et al. 2015b).**

Mutation rates are known to vary not only between species, but also between populations, individuals, sexes, and even different regions of a single genome, though the evolutionary processes governing this variation remain poorly understood. As a consequence, the spontaneous mutation rate reported here should be considered as an approximate representation of the species' mutation rate. Despite expected interindividual and interpopulation variation, all mutation rate estimates previously obtained from whole genome pedigrees in humans fall into a similar range of 0.96×10^{-8} – 1.28×10^{-8} per base pair per generation (95% CI: 0.67 x 10^{-8} –1.7 x 10^{-8}) (1000 Genomes Project Consortium 2010; Awadalla et al. 2010; Roach et al. 2010; Conrad et al. 2011; Campbell et al. 2012; Kong et al. 2012; Michaelson et al. 2012; Jiang et al. 2013; Besenbacher et al. 2015; Francioli et al. 2015; Yuen et al. 2015; Goldmann et al. 2016; Rahbari et al. 2016; Wong et al. 2016). Although limited in sample size to individuals from a single trio, this study represents the first direct estimate of mutation rate in an Old World monkey and—in contrast to the previously reported mutation rate of 5.9×10^{-9} in the species' close relative, the Rhesus macaque (indirectly calculated from sequence divergence based on alignments with baboon (Hernandez et al. 2007) and thus suffering from considerable uncertainty in the estimate)—this direct estimate is similar to the range observed in humans and is only slightly lower than the rate of 1.2×10^{-8} per base pair per generation reported from the only pedigree-based study in chimpanzees (Venn et al. 2014).

POPULATION GENETICS OF THE AFRICAN GREEN MONKEY

Previous work suggested the presence of genetic structure between different African green monkey populations (Pfeifer 2017b), potentially due to a number of factors such as physical barriers (e.g., mountains, rivers, or deserts) or past climate shifts, leading to isolated populations with limited contact among each other. The spontaneous mutation rate (μ) directly estimated from the pedigree obtained from this study can be used to more reliably date the time of divergence (t) between the different African green monkey populations (i.e., *C. a. aethiops*, *C. a. cynosurus*, *C. a. pygerythrus, C. a. sabaeus*, and *C. a. tantalus*; as classified by Grubb et al. 2003). Assuming a neutral nucleotide divergence $d = 2\mu t$, divergence time estimates range from ~19,300–23,000 generations for the basal split of *C. a. aethiops* from the other four populations, ~9000–12,300 generations for (*C. a. sabaeus* + *C. a. tantalus*) / (*C. a. cynosurus* + *C. a. pygerythrus*), ~8800 generations for *C. a. sabaeus* / *C. a. tantalus*, and ~5300 generations for *C. a. cynosurus* / *C. a. pygerythrus* (Fig. 3)—much younger than previously reported (Guschanski et al. 2013; Warren et al. 2015; Pfeifer 2017b). As a result, the history of African green monkey populations reflects a similar pattern than those of the

peopling of the African continent (see Campbell et al. 2014 for an in-depth review). The split of *C. a. aethiops* from the other four populations around 164–195 kya resembles the divergence estimate for the earliest population split in the modern human lineage on the African continent (i.e., between the ancestors of Khoesan-speaking San hunter-gatherers and other sub-Saharan Africans >100 kya (Schuster et al. 2010; Gronau et al. 2011; Lachance et al. 2012; Schlebusch et al. 2012; Veeramah et al. 2012; Excoffier et al. 2013; Schlebusch et al. 2013)). The split between *C. a. sabaeus* and *C. a. tantalus* ~75 kya is close to the divergence time estimates between the ancestors of Pygmy and non-Pygmy populations ~60–70 kya (Quintana-Murci et al. 2008; Batini et al. 2011a; Batini et al. 2011b). In contrast, both northwestern and southeastern Khoesan-speakers as well as Bantu and non-Bantu Niger-Kordofanian-speakers in western Africa have separated relatively recently, within the last 30,000 years (Tishkoff et al. 2009; Bryc et al. 2010; Pickrell et al. 2012)—similar to more recent divergence estimates for *C. a. cynosurus* and *C. a. pygerythrus* ~45 kya.

Levels of autosomal nucleotide diversity in the noncoding, nonrepetitive parts of the genome range from $\Theta_{\text{intergenic}} = 3.2 \times 10^{-4} - 4.7 \times 10^{-4}$ in the different African green monkey populations (Pfeifer 2017b), corresponding to estimated effective population sizes N_e of ~8500–12,500 under the assumption of neutrality—in the range of the effective population sizes $N_e =$ 11,000 for western chimpanzees and $N_e = 18{,}000$ for humans (assuming an underlying mutation rate of 1.8×10^{-8} and diversity levels between human individuals from Yoruba (Africa) and western chimpanzees; Fischer et al. 2006; Hernandez et al. 2011). It is important to note the assumption of neutrality in the above estimates, and the potential reduction in N_e owing to linked selection in these species (see Pfeifer and Jensen 2016).

Conclusions

Accurately estimating germ line mutation rates is challenging. Thus, despite being one of the most central parameters in evolutionary genetics, little is known about the rate at which mutations are introduced in to the genomes of different species. In primates, direct estimates from pedigree data are only available for humans and our closest extant evolutionary relative, the chimpanzee. Here, I present the first direct estimate of the spontaneous mutation rate in an Old World monkey. The estimated de novo mutation rate for the African green monkey of 0.94×10^{-8} per site per generation is similar to that of both humans ($0.96 \times 10^{-8} - 1.28 \times 10^{-8}$ per base pair per generation; 95% CI: $0.67 \times 10^{-8} - 1.7 \times 10^{-8}$) and chimpanzees ($1.2 \times 10^{-8}$ per base pair per generation). However, due to the limited sample size, the reported rate should be considered as an approximate representation of the species' mutation rate and additional pedigree studies will be required to determine an average mutation rate, to highlight interindividual variation, as well as to quantify male mutation bias and parental age effects in the species.

Evolutionary history has strongly influenced patterns of genetic diversity between different African green monkey populations and thus, the obtained estimate allows considerable refinement in the scaling of the recently inferred demographic history of this species. Knowledge of this population history is of fundamental importance given that vervets are widely employed in biomedical research, as traditional disease mapping strongly relies on accurate inference of population structure to avoid false associations in case/control studies.

More generally, this work offers unique insight on the constancy of mutation rates across the primate lineage. While it is perhaps not surprising that the two closely related great apes studied to date share a nearly similar rate, this work demonstrates this rate conservation over considerably deeper evolutionary time. This observation is in stark contrast to recent work surrounding another fundamental evolutionary parameter, recombination rate, in which recent estimates between humans, chimpanzees, and Rhesus macaque, have demonstrated considerable rate variation over relatively short evolutionary time (Xue et al. 2016)–perhaps indicative of weaker evolutionary constraint. Given the roughly similar effective population sizes between the species examined here, and thus similar expected effects of genetic drift, this work may lend further credence to the "drift-barrier hypothesis" of mutation rate evolution, in which natural selection operates simply to improve replication fidelity to the limit achievable given the effects of genetic drift (see Lynch et al. 2016). Indeed, this is an elegant explanation for the observed negative correlation between effective population size and mutation rate found across eukaryotes. However, it would be highly beneficial to increase the number of pedigree-based studies such as this to better evaluate these expectations, owing to the much finer resolution in mutation rate estimates afforded compared to other available approaches.

AUTHOR CONTRIBUTIONS

SPP conceived the study, conducted all analyses, and wrote the manuscript.

ACKNOWLEDGMENTS

I am grateful to Jeffrey Jensen for helpful comments and discussion. Computations were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the Swiss Institute of Bioinformatics (SIB).

DATA ARCHIVING

The data used in this study is available from NIH's Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under the BioProject accession number PRJNA240242 (individual identifiers: SRS578089, SRS578771, SRS578719, SRS578721, SRS578731, SRS578744, and SRS578886).

LITERATURE CITED

1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467:1061–1073.

Abee, C., K. Mansfield, S. Tardif, and T. Morris. 2012. Nonhuman primates in biomedical research. Elsevier, London, UK.

Amann, R. P. 2008. The cycle of the seminiferous epithelium in humans: a need to revisit? J. Androl. 29:469–487.

Amos, W. 2010. Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence? Proc. Biol. Sci. 277:1443–1449.

Amster, G., and G. Sella. 2016. Life history effects on the molecular clock of autosomes and sex chromosomes. Proc. Natl. Acad. Sci. USA 113:1588–1593.

Awadalla, P., J. Gauthier, R. A. Myers, F. Casals, F. F. Hamdan, A. R. Griffing, M. Côté, E. Henrion, D. Spiegelman, J. Tarabeux, et al. 2010. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. Am. J. Hum. Genet. 87:316–324.

Batini, C., G. Ferri, G. Destro-Bisol, F. Brisighelli, D. Luiselli, P. Sanchez-Diz, J. Rocha, T. Simonson, A. Brehm, V. Montano, et al. 2011a. Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. Mol. Biol. Evol. 28:2603–2613.

Batini, C., J. Lopes, D. M. Behar, F. Calafell, L. B. Jorde, L. van der Veen, L. Quintana-Murci, G. Spedini, G. Destro-Bisol, D. Comas. 2011b. Insights into the demographic history of African Pygmies from complete mitochondrial genomes. Mol. Biol. Evol. 28:1099–1110.

Behringer, M. G., and D. W. Hall. 2015. Genome-wide estimates of mutation rates and spectrum in Schizosaccharomyces pombe indicate CpG sites are highly mutagenic despite the absence of DNA methylation. G3 6:149–160.

Besenbacher, S., S. Liu, J. M. Izarzugaza, J. Grove, K. Belling, J. Bork-Jensen, S. Huang, T. D. Als, S. Li, R. Yadav, et al. 2015. Novel variation and de novo mutation rates in population-wide de novo assembled Danish trios. Nat. Commun. 6:5969.

Broussard, S. R., S. I. Staprans, R. White, E. M. Whitehead, M. B. Feinberg, and J. S. Allan. 2001. Simian immunodeficiency virus replicates to high levels in naturally infected African green monkeys without inducing immunologic or neurologic disease. J. Virol. 75:2262–2275.

Bryc, K., A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser, S. Williams, A. Froment, J. M. Bodo, C. Wambebe, S. A. Tishkoff, et al. 2010. Genome-wide patterns of population structure and admixture in West Africans and African Americans. Proc. Natl. Acad. Sci. USA 107:786–791.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula flycatchers*. Genome Res. 25:1656–1665.

Campbell, C. D., J. X. Chong, M. Malig, A. Ko, B. L. Dumont, L. Han, L. Vives, B. J. O'Roak, P. H. Sudmant, J. Shendure, et al. 2012. Estimating the human mutation rate using autozygosity in a founder population. Nat Genet. 44:1277–1281.

Campbell, M. C., J. B. Hirbo, J. P. Townsend, and S. A. Tishkoff. 2014. The peopling of the African continent and the diaspora into the new world. Curr. Opin. Genet. Dev. 29:120–132.

Chapman, C. A., S. Friant, K. Godfrey, C. Liu, D. Sakar, V. A. Schoof, R. Sengupta, D. Twinomugisha, K. Valenta, and T. L. Goldberg. 2016. Social behaviours and networks of monkeys are influenced by gastrointestinal parasites. PLoS One 11:e0161113.

Chimpanzee Seq. Anal. Consort. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87.

Conrad, D. F., J. E. Keebler, M. A. DePristo, S. J. Lindsay, Y. Zhang, F. Casals, Y. Idaghdour, C. L. Hartl, C. Torroja, K. V. Garimella, et al. 2011. Variation in genome-wide mutation rates within and between human families. Nat. Genet. 43:712–714.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. Genetics 148:1667–1686.

Duret, L., and N. Galtier. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu. Rev. Genomics Hum. Genet. 10:285–311.

Elango, N., J. W. Thomas, and S. V. Yi, NISC Comparative Sequencing Program. 2006. Variable molecular clocks in hominoids. Proc. Natl. Acad. Sci. USA 103:1370–1375.

Emborg, M. E. 2007. Nonhuman primate models of Parkinson's disease. ILAR J. 48:339–355.

Ewing, A. D., K. E. Houlahan, Y. Hu, K. Ellrott, C. Caloian, T. N. Yamaguchi, J. C. Bare, C. P'ng, D. Waggott, V. Y. Sabelnykova, et al. 2015a. Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. Nat. Methods 12:623–630.

Ewing, G., P. Reiff, and J. D. Jensen. 2015b. PopPlanner: visually constructing demographic models for simulation. Front Genet. 6:150.

Excoffier, L., I. Dupanloup, E. Huerta-Sanchez, V. C. Sousa, and M. Foll. 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9:e1003905.

Fischer, A., J. Pollack, O. Thalmann, B. Nickel, and S. Pääbo. 2006. Demographic history and genetic differentiation in apes. Curr. Biol. 16:1133–1138.

Francioli, L. C., P. P. Polak, A. Koren, A. Menelaou, S. Chun, I. Renkens, C. M. van Duijn, M. Swertz, C. Wijmenga, G. van Ommen, et al. 2015. Genome-wide patterns and properties of de novo mutations in humans. Nat. Genet. 47:822–826.

Gao, Z., M. J. Wyman, G. Sella, and M. Przeworski. 2016. Interpreting the dependence of mutation rates on age and time. PLoS Biol. 14:e1002355.

Goldmann, J. M., W. S. Wong, M. Pinelli, T. Farrah, D. Bodian, A. B. Stittrich, G. Glusman, L. E. Vissers, A. Hoischen, J. C. Roach, et al. 2016. Parent-of-origin-specific signatures of de novo mutations. Nat. Genet. 48:935–939.

Goodman, M. 1963. Serological analysis of the systematics of recent hominoids. Hum. Biol. 35:377–436.

Gronau, I., M. J. Hubisz, B. Gulko, C. G. Danko, and A. Siepel. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nat. Genet. 43:1031–1034.

Grubb, R., T. M. Butynski, J. E. Oates, S. K. Bearder, T. R. Disotell, C. P. Groves, and T. T. Struhsaker. 2003. Assessment of the diversity of African primates. Int. J. Primatol. 24:1301–1357.

Guschanski, K., J. Krause, S. Sawyer, L. M. Valente, S. Bailey, K. Finstermeier, R. Sabin, E. Gilissen, G. Sonet, Z. T. Nagy, et al. 2013. Next-generation museomics disentangles one of the largest primate radiations. Syst. Biol. 62:539–554.

Hahn, M. W., S. V. Zhang, and L. C. Moyle. 2014. Sequencing, assembling, and correcting draft genomes using recombinant populations. G3 4:669–679.

Haldane, J. B. S. 1935. The rate of spontaneous mutation of a human gene. J. Genet. 31:317–326.

Han, E., J. S. Sinsheimer, and J. Novembre. 2014. Characterizing bias in population genetic inferences from low-coverage sequencing data. Mol. Biol. Evol. 31:723–735.

Harris, K. 2015. Evidence for recent, population-specific evolution of the human mutation rate. Proc. Natl. Acad. Sci. USA 112:3439–3444.

Harris, K., and J. K. Pritchard. 2017. Rapid evolution of the human mutation spectrum. Elife. https://doi.org/10.7554/eLife.24284

Heller, C. H., and Y. Cermont. 1964. Kinetics of the germinal epithelium in man. Recent Prog. Horm. Res. 20:545–575.

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

Hernandez, R. D., J. L. Kelley, E. Elyashiv, S. C. Melton, A. Auton, G. McVean, G. Sella, and M. Przeworski. 2011. Classic selective sweeps were rare in recent human evolution. Science 331:920–924.

Huang, Y. S., V. Ramensky, S. K. Service, A. J. Jasinska, Y. Jung, O. W. Choi, R. M. Cantor, N. Juretic, J. Wasserscheid, J. R. Kaplan, et al. 2015. Sequencing strategies and characterization of 721 vervet monkey genomes for future genetic analyses of medically relevant traits. BMC Biol. 13:41.

Jiang, Y. H., R. K. Yuen, X. Jin, M. Wang, N. Chen, X. Wu, J. Ju, J. Mei, Y. Shi, M. He, et al. 2013. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. Am. J. Hum. Genet. 93:249–263.

Keightley, P. D., R. W. Ness, D. L. Halligan, and P. R. Haddrill. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. Genetics 196:313–320.

Keightley, P. D., A. Pinharanda, R. W. Ness, F. Simpson, K. K. Dasmahapatra, J. Mallet, J. W. Davey, and C. D. Jiggins. 2015. Estimation of the spontaneous mutation rate in *Heliconius melpomene*. Mol. Biol. Evol. 32:239–243.

Kim, S.-H., N. Elango, C. Warden, E. Vigoda, and S. V. Yi. 2006. Heterogeneous genomic molecular clocks in primates. PLoS Genet. 2:e163.

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

Kondrashov, A. S. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum. Mutat. 21:12–27.

Kondrashov, A. S., and J. F. Crow. 1993. A molecular approach to estimating the human deleterious mutation rate. Hum. Mutat. 2:229–234.

Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, A. Jonasdottir, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. Nature 488:471–475.

Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature 392:917–920.

Lachance, J., B. Vernot, C. C. Elbers, B. Ferwerda, A. Froment, J. M. Bodo, G. Lema, W. Fu, T. B. Nyambo, T. R. Rebbeck, et al. 2012. Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. Cell 150:457–469.

Lemere, C. A., A. Beierschmitt, M. Iglesias, E. T. Spooner, J. K. Bloom, J. F. Leverone, J. B. Zheng, T. J. Seabrook, D. Louard, D. Li, et al. 2004. Alzheimer's disease abeta vaccine reduces central nervous system abeta levels in a non-human primate, the Caribbean vervet. Am. J. Pathol. 165:283–297.

Li, H. 2011. Improving SNP discovery by base alignment quality. Bioinformatics 27:1157–1158.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li, W. H., D. L. Ellsworth, J. Krushkal, B. H. Chang, and D. Hewett-Emmett. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol. Phylogenet. Evol. 5:182–187.

Liu, H., Y. Jia, X. Sun, D. Tian, L. D. Hurst, and S. Yang. 2017. Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. Mol. Biol. Evol. 34:119–130.

Lynch, M. 2010a. Evolution of the mutation rate. Trends Genet. 26:345–352.

———. 2010b. Rate, molecular spectrum, and consequences of human mutation. Proc. Natl. Acad. Sci. USA 107:961–968.

Lynch, M., M. S. Ackerman, J. F. Gout, H. Long, W. Sung, W. K. Thomas, and P. L. Foster. 2016. Genetic drift, selection and the evolution of the mutation rate. Nat. Rev. Genet. 17:704–714.

Mathieson, I., and D. Reich. 2017. Differences in the rare variant spectrum among human populations. PLoS Genet. 13:e1006581.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next generation DNA sequencing data. Genome Res. 20:1297–1303.

Michaelson, J. J., Y. Shi, M. Gujral, H. Zheng, D. Malhotra, X. Jin, M. Jian, G. Liu, D. Greer, A. Bhandari, et al. 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell 151:1431–1442.

Moorjani, P., C. E. Amorim, P. F. Arndt, and M. Przeworski. 2016a. Variation in the molecular clock of primates. Proc. Natl. Acad. Sci. USA 113:10607–10612.

Moorjani, P., S. Sankararaman, Q. Fu, M. Przeworski, N. Patterson, and D. Reich. 2016b. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. Proc. Natl. Acad. Sci. USA 113:5652–5657.

Nachman, M. W., and S. L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics 156:297–304.

Parker, G. A. 1970. Sperm competition and its evolutionary consequences in the insects. Biol. Rev. 45:525–567.

Pfeifer, S. P. 2017a. From next-generation resequencing reads to a high-quality variant data set. Heredity 118:111–124.

———. 2017b. The demographic and adaptive history of the African green monkey. Mol. Biol. Evol. 34:1055–1065.

Pfeifer, S. P., and J. D. Jensen. 2016. The impact of linked selection in chimpanzees: a comparative study. Genome Biol. Evol. 8:3202–3208.

Pickrell, J. K., N. Patterson, C. Barbieri, F. Berthold, L. Gerlach, T. Guldemann, B. Kure, S. W. Mpoloka, H. Nakagawa, C. Naumann, et al. 2012. The genetic prehistory of southern Africa. Nat. Commun. 3: 1143.

Quintana-Murci, L., H. Quach, C. Harmant, F. Luca, B. Massonnet, E. Patin, L. Sica, P. Mouguiama-Daouda, D. Comas, S. Tzur, et al. 2008. Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers. Proc. Natl. Acad. Sci. USA 105:1596–1601.

Rahbari, R., A. Wuster, S. J. Lindsay, R. J. Hardwick, L. B. Alexandrov, S. Al Turki, A. Dominiczak, A. Morris, D. Porteous, B. Smith, et al. 2016. Timing, rates and spectra of human germline mutation. Nat. Genet. 48:126–133.

Roach, J. C., G. Glusman, A. F. A. Smit, C. D. Huff, R. Hubley, P. T. Shannon, L. Rowen, K. P. Pant, N. Goodman, M. Bamshad, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 328:636–639.

Scally, A. 2016. The mutation rate in human evolution and demographic inference. Curr. Opin. Genet. Dev. 41:36–43.

Scally, A., and R. Durbin. 2012. Revising the human mutation rate: implications for understanding human evolution. Nat. Rev. Genet. 13:745–753.

Schlebusch, C. M., M. Lombard, and H. Soodyall. 2013. MtDNA control region variation affirms diversity and deep sub-structure in populations from southern Africa. BMC Evol. Biol. 13:56.

Schlebusch, C. M., P. Skoglund, P. Sjodin, L. M. Gattepaille, D. Hernandez, F. Jay, S. Li, M. De Jongh, A. Singleton, M. G. Blum, et al. 2012.

Genomic variation in seven KhoeSan groups reveals adaptation and complex African history. Science 338:374–379.

Schrider, D. R., J. N. Hourmozdi, M. W. Hahn. 2011. Pervasive multinucleotide mutational events in eukaryotes. Curr. Biol. 21:1051–1054.

Schuster, S. C., W. Miller, A. Ratan, L. P. Tomsho, B. Giardine, L. R. Kasson, R. S. Harris, D. C. Petersen, F. Zhao, J. Qi, et al. 2010. Complete Khoisan and Bantu genomes from southern Africa. Nature 463:943–947.

Ségurel, L., M. J. Wyman, and M. Przeworski. 2014. Determinants of mutation rate variation in the human germline. Annu. Rev. Genomics Hum. Genet. 15:47–70.

Smeds, L., A. Qvarnström, and H. Ellegren. 2016. Direct estimate of the rate of germline mutation in a bird. Genome Res. 26:1211–1218.

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinformatics 14:178–192.

Tishkoff, S. A., F. A. Reed, F. R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J. B. Hirbo, A. A. Awomoyi, J. M. Bodo, O. Doumbo, et al. 2009. The genetic structure and history of Africans and African Americans. Science 324:1035–1044.

Van der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, et al. 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr. Protoc Bioinformatics 43:11.10.1–11.10.33.

Veeramah, K. R., D. Wegmann, A. Woerner, F. L. Mendez, J. C. Watkins, G. Destro-Bisol, H. Soodyall, L. Louie, and M. F. Hammer. 2012. An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. Mol. Biol. Evol. 29:617–630.

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

Wallberg, A., F. Han, G. Wellhagen, B. Dahle, M. Kawata, N. Haddad, Z. L. Simões, M. H. Allsopp, I. Kandemir, P. De la Rúa, et al. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. Nat. Genet. 46:1081–1088.

Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. Nucleic Acids Res. 38:e164.

Wang, H., and X. Zhu. 2014. De novo mutations discovered in 8 Mexican American families through whole genome sequencing. BMC Proc. 8(Suppl 1):S24.

Warren, W. C., A. J. Jasinska, R. García-Pérez, H. Svardal, C. Tomlinson, M. Rocchi, N. Archidiacono, O. Capozzi, P. Minx, M. J. Montague, et al. 2015. The genome of the vervet (*Chlorocebus aethiops sabaeus*). Genome Res. 25:1921–1933.

Wilson Sayres, M. A., C. Venditti, M. Pagel, and K. D. Makova. 2011. Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. Evolution 65:2800–2815.

Wong, W. S., B. D. Solomon, D. L. Bodian, P. Kothiyal, G. Eley, K. C. Huddleston, R. Baker, D. C. Thach, R. K. Iyer, J. G. Vockley, et al. 2016. New observations on maternal age effect on germline de novo mutations. Nat. Commun. 7:10486.

Xu, K., S. Oh, T. Park, D. C. Presgraves, and S. V. Yi. 2012. Lineage-specific variation in slow- and fast-X evolution in primates. Evolution 66:1751–1761.

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

Yang, S., L. Wang, J. Huang, X. Zhang, Y. Yuan, J.-Q. Chen, L. D. Hurst, and D. Tian. 2015. Parent-progeny sequencing indicates higher mutation rates in heterozygotes. Nature 523:463–467.

Yi, S. V. 2013. Morris Goodman's hominoid rate slowdown: the importance of being neutral. Mol. Phylogenet Evol. 66:569–574.

Yi, S., D. L. Ellsworth, W. H. Li. 2002. Slow molecular clocks in Old World monkeys, apes, and humans. Mol. Biol. Evol. 19:2191–2198.

Yuen, R. K., B. Thiruvahindrapuram, D. Merico, S. Walker, K. Tammimies, N. Hoang, C. Chrysler, T. Nalpathamkalam, G. Pellecchia, Y. Liu, et al. 2015. Whole-genome sequencing of quartet families with autism spectrum disorder. Nat. Med. 21:185–191.

Associate Editor: M. Wilson Sayres
Handling Editor: M. Servedio

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Number of autosomal heterozygous (het) sites for read depths (DP) of 1–20 (read depths larger than 20 were counted towards the read depth 20^{+} bin) used to generate empirical distributions of allele-balance.