J Mol Evol (1995) 40:70–77

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1995

Estimating the Intensity of Male-Driven Evolution in Rodents by Using X-Linked and Y-Linked *Ube 1* Genes and Pseudogenes

Benny Hung-Junn Chang, Wen-Hsiung Li

Center for Demographic and Population Genetics, University of Texas, P.O. Box 20334, Houston, TX 77225, USA

Received: 23 April 1994 / Accepted: 3 August 1994

Abstract. Using sequence data from the last introns of ZFX and ZFY genes, we previously estimated the male-to-female ratio (α) of mutation rate to be close to 6 in higher primates and 1.8 in rodents. As the mutation rate may vary among different regions of the mammalian genome, it is interesting to see whether sequence data from other regions will give similar estimates. In this study, we have determined the partial genomic sequences of the ubiquitin-activating enzyme E1 genes (*Ube 1x* and *Ube 1y* for the X-linked and Y-linked homologues, respectively) of mice and rats and two mouse *Ube 1y* pseudogenes. From the intron sequences of the *Ube 1* genes, we calculated the divergence of the Y-linked genes ($Y = 0.161$) and that of the X-linked genes ($X = 0.107$) between mouse and rat, and found the Y/X ratio to be 1.50. This ratio led to an estimate of $\alpha = 2.0$ with a 95% confidence interval of (1.0, 3.9). Similar estimates of α were obtained if mouse *Ube 1y* pseudogenes were used instead of the mouse *Ube 1y* functional gene. These estimates are consistent with our previous estimate for rodents and suggest that the sex ratio of mutation rate in rodents is approximately only one-third of that in higher primates. Our estimate of the divergence time between *Ube 1x* and *Ube 1y* supports the view that the two genes separated before the eutherian radiation.

Key words: Substitution rate — Mutation rate — Sex ratio — Mechanism of mutation

Correspondence to: W.-H. Li

Introduction

The larger number of germ-cell divisions in human males than in females has led to the suggestion of a higher mutation rate in males than in females (Haldane 1947). However, direct estimation of the sex ratio of mutation rate requires determination of the parental origin of mutations, which is difficult to do with classical approaches (Rosendaal et al. 1990). Direct genomic sequencing (Ketterling et al. 1993) provides a better means with which to trace the origin of mutation but is laborious, and the method is not readily applicable to nonhuman organisms. Miyata et al. (1987) proposed a method to estimate the male-to-female ratio of mutation rate (α) from the ratio of substitution rates in sequences on different chromosomes (i.e., autosomes, X chromosome, and Y chromosome). We and other authors obtained an estimate of $\alpha \approx 6$ in higher primates and $\alpha \approx 2$ in rodents by applying Miyata et al.'s method to the intron sequences of X-linked and Y-linked zinc finger protein genes (Pamilo and Bianchi 1993; Shimmin et al. 1993; Chang et al. 1994).

However, the mutation rate may vary among regions of the genome (Wolfe et al. 1989), and so the ratio estimated from one single pair of sequences may not be reliable. It is, therefore, desirable to use sequences from other regions of the genome to estimate the ratio. Toward this end, we have obtained partial genomic sequences of the ubiquitin-activating enzyme E1 (*Ube 1*) genes in mice and rats. *Ube 1* genes have been mapped to the X and Y chromosomes of many mammalian species (Mitchell et al. 1991, 1992), although no Y-linked homologue is found in higher primates (Mitchell et al. 1991). There is an autosomal *Ube 1* gene in humans,

```
1 Mmu Ube1x    c ctt ttc aag cag cca gca gaa aat gtt aat cag tac ctc ac GTGAGTAACTTCCACATCT
2 Rno Ube1x    c ctt ttc aag cag cca gca gaa aat gtt aat cag tac ctc ac GTGAGTAACTTCCACACCC
3 Mmu Ube1y-p1 g ctc ttc aag cag tca gct gaa aat att aac caa tac ctc ac GTAAGTAACAGACAAACC-
4 Mmu Ube1y-p2 g ctc ttc aaa cag tca gct gaa aat att aac caa tac ctc ac GTAAGTAACAGACAAACC-
5 Mmu Ube1y    g ctc ttc aag cag tca gct gaa aat gtt aac caa tac ctc ac GTAAGTAACAGACAAACC-
6 Rno Ube1y    a ctg ttc aag cag tca gcc gaa aat gtt aac caa tac ctc at GTAAGTAACAGACAAACCA

1 TAC-----CCTTCTGCTTGCCACCAAAGACATTC------------AGCTGTCA-TCCTCAATGCTAGGTGCACACTTTTCAAAGGAC
2 CAC-----CCTCTTGCCTGCCACCAAAGACATTCTCTGCTTTCTTCAGCTGTTACCCCTCAATGCTAGGTGTACACTTTTCAAAGGAT
3 AACAGTGGACAGAAGGCAGTCAACAAAGGCATCCT----TGTCTGCTCTCCTCA-GGCAC-ATCCACGGTGGATGCTTT-----GGAT
4 AACAGTGGACAGAAGGCAGTCAACAAAGACGTCCT----TGTCTGCTCTCCTCA-GGCAC-ATCCTCGGTGGATGCTTT-----GGAT
5 AACAGCCAACAGAAACCAATCAACAAAGGCGTCCT----TCTCTGCTCGCTTCAATGCAC-ATCCTCGGTAGATGCTTT-----GGAT
6 AACAGCAGACAGACGCCAACCAACAAAGGCATCCT----TGTCTGCTCTCCTCA-GGCAC-ATCCTTGGTGGATGAGTT-----GGGT

1 ACTAAATGTTTAGTT----CTCATGCTGTTG------------GGGGATCAAGATGAGGAG---------------CTTTCTGTGGAG
2 GCTAAATGTTCAGTG----TTCAAGCTGTTA------------CGGAATGAAGATGAGGAG---------------CTTTCTGTGGAG
3 AACTAGTGGTTGATTGAGCCTTATGCTGGTG---------------GAGGATGATGGAAAATGGTCAGGATTGACATTTTCCCTGGAG
4 AATTAATGGTTGATTGAGCCTTATGCTGGTG---------------GAGGTTGATAGGAAATGGTCAGGATTGACATTTTTCCTGGAG
5 AATTAATGGTTGATTGAGCCTCATGCTGGTGGTAGATGATGATGATGATGATGATGGGAAGTGGTCAGCATTGACATTTTCCCTGGAG
6 A----------------------------------------------------ATGGGAAATGGTCCGGGTTGACTTTTTCCATAGAG

1 AAAGTGGGATTGCCCTGGAGCCTAGTTCAGTGGCCACACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATAT
2 AAAGTGGGATTGCCCTGGAGCCTGGTTCAGTGGCCAC---------------------------------------------------
3 AAAGAAGGCTTGCACAGAAACTTTACTTCTAGGCCTC---------------------------------------------------
4 AAAGAAGGCTTGCCCAGAAACCTTACTTCTAGGCCTC---------------------------------------------------
5 AAAGAAGGCTTGCCCAGAAACCTTACTTCTAGGCCTC---------------------------------------------------
6 AAAGAAGGCTTGTCTAGAAACTTTACTTCTAGGCCTC---------------------------------------------------

1 ATATATATATATATATATATATATATATATATATATATATATATATATATATACATATACATATATATATACACTCTTGCCTCTTGGC
2 ---------------------------------------------------------------------TATATACTTGCCCCTTGGC
3 -----------------------------------------------------------------------TGAAGTAATCCA--GGT
4 -----------------------------------------------------------------------TGAAGTAATCCA--GGT
5 -----------------------------------------------------------------------TGAAGTAATCCA--GGT
6 -----------------------------------------------------------------------TGAAGTAATCCA--GGT

1 CCCCTCTACTCCCTAACAG   a gac tcc aaa ttt gtg gag cgg aca ttg cgg ctg gct ggt acc cag cca
2 CCCCTCTGCTCCCTAACAG   a gac tcc aaa ttt gtg gag cgg aca ttg cgg ctg gcg ggt acc cag cca
3 CTTTG-TATTG---GATAG   g gac ccc aag ttc atg gag cag aca ctg cag cta gct ggc acc cag cct
4 CTTTG-TATTT---GATAG   a gat ccc acg ttc atg gaa cag aca cag cag gta gct ggc acc cag cct
5 CTTTG-TATTG---GATAG   g gac ccc aag ttc atg gag cgg aca ctg cag cta gcc ggc acc cag cct
6 CTTTG-CATTG---GATAG   g gac ccc aag ttc atg gag cgg aca ctg cag cta gct ggc acc cag cct

1 ttg gag gtg ctg gag gct gtg cag cgc agc ctg gtg ttg cag cga cca cag act tgg gga --- gac
2 ttg gag gtg ctg gag gct gtg cag cgc agc ctg gtg ttg cag cga cca cag act tgg ggc --- gac
3 ttg taa gta ctg gag gcc ata cac tgc agc ctg gtc ctg cag agg cca cag act tgg gcc gac gac
4 ttg gaa ata ctg gag gcc ata cac tgt agt ctg gtc ctg cag agg cca cag act agg gcc --- gac
5 ttg gaa gta ctg gag gcc ata cac tgc agc ctg gtc ctg cag agg cca cag act tgg gcc --- gac
6 ttg gaa gta ctg gag gct ata cag tgc agc ctg gtc ctg cag agg cca cag act tgg gcc --- gac
          uy018                                       uy014
1 tgt gtg acc tgg gcc tgc cac cac tgg cac acc cag tac tgt aac aac atc cgg caa ctg ctg cac
2 tgt gtg acc tgg gcc tgc cac cac tgg cac acc cag tac tgt aac aac ata cgg cag ctg ctg cac
3 tgt gtg act tgg gcc tac cag cac tga cac gcc gag tat tct cac aac atc cag cag ttg ttg cac
4 tgt gtg tct tgg gcc tac cag cac tgg cac acc cag tat tct cac aac atc cag ccg tcg ctg cac
5 tgt gtg act tgg gcc tac cag cac cgg cac acc cag tat tct cac aac atc cag cag ttg ctg cac
6 tgt gtg act tgg gcc tac cag cac tgg cac acc cag tat tct cac aac atc cag cag ttg ctg cac
                                          ux023
1 aac ttt cct cct gac cag GTAACATAC-----------TTGTCTGGCAACATGGGTTTCCTTGCAACCAG---TTTGCTGCA
2 aac ttt cct ccc gac cag GTAACATGT-----------CTGTTTGGCAGCATGAGTTTCCTTGCAACCAG---TTTGCTGCA
3 aac ttt cct cca gct cag GTATTACATACTTGGTGGATTTATTTGGCTGAGCACATTTTCAGGAAATTGGATGCTTACTGTC
4 agc ttc cct cca gct cag GTGTTACATACTTGGTGGATGTATTTGGCTGAGCACATTTCCAGAAAATTGGAGACTTACTGTC
5 aac ttc cct cca gca cag GTATTACATACTTTGTGGATGTATTTGGCTGAGCAGATTTCCAGGAAATTGGAATCTTACTGTC
6 aac ttc cct cca gac cag GTATTATATACTCAGTGGGTATATTTGGCCCAACACATTTCCAGGAAATGGGATGCTTACTGTC
          ux028
1 GACTCTT---AAGTAGCTTCTCCCAGTTAACTAGTCCAG-----------ATGTGA-GGCAAAAGATGTCAGAAAATTG-----ACCT
2 GACTCTT---AAGTAGCTTCTCCCAGTT-ACTAGTCTAGATATGGCTGGTATCTGA-GGCAGGAGATGTCAGAAAAATAAA---AACT
3 CTCTGTCTGTACATGGCTTCTTCCAG----GCAGTCTAGCTGC-ACAGGCACCTCAAGGCAAGGTGTAGGCCAGTACTTGGTGCAAGG
4 CTCTGTC---------CTTCTGCCAG----GCAGCCTACCTGC-ACAGGTACCAGAATGCAAGGTGTAGGCTAGTACTTGGTGCAAGC
5 CTCTGTCTAGACATGGCTTCGTCCAG----GCAGTCTAGCTGC-ACAGGCACCACAAGGCAAGCTGTAGGCCAGTACTTGGTGCAAGC
6 CTCTGTCTGGACCTGGCTTCTTCCAG----GTAGCCTAAGTGCCACAAGCACCCCACGGCAAGGATTAGGCTAGTTCTCCGTGTAAGC
```

Fig. 1. Aligned partial genomic sequences of the *Ube 1* genes and pseudogenes in mouse (*Mmu: Mus musculus*) and rat (*Rno: Rattus norvegicus*). *Lowercase*, exon sequences; *uppercase*, intron sequences; *overlined sequences*, dinocleotide repeats; *underlined italic*, nucleotides different from the pY8/b sequence determined by Tucker et al. (1992); *underlined sequences*, sequencing primers for X-linked *Ube 1* genes (with prefix *ux-*) and Y-linked *Ube 1* genes (with prefix *uy-*). Primers uy018, uy019, uy015, ux023, and ux029 are for sequencing of sense strands, and the others for sequencing of antisense strands (complementary sequences are used in such case). The sequences reported in this paper have been deposited in the GenBank data base (accession No. U09051-U09056).

```
1 TTTTGTTTTGTCCTTCATTTCTTGAAGCAAGAGAA--------CTTGCTGTTCCTTTT-CACTGTGCTGTAGACTGTGTCTAACCTGT
2 TTTTGTTTCGTTCCTCATTCCTTGAAGCAAGAGAACCGGTTCTCTTGCTGTTCCTGGT-CACTGTGCTGTAGGCTATGTCTGACCTGA
3 ATTTGTGTTTTCCTAC--TTCAGTCAGCCAGGCAAGGGGCTTAGTTTCCTTCCCAGGTGCACT----TGTGGGCTGAGTCACAGCTGA
4 TTTTGTGTTTTCTTCC--TCCTGTCAGCCAGGGAAGGGGTTTAGTTTCCTTCCCAGCTGCACT----TGTGGGCTGAGTCATAGCTGA
5 ATTTGTGTTTTCCTTC--TTCTGTCAGCCAGGCAAGTTTCTTAGTTTCCTTCCCAGCTGCACC----TGTGGGCTGAGTCACAGCTGA
6 ATTTGCGTTTTCCTCC--TCCTGCCCACCAAGCAAGGGGCTTAGTTTCCTTTCCAGCTGCACT----TGTGGGCTGAGTCATATCTGA
                                              uy020      uy019
1 GCTGTAGACTGTGTCTAACCTGATTTTTTCAGGGTTGTAGTGAACCTTGAGGCCAAGATCCTCGACTTGAGAGCAAGGGAGTCAGCAT
2 ----------------------TTTTTATCAGGGTTGTAGTGAACCTCGAGGCAGAGATCCTTGACATGAGAGCAAAGGAGTCAGCAT
3 G--------------------TTCTCTTTGAGGGCTGTAGTTGTAAGTGAG----AAATTTTACAGTTAAGCATGCAGTGATCCGTGG
4 G--------------------TTCTCTTTGAGGGCTGTAGTT----GTGAG----AAATTTTATAGTTAAGCATAGAGTGATCTGTGT
5 G--------------------TTCCCTTTGAGGGCTGTAGTTGCAAGTGAG----AAATTTTATAGTTAAGCATACAGTGATCCAAGT
6 G--------------------TTCCCTCGGAGGGTTGCAGTTTTAAGTGAG----AAATATGATAGTTAAATATACAGCGGTCTATGG

1 TTATTCAAATCCCA----------GTTTTCTCTAAACTTGGG--CAAGTCATTTCTCTGTACCT-TTCTTTAGCTTGAAGGAGGAA--
2 TTATTCAAATCCCA----------GTTTTCTCTAAACTTGG---CAAGTCATTTCTCTGTACCT-TTCTTTAGCTTGAACGAGGAA--
3 GT-TTCAAGTCCCA----------GCTGTGTTTGACACTGGA-TTGTGACGTGCCTA---ATGTCTTCGTAAGCCATATCTGTGAAGT
4 G--TTCAAGTCCCA----------GCTGTGTTTGACACTGGG-TCGTGTCATGCCTA---ATGTCTTGCTAAGCCATATCTGGGAAGT
5 GG-TTCAAGTCCCA----------GCTGTGTTTGACACTGGG-TCCTGTCATGCCTA---ATGTCTTGGTAAGCCATATTTGTGAAGT
6 GT-TCATGGGTCTATGGGTCTGTGGCTGTGTTTGACAGTGGGGTCATGTCAAGCCTAATGATGTCTTGGCAAGCCATATCTGTGAAGT
                 ux029
1 CAGTTAGATATGTTTCACAGGGC----CTTCTGTGA----GGATTCATTGGTATTTATTAATACT---------------AAGTACA
2 CAGTTAG--ATGTTCCACAGGGC----CTTCTGTGAATCCACATTCGTTGGTATTTATTAATACT---------------AAGTACA
3 AGGT--------GGCTTGTGGGT----GTTTAGAAA----GGACTAACTTGCATTTCTTAGTGTT----------------GCCTCA
4 AGGTATGACAATGGC-TGTGGGC----ATTTAGAAA----GGACTAACTTTCATTTCTTAATTTT----------------GCCTCA
5 AGGTATGATAATGGCTTATGGGC----ATTTAGAAA----GGACTAACTTGCATTTCTTATTATT----------------GACTCA
6 AGGTATGATAATGGCTTGTGGGCCATTATTTAGAAA----GGAGTAATTTGCATTTCTTAGTATTGCATTGGTTTAGCCTAAGCCTCA
                                                                                 ux022
1 TAC---CTGGCGATCCAGTTCTAGACTAGCAAGTAAAGCAAGTGT-----------TAGTTTT--CTCAGCTAGGACCTTGCAGTGTA
2 TAC---CTGGCAATCTAGTAC-AGACCAGTAAGTAAAGCAAATGTCTAAGAATTGATGGTTTT--TTCAGCCAGGACCTTGCAGTGTA
3 TATAAATTGGTAAGTT-GAACT-----------------------------TGCCGTGTGT--TTGACTC---ATCGTG-AATTTC
4 TATGAATTGGTAAGTT-GAACT-----------------------------TGCCATGTGT--TTGGCTC---ATCATG-AATTTC
5 TATACATTGGTAAGTT-GAACT-----------------------------TGCCGTGTGT--TTGGCTC---ATTGTG-AATTTC
6 TACAAATTGGTAAGTT-GAACT-----------------------------TGCTGTGTGTTTTTGGCTC---ATCATG-GGT---
        uy016                                      uy015
1 GGTCTTAGATGACTGATTCCAGCCCTCTTTGGGACCCAGCCAAGGGTAGTGAGTTGGGGTGCATCTAGAATTGTCAGAGATCTTACTC
2 GGTCTTAGATGATTGACTCCAGCCCTCTTCAGGACCCAGCCAAGGGTAGTGAGTTGGGGTGCATCTAGGATTGTC--AGATCTTGCTC
3 CCA---AGAATGGTGAACCTAGGCTTCTT---------GGCCTGAGGACTGCCTGAGAGTGCTTCTCCCTGGGTC----------CTC
4 CCA---AGAATGGTGAACCTAGGCTTCTT---------GGCATGAGGGTTTCCTGAGAGTGCTTCTCCCTCGGTC----------CTC
5 CCA---AGAATGGTGAACCTAGGCTTCTT--------------------GGCCTGAGAGTGCTTCTCCCTCGGTG----------CTG
6 CCA---AGAATGGTGAACTTAGGCTTCTT---------AGTCTGAGACTTGCCTGAGAGTGCTTCTCCTTCAGTC----------CTC

1 CTGAAACATTCTTATTTTGTTTTAG  ctc acc agc tca ggg gcc cct ttc tgg tct gga ccc aaa cgc tgt
2 CTGAAAC--TCTTTCTTTGTTTTAG  ctc acc agc tca ggg gcc cct ttc tgg tct gga ccc aaa cga tgt
3 TTCATGCGT--TTGCATTCTCCTAG  ctt aca agc tct gga tca ctt ttc tgg tca gga cca aaa cgc tgt
4 TTCATATCT--TTGCATTCTCCTGG  ctt act agc tct gaa gca ctg ttt tga tca gga cca aaa cac tgt
5 TTCATATCT-------TTCTCTTAG  ctt acg agc tct gga gca ctt ttt tgg tca gga cca aaa cgc tgt
6 TTCATACCT--TTGCATTCTTCTAG  ctt aca agc tct ggg gta ctt ttt tgg tca gga cct aaa cgc tgt

1 cca cac cca ctt act ttt gat gtt aac aat GTAAGTCTTCTTACTGGGTTTTCCTGGGGTCAAATACAGAGAAGATGG
2 cca cac cca ctt act ttt gat gtt aac aat GTAAGTCTTCTTA-TTGGATTTCCTGGGGTC--ATATAGAAAGGATGG
3 cca cat ctg ctc acc ttt gac ata aac aat GTAAGTGTCCT---TGGGATCTCCCAGG-----------------TGG
4 cca cat cca ctc acc ttt gac ata aac aat GTAAGTGTCCT---TGGGATCTCCTAGG-----------------TGG
5 cca cat cct ctc acc ttt gac ata aac aat GTAAGTGTCCT---TGGGATCTCCCAGG-----------------TTG
6 cca cat cca ctc acc ttt gac aca aac aat GTAAGTGTCCT---TGGGATCTCCCAGG-----------------TGG

1 ACAGGTGGGAAGGAGGTGGTGGCTTTCCATCCAAGGGAAGACGTGTTAACCCTACCATGC-ATGCCCTG-CCTTCCTCTAG  aca
2 GTAGGTGGGAAGGAGGTGGTGGCTTCCTGTCTAAGGGCAGATGTGTTAACCCTACCAT-----GCCCTG-CCTTTGCCTAG  aca
3 GTATGTGGGAAG--TG-GGTCTCTCTACTCTTAACCTGTGGT-TGTTAAGCTGACTTTCCTTTGCCATGGCTTTCTTTCAG  ccc
4 TTGTGTGGGAAT--TA-GGTCTCTCTACTCTTAACCTGTGGT-TGTTAAGCTGACTTTCCTTTGCCATGCCATTCTTTCAG  tcc
5 GTGTGTGGGGAG--TG-GGTCTCTCTACTCTTAACCTGTGGT-TGTTAAGCTGACTTTCCTTTGCCATGCCATTCTTTCAG  ccc
6 TTGGGTGGGGGT--GGTGGAGTCTCTACTCTTACCCTGTGGT-TGTTAAGCTAACTTTTCTTTGCCATACCTTTCTTTCAG  ccc

1 ttg cat ctg gat   1403 bp
2 ttg cat ctg gat   1294 bp
3 ctg cat ctg gat   1220 bp
4 ctg cat ctg gat   1236 bp
5 ctg cat ctg gat   1234 bp
6 cta cat ctg gat   1231 bp
```

Fig. 1. Continued.

which is quite divergent from the X-linked homologue (Kok et al. 1993). It is not known whether this autosomal gene is homologous to the autosomal gene in marsupials (Mitchell et al. 1992) or whether it is also present in other mammalian species. There seems to be only one functional gene in each of the X and Y chromosomes of the mouse, but several nonprocessed pseudogenes are linked to the Y chromosome (Mitchell et al. 1991). In this study

we sequenced two Y-linked pseudogenes in the mouse as well as the X-linked and Y-linked functional genes in the mouse and the rat. These sequence data were used to estimate α and to infer the evolutionary history of the *Ube 1* genes. In addition, we used the pseudogene sequences to examine the assumption that introns are largely free of selective constraints.

Materials and Methods

DNA Sources. Freshly dissected liver tissues from inbred mice (strain BALB/cAncr) were gifts from Dr. B.W. McIntyre, M.D. Anderson Cancer Center at Houston, and those from rats (*Rattus norvegicus*, strain Lewis) were gifts from Dr. Rafik Ghobrial, the University of Texas Medical School at Houston. Genomic DNA was extracted from 0.1 g of tissue as described in Ellsworth et al. (1993).

PCR Amplification. Two consensus PCR primers (u011: TGGGCTCGGGATGAGTTTGAAGG at the 5′ end and u012: GCACCTAGGTTGGCAGCAGCCATGACATA at the 3′ end) were designed from conserved regions of the aligned sequences of mouse and human cDNA (Mitchell et al. 1991; Handley et al. 1991), a mouse pseudogene (Tucker et al. 1992), and a kangaroo gene (Mitchell et al. 1992). The amplified region encompasses three complete introns, two complete exons, and two partial exons. Amplification was done by using 1 μg of genomic DNA through 30 cycles of reaction under a condition of 95°C for 1 min, 58°C for 1 min, and 72°C for 2 min per cycle in a reaction mixture containing 10 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (w/v) geletin, 1 U of Taq DNA polymerase, and 50 pmol of each of the primers.

Cloning and Sequencing. The PCR products were ligated into pBluescript SK+ vector (Stratagene), and for each fragment, three clones each derived from an independent PCR and cloning reaction were sequenced. The procedures used were the same as in Chang et al. (1994).

Southern Blot Analysis. Ten micrograms of genomic DNA was digested overnight with 20 units of *Bam*HI (New England BioLab). The digested DNAs were then extracted with phenol and chloroform, precipitated with ethanol, resuspended in TE buffer, and then run through 0.8% agarose gel at 22 V for 20 h. The size-fractionated DNAs in the gel were depurinated, denatured, renatured, and capillarily transferred to Nytran nylon membrane (S&S). A 200-bp PCR-derived DNA fragment belonging to one of the Y-linked *Ube 1* genes was labeled using random priming kits (USB) to study the hybridization profile, which was compared with that of Mitchell et al. (1991).

Sequence Analysis. Sequence data editing and aligning were handled using the GCG package in the M.D. Anderson Cancer Center at Houston. The aligned sequences were then adjusted manually. The divergence of intron sequences was analyzed using Tajima and Nei's (1984) method. The exon sequences were analyzed using Li's (1993) method. The estimation of α, the sex ratio of mutation rate, was as described in Shimmin et al. (1993) but, briefly, sequence divergences between Y-linked genes (*Y*) and between X-linked genes (*X*) are used to estimate α according to Miyata et al.'s (1987) formula: $Y/X = 3\alpha/(\alpha + 2)$.

Results

Nucleotide Sequences

In the amplification of the X-linked *Ube 1* gene in mice, we encountered a technical problem—the fragment of our interest was difficult to amplify and clone. When we did succeed in obtaining one clone and using sequencing primers (ux022-ux029) to determine the entire sequence, we found that the fragment contained two consecutive regions of dinucleotide repeats: $(CT)_{23}$ and $(AT)_{37}$ (Fig. 1). It was unclear whether the dinucleotide repeats or other physical structures of the region hindered the amplification and/or cloning. In attempting to obtain more PCR products and clones, we avoided the difficulty by replacing the 3′ PCR primer u012 with a primer (ux028, Fig. 1) that is only 213 nucleotides downstream from the repeats; with this and the 5′ PCR primer (u011) we successfully amplified the subfragment containing the repeats. Other primers (ux023 and ux022, ux029 and u012, the 3′ PCR primer; Fig. 1) were then used to generate, by PCR, two overlapping subfragments of this region of the X-linked *Ube 1* gene. These altogether comprise three overlapping subfragments that cover the region of our interest.

In the amplification of the mouse Y-linked *Ube 1* gene, we also encountered a problem in that there are at least five copies of the *Ube 1* gene and pseudogenes on the mouse Y chromosome (Mitchell et al. 1991). Although only one PCR product was noticeable under preparative agarose gel electrophoresis, several different sequences were obtained after cloning and sequencing, and some of them were obviously recombinants during PCR. To avoid these artifacts, we generated, by the following strategy, the male-specific and locus-specific genomic DNA fragments as templates for PCR amplification. Mitchell et al. (1991) have shown, by Southern analysis, six male-specific *Bam*HI restricted genomic fragments and have assigned each fragment to one or two unique loci. Using a PCR-generated *Ube 1y* gene fragment as probe we were able to reproduce the same Southern hybridization profile as in Mitchell et al. (1991) (data not shown). We chose to use a pool of ~5 kb *Bam*HI restricted genomic DNA (where the only expressed mouse *Ube 1y* gene is located in the T5 locus as determined in Mitchell et al. 1991), a pool of ~4.5 kb *Bam*HI restricted genomic DNA (where a *Ube 1y* gene in the B3 locus is located), and a pool of ~9 kb *Bam*HI restricted genomic DNA (where a *Ube 1y* gene in the Sxr^b region is located) as templates in separate PCR amplifications. This strategy allowed us to amplify a single sequence each from the B3 locus, the T5 locus, and the Sxr^b region with no signs of PCR recombination after sequencing eight to 12 clones from each PCR cloning; the results were reproducible.

The amplified *Ube 1* partial genomic sequences each comprise of three introns, two complete exons, and two partial exons. The sizes of the amplified products were 1,403, 1,294, 1,236, 1,231, 1,221, and 1,236 base pairs (bp), respectively, from the X-linked mouse and rat homologues (*Ube 1x*), the mouse Y-linked functional gene (*Ube 1y*) and two Y-linked nonprocessed pseudogenes

Table 1. The mean (below diagonal) and standard error (above diagonal) of the number of nucleotide substitutions per site in the introns between rodent *Ube1* genes and pseudogenes[a]

	MmuX	RnoX	MmuY-p1	MmuY-p2	MmuY	RnoY
MmuX		0.013	0.047	0.047	0.047	0.052
RnoX	0.107		0.044	0.045	0.045	0.048
MmuY-p1	0.678	0.639		0.011	0.011	0.017
MmuY-p2	0.679	0.641	0.085		0.011	0.016
MmuY	0.674	0.645	0.079	0.082		0.017
RnoY	0.739	0.689	0.163	0.157	0.161	

[a] Mmu: *Mus musculus;* Rno: *Rattus norvegicus;* X: *Ube1x;* Y: *Ube1y;* Y-p1: *Ube1y-p1;* Y-p2: *Ube1y-p2. Ube1x* and *Ube1y* are functional genes, whereas *Ube1y-p1* and *Ube1y-p2* are pseudogenes

(*Ube 1y-p1* in B3 locus, and *Ube 1y-p2* in Sxr[b] region), and rat Y-linked *Ube 1y.* (See the end of Fig. 1.) The exon parts of the amplified fragments were all the same in length (328 bp) with the exception of *Ube 1y-p1* (311 bp). The intronic regions of the genes were variable in length. These sequences were aligned as shown in Fig. 1.

The exon sequences of the mouse X-linked and Y-linked functional *Ube 1* genes determined in this study were the same as the published cDNA sequences (Imai et al. 1992). Both of the two pseudogenes determined in this study contain nonsense mutations in their coding regions. One of the pseudogenes (*Ube 1y-p1,* from the B3 locus) was the same as pY8/b (also from the B3 locus) determined in Tucker et al. (1992) except for five differences (underlined italic nucleotides in Fig. 1): four in intronic regions and one in an exonic region. These differences might be due to polymorphisms in different mouse strains used. No rat cDNA sequence was available for comparison, but the sequence divergence between the rat and mouse sequences determined in this study are as expected from the evolutionary divergence between mice and rats (see later).

Sequence Divergence

From the aligned sequences we computed the divergence between each pair of sequences. We treated introns and exons separately. Table 1 shows the matrix of pairwise distances between intron sequences. The shortest distances are found between the three mouse Y-linked homologues (*Ube 1y-p1, Ube 1y-p2,* and *Ube 1y*), ranging from 0.079 to 0.085. These suggest that the two pseudogenes *Ube 1y-p1* and *Ube 1y-p2* were derived from duplication of the functional gene *Ube 1y* after the separation of the mouse and rat lineages. The divergence between the mouse and rat X-linked intron sequences is estimated to be 0.107, which is smaller than the divergence between the mouse and rat Y-linked homologous sequences, 0.161 (Table 1). The distances between X-linked and Y-linked sequences are quite large, ranging from 0.641 to 0.739.

In the coding regions 6 (0.086) synonymous and 0 nonsynonymous changes between mouse and rat *Ube 1x* genes were found, while 9.5 (0.144) synonymous and 6.5 (0.027) nonsynonymous changes between mouse and rat *Ube 1y* genes were estimated (Table 2). The observations that no nonsynonymous difference was found between mouse and rat *Ube 1x* genes and that the nonsynonymous distances between human and rodent *Ube 1x* genes are very small (0.016 and 0.016) suggest that the enzyme Ube 1x is very conservative in evolution. The nonsynonymous divergence between mouse and rat *Ube 1y* genes is 0.027 and so *Ube 1y* is less conservative than *Ube 1x.* The two mouse pseudogenes are not equally distant from the functional *Ube 1y* genes of the mouse and rat in exon sequences. For synonymous substitutions, the divergence between mouse *Ube 1y-p1* and *Ube 1y* is 0.097 whereas that between mouse *Ube 1y-p2* and *Ube 1y* is 0.145, and the divergence between mouse *Ube 1y-p1* and rat *Ube 1y* is 0.131 whereas that between mouse *Ube 1y-p2* and mouse *Ube 1y,* is 0.188. For nonsynonymous substitutions, the differences are even more pronounced: 0.035 between *Ube 1y-p1* and mouse *Ube 1y* vs 0.072 between *Ube 1y-p2* and mouse *Ube 1y* , and 0.059 between *Ube 1y-p1* and rat *Ube 1y* vs 0.093 between *Ube 1y-p2* and rat *Ube 1y.* A simple explanation for these observations is that the *Ube 1y-p2* pseudogene became nonfunctional earlier than did the *Ube 1y-p1* pseudogene. Note that since the nonsynonymous divergence between mouse and rat *Ube 1y* is only 0.027, pseudogenes *Ube 1y-p1* and *Ube 1y-p2* have evolved at least two times faster than mouse *Ube 1y* in terms of nonsynonymous substitution.

The synonymous divergences between X-linked and Y-linked *Ube 1* genes are relatively uniform among pairwise comparisons (from 0.695 to 0.882) with the pseudogenes showing slightly higher divergences from the X-linked genes (0.763 to 0.882). More pronounced differences were found in nonsynonymous substitutions—for example, 0.084–0.089 between the Y-linked and X-linked functional genes in contrast to 0.119–0.153 between the pseudogenes and the X-linked genes (Table 2), in agreement with the above observation that the

Table 2. Number of substitutions per synonymous site (below diagonal) and per nonsynonymous site (above diagonal) between *Ube1* genes and pseudogenes of rodents and humans[a]

	MmuX	RnoX	MmuY-p1	MmuY-p2	MmuY	RnoY	HsaX
MmuX		0.000 ± 0.000	0.119 ± 0.025	0.152 ± 0.029	0.084 ± 0.020	0.089 ± 0.021	0.016 ± 0.008
RnoX	0.086 ± 0.039		0.119 ± 0.025	0.153 ± 0.029	0.084 ± 0.020	0.089 ± 0.021	0.016 ± 0.008
MmuY-p1	0.763 ± 0.165	0.811 ± 0.178		0.085 ± 0.020	0.035 ± 0.012	0.059 ± 0.017	0.101 ± 0.023
MmuY-p2	0.853 ± 0.175	0.882 ± 0.181	0.131 ± 0.044		0.072 ± 0.019	0.093 ± 0.022	0.139 ± 0.028
MmuY	0.811 ± 0.169	0.824 ± 0.171	0.097 ± 0.038	0.145 ± 0.052		0.027 ± 0.011	0.067 ± 0.018
RnoY	0.695 ± 0.143	0.737 ± 0.154	0.131 ± 0.044	0.188 ± 0.058	0.144 ± 0.051		0.071 ± 0.019
HsaX	0.422 ± 0.094	0.458 ± 0.106	0.651 ± 0.141	0.714 ± 0.149	0.719 ± 0.170	0.756 ± 0.208	

[a] Hsa: *Homo sapien;* other species and gene names, see Table 1

pseudogenes have evolved faster than the functional genes.

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

In the introns the divergence between the mouse and rat *Ube 1y* genes is $Y = 0.161$ and that between the mouse and rat *Ube 1x* is $X = 0.107$. From these we compute $Y/X = 1.50$ and obtain an estimate of $\alpha \approx 2.0$ with a 95% confidence interval from 1.0 to 3.9. Similar α values are obtained if we use mouse *Ube 1y* pseudogenes (i.e., *Ube 1y-p1* and *Ube 1y-p2*) instead of the mouse functional gene *Ube 1y*. Moreover, a similar estimate is obtained using synonymous substitutions ($\alpha = 2.37$), although the 95% confidence interval is larger (–0.21 to 5.10), perhaps due to a smaller data set.

Discussion

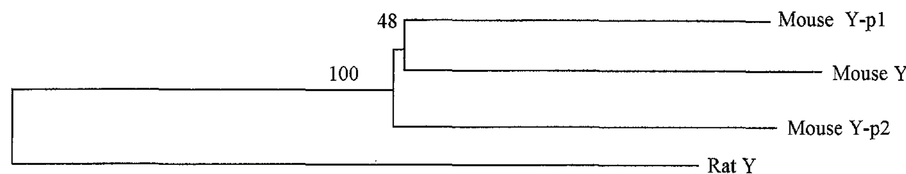
Although the mouse *Ube 1y* gene and pseudogenes are physically very close to the mouse *Zfy* gene in the Sxr region (Mitchell et al. 1991), the mouse *Ube 1x* gene is quite a distance from the mouse *Zfx* gene (~30 centiMorgans, according to Brown et al. 1992). Different regions of the genome have been found to have different rates of substitution (Wolfe et al. 1989). The divergence ratio (*Y/X*) of the Y- and X-linked homologues in mouse and rat is 0.161/0.107 in *Ube 1* introns, while it is 0.182/0.128 in the last intron of the zinc finger protein genes (Chang et al. 1994). The absolute magnitudes of divergence are indeed somewhat different in these introns, but the *Y/X* ratios (1.50 vs 1.42) are quite similar. These lead to similar estimates of α (1.8 vs 2.0). Thus, the present study supports our previous suggestion that the sex ratio of mutation rate in mice and rats is relatively low (i.e., only 2) compared to that in higher primates (i.e., ~6). As explained in Chang et al. (1994), the estimated αs are very close to the male-to-female ratio of the number of germ-cell divisions, and this suggests that substitution mutation is largely replication-dependent. A study based on in vitro assay, at the protein level, of the substitution mutation rate in the human lymphoblastoid cell line also supports this notion (Kuick et al.1992).

In using intron sequences to estimate α we have implicitly assumed that the effect of selective constraints on intron divergence is negligible. The present data provide some support for this assumption. While the two pseudogenes *Ube 1y-p1* and *Ube 1y-p2* have evolved at least two times faster than mouse *Ube 1y* in terms of nonsynonymous substitution, they evolved at the same rate as mouse *Ube 1y* in introns—the divergences between the two pseudogenes and rat *Ube 1y* (0.163 and 0.157) are very close to the divergence between mouse *Ube 1y* and rat *Ube 1y* (0.161, Table 1).

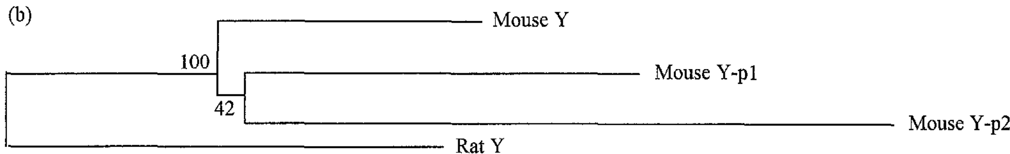
Almost all the mammalian species so far tested have *Ube 1* homologous genes on both X and Y chromosomes, with the exception of the primates in which only a single X-chromosome homologue was detected (Mitchell et al. 1991, 1992). It is suggested that the primate Y homologue was lost after the separation of the primate lineage from other mammalian lineages (Mitchell et al. 1992). There is some support for this hypothesis. In Table 2 the synonymous divergence between the mouse and rat *Ube 1y* genes is 0.14 whereas those between rodent *Ube 1y* genes and rodent or human *Ube 1x* genes are ⩾0.70, the latter being at least five times larger than the former. Since an X-linked sequence would have evolved more slowly than a Y-linked sequence, the above ratio implies that the divergence time between *Ube 1x* and *Ube 1y* is at least five times earlier than the mouse–rat split. According to O'hUigin and Li's (1992) estimate, the mouse–rat split occurred more than 20 million years (Myr) ago. So, we may conclude that *Ube 1x* and *Ube 1y* diverged more than 100 Myr ago, i.e., before the rodent–primate split, which is commonly assumed to be about 80 Myr ago. Of course, these arguments involve many assumptions, and so the conclusion is rather tentative.

The two nonprocessed pseudogene sequences determined in this study seem to have diverged from the functional gene at similar times because the distance from *Ube 1y-p1* and *Ube 1y-p2* to mouse *Ube 1y* in the introns are very similar (0.079 vs 0.082; Table 1). A phyloge-

(a)



(b)



(c)

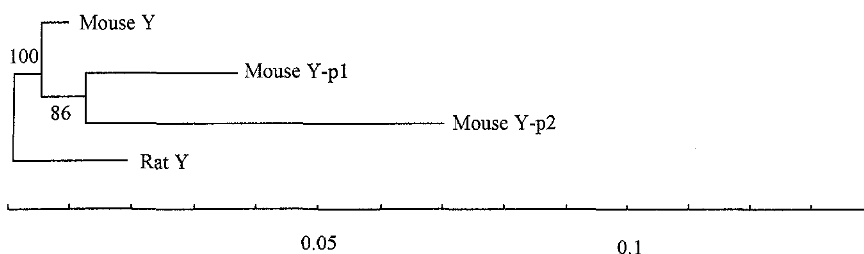


Fig. 2. Phylogenetic trees of rodent *Ube 1y* genes and pseudogenes constructed by the neighbor-joining method (Saitou and Nei 1987) using **(a)** intron sequences, **(b)** synonymous substitutions, and **(c)** nonsynonymous substitutions. *Y, Ube 1y* gene; *Y-p1,* mouse *Ube 1y-p1* pseudogene; *Y-p2,* mouse *Ube 1y-p2* pseudogene.

netic analysis of the sequence data supports this view. Indeed, the neighbor-joining tree based on the intron sequences suggests a near trichotomy for mouse *Ube 1y, Ube 1y-p1,* and *Ube 1y-p2* (Fig. 2a). The two trees based on synonymous substitutions and nonsynonymous substitutions, respectively, both suggest a closer relation between *Ube 1y-p1* and *Ube 1y-p2* than either of them is to mouse *Ube 1y* (Fig. 2b,c). In both trees, *Ube 1y-p1* and *Ube 1y-p2* would have been derived from a duplication of an ancestral gene that was derived from a duplication of mouse *Ube 1y.* Under any of the trees in Fig. 2, *Ube 1y-p2* would have evolved faster than *Ube 1y-p1* in terms of synonymous and nonsynonymous substitutions. As mentioned above, a simple explanation for this observation is that *Ube 1y-p2* became nonfunctional earlier than *Ube 1y-p1.*

Acknowledgments. We thank David Hewett-Emmett and Lawrence Simmin for suggestions. This study was supported by NIH grants. We thank the University of Texas M.D. Anderson Cancer Center (UTMDACC NCI-CA-16672) for access to the GCG software package.

References

Brown SDM, Avner P, Herman GE (1992) Mouse X chromosome. Mamm Genome 3:S274–S288

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 U S A 91:827–831

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

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

Handley PM, Mueckler M, Siegel NR, Ciechanover A, Schwartz AL (1991) Molecular cloning, sequencing, and tissue distribution of the human ubiquitin-activating enzyme E1. Proc Natl Acad Sci U S A 88:258–261

Imai N, Kaneda S, Nagai Y, Seno T, Ayusawa D, Honaoka F, Yamao F (1992) Cloning and sequence of a functionally active cDNA encoding the mouse ubiquitin-activating enzyme E1. Gene 118: 279–282

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. Am J Hum Genet 52:152–166

Kok K, Hofstra R, Pilz A, Berg AVD, Terpstra P, Buys CHCM, Carritt B (1993) A gene in the chromosomal region 3p21 with greatly reduced expression in lung cancer is similar to the gene for ubiquitin-activating enzyme. Proc Natl Acad Sci U S A 90:6071–6075

Kuick RD, Neel JV, Strahler JR, Chu EHY, Bargal R, Fox DA, Hanash SM (1992) Similarity of spontaneous germinal and in vitro somatic cell mutation rates in humans: implications for carcinogenesis and for the role of exogenous factors in "spontaneous" germinal mutagenesis. Proc Natl Acad Sci U S A 89:7036–7040

Li W-H (1993) Unbiased estimation of the rates of synonymous substitutions. J Mol Evol 36:96–99

Mitchell MJ, Woods DR, Tucker PK, Opp JS, Bishop CE (1991) Homology of a candidate spermatogenic gene from the mouse Y chro-

mosome to the ubiquitin-activating enzyme E1. Nature 354:483–486

Mitchell MJ, Woods DR, Wilcox SA, Graves JAM, Opp JS, Bishop CE (1992) Marsupial Y chromosome encodes a homologue of the mouse Y-linked candidate spermatogenesis gene Ube 1y. Nature 359:528–533

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

O'hUigin C, Li W-H (1992) The molecular clock ticks regularly in muroid rodents and hamsters. J Mol Evol 35:377–384

Pamilo P, Bianchi NO (1993) Evolution of the ZFX and ZFY genes: rates and interdependence between the genes. Mol Biol Evol 10: 271–281

Rosendaal FR, Brocker-Vriends AHJT, van Houwilingen JC, Smit C, Varekamp H, van Kijck H, Suurmeijer TPBM, Vandenbroucke JP, Briet E (1990) Sex ratio of the mutation frequencies in haemophilia A: estimation and meta-analysis. Hum Genet 86:139–146

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

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

Tucker PK, Philips KS, Lundrigan B (1992) A mouse Y chromosome pseudogene is related to human ubiquitin-activating enzyme E1. Mamm genome 3:28–35

Wolfe KH, Sharp PM, Li W-H (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285