Reduced Representation Genome Sequencing Suggests Low Diversity on the Sex Chromosomes of Tonkean Macaque Monkeys

Ben J. Evans,*[,1] Kai Zeng,[2] Jacob A. Esselstyn,[3] Brian Charlesworth,[4] and Don J. Melnick[5]

[1]Biology Department, McMaster University, Hamilton, ON, Canada

[2]Department of Animal and Plant Sciences, Alfred Denny Building, University of Sheffield, Sheffield, United Kingdom

[3]Department of Biological Sciences and Museum of Natural Science, Louisiana State University

[4]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom

[5]Department of Ecology, Evolution, and Environmental Biology, Columbia University

***Corresponding author:** E-mail: evansb@mcmaster.ca.

Associate editor: Rasmus Nielsen

Abstract

In species with separate sexes, social systems can differ in the relative variances of male versus female reproductive success. Papionin monkeys (macaques, mangabeys, mandrills, drills, baboons, and geladas) exhibit hallmarks of a high variance in male reproductive success, including a female-biased adult sex ratio and prominent sexual dimorphism. To explore the potential genomic consequences of such sex differences, we used a reduced representation genome sequencing approach to quantifying polymorphism at sites on autosomes and sex chromosomes of the tonkean macaque (*Macaca tonkeana*), a species endemic to the Indonesian island of Sulawesi. The ratio of nucleotide diversity of the X chromosome to that of the autosomes was less than the value (0.75) expected with a 1:1 sex ratio and no sex differences in the variance in reproductive success. However, the significance of this difference was dependent on which outgroup was used to standardize diversity levels. Using a new model that includes the effects of varying population size, sex differences in mutation rate between the autosomes and X chromosome, and GC-biased gene conversion (gBGC) or selection on GC content, we found that the maximum-likelihood estimate of the ratio of effective population size of the X chromosome to that of the autosomes was 0.68, which did not differ significantly from 0.75. We also found evidence for 1) a higher level of purifying selection on genic than nongenic regions, 2) gBGC or natural selection favoring increased GC content, 3) a dynamic demography characterized by population growth and contraction, 4) a higher mutation rate in males than females, and 5) a very low polymorphism level on the Y chromosome. These findings shed light on the population genomic consequences of sex differences in the variance in reproductive success, which appear to be modest in the tonkean macaque; they also suggest the occurrence of hitchhiking on the Y chromosome.

Key words: **X chromosome, Y chromosome, molecular polymorphism, GC-biased gene conversion, demography, male driven evolution.**

Introduction

In species with separate sexes, social systems can differ in the degree to which females and males differ in their variances of reproductive success. If there is a 1:1 sex ratio and no sex difference in the variances of reproductive success, the relative effective population sizes of autosomal DNA (A) and sex chromosome DNA match "null" Wright–Fisher expectations, such as a 3:4 ratio of effective population sizes of the X chromosome (X) and autosomes (hereafter the X:A ratio; Hedrick 2007; Charlesworth 2009). However, if there is a higher male than female variance in reproductive success, the equilibrium X:A ratio is greater than 0.75, with a maximum value of 1.125 reached when there is effectively only one reproducing male (Charlesworth 2001). High variance in male reproductive success compared with females also drives down the ratio of the effective population sizes of the Y chromosome (Y) to autosomes (Y:A ratio), with a lower limit of 0.125 (Charlesworth 2001). If the variance in female reproductive success is higher than for males, the equilibrium X:A ratio is less than 0.75, with a minimum value of 0.5625, corresponding to effectively only one reproducing female (Charlesworth 2001).

Importantly, the X:A ratio is also affected by demographic changes such as recent bottlenecks of population size (Pool and Nielsen 2007, 2008). Hitchhiking effects of either positive or negative selection at linked sites may also have greater effects on neutral diversity levels at sites on the X than the A in mammals, because the X does not recombine with most of the Y in males. These effects are likely to be most pronounced near genes and other functionally significant components of the genome (Charlesworth 2012a). Additionally, even at putatively neutral sites, selection-like processes such as GC-biased gene conversion (gBGC) can affect diversity levels; if not properly handled, these factors can compromise estimates of demographic change (Zeng 2012) and, hence, of

Article

the X:A ratio of effective population sizes. These considerations mean that the precision of X:A ratio estimates may be improved by accommodating these factors.

Macaque monkeys (and papionins in general) are considered to be prime examples of species where the variance in reproductive success is higher in males than in females. Field studies suggest that the adult sex ratio of macaques is frequently female biased (Dittus 1975; Hsu and Lin 2001; Beisner et al. 2012; Hasan et al. 2013), and sexual dimorphism in size and craniofacial morphology is marked (Albrecht 1978; Schillaci et al. 2007). Incongruence between mitochondrial gene trees and species relationships (Tosi et al. 2002, 2003; Evans et al. 2010) is also potentially consistent with a higher variance in male than female reproductive success, which increases the relative lineage sorting period of the maternally inherited mitochondrial DNA, compared with expectations for other genomic regions that are partially or fully paternally inherited. The timescale for lineage sorting of mitochondrial DNA is also potentially increased by female philopatry (Hoelzer et al. 1998; Hedrick 2007), a characteristic of macaque societies (Dittus 1975; Van Noordwijk and Van Schaik 1985; Melnick and Hoelzer 1992).

In order to quantify the genomic effects of sex-biased variance in reproductive success, we used reduced representation genome sequencing (RADseq; Baird et al. 2008) to collect genome-wide molecular polymorphism data from genetic samples of a wild caught papionin monkey, the tonkean macaques (*Macaca tonkeana*). Tonkean macaques are endemic to the Indonesian island of Sulawesi, which is home to seven macaque species (Riley 2010). The Sulawesi macaques are thought to be derived from a single dispersal event from Borneo by an ancestor of the pig-tailed macaque *M. nemestrina* (Fooden 1969; Evans et al. 2010). The tonkean macaque is known to hybridize with parapatric macaque species (Watanabe, Lapasere, et al. 1991; Watanabe, Matsumura, et al. 1991; Froehlich and Supriatna 1996; Bynum et al. 1997; Evans et al. 2001), and the population in the eastern portion of its range is genetically differentiated from the one in the western portion of its range (Evans et al. 2001).

We quantified the X:A and the Y:A ratios of effective population sizes of the tonkean macaque by measuring pairwise diversity for each chromosomal category. Because mutation rates may differ between males and females (Makova and Li 2002), we quantified divergence from homologous sequences obtained from outgroup species and used these values to standardize the diversity estimates. We also evaluated the X:A ratio with a new approach that allows for a dynamic demography and natural selection on GC content (or a selection-like process such as gBGC).

Results

Molecular Polymorphism and Evidence of Natural Selection Near Genes

We used a reduced representation genome sequencing approach called RADseq (Baird et al. 2008) to identify nucleotide polymorphisms from nine wild-caught tonkean macaque individuals. These data were aligned to outgroup genomes (rhesus macaque, olive baboon, and human), and genotype calls and data filtering were performed using rigorous criteria detailed in the Materials and Methods section and in the supplementary material S1, Supplementary Material online. Filtering of positions near insertions/deletions resulted in the removal of 479,495 positions. Single nucleotide polymorphism (SNP) quality filtering led to the removal of 15,894,391 positions, with most (15,034,607) due to the filtering of autosomal sites with a Phred-scaled genotype quality score below 50. Sex chromosome filtering identified 783 positions on the nonpseudoautosomal region of the X chromosome with a heterozygous genotype call in one or more males. After excluding a 10,000-bp buffer around these positions to avoid putative pseudoautosomal regions in the tonkean macaque that differed from the rhesus macaque, a total of 260,150 and 7,404,310 high-quality genotype calls in at least one individual's X chromosome DNA and autosomal DNA, respectively, were retained for analysis. A total of 13,044 high quality genotype calls in at least one individual's Y chromosome DNA were recovered. After additional individual genotypes were removed that had less than 10-fold coverage or greater than 45-fold coverage (to minimize potential problems related to low coverage or copy number variation; supplementary material S1, Supplementary Material online), high-quality genotypes were obtained for all alleles for all individuals (i.e., sites that had no missing data) for 7,070,587, 118,337, and 6,233 sites for A, X, and Y, respectively (table 1). The average coverage for individual genotypes was 20.1-fold with a standard deviation of 7.7-fold.

Tables 1 and 2 summarize molecular polymorphism and divergence patterns for our samples of the tonkean macaque, using humans and olive baboon, respectively, as outgroups. Polymorphism data for the tonkean macaque vary slightly between these tables due to assembly quality and possible linkage differences between the outgroups. The X chromosome comprises approximately 5.3% of the "rhemac2" rhesus genome assembly. However, only approximately 1.7% of our high-quality genotypes mapped to X chromosome DNA. This suggests that our less stringent cutoff value for genotype scores on the X chromosome DNA (a Phred-scaled quality score of 30 instead of 50 was required because most of the X chromosomes were haploid in our sample; see supplementary material S1, Supplementary Material online) was probably still more stringent than our criterion for autosomal DNA in terms of number of reads and read quality per genotyped allele.

Autosomal regions that include genes and 1,000-bp upstream and downstream of genes (partition *a*) have lower pairwise diversity divided by divergence (π/d), a higher proportion of singleton segregating sites (S_e/S), and a more negative Tajima's *D* than the other partitions that include data which are farther from genes (fig. 1, tables 1 and 2). This is based on nonoverlapping 95% confidence intervals (CIs) of partitions (*a*) and (*e*), and trends of the intervening partitions. Although we note the caveat that physical distance does not correspond precisely with genetic distance, these patterns are generally consistent with stronger background selection

Table 1. Information about Autosomal, X Chromosome, and Y Chromosome Data Using Humans as an Outgroup, and Based on Positions with High Confidence Genotype Calls in All Individuals.

	Partition (*a*)	Partition (*b*)	Partition (*c*)	Partition (*d*)	Partition (*e*)	Partitions (*c*)–(*e*)
Autosomal DNA						
# Sites	834,191	1,572,163	877,929	610,603	3,175,701	4,664,233
# RAD tags	4,202	7,078	4,369	3,172	17,958	25,375
S	7,884 (7,701 to 8,066)	15,816 (15,572 to 16,059)	9,672 (9,487 to 9,866)	6,901 (6,737 to 7,056)	37,269 (36,891 to 37,639)	53,842 (53,388 to 54,283)
π	0.00219 (0.00213 to 0.00225)	0.00240 (0.00236 to 0.00244)	0.00263 (0.00257 to 0.00269)	0.00268 (0.00260 to 0.00275)	0.00275 (0.00272 to 0.00278)	0.00272 (0.00269 to 0.00274)
θ_W	0.00275 (0.00268 to 0.00281)	0.00292 (0.00288 to 0.00297)	0.00320 (0.00314 to 0.00327)	0.00329 (0.00321 to 0.00336)	0.00341 (0.00338 to 0.00345)	0.00336 (0.00333 to 0.00338)
D	−0.871 (−0.919 to −0.822)	−0.770 (−0.802 to −0.738)	−0.771 (−0.815 to −0.729)	−0.793 (−0.843 to −0.742)	−0.835 (−0.857 to −0.813)	−0.818 (−0.836 to −0.800)
d	0.05556 (0.05504 to 0.05613)	0.05887 (0.05848 to 0.05927)	0.06094 (0.06039 to 0.06152)	0.06236 (0.06170 to 0.06303)	0.06273 (0.06243 to 0.06302)	0.06234 (0.06209 to 0.06257)
π/d	0.039 (0.03822 to 0.04048)	0.041 (0.04002 to 0.04158)	0.043 (0.04209 to 0.04411)	0.043 (0.04163 to 0.04427)	0.044 (0.04323 to 0.04437)	0.044 (0.04308 to 0.04401)
X chromosome						
# Sites	12,575	28,096	11,494	10,065	56,107	77,666
# RAD tags	92	160	88	81	475	641
S	48 (35 to 62)	73 (57 to 89)	48 (35 to 63)	38 (27 to 50)	200 (173 to 228)	286 (252 to 319)
π	0.00120 (0.00087 to 0.00157)	0.00083 (0.00063 to 0.00103)	0.00122 (0.00088 to 0.00162)	0.00106 (0.00071 to 0.00142)	0.00107 (0.00091 to 0.00123)	0.00109 (0.00096 to 0.00124)
θ_W	0.00135 (0.00098 to 0.00174)	0.00092 (0.00072 to 0.00112)	0.00148 (0.00108 to 0.00194)	0.00133 (0.00095 to 0.00176)	0.00126 (0.00109 to 0.00144)	0.00130 (0.00115 to 0.00145)
D	−0.547 (−1.059 to 0.011)	−0.472 (−0.874 to −0.081)	−0.835 (−1.319 to −0.336)	−1.006 (−1.514 to −0.496)	−0.749 (−0.95 to −0.505)	−0.806 (−1.002 to −0.605)
d	0.04285 (0.03906 to 0.04665)	0.04818 (0.04556 to 0.05103)	0.04766 (0.04341 to 0.05188)	0.04314 (0.03913 to 0.04745)	0.04908 (0.04717 to 0.05109)	0.04810 (0.04654 to 0.04966)
π/d	0.028 (0.01996 to 0.03758)	0.017 (0.01301 to 0.02174)	0.026 (0.01796 to 0.03466)	0.025 (0.01597 to 0.03360)	0.022 (0.01841 to 0.02533)	0.023 (0.01987 to 0.02570)
Y chromosome						Partitions (a)–(e)
# Sites	1520	1184	178	611	2740	6233
# RAD tags	9	8	3	4	18	42
S	2 (0 to 4)	1 (0 to 3)	0	0	5 (1 to 10)	8 (2 to 13)
π	0.00033 (0 to 0.00098)	0.00021 (0 to 0.00084)	0.00000	0.00000	0.00057 (0.00009 to 0.00119)	0.00037 (0.00011 to 0.00069)
θ_W	0.00051 (0 to 0.00102)	0.00033 (0 to 0.00098)	0.00000	0.00000	0.00070 (0.00014 to 0.00141)	0.00050 (0.00012 to 0.00081)
D	−1.310 (−1.448 to 1.1665)	−1.055 (−1.310 to 1.166)	—	—	−0.839 (−1.674 to 1.008)	−1.197 (−1.701 to 0.839)
d	0.10197 (0.08427 to 0.11967)	0.09225 (0.07336 to 0.11114)	0.11514 (0.05946 to 0.17082)	0.14894 (0.11353 to 0.18434)	0.17920 (0.16023 to 0.19817)	0.14156 (0.13084 to 0.15228)
π/d	0.003	0.002	0.000	0.000	0.003	0.003

NOTE.—Data are divided into partitions based on distance from genes (partitions *a*–*e* and combined partitions *c*–*e*; see text). Information includes the number of positions genotyped (# sites), the number of RAD tags (# RAD tags), the number of polymorphic sites (*S*), pairwise nucleotide diversity per site (π), Watterson's θ (θ_W), Tajima's *D* (*D*), pairwise divergence per site compared with humans with correction for multiple substitutions and ancestral polymorphism (*d*), and π/d. The 95% CIs estimated from bootstrapping or (for the Y chromosome) coalescent simulations are indicated in parentheses.

Table 2. Information about Autosomal and X Chromosome Data Inferred Using Olive Baboons as an Outgroup.

	Partition (*a*)	Partition (*b*)	Partition (*c*)	Partition (*d*)	Partition (*e*)	Partitions (*c*)–(*e*)
Autosomal DNA						
# Sites	782,137	1,490,588	837,362	579,205	3,009,601	4,426,186
# RAD tags	4,203	7,078	4,370	3,172	17,959	25,377
S	7,458 (7,286 to 7,629)	15,145 (14,902 to 15,382)	9,303 (9,112 to 9,497)	6,631 (6,482 to 6,796)	35,856 (35,492 to 36,234)	51,790 (51,361 to 52,232)
π	0.00220 (0.00214 to 0.00226)	0.00243 (0.00239 to 0.00248)	0.00265 (0.00258 to 0.00270)	0.00271 (0.00264 to 0.00278)	0.00279 (0.00276 to 0.00283)	0.00275 (0.00273 to 0.00278)
θ_W	0.00277 (0.00271 to 0.00284)	0.00295 (0.00291 to 0.00300)	0.00323 (0.00316 to 0.00330)	0.00333 (0.00325 to 0.00341)	0.00346 (0.00343 to 0.00350)	0.00340 (0.00337 to 0.00343)
D	−0.883 (−0.933 to −0.837)	−0.760 (−0.795 to −0.725)	−0.776 (−0.821 to −0.736)	−0.794 (−0.843 to −0.745)	−0.830 (−0.852 to −0.810)	−0.816 (−0.834 to −0.799)
d	0.01327 (0.01298 to 0.013539)	0.01448 (0.01426 to 0.01469)	0.01486 (0.01457 to 0.01513)	0.01606 (0.01571 to 0.01640)	0.01609 (0.01593 to 0.01625)	0.01586 (0.01573 to 0.01598)
π/d	0.166 (0.16030 to 0.17181)	0.168 (0.16377 to 0.17203)	0.178 (0.17204 to 0.18367)	0.169 (0.16286 to 0.17536)	0.174 (0.17093 to 0.17629)	0.174 (0.17151 to 0.17603)
X chromosome						
# Sites	1,1233	2,6161	10,115	8,714	51,740	70,569
# RAD tags	90	163	88	79	467	631
S	42 (30 to 56)	70 (54 to 86)	42 (30 to 55)	30 (20 to 40)	189 (163 to 219)	261 (230 to 293)
π	0.00113 (0.00077 to 0.00152)	0.00085 (0.00063 to 0.00107)	0.00118 (0.00079 to 0.00157)	0.00093 (0.00058 to 0.00127)	0.00110 (0.00094 to 0.00128)	0.00109 (0.00095 to 0.00125)
θ_W	0.00132 (0.00094 to 0.00176)	0.00095 (0.00073 to 0.00116)	0.00147 (0.00105 to 0.00192)	0.00122 (0.00081 to 0.00162)	0.00129 (0.00111 to 0.00150)	0.00131 (0.00115 to 0.00147)
D	−0.699 (−1.201 to −0.171)	−0.493 (−0.932 to −0.080)	−0.947 (−1.471 to −0.429)	−1.139 (−1.666 to −0.480)	−0.737 (−0.992 to −0.485)	−0.828 (−1.022 to −0.616)
d	0.00990 (0.00797 to 0.01200)	0.01005 (0.00884 −0.01145)	0.00857 (0.00657 to 0.01069)	0.00796 (0.00612 to 0.00994)	0.00974 (0.00988 to 0.01169)	0.00935 (0.00964 to 0.01117)
π/d	0.114 (0.07344 to 0.16971)	0.085 (0.06210 to 0.10961)	0.138 (0.08729 to 0.21000)	0.117 (0.06633 to 0.17788)	0.113 (0.00888 to 0.01068)	0.117 (0.00863 to 0.01017)

NOTE.—Abbreviations and symbols follow Table 1.

(Charlesworth et al. 1993) or selective sweep effects (Maynard Smith and Haigh 1974) due to proximity to genes. These effects are not apparent on the X, either due to the distinct nature of molecular evolution and adaptation on this chromosome, or (more probably) due to the considerably smaller data set and much noisier estimates of the polymorphism statistics. As a result there is no obvious relationship between the X:A ratio and distance from genes after scaling for possible differences in mutation rate (fig. 1*c*). When we required 15× coverage instead of 10× coverage for each diploid genotype call (and half this for haploid genotype calls), polymorphism statistics showed similar trends with the exception that the X:A ratio appeared to increase with distance from genes (supplementary fig. S1, Supplementary Material online). In view of the clear effect of natural selection on loci near genes in autosomal DNA, for subsequent analyses we combined partitions (*c*), (*d*), and (*e*) for autosomal and X chromosome DNA. We thus focus the remaining discussion on sites greater than 51,000 bp from genes, and with a minimum of 10× coverage required per diploid genotype call.

Population Structure

Using the combined data from partitions (*c*) to (*e*), Mantel tests did not recover evidence for isolation by distance in either autosomal DNA or X chromosome DNA. For A loci, the correlation between genetic similarity and geographic distance was actually positive ($r = 0.44$, $P = 0.932$, one-tailed test), which is the opposite relation to that expected under isolation by distance. For X loci, after excluding the one female sample to make the pairwise similarity calculations consistent across all samples, the correlation was negative but not significant ($r = -0.215$, $P = 0.154$). When the female sample was included, the correlation was positive but still not significant ($r = 0.146$, $P = 0.791$).

However, examination of the site frequency spectrum (SFS) suggests the possibility that some more subtle form of population structure could be present despite the lack of evidence for isolation by distance. In particular, intermediate frequency autosomal variants appear to be more prevalent than expected for a panmictic population. Figure 2 shows the SFS for GC and AT variants for sites with high confidence genotype calls in all individuals as detailed below; the SFS for all 4 nt, not shown, also has a peak at intermediate frequency variants. To examine the possibility of systematic biases in genotype calls, we compared this to the SFS of positions with missing data and also quantified the number of heterozygotes across polymorphic sites. This peak was also apparent in the SFS of autosomal sites for which one high confidence genotype call was not possible for one individual. The number of heterozygous genotypes did not differ significantly from the expectation when alleles paired randomly for each polymorphism frequency category, though for intermediate frequency polymorphisms (where half of the nucleotides have a derived polymorphism) this expectation did fall within the first quartile (supplementary fig. S2, Supplementary Material online). Although not statistically significant, these results open the possibility that either

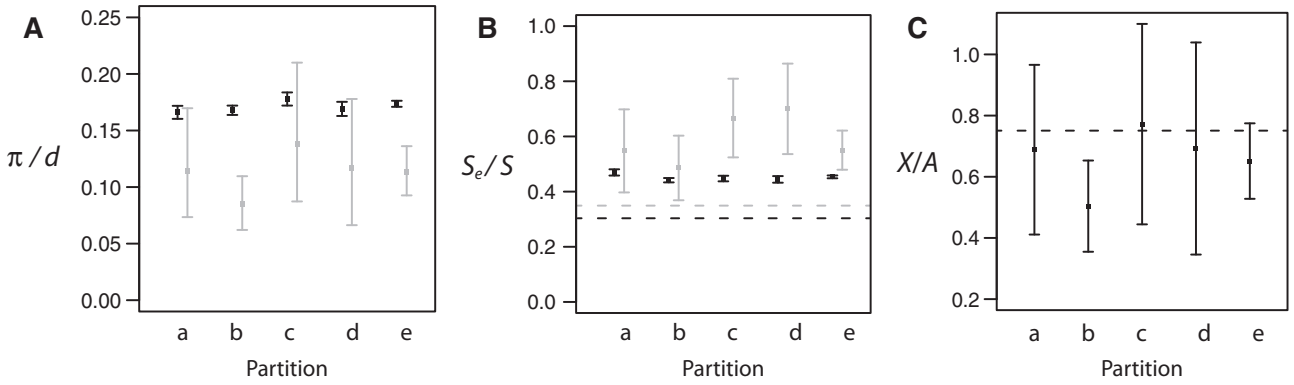

Fig. 1. Polymorphism statistics as a function of distance from genes, including pairwise nucleotide diversity divided by divergence from the olive baboon after correction for multiple substitutions and ancestral polymorphism (π/d; panel *A*), the proportion of polymorphisms that are singletons (S_e/S; panel *B*), and the scaled X:A ratio of diversities (*X/A*; panel *C*). Partitions include data that (*a*) spans genes plus 1,000-bp upstream and downstream, (*b*) are 1,001–51,000 bp from genes, (*c*) are 51,001–101,000 bp from genes, (*d*) are 101,001–151,000 bp from genes, or (*e*) are greater than 151,000 from genes. Autosomal DNA and X chromosome DNA are indicated in black and gray, respectively; bars indicate 95% CIs estimated as described in Materials and Methods. In (*B*), gray and a black dotted lines indicate the neutral expectation for the X chromosome and autosomes, respectively. In (*C*), a black dotted line indicates the expectation with no sex difference in the variance in reproductive success.

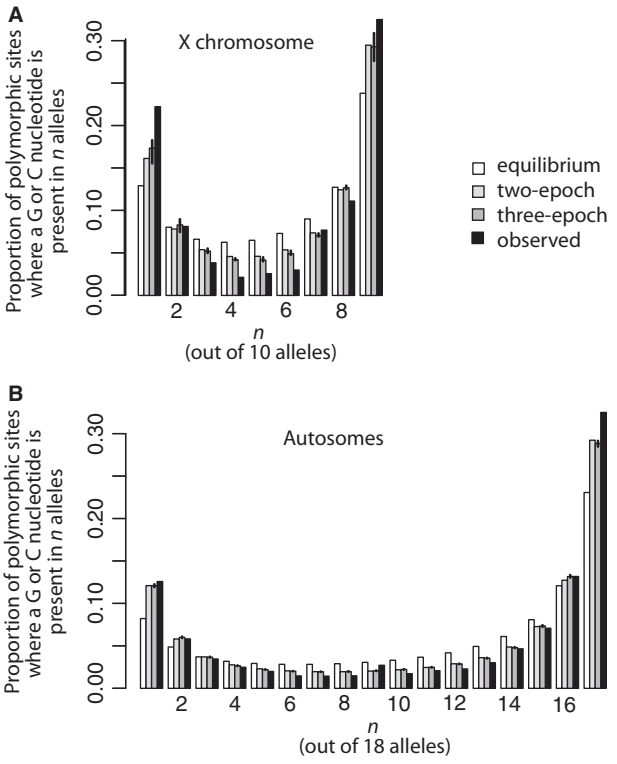

Fig. 2. SFSs of nucleotide polymorphisms where a G or a C is segregating with an A or a T in (*A*) X chromosome DNA and (*B*) autosomal DNA, based on the predictions from the equilibrium (white), two-epoch $\gamma_X = \lambda\gamma_A$ model (light gray), and three-epoch $\gamma_X = \lambda\gamma_A$ model (dark gray), and observed data (black). Only the SFSs for positions with genotype calls in all individuals is shown. For aDNA, the observed data has a peak at the intermediate value ($n = 9$). Bars indicate 95% CIs for the predicted SFS based on bootstrap analysis of the three-epoch $\gamma_X = \lambda\gamma_A$ model.

population structure or biases in genotype calling at nearly intermediate frequency polymorphisms could have modest effects. However, Zeng and Charlesworth (2010a) found that population structure did not compromise the ability of our model-based approach to detect selection.

The X:A Ratio and Y:A Ratio of *M. tonkeana* Nucleotide Site Diversities

After scaling π by the corrected pairwise divergence from humans, the estimate of the X:A ratio was 0.519 (95% CI: 0.446–0.593). After scaling by the corrected pairwise divergence from the olive baboon, the estimate of the X:A ratio was 0.672 (95% CI: 0.562–0.782). The *z*-test described in the Materials and Methods section shows that these point estimates are significantly lower or not significantly lower, respectively, than the expectation (i.e., 0.75) for equal variances in reproductive success in males and females ($P < 0.0001$ and $P = 0.168$, respectively, two-tailed *z*-tests). A power analysis indicated that, given the data, a two-tailed *z*-test should be able to detect a significant difference from the null hypothesis that the X:A ratio is equal to 0.75 when the true X:A ratio is higher than 0.813, or lower than 0.695.

We also explored the effects of various quality control criteria used in our analysis by calculating the X:A ratio using alternative settings and found that the results were consistent. For example, when a buffer of 5,000 instead of 10,000 bp was discarded around sites on the male-hemizygous portion X chromosome that were heterozygous in one or more males (a possible indication that the region was actually pseudoautosomal or autosomal), the X:A ratio after scaling by the corrected pairwise divergence from the olive baboon was 0.672 (95% CI: 0.564–0.781). Similarly, when a more stringent Phred-scaled genotype quality score of 40 was enforced for the sex chromosomes (instead of 30), this ratio was 0.670 (95% CI: 0.561–0.781).

The combined Y chromosome data, after scaling by pairwise divergence from humans, and compared with partitions (*c*) to (*e*) from the autosomes, yielded a Y:A ratio of 0.060. We did not use a *z*-test to evaluate this ratio because π for the nonrecombining portion of the Y chromosome is the outcome of a single coalescent process, rather than representing the mean over a large number of independent coalescent processes. However, the upper 95% CI for π_Y obtained from

coalescent simulations, divided by the lower 95% CI for d_Y (0.0053), was lower than the lower 95% CIs for $R_Y\pi_A/d_A$ (0.011 or 0.0055), where π_A/d_A is the pairwise diversity of the autosomes scaled by divergence to humans, and R_Y is 0.25 or 0.125, which correspond, respectively, to the expected scaling factor for equal variances in reproductive success in males and females, and the expected scaling factor associated with the maximal variance in male reproductive success (see Materials and Methods). This suggests that the Y:A ratio may indeed be much smaller than 0.125.

Model-Based Estimation of the X:A Ratio in *M. tonkeana*

Use of the new model-based approach described in the Materials and Methods section and in supplementary material S1, Supplementary Material online, resulted in a significantly better fit to the data when population size changes were included, providing new insights into the demography and genome evolution of the tonkean macaque. After accounting for a dynamic demography and natural selection by fitting the evolutionary model to the combined partitions (*c*)–(*e*) data, the three-epoch model was preferred over a two-epoch and the equilibrium models (table 3); for the comparison between the full three-epoch model and the equilibrium model, the difference in the Akaike Information Criterion (ΔAIC) = 5516.6 and $P < 1 \times 10^{-15}$ (χ^2 test, degrees of freedom [df] = 5); for the comparison between the full three- and two-epoch models ΔAIC = 38.6 and $P = 5.6 \times 10^{-10}$ (χ^2 test, df = 2).

Within the three-epoch model, we explored variants of the "full" model by constraining some parameters to either be equal to other parameters or be equal to some constant value expected under some circumstance (e.g., zero selection or equivalent variance in male and female reproductive success). A model in which the X:A ratio (denoted by λ in the model) was fixed at 0.75 was not significantly worse than the full model in which λ was treated as a free parameter (table 3; $P = 0.59$, χ^2 test, df = 1). The model also includes parameters γ_A and γ_A which are equal to $4N_{eX}s_X$ and $4N_{eA}s_A$, where N_{eX} and N_{eA} are the effective population sizes of the X and the A, respectively; and s_X and s_A are the GC-bias parameters for the X and A, respectively. Because our interests centered on the X:A ratio, we performed a bootstrap analysis on the model in which λ was estimated but the γ_X parameter was fixed at a value equal to $\lambda\gamma_A$, even though this model did not provide the best fit to the data. The 95% CI based on the bootstrapped data for the λ estimate is 0.567–0.860, which is consistent with the inference from the chi-square test and also similar to the conclusions based on nucleotide diversities using divergence from baboon to control for variation in mutation rate between the X and A. Therefore, even though the maximum-likelihood (ML) estimate of λ is lower than 0.75 for almost all of the three-epoch models employed, which suggests that the variance in female

Table 3. Models and Estimated Parameter Value; Further Details Are Available in Text.

Model	$\theta_{01\ X}$	$\theta_{10\ X}$	γ_X	$\theta_{01\ A}$	$\theta_{10\ A}$	γ_A	λ	ρ_1	τ_1	ρ_2	τ_2	ln *L*	ΔAIC
Equilibrium	0.00167	0.00070	0.77237	0.00612	0.00217	1.182	—	—	—	—	—	−5,945,825.43	—
Two epoch													
Full	0.00111	0.00052	0.65495	0.00375	0.00159	0.995	0.942	2.655	0.234	—	—	−5,943,083.43	1.3
$\lambda = 0.75$	0.00106	0.00049	0.66452	0.00376	0.00160	0.994	0.75 (fixed)	2.655	0.231	—	—	−5,943,083.79	0.0
$\gamma_A = 0$	0.00104	0.00049	0.64500	0.00220	0.00250	0 (fixed)	1.010	3.110	0.235	—	—	−5,950,000.00	13832.4
$\gamma_X = 0$	0.00082	0.00073	0 (fixed)	0.00375	0.00159	0.994	0.964	2.656	0.234	—	—	−5,943,096.44	25.3
$\gamma_X = \lambda\gamma_A$	0.00109	0.00047	$\lambda\gamma_A$ (fixed)	0.00376	0.00160	0.994	0.742	2.655	0.231	—	—	−5,943,083.98	0.4
$\theta_{01\,A} = \theta_{10\,A}$	0.00105	0.00049	0.64614	$\theta_{10\,A}$ (fixed)	0.00244	0.131	0.905	3.041	0.213	—	—	−5,947,129.99	8092.4
$\theta_{01\,X} = \theta_{10\,X}$	$\theta_{10\,X}$ (fixed)	0.00077	−0.11128	0.00375	0.00159	0.994	0.956	2.656	0.234	—	—	−5,943,101.28	35.0
$\theta_{01\,X} = \lambda\theta_{01\,A}$	$\lambda\theta_{01\,A}$ (fixed)	0.00033	0.97190	0.00379	0.00161	0.993	0.255	2.654	0.223	—	—	−5,943,096.12	24.7
$\theta_{10\,X} = \lambda\theta_{10\,A}$	0.00079	$\lambda\theta_{10\,A}$ (fixed)	0.55496	0.00380	0.00162	0.995	0.254	2.652	0.220	—	—	−5943097.02	26.5
Three epoch													
Full[a]	0.00104	0.00046	0.71394	0.00407	0.00162	1.061	0.686	50.994	0.010	1.441	0.077	−5,943,062.13	3.3
$\lambda = 0.75$	0.00107	0.00048	0.70045	0.00406	0.00162	1.060	0.75 (fixed)	50.797	0.010	1.466	0.077	−5,943,062.27	1.5
$\gamma_A = 0$[a]	0.00098	0.00042	0.75276	0.00237	0.00268	0 (fixed)	0.713	69.209	0.010	1.188	0.073	−5,948,417.98	10713.0
$\gamma_X = 0$	0.00075	0.00067	0 (fixed)	0.00407	0.00162	1.061	0.696	51.144	0.010	1.427	0.076	−5,943,074.94	26.9
$\gamma_X = \lambda\gamma_A$	0.00104	0.00045	$\lambda\gamma_A$ (fixed)	0.00407	0.00162	1.061	0.681	51.010	0.010	1.439	0.076	−5,943,062.13	1.3
$\theta_{01\,A} = \theta_{10\,A}$	0.00108	0.00051	0.64594	$\theta_{10\,A}$ (fixed)	0.00251	0.131	0.901	3.933	0.130	0.969	0.024	−5,947,119.62	8116.2
$\theta_{01\,X} = \theta_{10\,X}$	$\theta_{10\,X}$ (fixed)	0.00070	−0.11132	0.00407	0.00162	1.062	0.677	51.426	0.010	1.405	0.076	−5,943,079.28	35.6
$\theta_{01\,X} = \lambda\theta_{01\,A}$	$\lambda\theta_{01\,A}$ (fixed)	0.00019	1.82820	0.00412	0.00162	1.075	0.317	55.144	0.010	1.033	0.068	−5,943,073.91	24.8
$\theta_{10\,X} = \lambda\theta_{10\,A}$	0.00081	$\lambda\theta_{10\ A}$ (fixed)	0.30619	0.00396	0.00166	1.008	0.320	6.150	0.054	1.520	0.058	−5,943,083.24	43.5
$\gamma_X = 0.75\gamma_A$	0.00112	0.00045	$\lambda\gamma_A$ (fixed)	0.00406	0.00162	1.059	0.75 (fixed)	50.776	0.010	1.470	0.077	−5,943,062.50	0.0

NOTE.—ΔAIC refers to the difference in the Akaike Information Criterion statistic of each model and the best model in the two-epoch or the three-epoch categories; larger ΔAIC values indicate a poorer fit compared with the best fitting model. ΔAIC comparison among model categories (not shown) indicates that the three-epoch model is preferred over the two-epoch model, and both of these models are preferred over the equilibrium model. For all analyses, multiple runs reached convergence within a relative functional tolerance of 1×10^{-15} except where noted.

[a]Multiple runs converged within a relative functional tolerance of 1×10^{-12}.

reproductive success is higher than that of males and/or that hitchhiking effects are more intense on X than A, we cannot reject the hypothesis that λ is equal to 0.75 after taking demography, selection, and mutational bias into account. To explore the effects of missing data from RADseq (Arnold et al. 2013; Davey et al. 2013), model fitting was also performed after excluding sites with missing data, and the results were similar (supplementary table S2, Supplementary Material online). We anticipate that the power of this ML approach is somewhat lower than the *z*-test implies because the ML approach does not consider divergence information. However, because the ML estimate of λ is below 0.75, our failure to detect a value that is significantly higher than 0.75 does not appear to be related to a lack of statistical power.

GC Content; Departures from Stationarity

The expected equilibrium autosomal diversity prior to population size changes (π_A) is equal to $2\kappa\theta_{10A}/(1+\kappa)$ where κ is the ratio of the GC $\rightarrow$ AT mutation rate (μ_{01A}) to the AT $\rightarrow$ GC mutation rate (μ_{10A}) (Charlesworth and Charlesworth 2010, p. 237). Using the parameter values estimated for the three-epoch model with $\gamma_X = \lambda\gamma_A$, κ is equal to 2.51 for the autosomes, indicating that the mutation rate to AT nucleotides is higher than the mutation rate to GC nucleotides. The equilibrium diversity for autosomal loci obtained from the model fitting is therefore 0.00232, which is similar to the estimates from the pairwise nucleotide diversity (tables 1 and 2).

An interesting characteristic of the data is revealed by examining the frequency spectra of polymorphic positions in which a G or C nucleotide is segregating with an A or T nucleotide: In X chromosome and autosomal DNA, high frequency G or C polymorphisms are more common than high frequency A or T polymorphisms (fig. 2). Consistent with this observation, the model fitting found significant evidence for gBGC or selection favoring increased GC content in that the models where the selection parameters γ_X or γ_A were set to zero were significantly less likely than the models in which these parameters were estimated as free parameters, or when they were fixed by the relation $\gamma_X = \lambda\gamma_A$ (table 3). Moreover, in the three-epoch model with $\gamma_X = \lambda\gamma_A$, the 95% CIs for γ_A were greater than zero (0.99–1.08). The likelihood of a model in which the γ_X parameter was fixed to be equal to $\lambda\gamma_A$ was almost identical to that of the full model in which γ_X and γ_A were estimated separately, and not significantly worse than the full model (table 3; $P = 0.95$, χ^2 test, df = 1). This indicates that there is no significant difference in the effects of gBGC or selection for GC content between X and A (table 3). To explore whether a combination of two models was preferred, we tested a model in which the γ_X parameter was fixed at a value of $0.75\gamma_A$; this model was also not significantly worse than the full model (table 3; $P = 0.69$, χ^2 test, df = 2) and it had the lowest Akaike Information Criterion value overall.

Counts of divergent sites between the tonkean macaque and the outgroup species suggest significant departure from stationarity for A for both outgroup comparisons. This result involved only a small difference, and could be due to slight departures from equilibrium in the outgroup lineages, the tonkean macaque, or a combination of the two possibilities. Sites that were G or C in the tonkean macaque and A or T in the outgroup, hereafter GC:AT sites, were more common than sites with the opposite pattern, hereafter AT:GC sites, except for the comparison with human X chromosome DNA. For autosomal DNA, GC:AT sites comprised 51.0% out of 428,990 sites for the comparison with humans ($P < 0.00001$; binomial test) and 50.8% out of 128,267 sites for the comparison with baboons ($P < 0.00001$; binomial test). For X chromosome DNA, GC:AT sites comprised 47.9% out of 12,993 sites for the comparison with humans ($P < 0.00001$; binomial test), and 51.5% out of 2,735 sites for the comparison with baboons ($P = 0.058$; binomial test).

Male to Female Mutation Rate Ratio

We estimated the mutation rate of X chromosome DNA to autosomal DNA (μ_X/μ_A) based on Jukes–Cantor (1969) corrected pairwise divergence between the tonkean macaque and an outgroup, and from the model fitting, where $\mu_X/\mu_A = (\theta_X/\theta_A)(1/\lambda)$ (Materials and Methods described in supplementary material S1, Supplementary Material online). There was a substantial difference between each estimate: μ_X/μ_A based on divergence from humans was 0.772 (95% CI: 0.746–0.797), μ_X/μ_A based on divergence from baboons was 0.590 (95% CI: 0.544–0.635), and μ_X/μ_A based on the model fitting was 0.398 (95% CI: 0.392–0.405). When μ_X/μ_A was instead estimated from the AT to GC mutation rate (supplementary material S1, Supplementary Material online), the estimate of this ratio was almost identical to the outgroup-based estimates above. The disparity between the outgroup-based estimates and the model-based estimate could be related to the latter taking natural selection for GC or gBGC into account, or possibly a change in the rate of mutation subsequent to the the divergence of humans and macaques. It is also possible that this disparity is related to the simplifying assumptions used to construct the model, which is associated with an imperfect fit to the observed data (see Discussion). These values also may be uniquely affected by a possible reduction in the autosomal mutation rate in the tonkean macaque lineage (supplementary material S1, Supplementary Material online). In any case, all of these estimates yield a ratio that is significantly less than 1, which is consistent with faster male evolution.

A Dynamic Demography

Because of uncertainty surrounding variation in the rate of mutation in the tonkean macaque lineage (supplementary table S1, Supplementary Material online), we provide "ballpark" interpretations of demographic parameters from the model fitting, using the mutation rate inference from AT to GC divergence (μ_{10}) of X chromosome DNA using humans as an outgroup to estimate N_e and time in generations (supplementary material S1, Supplementary Material online), and we intentionally do not include CIs. The diversity estimates from the three-epoch model with $\gamma_X = \lambda\gamma_A$ translate to ancestral N_e equal to approximately 110,000 for autosomes. If the

ancestral population size is considered to be Epoch 0, going forward in time, the estimated population size in Epoch 1 is approximately 5.5 million individuals, with a duration of approximately 110,000 generations. The estimated effective population size in Epoch 2 is approximately 160,000 individuals, with a duration of approximately 24,000 generations. Given a macaque generation time of 5 years (Dittus 1975; Lindburg and Harvey 1996), the model fitting thus supports a substantial population expansion relative to the ancestral population approximately 550,000 years ago followed by a return to near the ancestral population size approximately 120,000 years ago. Because of the sensitivity of these demographic parameters to the value of the mutation rate, we prefer to interpret them in relative terms, as presented in table 3.

Discussion

Purifying Selection in Genic Regions

In humans and several other ape species, molecular polymorphism is higher in nongenic than genic regions, with the diversity at X chromosome sites increasing at a faster rate with their distance from genes than that for autosomal sites, resulting in an increase in the X:A ratio with distance from genes (Hammer et al. 2010; Gottipati et al. 2011; Prado-Martinez et al. 2013). In the tonkean macaque, multiple aspects of molecular polymorphism in or near autosomal genic regions on the autosomes are consistent with a stronger effect of purifying selection as compared with regions that are more distant from genes (fig. 1) but this pattern was not observed for the X chromosome. This probably reflects the smaller number of high confidence genotype calls that were made on the X, and the resulting wider CIs (fig. 1). Even with substantially more data, a strong positive correlation with distance from genic regions was not observed for the Eastern lowland gorilla and the Bornean orangutan, and the autosomes also do not exhibit this trend in bonobos (Prado-Martinez et al. 2013). This suggests a similar level of genetic drift in genic and nongenic regions within these species, possibly related to small N_e or other demographic effects.

The X:A Ratio

The X:A ratio of pairwise coalescence time for the tonkean macaque obtained from nucleotide site diversities is lower than 0.75, with the significance of this disparity depending on which outgroup is used to scale for differences in mutation rate. An X:A ratio below 0.75 could arise if the variance in reproductive success were higher in females than in males, or it could be caused by a recent reduction in population size, such that the population size for the X chromosome has approached its new equilibrium faster than that for the autosomes (Pool and Nielsen 2007). This latter interpretation is qualitatively consistent with the evidence for a history of past expansion in population size, followed by a reduction (table 3). However, when we controlled for demographic effects (and also selection for GC content/gBGC), we found that the ML estimate of the X:A ratio was 0.68 but that this value was not significantly different from 0.75. This suggests that the variances in male and female reproductive successes are in fact rather similar. The fact that similar results are obtained using two-epoch and three-epoch models suggests that this result is robust to demographic assumptions.

The findings that the X:A ratio is not significantly greater than 0.75, and that the most likely estimate is lower than 0.75, are surprising in light of the female-biased adult sex ratio that has been widely observed in natural populations, and also the pervasive reports of genealogical discordance between phylogenetic relationships in mitochondrial DNA and autosomal DNA (Hoelzer et al. 1992; Melnick et al. 1993; Morales and Melnick 1998; Tosi et al. 2002, 2003; Evans et al. 2010), both of which suggest high variance in male reproductive success. Interestingly, other studies also report an X:A ratio that is at or below 0.75 in other species of macaque monkey, which presumably have a social system similar to that of the tonkean macaque. A study of the pig-tailed macaque (*M. nemestrina*) by Evans et al. (2010), for example, reported an X:A ratio of 0.62 based on molecular variation at synonymous sites within genes. Likewise, after accounting for demographic effects to some degree, estimates of the X:A ratio for three populations of the long-tailed macaque *M. fascicularis* (Indonesia/Malaysia, Philippine, and Vietnam) each were less than 0.75 (Osada et al. 2013). Whole-genome shotgun sequencing of 16 rhesus macaques also suggested that SNPs on the X chromosome were only half as frequent as in the autosomes (Rhesus Macaque Genome Sequencing and Analysis Consortium et al. 2007), although it is not clear how this translates to an X:A ratio of N_e because divergence information was not reported in that study.

At least five, not mutually exclusive, factors could contribute to an X:A ratio below 0.75 in macaque monkeys:

i) When generations overlap, the X:A ratio is influenced by the degree to which the same individual males monopolize breeding opportunities across breeding seasons. Within-sex variation in reproductive success among breeding seasons may counteract to a large extent any elevation of X:A associated with a sex difference in the variance of reproductive success within breeding seasons (Evans and Charlesworth 2013). Even social systems with very high within-breeding season variance in male reproductive success, such as a harem social system, can have an X:A ratio near 0.75 if turnover in male tenure of harems among breeding seasons is high (Evans and Charlesworth 2013). This could therefore contribute to an X:A ratio in the tonkean macaque being close to 0.75.

ii) The variance in female reproductive success could in fact be quite high in tonkean macaques. Life history data from Kenyan baboons, another species of papionin monkey, suggested that the variance in female reproductive success due to stochastic survivability during the reproductive (adult) phase was sufficiently high to offset the effects of high variance in male reproductive success on the X:A ratio (Storz et al. 2001).

iii) Male-biased dispersal during the initial colonization of Sulawesi Island could have introduced more autosomal than X chromosome diversity. Male-biased migration of humans into Europe, for example, has been proposed to account for an X:A ratio below 0.75 in non-African populations (Keinan and Reich 2010). However, in contrast to the hypothesized evolutionary scenarios in non-African humans, a demographic model with ongoing migration between Borneo and Sulawesi was not previously preferred over a model with a single dispersal event to Sulawesi (Sulawesi macaques originated from an ancestor of pig-tailed macaques on Borneo; Evans et al. 2010). This suggests that male-biased dispersal from Borneo to Sulawesi was limited in time to the initial colonization of Sulawesi, and is therefore unlikely to have affected X chromosome diversity in contemporary tonkean macaques.

iv) Asymmetrical male-biased dispersal across Sulawesi hybrid zones could decrease the X:A ratio in the tonkean macaque, by introducing molecular polymorphism into this species to a greater extent in autosomal DNA than X chromosome DNA. The tonkean macaque is known to hybridize with *M. maura*, *M. ochreata*, and *M. hecki* (Watanabe, Lapasere, et al. 1991; Watanabe, Matsumura, et al. 1991; Froehlich and Supriatna 1996; Bynum et al. 1997; Evans et al. 2001) and the differentiated populations west and east of the Bongka River (Evans et al. 2001) may also have male-biased migration. Analysis of sex-biased dispersal at the *M. maura*/*M. tonkeana* hybrid zone, however, supports male-mediated gene flow from *M. tonkeana* to *M. maura* within their hybrid zone (Evans et al. 2001), which is the opposite direction from that needed to reduce the X:A ratio in the tonkean macaque. At this time it is not clear what patterns of sex-specific dispersal have occurred at the other hybrid zones, or to what degree hybridization actually results in genetic exchange beyond these hybrid zones.

v) There could be a high frequency of selective sweeps of recessive beneficial mutations on the X chromosome compared with autosomes due to male hemizygosity on the X chromosome (Charlesworth et al. 1987). In chimpanzees from central Africa, for example, synonymous diversity on the X chromosome is dramatically lower than on the autosomes, and the patterns of molecular polymorphism in coding regions are consistent with more frequent selective sweeps and stronger purifying selection on the X chromosome than autosomal DNA (Hvilsom et al. 2012). If many adaptive mutations are recessive (Hvilsom et al. 2012), X chromosome hemizygosity in males and an associated lower effective rate of recombination on the X chromosome could underlie this pattern. Interestingly, in *Drosophila* the absence of recombination in males causes the effective recombination rate to be lower on the autosomes than on the X chromosome, a factor that may increase the influence of background selection or selective sweeps on the autosomes and contribute to an observed X:A ratio for African *Drosophila melanogaster* populations that is higher than 0.75 (Charlesworth 2012b).

Low Polymorphism on the Y Chromosome

Very little polymorphism was observed on the Y chromosome and the most likely value of the Y:A ratio was significantly less than the expectation with an equal variance in reproductive success in each sex. At first sight, this suggests higher variance in male than female reproductive success. But this interpretation is inconsistent with the X:A ratio discussed above, and the Y:A ratio is also significantly smaller than the minimum value that male competition can generate. However, Y chromosome polymorphism in *M. nemestrina* from Borneo is also very low (Evans et al. 2010). Thus, low polymorphism in *M. tonkeana* could be related in part to nonequilibrium conditions associated with reduced variation in an ancestor before colonization of Sulawesi. Another possibility is that founder effects during colonization of Sulawesi may have caused reduced diversity of the Y chromosome in the tonkean macaque. However, genomic regions with smaller effective population sizes than the autosomes, such as the X and Y chromosomes, are expected to reach equilibrium after a bottleneck more quickly than the autosomes (Hutter et al. 2007). This would elevate the X:A or Y:A ratios in the period before the autosomes reach equilibrium, which is the opposite of what we observed; founder effects associated with the colonization of Sulawesi therefore seem to be an unlikely explanation for low diversity on the Y compared with the A (or the X compared with the A).

Also of relevance is that the nonrecombining region of the Y chromosome is probably strongly influenced by background selection (Kaiser and Charlesworth 2009; Wilson Sayres et al. 2014), which would reduce variation on this chromosome compared with other parts of the genome. Variation on the Y chromosomes of humans is also significantly lower than the neutral expectation, even after correcting for the higher mutation rate in males, and is consistent with the expected effects of purifying selection via background selection (Wilson Sayres et al. 2014). Selective sweeps on the Y chromosome would also produce similar effects.

GC Content

Mammalian genomes typically have high heterogeneity in GC content among genomic regions, with GC content being positively correlated with functional regions and genomic features, including gene density, repeat elements, recombination, and the level of gene expression (Eyre-Walker and Hurst 2001; Fullerton et al. 2001; Galtier et al. 2001; Kudla et al. 2006; Romiguier et al. 2010). If base composition is at equilibrium under drift and mutational bias alone, the mean allele frequency of G or C variants segregating from an AT → GC or a GC → AT mutation must equal the mean allele frequency of A or T variants (Eyre-Walker 1999). However, our results indicate that the mean frequency of G or C variants was higher

than that of A or T variants (fig. 2 and table 3). At sites for which high-confidence genotype calls were made for all individuals, for example, 1.6 or 2.2 times more G or C variants (for the X chromosome and autosomes, respectively) were segregating at greater than 50% frequency than A or T variants.

As pointed out by Eyre-Walker (1999), there are various explanations for these patterns, including nonstationarity of base composition, gBGC, and natural selection. The latter two cannot be distinguished with polymorphism data alone, because gBGC has the same effect as natural selection favoring GC content (Naglaki 1983). Using the test of Eyre-Walker (1994), we recovered evidence for nonstationarity based on a small overrepresentation of G or C nucleotides at divergent sites in *M. tonkeana*, so this explanation could contribute to our results to some degree, although the effect was slight. In addition, this observation does not necessarily mean that nonstationarity was present in the ancestral population; overall, it does not conclusively negate the significant evidence for gBGC or selection favoring GC content (table 3).

Recent surveys of genome sequences and exomes of mammals have also recovered support for gBGC, possibly combined with natural selection for GC content in mammals (Romiguier et al. 2010; Lartillot 2012; Pessia et al. 2012). Our model-based estimates of γ, the intensity of selection/gBGC, were approximately 1, which is comparable in magnitude to estimates derived previously in other primates. For instance, in humans, Spencer et al. (2006) estimated that $0.5 < \gamma < 1.3$, whereas Lynch (2010) suggested that $\gamma \approx 0.99$. De Maio et al. (2014) inferred that γ in the orangutan lineage was about 0.35, which was significantly lower than that inferred in the human–chimpanzee lineage (≈ 0.7), but noted that simulations suggested that their method tends to underestimate γ. Overall then, it seems plausible that gBGC or natural selection for GC content also operates in the genome of the tonkean macaque, and accounts for the poor performance of the models with no selection for GC content or gBGC (table 3).

A feature of primate genomes that was omitted from our new model was variation in GC content across different genomic regions (the isochore structure; Bernardi 2000; Eyre-Walker and Hurst 2001). There is evidence that γ increases with GC content (e.g., ranging from ≈ 0.2 to ≈ 1.2 in humans; De Maio et al. 2014). We opted not to incorporate this extra factor into our model, because we had relatively few polymorphic sites from the X chromosome, and further increasing the number of parameters may result in difficulty in parameter estimation. This and other simplifications of our model may account for the imperfect correspondence between the observed site frequency and that predicted by the model (fig. 2).

Male to Female Mutation Rate Ratio

McVean and Hurst (1997) proposed that natural selection could favor a lower mutation rate on X chromosomes due to exposure to deleterious recessives in hemizygous males, and that this, rather than differences in male and female mutation rate, could cause different neutral evolutionary rates on the autosomes and X chromosomes in rodents. However, we observed approximately 2-fold higher level of divergence in Y chromosome DNA than autosomal DNA (table 1 and supplementary table S1, Supplementary Material online); collectively these observations therefore are better explained by faster male evolution. Indeed, other studies have found faster rate of male evolution in macaques: The male:female mutation rate ratio (α) of the long tailed macaque (*M. fascicularis*) was estimated to be 3.8 (Osada et al. 2013) and α of the rhesus macaque was estimated to be 2–2.87 (Rhesus Macaque Genome Sequencing and Analysis Consortium et al. 2007; Elango et al. 2009).

Our population genetic model-based approach also suggests that autosomal DNA mutates at a higher rate than X chromosome DNA, but this mutation rate difference failed to yield an interpretable difference in male versus female mutation rate using the approach of Miyata et al. (1987), because the estimated μ_X/μ_A is lower than the theoretical minimum of 2/3. This shortcoming of the model-based approach is probably related to the two-allele simplification (i.e., the data were reduced from 4-nt states to only 2). Because of this simplification, the model did not account for selection and mutation biases associated with G $\leftrightarrow$ C or A $\leftrightarrow$T transversions which could affect the μ_X/μ_A (and thus α) estimate.

Conclusions

These data and analyses extend the efforts by Evans et al. (2010) to quantify the X:A ratio in macaques by analyzing molecular polymorphism in nongenic regions. These regions presumably are influenced to a lesser degree by background selection and selective sweeps, and the methods deployed in this study also better accommodate a dynamic demography and selection on GC content and gBGC. Nonetheless, the results are strikingly similar between the two studies, despite use of different species, differently sized data sets, and distinct analytical approaches. In both studies, the estimated X:A ratio was less than or equal to the expectation if variance in reproductive success is equal in males and females, and very low polymorphism was observed on the Y chromosome. This first finding is 1) surprising in light of the observed adult female-biased sex ratio in many macaque populations but is consistent with independent results based on molecular polymorphism from other macaque species, and 2) unlikely to be attributed to a pulse of male-biased dispersal during the initial colonization of Sulawesi. The second finding of low polymorphism on the Y is potentially contradictory to the low X:A ratio, but the two can be reconciled by invoking background selection/recurrent selective sweeps on the nonrecombining portion of the Y chromosome. Another possible contributing factor to the low polymorphism on the Y chromosome of *M. tonkeana* is that nonequilibrium conditions were present in the macaque ancestor prior to dispersal to Sulawesi.

In summary, it appears that female-specific factors counteract or overshadow some of the genomic consequences of high variance in male reproductive success in macaque monkeys. This observation is not typical of great apes (Prado-Martinez et al. 2013), including at least some human populations (Wilson Sayres et al. 2014), and raises the

question of whether and to what degree this occurs in other primates.

Materials and Methods

Genetic Samples, "RAD Tags," and Genotyping

To evaluate the relative effective population sizes (N_e) of autosomal DNA and sex chromosome DNA, we collected molecular polymorphism data from nine (one female and eight male) tonkean macaques, with sampling locations on Sulawesi Island, Indonesia, as detailed in Evans et al. (2003), for the following samples: PF515, PM561, PM565, PM566, PM567, PM582, PM584, PM592, and PM602. Because all individuals originated from habitat on only one side (west) of the Bongka River, the effects of large-scale population structure in this species (Evans et al. 2001, 2003) were expected to be minimal, although gene flow with populations east of the Bongka River, or with other macaque species on Sulawesi, is possible. To further explore the possibility of isolation by distance (Wright 1943), we tested for evidence of a correlation between genetic and geographic distance as described below.

We collected polymorphism data using restriction enzyme associated DNA sequencing (RADseq; Baird et al. 2008). RADseq involves adding linkers to genomic DNA ends that are cut with a rare-cutting restriction enzyme (SbfI) and also sheared at positions near the restriction sites, resulting in a large sequence data set for multiple individuals that consists of homologous portions of the genome near the restriction enzyme sites. RAD tag libraries were prepared by Floragenex (OR), and multiplexed paired end sequencing was performed using one Illumina lane. These data have been deposited in the NCBI (short read accession number SRP041222).

We first mapped the tonkean macaque sequences to the rhesus macaque genome (rhemac2), and performed genotype calling using BWA, Samtools, the Genome Analysis Toolkit, and vcftools (Li et al. 2009; Li and Durbin 2010; McKenna et al. 2010), as described in supplementary material S1, Supplementary Material online. Data that passed rigorous quality control filters (supplementary material S1, Supplementary Material online) were retained for analysis. Based on lineage-specific estimates of divergence discussed in supplementary material S1, Supplementary Material online, the time since divergence of the rhesus and tonkean macaque in units of two times the autosomal effective population size (N_e) of *M. tonkeana* generations is approximately 2–3. Thus, the rhesus macaque is sufficiently closely related to the tonkean macaque that genetic distance is likely to be heavily influenced by ancestral polymorphism (Charlesworth et al. 2005). For this reason, the rhesus macaque genome sequence was used to inform the genomic locations of the tonkean macaque data, but not as an outgroup for these genomic regions.

For autosomal data, we considered only positions that did not map to either the X or the Y chromosome of the rhesus macaque (genome assembly rhemac2 and Y chromosome data from Hughes et al. [2012]) or X chromosome data of an outgroup species, which was either the olive baboon (genome assembly papAnu2) or human (genome assembly hg19). For X chromosome data, we considered only positions that mapped to the nonpseudoautosomal portion of the X chromosome in the rhesus macaque and an outgroup species, and we discarded sites within and including 10,000 bp from any heterozygous genotype call in a male. This latter measure was taken in an attempt to exclude putatively species-specific pseudoautosomal regions of the tonkean macaque that would not be identified using annotations from the rhesus macaque. For Y chromosome data, we only considered positions that mapped to the Y chromosome of humans and rhesus macaques.

Alignments of the rhemac2 and hg19 genomes were obtained from the UCSC genome browser. We aligned the genome sequences of the rhesus and olive baboon using LASTZ, release 1.02.00 (Harris 2010), with the same settings used for the human–rhesus alignment as described on the UCSC browser (http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/vsRheMac2/, last accessed June 30, 2014). Our analyses in part focused on positions for which genotype calls were made for all nine tonkean macaque individuals, in order to minimize biases related to missing haplotypes (Arnold et al. 2013; Davey et al. 2013). The main analysis with the newly developed software, however, used the full postfiltering data set, including sites with missing genotypes in some individual tonkean macaques. We repeated this analysis including only sites with genotype calls in all individuals, with similar results, as detailed in supplementary material S1, Supplementary Material online.

Evidence for Isolation by Distance or Other Forms of Population Structure

Isolation by distance is a pervasive population genetic pattern that is not well approximated by the epoch model deployed here (see below). To test for evidence of isolation by distance we performed Mantel tests using genetic similarity and geographic distances estimated from the sampling localities (Evans et al. 2003). We estimated the percent similarity between each individual and all others for autosomal DNA and for X chromosome DNA, using positions where high confidence genotype calls were made in all individuals. Significance was assessed with 10,000 permutations using the program zt (Bonnet and Van de Peer 2002). The null hypothesis for this test is that there is no correlation between genetic and geographic distance; the alternative hypothesis is that genetic similarity decreases as geographic distance increases (a one-tailed test for a negative correlation).

Testing for Effects of Natural Selection Near Genes

To explore genome-wide effects of natural selection, we divided the data into five categories based on locations with respect to annotated genes in the rhesus macaque genome. These five categories included: a) Regions containing genes, including introns, exons, and 5′- and 3′-untranslated regions, plus 1,000-bp upstream and downstream; b) regions that were up to 50,000 bp from (a); c) regions that were 50,001–100,000 bp from (a); d) regions that were

100,001–150,000 bp from (a); and e) regions that were greater than 150,000 bp from (a).

We then tested for variation in the nature or strength of natural selection among categories by examining how polymorphism varied with respect to the genomic distance from genes. In particular, for each partition we calculated pairwise nucleotide diversity per site (π) divided by pairwise divergence from the outgroup species (olive baboons or humans), with corrections for multiple substitutions and ancestral polymorphism (π/d). Corrections for multiple substitutions used the method of Jukes and Cantor (1969). Corrections for ancestral polymorphism were done by subtracting π (averaged across the autosomes, X chromosome, or Y chromosome) from the pairwise distance for the corresponding genomic region (Makova and Li 2002; Charlesworth B and Charlesworth D 2010, p. 258–259). This correction assumes that the effective population sizes and mutation rates of the homologous sequences in the most recent common ancestor of the outgroup and the tonkean macaque were similar to those of the homologous sequences of the tonkean macaque.

Following Zeng and Charlesworth (2011), we also estimated the proportion of polymorphisms (S_e) that are singletons (that is, present as only one segregating variant in the sample) among the total number of segregating sites in the region (S) (S_e/S). The expectation under neutral equilibrium is approximately 30% for our sample size of autosomes and approximately 35% for our sample size of X chromosomes (Charlesworth B and Charlesworth D 2010, p. 273). Finally, we calculated Tajima's D (Tajima 1989) on a per sequence basis for each of these categories. Approximate 95% CIs for S_e/S were obtained by assuming that the variance in the number of singletons followed binomial expectations. For π/d, D, and other parameters for the X chromosome and autosomes in tables 1 and 2, 95% CIs were approximated by generating 1,000 pseudoreplicate data sets by bootstrapping by site and identifying the 2.5th and 97.5th percentiles among these pseudoreplicates. We note that this approach provides only a rough estimate of CIs because it does not account for nonindependence of mutation rates among linked sites. For analyses of π/d, S_e/S, and D, we restricted the focus to nucleotide positions for which high-quality genotype calls (supplementary material S1, Supplementary Material online) were made for all nine individuals. If natural selection and hitchhiking are weaker in genomic regions that are far from genes, π/d should be higher, S_e/S should be lower, and D should be nearer zero in these regions than in those genomic regions near genes. These expectations remain qualitatively unchanged in the presence of recent demographic changes (Zeng 2013).

For Y chromosome data, 95% CIs for most of the parameter estimates in table 1 were estimated from coalescent simulations using the program ms (Hudson 2002). For these simulations, the polymorphism parameter θ was set to the pairwise nucleotide diversity estimate (π) from the data, and simulations were performed under the demographic parameters inferred from the $\gamma_X = \lambda\gamma_A$ version of the three-epoch model (see below). Because sequence from the baboon Y chromosome was not available, we calculated pairwise divergence only from the human Y chromosome, and applied the same corrections for multiple substitutions and ancestral polymorphism detailed above. We obtained 95% CIs based on the variance calculated with equation (5) of Kimura and Ohta (1972).

Estimating the X:A Ratio from Nucleotide Site Diversities

For neutral or nearly neutral variants, π (Tajima 1983) provides an estimate of the mean coalescent time for a pair of alleles, which is equal to two times the effective population size (N_e) (Hudson 1990). Thus the relative diversity values, after adjusting for differences in mutation rate, provide an estimate of the X:A ratio of effective population size values. We note that diversity values estimated in this way will typically be lower than the neutral value with free recombination because of hitchhiking effects due to selection at other sites in the genome, and that these factors are probably not uniform across the autosomes and X chromosomes due to differences in recombination rates, selection intensities, mutation rates, and modes of inheritance (Charlesworth 2012a). Differences in mutation rate between the X chromosome and the autosomes can be controlled for by scaling these estimates by divergence from an outgroup after correcting divergence for multiple substitutions and ancestral polymorphism. Thus the X:A ratio = $(d_A/d_X)(\pi_X/\pi_A)$, where d_A and d_X are corrected divergence estimates for autosomal and X chromosome DNA, respectively, and π_A and π_X are π for A and X, respectively. Estimates of the X:A ratio using Watterson's (1975) estimator based on the number of segregating sites in a sample (θ_W) and π were similar; we report those from π because of its more direct theoretical link to coalescent time (Hudson 1990). The variances of π were calculated following Tajima (1983; Charlesworth B and Charlesworth D 2010 formula B5.6.3, p. 212–213). For A and X (but not the Y chromosome), we assumed that sites were independent, and divided the second term in these equations by the number of RAD tags in each chromosomal category, where a RAD tag was defined as one or more high-quality genotypes that is at least 5,000 bp away from another such genotype.

We also estimated the ratio of the effective population sizes of the male-specific portion of the Y chromosome to autosomal DNA (the Y:A ratio), using a similar approach. Because orthologous sequences from the baboon Y chromosome are not yet available, this analysis was only performed using only human divergence to correct the mutation rate.

The variance of the X:A ratio was estimated using the delta method to estimate squared standard errors of ratios Var $\left[\overline{X}/\overline{Y}\right]$ and products Var$\left[\overline{X}\overline{Y}\right]$:

$$\mathrm{Var}\left[\overline{X}/\overline{Y}\right] = \left(\hat{X}/\hat{Y}\right)^2\left[\mathrm{CV}\left(\hat{X}\right)^2 + \mathrm{CV}\left(\hat{Y}\right)^2\right] \quad (1)$$

and

$$\mathrm{Var}\left(\overline{X}\overline{Y}\right) = \hat{X}^2\mathrm{Var}\left(\hat{Y}\right) + \hat{Y}^2\mathrm{Var}\left(\hat{X}\right), \quad (2)$$

where $\overline{X}$ and $\overline{Y}$ are the expected values of the random

variables X and Y, respectively; $\hat{X}$ and $\hat{Y}$ are the mean estimates of random variables X and Y, respectively; CV() is the standard error of the random variable given in parentheses, divided by its mean (the coefficient of variation); and Var() is the variance of the random variable given in parentheses. For this calculation, the variance of π was calculated as described above and the variance of divergence was obtained with formula (5) in Kimura and Ohta (1972). Approximate 95% confidence limits for the X:A ratio was then calculated under the assumption of normality. Because we assume independence of sites, and do not take into account differences in the effective recombination rates between the X and A, including differences related to sex-specific rates of recombination (Mank 2009), these CIs (and those estimated from the bootstrap analysis below) may be anticonservative to some degree.

Departure of the X:A ratio from the expectations under equal variance in reproductive success in each sex was tested using a z-test as described and justified in Evans and Charlesworth (2013). In particular, to test whether the X:A ratio departed from the null expectation (R_{XA}) of 0.75, we tested whether the absolute value of the difference between $(1/R_{XA})$ $(\theta_X/d_X) - (\theta_A/d_A)$ was significantly greater than zero. The variance of this difference (Evans and Charlesworth 2013; eq. 3) was calculated from the variances of (θ_X/d_X) and (θ_A/d_A) obtained from the delta method described above.

A power analysis was performed by fixing π_A, d_A, d_X, and their associated variances at the observed values, and then evaluating the degree to which π_X would have to depart from the expectation of an X:A ratio of 0.75 in order for the z-test to detect a significant difference from this expectation 95% of the time.

Because the nonrecombining portion of the Y chromosome is inherited as a unit, the coalescent process gives a nonnormal distribution for π_Y and we therefore avoided the assumption that this statistic was normally distributed. Instead we compared the upper 95% CI for π_Y obtained from simulations (described earlier), divided by the lower 95% CI of the corrected pairwise Y divergence from humans (d_Y), with the lower 95% CI of π_A/d_A multiplied by the expected value (R_{YA}) of 0.25 or 0.125 associated, respectively, with no sex bias in variance in reproductive success, or with maximal male variance in reproductive success (Charlesworth 2001).

A New Model for Estimating the X:A Ratio from Polymorphism Data

The diversity-based metrics presented above only relate to the relatively recent past, involving a timescale of the order of twice the effective population size; the X:A ratio of N_e estimated in this way can be strongly influenced by factors such as recent population bottlenecks, and may not reflect long-term patterns if the population is far from equilibrium under drift and mutation (Pool and Nielsen 2007). In addition, these estimates may be influenced by selection for GC content or gBGC (Zeng 2012). We thus sought to estimate the X:A ratio of *M. tonkeana* while taking into account the potential influence of these factors. To this end, we have extended the method of Haddrill et al. (2011) (see also Zeng and Charlesworth 2009, 2010b, 2011) by making two changes: 1) We let the autosomes and the X chromosome have independent mutation rates, and 2) the model is implemented by using a more efficient numerical method based on Zhao et al. (2013) (see supplementary material S1, Supplementary Material online, for full details).

As described in Zeng and Charlesworth (2009), the model includes parameters for mutational bias, gBGC/selection, and demography. The model reduces the four states of nucleotide polymorphism to two states (A_0 and A_1), with reversible mutation between these states. The mutation parameters include a mutation rate μ_{01} from A_0 to A_1, and μ_{10} for the reverse mutation. The model allows the separate μ_{01} and μ_{10} parameters to be estimated for X and A, and permits either A_0 or A_1 to be favored by natural selection or by gBGC. Because the sequence data are nucleotides with four possible states, we collapsed the data into two states—G or C nucleotides (A_0) and A or T nucleotides (A_1). A consequence of using this model is that information is lost, for example, because a segregating polymorphism of A and T (or of G and C) is considered invariant, and does not influence the likelihood of the data. However, this approach allowed us to include data from positions that violated the infinite sites model by having three segregating polymorphisms; these (rare) positions were excluded from the calculations discussed earlier. The model is also able to distinguish selection from demography and is robust to linkage between sites (Zeng and Charlesworth 2010a; Zeng 2012).

Prior to postulated changes in population size, we assumed that the population was at mutation–selection–drift equilibrium. Under this scenario, the X and A levels of polymorphism are a combined result of potentially different selection coefficients and effective population sizes. Nonequilibrium conditions are triggered by changes in population size, and subsequent polymorphism dynamics of X and A are affected by the relative sizes of their equilibrium effective population sizes (the X:A ratio). The nonequilibrium model thus has a parameter λ, which is the X:A ratio, such that $N_{eX} = \lambda N_{eA}$. Changes in population size are assumed to involve a series of epochs with different population sizes and durations, with transitions in population size occurring instantaneously between epochs, and with a constant population size within an epoch. Each of I epochs thus has a duration (τ_i) and population size (ρ_i) parameter, with $\rho_i = N_{e\_i}/N_{e\_ancestral}$ where N_{e_i} and $N_{e\_ancestral}$ refer to the effective population sizes of the ith epoch and the ancestral epoch (before any change in population size occurred), respectively, and τ_i is in units of $2N_{e\_i}$ generations. Population size changes are assumed to affect X and A identically, such that $N_{e\_i\_X}/N_{e\_ancestral\_X} = N_{e\_i\_A}/N_{e\_ancestral\_A} = \rho_i$. Because the number of parameters increases with each epoch considered, complicating the likelihood estimation procedure, we restricted our analysis to between one and three epochs.

The model also assumes codominance of any fitness effects. Following Haddrill et al. (2011), we define s_A and s_X to be the heterozygous selection coefficients for A and X,

respectively, such that A_0 is favored or selected against on autosomes when s_A is greater or less than zero (s_X is similarly defined). γ_A and γ_X are equal to 4 $N_{e\_ancestral}s_A$ and 4 $N_{e\_ancestral}s_X$, respectively; and θ_{01} (= 4 $N_{e\_ancestral}\mu_{01}$) and θ_{10} (= 4 $N_{e\_ancestral}\mu_{10}$) are the mutation parameters for A_0 (G or C nucleotides) and A_1 (A or T nucleotides). Thus, a positive value of γ implies that there is gBGC and/or selection favoring increased GC content. Additional mathematical details of this model are provided in supplementary material S1, Supplementary Material online.

The 95% CIs for the parameter estimates were estimated from 100 bootstrap replicates of the data (including polymorphic and nonvariant sites) within each level of missing data (2–18 even numbers of alleles for autosomal DNA and 1–10 alleles for X chromosome DNA alleles). At least 100 searches for the ML parameter values were performed on each bootstrap replicate, and we imposed a stringent convergence threshold with a relative tolerance of 10^{-13} on each replicate to ensure that the ML estimates we reported were reached multiple times by the search algorithm in independent runs, as described in supplementary material S1, Supplementary Material online.

Alternative models were developed by fixing various parameters to be equal to each other or to a constant, and calculating the likelihood of the data under these restrictions. Models with different numbers of parameters were then compared by means of likelihood ratio tests with appropriate degrees of freedom. For instance, to test whether there is support for a dynamic demography, we compared the likelihoods of a model with one epoch to a model with two or three epochs. To test whether an estimated value of λ was significantly different from the null expectation of 0.75, we compared the likelihood of a model where this parameter was fixed at this value to the likelihood of a model where λ was estimated. Using this strategy, we were also able to test whether s_A and s_X were significantly different from zero or from each other, whether rates of mutation between each class of mutation (from A_0 to A_1, and vice versa) differed significantly from each other within X or A, and whether each of these rates differed significantly between X or A. Software implementing these analyses is available at zenglab.group.shef.ac.uk.

Testing for Stationarity of Base Composition

If selection for GC content is present, or equivalently if gBGC exists, equilibrium (stationarity) of base composition exists when the GC content of a genome is constant and the number of mutations that convert an A or a T to a G or a C is equal to the number of mutations that convert a G or a C to an A or a T. To test for stationarity of base composition, we used a test proposed by Eyre-Walker (1994). Eyre-Walker (1994) pointed out that, under stationarity, the number of bases that are A or T in one species but G or C in a second species should be equal to the number of bases that are A or T in the second species but G or C in the first. We thus counted these categories of divergent sites in comparisons between the tonkean macaque and each of the two outgroup species (olive baboon and humans). For polymorphic positions in the tonkean macaque, the most common base was considered (Eyre-Walker 1999); if two alleles were segregating with equal frequency, a randomly chosen allele was compared with the outgroup. A binomial test was used to compare counts of each mutational category.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank John Davey, Alon Keinan, Gabor Marth, and Stephen Wright, and two anonymous reviewers for helpful discussions and comments. This work was supported by the National Science and Engineering Research Council of Canada (RGPIN/283102-2012 to B.J.E.). Data from this study have been deposited in the NCBI (short read accession number SRP041222).

References

Albrecht GH. 1978. The craniofacial morphology of the Sulawesi macaques, multivariate approaches to biological problems. In: Szalay FS, editor. Contributions to primatology. Basel: Karger. p. 1–151.
Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Mol Ecol.* 22:3179–3190.
Baird NS, Etter PD, Atwood TS, Currey MC, Schriver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3:e3376.
Beisner BA, Jackson ME, Cameron A, McCowan B. 2012. Sex ratio, conflict dynamics, and wounding in rhesus macaques (*Macaca mulatta*). *Appl Anim Behav Sci.* 137:137–147.
Bernardi G. 2000. Isochores and the evolutionary genomics of vertebrates. *Gene* 241:3–17.
Bonnet E, Van de Peer Y. 2002. zt: a software tool for simple and partial Mantel tests. *J Stat Softw.* 7:1–12.
Bynum EL, Bynum DZ, Supriatna J. 1997. Confirmation and location of the hybrid zone between wild populations of *Macaca tonkeana* and *Macaca hecki* in Central Sulawesi, Indonesia. *Am J Primatol.* 43: 181–209.
Charlesworth B. 2001. The effect of life-history and mode of inheritance on neutral genetic variability. *Genet Res.* 77:153–166.
Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* 10:195–205.
Charlesworth B. 2012a. The effects of deleterious mutations on evolution at linked sites. *Genetics* 190:5–22.
Charlesworth B. 2012b. The role of background selection in shaping patterns of molecular evolution and variation: evidence from variability on the *Drosophila* X chromosome. *Genetics* 191:233–246.
Charlesworth B, Bartolomé C, Noël V. 2005. The detection of shared and ancestral polymorphisms. *Genet Res.* 86:149–157.
Charlesworth B, Charlesworth D. 2010. Elements of evolutionary genetics. Greenwood Village (CO): Roberts and Company Publishers.
Charlesworth B, Coyne JA, Barton NH. 1987. The relative rates of evolution of sex chromosomes and autosomes. *Am Nat.* 130:113–146.
Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* 134: 1289–1303.
Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML. 2013. Special features of RAG sequencing data: implications for genotyping. *Mol Ecol.* 22:3151–3164.
De Maio N, Schlötterer C, Kosiol C. 2014. Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. *Mol Biol Evol.* 30:2249–2262.

Dittus W. 1975. Population dynamics of the toque macaque, *Macaca sinica*. In: Tuttle RH, editor. Socioecology and psychology of primates. Mounton (The Hague). p. 125–152.
Elango N, Lee J, Peng Z, Loh Y-HE, Yi SV. 2009. Evolutionary rate variation in Old World monkeys. *Biol Lett.* 5:405–408.
Evans BJ, Charlesworth B. 2013. The effect of nonindependent mate pairing on the effective population size. *Genetics* 193:545–556.
Evans BJ, Pin L, Melnick DJ, Wright SI. 2010. Sex–linked inheritance in macaque monkeys: implications for effective population size and dispersal to Sulawesi. *Genetics* 185:923–937.
Evans BJ, Supriatna J, Andayani N, Melnick DJ. 2003. Diversification of Sulawesi macaque monkeys: decoupled evolution of mitochondrial and autosomal DNA. *Evolution* 57:1931–1946.
Evans BJ, Supriatna J, Melnick DJ. 2001. Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution* 55: 1685–1702.
Eyre-Walker A. 1994. DNA mismatch repair and synonymous cofon evolution in mammals. *Mol Biol Evol.* 11:88–98.
Eyre-Walker A. 1999. Evidence of selection on silent site base compositions in mammals: potential implications for the evolution of isochores and junk DNA. *Genetics* 152:675–683.
Eyre-Walker A, Hurst LD. 2001. The evolution of isochores. *Nat Rev Genet.* 2:549–555.
Fooden J. 1969. Taxonomy and evolution of the monkeys of Celebes (Primates: Cercopithecidae). Switzerland: S. Karger.
Froehlich J, Supriatna J. 1996. Secondary intergradation between *Macaca maurus* and *M. tonkeana* in South Sulawesi, and the species status of *M. togeanus*. In: Fa J, Lindburg DG, editors. Evolution and ecology of macaque societies. Cambridge: Cambridge University Press. p. 43–70.
Fullerton SM, Carvalho AB, Clark AG. 2001. Local rates of recombination are positively correlated with GC content in the human genome. *Mol Biol Evol.* 18:1139–1142.
Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159:907–911.
Gottipati S, Arbiza L, Siepel A, Clark AG, Keinan A. 2011. Analyses of X-linked and autosomal genetic variation in population scale whole genome sequencing. *Nat Genet.* 43:741–743.
Haddrill PR, Zeng K, Charlesworth B. 2011. Determinants of synonymous and nonsynonymous variability in three species of *Drosophila*. *Mol Biol Evol.* 28:1731–1743.
Hammer MF, Woerner AE, Mendez FL, Watkins JC, Cox MP, Wall JD. 2010. The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat Genet.* 42: 803–831.
Harris B. 2010. LASTZ Release 1.02.00, built January 12, 2010.
Hasan MK, Aziz MA, Alam SMR, Kawamoto Y, Jones-Engel L, Kyes RC, Akhtar S, Begum S, Feeroz MM. 2013. Distribution of Rhesus Macaques (*Macaca mulatta*) in Bangladesh: interpopulation variation in group size and composition. *Primate Conserv.* 26: 125–132.
Hedrick PW. 2007. Sex: differences in mutation, recombination, selection, gene flow, and genetic drift. *Evolution* 61:2750–2771.
Hoelzer GA, Hoelzer MA, Melnick DJ. 1992. The evolutionary history of the *sinica*-group of macaque monkeys as revealed by mtDNA restriction site analysis. *Mol Phylogenet Evol.* 1:215–222.
Hoelzer GA, Wallman J, Melnick DJ. 1998. The effects of social structure, geographical structure, and population size on the evolution of mitochondrial DNA: II. Molecular clocks and the lineage sorting period. *J Mol Evol.* 47:21–31.
Hsu MJ, Lin J-F. 2001. Troop size and structure in free-ranging Formosan macaques (*Macaca cyclopis*) at Mt. longevity, Taiwan. *Zool Stud.* 40: 49–60.
Hudson RR. 1990. Gene genealogies and the coalescent process. In: Futuyma D, Antoaovics J, editors. Vol. 7. Oxford surveys in evolutionary biology. New York: Oxford University Press. p. 144.
Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremitzki C, et al. 2012. Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* 483:82–86.
Hutter S, Li H, Beisswanger S, De Lorenzo D, Stephan W. 2007. Distinctly different sex ratios in African and European populations of *Drosophila melanogaster* inferred from chromosomewide single nucleotide polymorphism data. *Genetics* 177:469–480.
Hvilsom C, Qian Y, Bataillon T, Li Y, Mailund T, Sallé B, Carlsen F, Li R, Zheng H, Jiang H, et al. 2012. Extensive X-linked adaptive evolution in central chimpanzees. *PNAS* 109:2054–2059.
Jukes TH, Cantor CR. 1969. Evolution of protein molecules. In: Munro HN, editor. Mammalian protein metabolism. New York: Academic Press. p. 21–132.
Kaiser VB, Charlesworth B. 2009. The effects of deleterious mutations on evolution in non-recombining genomes. *Trends Genet.* 25:9–12.
Keinan A, Reich D. 2010. Can a sex-biased human demography account for the reduced effective population size of Chromosome X in Non-Africans? *Mol Biol Evol.* 27:2312–2321.
Kimura M, Ohta T. 1972. On the stochastic model for estimation of mutational distance between homologous proteins. *J Mol Evol.* 2:87–90.
Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. 2006. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biol.* 4:e180.
Lartillot N. 2012. Phylogenetic patterns of GC-biased gene conversion in placental mammals and the evolutionary dynamics of recombination landscapes. *Mol Biol Evol.* 30:489–502.
Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.
Lindburg DG, Harvey NC. 1996. Reproductive biology of captive lion-tailed macaques. In: Fa J, Lindburg DG, editors. Evolution and ecology of macaque societies. Cambridge: Cambridge University Press. p. 318–341.
Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. *PNAS* 107:961–968.
Makova KD, Li W-H. 2002. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* 416:624–626.
Mank JE. 2009. The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. *Genet Res.* 91:355–363.
Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res.* 23:23–25.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20:1297–1303.
McVean GT, Hurst LD. 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature* 386: 388–392.
Melnick DJ, Hoelzer GA. 1992. Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial DNA variation. *Int J Primatol.* 13:379–393.
Melnick DJ, Hoelzer GA, Absher R, Ashley MV. 1993. MtDNA diversity in rhesus monkeys reveals overestimates of divergence time and paraphyly with neighboring species. *Mol Biol Evol.* 10:282–295.
Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T. 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb Symp Quant Biol.* 52:863–867.
Morales JC, Melnick DJ. 1998. Phylogenetic relationships of the macaques (Cercopithecidae: *Macaca*), as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. *J Hum Evol.* 34:1–23.
Naglaki T. 1983. Evolution of a finite population under gene conversion. *PNAS* 80:6278–6281.
Osada N, Nakagome S, Mano S, Kameoka Y, Takahashi I, Terao K. 2013. Finding the factors of reduced genetic diversity on X chromosomes

of *Macaca fascicularis*: male-driven evolution, demography, and natural selection. *Genetics* 195:1027–1035.
Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GAB. 2012. Evidence for widespread GC-biased gene conversion in eukaryotes. *Genome Biol Evol.* 4:787–794.
Pool JE, Nielsen R. 2007. Population size changes reshape genomic patterns of diversity. *Evolution* 61:3001–3006.
Pool JE, Nielsen R. 2008. The impacty of founder events on chromosomal variability in multiply mating species. *Mol Biol Evol.* 25: 1728–1736.
Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* 499: 471–475.
Rhesus Macaque Genome Sequencing and Analysis Consortium, Gibbs Richard A, Rogers Jeffrey, Katze Michael G, Bumgarner Roger, Weinstock George M, Mardis Elaine R, Remington Karin A, Strausberg Robert L, Craig Venter J, et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316:222234.
Riley EP. 2010. The endemic seven: four decades of research on the Sulawesi macaques. *Evol Anthropol.* 19:22–36.
Romiguier J, Ranwez V, Douzery EJP, Galtier N. 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome size. *Genome Res.* 20:1001–1009.
Schillaci MA, Froehlich JW, Supriatna J. 2007. Growth and sexual dimorphism in a population of hybrid macaques. *J Zool.* 271:328–343.
Spencer CCA, Deloukas P, Hunt S, Mullikin JC, Myers S, Silverman B, Donnelly P, Bentley D, McVean G. 2006. The influence of recombination on human genetic diversity. *PLoS Genet.* 2:e148.
Storz JF, Ramakrishnan U, Alberts SC. 2001. Determinants of effective population size for loci with different modes of inheritance. *J Hered.* 92:497–502.
Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–460.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
Tosi AJ, Morales JC, Melnick DJ. 2002. *Macaca fascicularis* Y-chromosome and mitochondrial markers indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the Isthmus of Kra. *Int J Primatol.* 23:161–178.
Tosi AJ, Morales JC, Melnick DJ. 2003. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution* 57:1419–1435.
Van Noordwijk MA, Van Schaik CP. 1985. Male migration and rank acquisition in wild ling-tailed macaques (*Macaca fascicularis*). *Anim Behav.* 33:849–861.
Watanabe K, Lapasere K, Tantu R. 1991. External characteristics and associated developmental changes in two species of Sulawesi macaques, *Macaca tonkeana* and *M. hecki*, with special reference to hybrids and the borderland between the species. *Primates* 32:61–76.
Watanabe K, Matsumura S, Watanabe T, Hamada Y. 1991. Distribution and possible intergradation between *Macaca tonkeana* and *M. ochreata* at the borderland of the species in Sulawesi. *Primates* 32: 385–389.
Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.
Wilson Sayres MA, Lohmueller KE, Nielsen R. 2014. Natural selection reduced diversity on human Y chromosomes. *PLoS Genet.* 10: e1004064.
Wright S. 1943. Isolation by distance. *Genetics* 28:114–138.
Zeng K. 2012. The application of population genetics in the study of codon usage bias. In: Cannarozzi GM, Schneider A, editors. Codon evolution: mechanisms and models. Oxford: Oxford University Press.
Zeng K. 2013. A coalescent model of background selection with recombination, demography and variation in selection coefficients. *Heredity* 110:363–371.
Zeng K, Charlesworth B. 2009. Estimating selection intensity on synonymous codon usage in a non-equilibrium population. *Genetics* 183: 651–662.
Zeng K, Charlesworth B. 2010a. The effects of demography and linkage on the estimation of selection and mutation parameters. *Genetics* 186:1411–1424.
Zeng K, Charlesworth B. 2010b. Studying patterns of recent evolution at synonymous sites and intronic sites in *Drosophila melanogaster*. *J Mol Evol.* 70:116–128.
Zeng K, Charlesworth B. 2011. The joint effects of background selection and genetic recombination on local gene genealogies. *Genetics* 189: 251–266.
Zhao L, Yue X, Waxman D. 2013. Complete numerical solution of the diffusion equation of random genetic drift. *Genetics* 194:973–985.