J Mol Evol (1997) 44:463–465

JOURNAL OF MOLECULAR EVOLUTION

Sex Differences in Mutation Rate in Higher Primates Estimated from AMG Intron Sequences

Wei Huang, Benny H.-J. Chang, Xun Gu, David Hewett-Emmett, W.-H. Li

Human Genetics Center, SPH, University of Texas, P.O. Box 20334, Houston, TX 77225, USA

Received: 19 August 1996 / Accepted: 22 November 1996

Abstract. To study sex differences in mutation rate in primates, we sequenced the third introns of the AMGX and AMGY genes from humans, orangutans, and squirrel monkeys and estimated that the male-to-female ratio of mutation rate is $\alpha = 5.14$ with the 95% confidence interval (2.42, 16.6). Combining this data set and the data sets from ZFX/ZFY and SMCX/SMCY introns, we obtained an estimate of $\alpha = 5.06$ with the 95% confidence interval reduced to (3.24, 8.79). The α value is significantly higher in higher primates than in rodents.

Key words: Sex difference in mutation rate — Male-driven evolution — Amelogenin gene — Intron sequences — Higher primates

Introduction

If mutations in an organism arise mainly from DNA replication errors during germ-cell divisions, the mutation rate should be proportional to the number of germ-cell divisions per generation (see Li et al. 1996). Since the number of germ-cell divisions per generation in human males is much larger than that in females, the mutation rate in males should be higher than that in females (Haldane 1947). In other words, the male-to-female ratio of mutation rate (α) should be larger than 1 and approximately equal to the male-to-female ratio of the number of germ-cell divisions (c), i.e., $\alpha \approx c$ (Miyata et al. 1987). Under this assumption, Miyata et al. (1987) provided a simple method for estimating α from DNA sequences that satisfy two criteria: (1) They are subject to only weak or no functional constraints so that most mutations are selectively neutral, and (2) they have homologous copies in the nonrecombining regions of the X and Y chromosomes so that they are suitable for obtaining an accurate estimate of the ratio of mutation rates in X- and Y-linked sequences (Y/X). Then α can be estimated from the equation $Y/X = 3\alpha/(2 + \alpha)$.

From the last introns of ZFX and ZFY, Shimmin et al. (1993) estimated that the mean value of α in higher primates is about 6 but the 95% confidence interval is broad (2, 84). By adding two new species, Chang et al. (1996) obtained a narrower 95% confidence interval (3, 33). Ketterling et al. (1993) obtained an estimate of $\alpha = 3.5$ by using genomic sequencing to trace the origin of mutation in patients with hemophilia B. Recently, Chang et al. (1996) reported $\alpha \approx 4.2$ from SMCX and SMCY introns. To obtain a more reliable estimate of α in higher primates, more sequence data are needed.

Amelogenin is the major protein in the extracellular matrix of the dental enamel. In humans, the amelogenin genes on the X and Y chromosomes have been cloned, and they are named AMGX and AMGY (Lau et al. 1989; Nakahori et al. 1991). AMGX is located in the p22.1–p22.3 region on the X chromosome, and AMGY is located near the centromere of the Y chromosome (Salido et al. 1992). As a result of comparing the similarity between human AMGX and AMGY and their chromosomal locations, Lau et al. (1989) proposed that, about 45 Myr ago, an inversion event in the Y chromosome including the region of AMGY occurred. As a result, recombination between AMGX and AMGY was no longer

Correspondence to: W.-H. Li; e-mail: li@hgc.sph.uth.tmc.edu.

possible and divergence began. Although AMGX has a slightly higher GC content than AMGY (Eyre-Walker 1993), this difference may not have a strong effect on the mutation rate. Therefore, the AMGX and AMGY genes are suitable for studying the sex difference in mutation rate in higher primates. In this paper, we report the results obtained from the third introns of the AMGX and AMGY genes in humans and other higher primates.

Materials and Methods

DNA Sample Sources. Male and female human genomic DNA samples were each extracted from peripheral blood. The male and female genomic DNA of orangutans (*Pongo pygmaeus*), baboons (*Papio cynocephalus*), and squirrel monkeys (*Saimiri boliviensis*) were extracted from frozen liver tissues. The extraction procedure of Ellsworth et al. (1993) was followed. The frozen liver tissue of orangutan was purchased from the Yerkes Primate Center and that of baboon was a gift from the Southwest Foundation for Biomedical Research, San Antonio, Texas.

PCR Amplification, Cloning, and Sequencing. According to the human AMGX and AMGY cDNA sequences and the third intron position (Salido et al. 1992), one common 5′ PCR primer was designed (AMG13: 5′-TCATCCTGGGCACCCTGGTT-3′, positions 63–82 in the coding region of human AMGX and AMGY). Two 3′ primers were also designed: One was X-specific (AMG10: 5′-TACCACTTCAAAGGGGTAAGCA-3′, positions 125–103 in the coding region of human AMGX), and the other was Y-specific (AMG14: 5′-CTACATACTGGTGGTCTTATCAT-3′, the last 15 nucleotides were from positions 147–133 in the coding region of human; while the first eight nucleotides were located at the 5′ end of intron 4). The third intron of each of the four primate species (human, orangutan, baboon, squirrel monkey) was amplified by PCR using 400 ng genomic DNA as template; the reaction mixture contained 50 mM KCl, 10 mM Tris (pH 9.0), 0.1% Triton X-100, 0.2 mM dNTP, 1.5 mM $MgCl_2$, 40 pmol of each primer, and 1 unit of Taq DNA polymerase (Promega), with final volume of 40 μl. Each intron was amplified in an Omnigene thermal cycler (Hybaid): denaturation, 94°C for 45 s; annealing, 59°C for 1 min; and extension, 72°C for 2 min +2 s of autoextension for a total of 30 cycles, and a single final extension of 10 min at 72°C. However, the third intron of AMGY of baboon could not be amplified. According to the Southern analysis of Nakahori et al. (1991), no AMGY-specific band was found in baboon genomic DNA when hybridized with a human AMGY-specific probe. So, this AMGY gene might have been lost in the baboon lineage during evolution.

PCR products were separated by electrophoresis in 1% SEA PLAQUE GTG (FMC) agarose gels. The expected DNA fragments (about 1.5 kb for AMGX in four species and for AMGY in squirrel monkey; about 1.3 kb for AMGY in human and orangutan) were cut out and purified with the Wizard PCR preps DNA purification system (Promega).

The third introns of the amelogenin gene in human, orangutan, baboon, and squirrel monkey were mostly sequenced by using the *fmol* Cycle Sequencing System (Promega) with purified PCR products. Because the PCR primers needed for the amplification of intron 3 were not far enough from the 5′ end of intron 3, it was difficult to obtain the sequences in this part by direct sequencing. So, the 5′ end of intron 3 was also sequenced as cloned plasmid DNA using the Sequenase Sequencing System (Amersham). For each intron, three clones were derived from three independent PCRs and they were sequenced from both directions.

Table 1. Mean and standard error of the number of nucleotide substitutions per 100 sites between intron 3 sequences of the primate AMGX and AMGY genes

	AMGX	AMGY
Human vs orangutan	1.61 ± 0.38	4.93 ± 0.68
Human vs squirrel monkey	7.92 ± 0.87	15.5 ± 1.27
Orangutan vs squirrel monkey	7.62 ± 0.85	16.6 ± 1.33

Results and Discussion

Nucleotide Sequences

The nucleotide sequences of the third introns of AMGX and AMGY were determined. The sequence lengths of intron 3 of AMGX are 1,409 bp, 1,409 bp, and 1,387 bp, and those of AMGY are 1,227 bp, 1,199 bp, and 1,395 bp in human, orangutan, and squirrel monkey, respectively. The major reason for the size difference is that one large deletion (179 bp) occurred in intron 3 of human and orangutan AMGY. In baboon, the third intron of AMGX is 1,405 bp long. The sequences were aligned by the Pileup program in the GCG package (Devereux et al. 1984) and then the alignment was adjusted manually. By excluding all gaps, we obtained 1,133-bp nucleotide sites. The evolutionary distances were estimated by Tajima and Nei's (1984) method (Table 1). Sequences were deposited in Genbank under accession numbers U88979–U88983.

The Male-to-Female Ratio (α) of Mutation Rate

Since the third intron of AMGY in baboon cannot be amplified by PCR, α was estimated from three species: human, orangutan, and squirrel monkey. Table 1 shows the pairwise evolutionary distances and standard errors. For each pair of species, the distance of the AMGY intron is always significantly larger than that of the AMGX intron. For example, in the case of human vs orangutan , the distance difference between AMGY and AMGX is $D = 4.93 - 1.61 = 3.32$ and the sampling variance of D is $V(D) = 0.68^2 + 0.38^2 = 0.61$. Thus, by computing $Z = D/\sqrt{V(D)} = 4.26$, which approximately follows a normal distribution, the null hypothesis that AMGX and AMGY have the same evolutionary rate can be rejected at the 1% level.

The ratio Y/X can be estimated from the total branch lengths of AMGX (denoted by X) and AMGY (denoted by Y) introns (Shimmin et al. 1993). In the case of three species, X and Y can be easily estimated from pairwise distances (Table 1), i.e., $X = (1.61 + 7.92 + 7.62)/2 = 8.58$ and $Y = (4.93 + 15.5 + 16.6)/2 = 18.52$. The standard errors of X and Y are $s_X = 0.85$ and $s_Y = 1.33$, respectively, by the method of Li (1989). Therefore, we obtain $R = Y/X = 2.16 \pm 0.16$; the approximate standard

Table 2. Estimates of α in higher primates with the 95% confidence interval

Gene pair	*R*	α	95% confidence
AMGX/AMGY	2.16 ± 0.26	5.14	2.42–16.6
ZFX/ZFY	2.27 ± 0.29	6.26	2.63–32.4
SMCX/SMCY	2.03 ± 0.24	4.20	2.20–10.0
Combined data	2.15 ± 0.15	5.06	3.24–8.79

error ($s = 0.16$) is estimated by the method of Shimmin et al. (1993). Then, α is estimated by $\alpha = 2R/(3 - R) = 5.14$; for the formula, see Miyata et al. (1987). The 95% confidence interval can be approximately estimated as follows. By the normal distribution approximation, the lower bound of 95% confidence is $R^- = R - 1.96s$, and the upper bound in $R^+ = R + 1.96s \approx 2.67$; the 95% confidence should be taken with caution because *R* is not normally distributed. Thus, the 95% confidence interval of α is between $\alpha^- = 2R^-/(3 - R^-) \approx 2.42$ and $\alpha^+ = 2R^+/(3 - R^+) \approx 16.6$.

Our estimate of α from the AMGX/AMGY introns is similar to those from the ZFX/ZFY introns and the SMCX/SMCY introns (Shimmin et al. 1993; Chang et al. 1996). Indeed, as shown in Table 2, these estimates do not differ significantly. Thus, α can be more accurately estimated by combining the data sets. Let, R_i and $V_i (i = 1, 2, 3)$ be the mean and variance from the *i*th data set. The average *R* over the three data sets is weighted by the reciprocal of the sampling variance, i.e.,

$$w_i = \frac{1/V_i}{1/V_1 + 1/V_2 + 1/V_3}$$

$(i = 1, 2, 3)$. Therefore, the average is $\tilde{R} = w_1R_1 + w_2R_2 + w_3R_3 \approx 2.15$, and the standard error is

$$s_R = \sqrt{w_1^2V_1 + w_2^2V_2 + w_3^2V_3} \approx 0.15$$

Thus, the average ration $\tilde{\alpha}$ can be estimated by $\tilde{\alpha} = 2\tilde{R}/(3 - \tilde{R}) \approx 5.06$. By using the same approach as above, the 95% confidence interval of $\tilde{\alpha}$ is estimated to be from 3.24 to 8.79 (Table 2).

This 95% confidence interval does not overlap with the 95% confidence interval of α estimated from rodents (1.0, 3.2) (Chang et al. 1994). Therefore, the α value is significantly higher in higher primates than in rodents. This result indicates that the longer generation time in humans has a significant effect on the rate of mutation.

Our estimate of $\alpha = 5.06$ is similar to the sex ratio of the number of germ-cell divisions (*c*), which is between 3 and 6, if the average age of the reproductive males is 15 to 20 (Change et al. 1994; Li et al. 1996). This result implies that point mutation in nuclear DNA is largely DNA-replication dependent (see Li et al. 1996).

Acknowledgments. This study was supported by NIH grants.

References

Chang BH-J, Shimmin LC, Shyue S-K, Hewett-Emmett D, Li W-H (1994) Weak male-driven molecular evolution in rodents. Proc Natl Acad Sci USA 91:827–831

Change BH-J, Hewett-Emmett, Li W-H (1996) Male-to-female ratios of mutation rate in higher primates estimated from intron sequences. Zool Studies 35:36–48

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis program for the VAX. Nucleic Acids Res 12: 387–395

Ellsworth DL, Rittenhouse KD, Honeycutt RL (1993) Artificial variation in randomly amplified polymorphic DNA binding patterns. Biotechniques 14:214–217

Eyre-Walker A (1993) Recombination and mammalian genome evolution. Proc R Soc Lond [Biol] 252:237–243

Haldane JBS (1947) The mutation rate of the gene for hemophilia and its segregation ratios in males and females. Ann Eugen 13:262–271

Ketterling RP, Vielhaber E, Bottema CDK, Schaid DJ, Cohen MP, Sexauer CL, Sommer SS (1993) Germ-line origins of mutation in families with hemophilia B: the sex ratio varies with the type of mutation. Ann J Hum Genet 52:152–166

Lau EC, Mohandas TK, Shapiro LJ, Slvkin HC, Snead ML (1989) Human and mouse amelogenin gene loci are on the sex chromosomes. Genomics 4:162–168

Li W-H (1989) A statistical test of phylogenies estimated from sequence data. Mol Biol Evol 6:424–435

Li W-H, Ellsworth DL, Krushkal J, Chang HJ, Hewett-Emmett D (1996) Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. Mol Phylogenet Evol 1:211–214

Miyata T, Hayashida H, Kuma K, Mitsuyasa K, Yasunaga T (1987) Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harb Symp Quant Biol 52:863–867

Nakahori Y, Takenaka O, Nakagome Y (1991) A human X-Y homologous region encodes amelogenin. Genomics 9:264–269

Salido EC, Yen PH, Koprivnikar K (1992) The human enamel protein gene amelogenin is expressed from both the X and the Y chromosome. Am J Genet 50:303–316

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

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269–285