DOI: 10.1534/genetics.103.025072

Rapid Evolution of Mammalian X-Linked Testis-Expressed Homeobox Genes

Xiaoxia Wang and Jianzhi Zhang[1]

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109

Manuscript received November 26, 2003
Accepted for publication February 11, 2004

ABSTRACT

Homeobox genes encode transcription factors that function in various developmental processes and are usually evolutionarily conserved in their sequences. However, two X-chromosome-linked testis-expressed homeobox genes, one from rodents and the other from fruit flies, are known to evolve rapidly under positive Darwinian selection. Here we report yet another case, from primates. *TGIFLX* is an X-linked homeobox gene that originated by retroposition of the autosomal gene *TGIF2*, most likely in a common ancestor of rodents and primates. While *TGIF2* is ubiquitously expressed, *TGIFLX* is exclusively expressed in adult testis. A comparison of the *TGIFLX* sequences among 16 anthropoid primates revealed a significantly higher rate of nonsynonymous nucleotide substitution (d_N) than synonymous substitution (d_S), strongly suggesting the action of positive selection. Although the high d_N/d_S ratio is most evident outside the homeobox, the homeobox has a d_N/d_S of ∼0.89 and includes two codons that are likely under selection. Furthermore, the rate of radical amino acid substitutions that alter amino acid charge is significantly greater than that of conservative substitutions, suggesting that the selection promotes diversity of the protein charge profile. More interestingly, an analysis of 64 orthologous homeobox genes from humans and mice shows substantially higher rates of amino acid substitution in X-linked testis-expressed genes than in other genes. These results suggest a general pattern of rapid evolution of mammalian X-linked testis-expressed homeobox genes. Although the physiological function of and the exact selective agent on *TGIFLX* and other rapidly evolving homeobox genes are unclear, the common expression pattern of these transcription factor genes led us to conjecture that the selection is related to one or more aspects of male reproduction and may contribute to speciation.

HOMEOBOX genes are characterized by the presence of an ∼60-codon sequence motif known as the homeobox, which encodes a helix-turn-helix DNA-binding domain named the homeodomain (Gehring *et al.* 1994a). Initially identified from fruit flies (McGinnis *et al.* 1984; Scott and Weiner 1984), homeobox-containing genes have now been found in fungi, plants, and animals and form a large gene superfamily (Kappen *et al.* 1993; Bharathan *et al.* 1997; Kappen 2000; Banerjee-Basu and Baxevanis 2001). Homeobox genes function as transcription factors that regulate the expressions of their target genes in various developmental processes such as body-plan specification, pattern formation, and cell fate determination (Gehring *et al.* 1994a). Because of their fundamental importance in development, homeobox genes are of substantial interest to evolutionary biologists as they may provide key information on the evolution of development (Shepherd *et al.* 1984; Garcia-Fernandez and Holland 1994; Zhang and Nei 1996; Carroll *et al.* 2001). Earlier studies showed that homeobox genes, particularly the homeobox region, are conserved in evolution (McGinnis *et al.* 1984; Gehring *et al.* 1994a), although two notable exceptions, *Pem* in rodents and *OdsH* in Drosophila, have been reported (Sutton and Wilkinson 1997; Ting *et al.* 1998). In both cases, high rates of amino acid substitution were found in the homeodomain and the action of positive selection was suggested. Interestingly, both genes are located on X chromosomes and are expressed in testis, although *Pem* is also expressed in female reproductive tissues. *OdsH* is in part responsible for the hybrid male sterility between *Drosophila simulans* and *D. mauritiana* (Ting *et al.* 1998). These intriguing findings suggest that homeobox genes may also be involved in developmental processes that vary among closely related species. Because such developmental variations may lead to reproductive isolation and speciation (Ting *et al.* 1998), it is of interest to identify new cases of rapidly evolving homeobox genes. Here we describe the identification of such a rapidly evolving homeobox gene, *TGIFLX* [TG-interacting factor (TGIF)-like X], from primates. TGIFLX is a member of TGIFs, a group of transcription factors of the three amino-acid loop extension (TALE) superclass of the homeodomain protein family (Bertolino *et al.* 1995; Blanco-Arias *et al.* 2002). Earlier evolutionary analyses suggested that the X-chromosome-linked *TGIFLX* gene

Sequence data from this article have been deposited in the EMBL/GenBank Data Libraries under accession nos. AY449635–AY449641.

[1] *Corresponding author:* Department of Ecology and Evolutionary Biology, University of Michigan, 3003 Nat. Sci. Bldg., 830 N. University Ave., Ann Arbor, MI 48109. E-mail: jianzhi@umich.edu

originated by retroposition of the autosomal *TGIF2* gene, a member of TGIFs (Blanco-Arias *et al.* 2002). In contrast to *TGIF2*, which is ubiquitously expressed, *TGIFLX* is specifically expressed in the germ cells of adult testis (Blanco-Arias *et al.* 2002; Lai *et al.* 2002). In this report, we show that (1) the retroposition event predated the divergence of primates and rodents, (2) *TGIFLX* evolved rapidly in primates under positive selection, and (3) mammalian X-linked testis-expressed homeobox genes evolve rapidly in general.

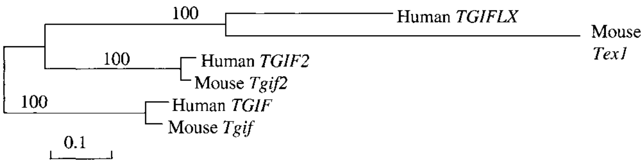

Figure 1.—Phylogenetic tree of *TGIFLX*, *TGIF2*, and *TGIF* genes. The tree is reconstructed with the neighbor-joining method with the protein Poisson distances. Bootstrap percentages from 1000 replications are shown on tree branches. Branch lengths show the numbers of amino acid substitutions per site. *TGIF* genes are used as outgroups.

MATERIALS AND METHODS

DNA amplification and sequencing: The *TGIFLX* coding region does not contain introns. The coding region was amplified from genomic DNAs of two Old World (OW) monkeys (green monkey *Cercopithecus aethiops* and douc langur *Pygathrix nemaeus*) and five New World (NW) monkeys (marmoset *Callithrix jacchus*, tamarin *Saguinus oedipus*, owl monkey *Aotus trivirgatus*, squirrel monkey *Saimiri sciureus*, and woolly monkey *Lagothrix lagotricha*), using polymerase chain reaction (PCR). For green monkey and douc langur, primers 2XL (5′-TTT GAATATGGAGGCCGCTG) and 2XR (5′-CATCATCAATCA TGGATTAG) were used; for tamarin, woolly monkey, and marmoset, primers 2XL and XIA1 (5′-GGATTAGACTC TTGCTTCTTCT) were used; for owl monkey and squirrel monkey, primers X2 (5′-ATATGGAGGCCGCTGCAgAAGAC) and X3 (5′-GGCTCTTGCTTCTTCTCTAGC) were used. PCRs were performed with MasterTaq under conditions recommended by the manufacturer (Eppendorf, Hamburg, Germany). The products were then purified and sequenced from both directions, using the dideoxy chain termination method with an automated sequencer.

Analysis of TGIFLX gene sequences: The DNA sequences of the *TGIFLX* coding region from five hominoids (humans and apes) and four OW monkeys (Blanco-Arias *et al.* 2002) were obtained from GenBank. The accession numbers are: human (*Homo sapiens*), AJ427749; chimpanzee (*Pan troglodytes*), AJ345073; gorilla (*Gorilla gorilla*), AJ345074; orangutan (*Pongo pygmaeus*), AJ345075; gibbon (*Hylobates lar*), AJ345076; talapoin (*Miopithecus talapoin*), AJ345077; rhesus monkey (*Macaca mulatta*), AJ345078; crab-eating macaque (*M. fascicularis*), AJ345079; and baboon (*Papio hamadryas*), AJ345080. These publicly available sequences are analyzed together with those determined in this study. Seven amino acids at the N terminus and 10 amino acids at the C terminus of the sequences are encoded by the primer sequences and were not included in data analysis. A total of 16 TGIFLX protein sequences were aligned using the software DAMBE (Xia and Xie 2001) followed by manual adjustments. The DNA sequence alignment was then made following the protein alignment. The MEGA2 program (Kumar *et al.* 2001) was used for phylogenetic analysis. The number of synonymous nucleotide substitutions per synonymous site (d_S) and that of nonsynonymous substitutions per nonsynonymous site (d_N) were computed using the modified Nei-Gojobori method (Nei and Gojobori 1986; Zhang *et al.* 1998), with an estimated transition/transversion ratio of 1.6. On the basis of the phylogeny of the 16 primates, we inferred ancestral *TGIFLX* sequences at all interior nodes of this tree, using the distance-based Bayesian method (Zhang and Nei 1997). The numbers of synonymous (s) and nonsynonymous (n) substitutions on each branch of the tree were then counted. Radical and conservative nonsynonymous substitutions with regard to amino acid charge and polarity were computed following Zhang (2000). Positive selection at individual codons was tested using the likelihood-based (Yang *et al.* 2000) and parsimony-based (Suzuki and Gojobori 1999) methods.

Analysis of other homeobox genes of human and mouse: We searched for homeobox genes from the human genome resources (http://www.ncbi.nlm.nih.gov/genome/guide/human/) and then found their mouse orthologs using the UniGene tool (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene). We downloaded the human and mouse protein sequences, aligned them using DAMBE, and computed protein *p*-distances (proportional differences; Nei and Kumar 2000) between human and mouse orthologs. The information on gene location and expression pattern was found using human genome resources and the LocusLink tool (http://www.ncbi.nlm.nih.gov/LocusLink/).

RESULTS

Retroposition predated the human-mouse separation: To determine when the retroposition that generated *TGIFLX* occurred in evolution, we conducted a BLAST search in the GenBank for homologous sequences to *TGIFLX* and its mother gene *TGIF2*. We identified a homeobox gene *Tex1* (also known as *Tgifx1-pending*) in the mouse that is mapped to a region of the X chromosome that is syntenic with human Xq21.3, where *TGIFLX* is located. *Tex1* is also specifically expressed in the germ cells of mouse testis (Lai *et al.* 2002). These facts suggest that mouse *Tex1* is orthologous to human *TGIFLX*. Furthermore, we obtained the gene sequences of human and mouse *TGIF2* from GenBank and conducted a phylogenetic analysis of these sequences. The human and mouse *TGIF* sequences are used as outgroups. The gene tree shows high bootstrap support for the retroposition that gave birth to *TGIFLX* occurring in a common ancestor of primates and rodents (Figure 1).

Although retroposition usually generates pseudogenes, a number of retroposition-mediated functional genes have been identified (Long 2001). *TGIFLX* is apparently a functional gene as its open reading frame has been maintained throughout mammalian evolution. Retroposition is a mutation-prone process due to a high error rate in retrotranscription. Also, newly duplicated genes often have elevated rates of evolution due to relaxation of functional constraints and/or positive selection (Zhang 2003). Thus, one may expect to see a burst of substitutions in the *TGIFLX* branch immediately following the retroposition. Interestingly, Figure 1 shows that *TGIFLX* evolves more rapidly than *TGIF2* not only in this branch, but also throughout its evolutionary history. We found that the number of amino acid substitutions per site (Poisson distance) between the orthologous human and mouse *TGIFLX* genes is 0.814 ± 0.080, and the corresponding number for *TGIF2* is 0.031 ± 0.013, their difference being statistically significant ($P < 0.001$). Of 1880 orthologous human and rodent genes analyzed by Makalowski and Boguski (1998), only 6 have substitution rates greater than that of *TGIFLX*, suggesting that it is evolving at an exceptionally high rate. To further characterize the substitution rate of *TGIFLX*, we conducted a detailed evolutionary study of this gene in primates.

Positive selection on primate *TGIFLX*: The *TGIFLX* coding sequences from five hominoids and four OW monkeys were reported by Blanco-Arias *et al.* (2002). We here determined the orthologous sequences in two additional OW monkeys and five NW monkeys. Thus, a total of 16 primate sequences are analyzed here. The alignment of these 16 protein sequences shows that they are highly variable (Figure 2). The nonhomeodomain regions show the highest variability, although 25 of the 63 amino acid positions in the homeodomain are also variable among the 16 primates. Hydrophobic amino acids are usually conserved in homeodomains; in the present case 22 of the 29 hydrophobic sites are completely conserved among the primate sequences, and in the rest 7 also involve only hydrophobic amino acid changes. In the third helix of the homeodomain, four amino acids (W51, F52, N54, and R56; positions in the homeodomain) are known to be conserved (Banerjee-Basu and Baxevanis 2001), which is also the case here. Position 53 is usually occupied by a polar amino acid in homeodomains, but was found to have a small, nonpolar amino acid in a previous analysis of TALE homeodomains (Burglin 1997). In our sequences, position 53 is variable with either polar or nonpolar amino acids. To examine whether the high sequence variability is a result of positive selection, we computed the synonymous (d_S) and nonsynonymous (d_N) distances between each pair of the sequences. For the entire coding region, higher d_N than d_S is observed in 93 of 120 pairwise comparisons (Figure 3A), suggesting the possible action of positive selection. This pattern is more apparent when only the nonhomeodomain regions are analyzed, as 98 of the comparisons show $d_N > d_S$ (Figure 3B). For the homeodomain, however, only 39 of the comparisons show $d_N > d_S$ (Figure 3C). These results indicate that the substitution rate and pattern may be different between amino acid positions inside and outside the homeodomain.

To test the hypothesis of positive selection more rigorously, we used a phylogeny-based approach (Zhang *et al.* 1997). The phylogentic relationships of the 16 primates are assumed to follow the tree in Figure 4. This phylogeny is relatively well established, especially for the major divisions (Goodman *et al.* 1998; Page and Goodman 2001; Singer *et al.* 2003; Steiper and Ruvolo 2003), and use of alternative trees does not affect our main conclusion. On the basis of this tree, we inferred the ancestral *TGIFLX* gene sequences at all interior nodes of the tree and counted the numbers of synonymous (s) and nonsynonymous (n) substitutions on each tree branch (Figure 4). We found that the sums of n and s for all branches are 195.5 and 58.5, respectively, for the nonhomeodomain regions. The potential numbers of nonsynonymous (N) and synonymous (S) sites are 322 and 128, respectively. Thus $n/s = 3.34$ is significantly greater than $N/S = 2.51$ ($P = 0.031$, binomial test). The binomial test used here is more conservative than Fisher's exact test used in Zhang *et al.* (1997) and is more appropriate here because of multiple substitutions that may have occurred at individual sites (Zhang and Rosenberg 2002). Fisher's exact test would have given a P value of 0.002 here. We also analyzed n/s in hominoids, OW monkeys, and NW monkeys separately, but did not find significant differences (Figure 4). The average number of synonymous substitutions per site is 0.155 between hominoids and New World monkeys and 0.0819 between hominoids and Old World monkeys. These values are virtually identical to the corresponding numbers obtained from multiple intron and noncoding sequences of primate genomes (0.149 and 0.079, respectively; Li 1997, pp. 221–224), suggesting that the synonymous substitution rate in *TGIFLX* is normal. Thus, our results strongly suggest that positive selection is responsible for the rapid evolution at nonsynonymous sites of the nonhomeodomain regions.

For the homeodomain, we found that n/s (2.42) is slightly lower than N/S (2.73) and that the null hypothesis of $n/s = N/S$ cannot be rejected. This may suggest that the homeodomain is under no functional constraints. It may also suggest that some sites in the homeodomain are under positive selection while other sites are under purifying selection, giving an overall pattern of similar average substitution rates at synonymous and nonsynonymous sites (see below). When we examine the substitution patterns of hominoids, OW monkeys, and NW monkeys separately, we find that the n/s ratio is higher among hominoids and OW monkeys (23.5/4.5 = 5.22) than among NW monkeys (25/12 = 2.08; Figure 4). However, this difference is not significant ($P = 0.132$). The n/s ratio is not significantly different from N/S for hominoids and OW monkeys ($P = 0.150$).

Statistical methods for identifying individual codons that are under positive selection have been developed in recent years (Suzuki and Gojobori 1999; Yang *et al.* 2000). We first applied the likelihood method (Yang *et al.* 2000) to the *TGIFLX* data and compared the likelihoods under models 7 and 8. Here model 7 assumes that the d_N/d_S ratio for individual sites follows a β-distribution between 0 and 1, while model 8 adds an extra

```
                                                                                             Homeodomain
                       ________________________________________________________________________
Human                 MEAAADGPAE TQSPVEKDS- ---------- ---PAKTQSP AQDTSIMSRN NADTGRVLAL PEHKKKRKGN LPAESVKILR DWMYKHRFKA YPSEEEKQML SEKTNLSLLQ ISNWFINARR RILPDMLQQR RNDPIIGHKT
Chimpanzee            .......... .....Q...- ---------- ---....... .......... .......... .......... .......V.. .......... .......... .......... .......... .........H ..........
Gorilla               .......... .....K...- ---------- ---....... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Orangutan             ...T...... .........- ---------- ---.V..... ......T... ...R...... .......H.. ...K...... ........R. .......... .......... V......... .........C G...V....M
Gibbon                .......S.. .........- ---------- ---....... .....TV... S....K.... ...T..P..Y .......... .......... .......... ........S. .......... .......KRH G...N...E.
Rhesus monkey         .......... .R.R.....- ------RRAI KDS....... .......L.. .....K.P.. .........Y S......... ........R. ....A..R.. .K......S. .......... ........R. G....V....
Crab-eating monkey    .......... .R.R.....- ------RRAK KDS....... .......L.. .....K.P.. .........Y .......... ........R. ....A..R.. .K......S. .......... ........R. G....V....
Baboon                ......S... .R.R.....R RVEKDSRRPK KDS....... .......L.. .....K.... .........Y .......... R.......R. ....A..R.. .K......S. .......... ........R. G..R.V....
African green monkey  .......... .R.R.....- ------RRAK KDS....... .......L.. .....K.... ........T.Y .......... N.......R. ....A..R.. .R......S. .......... ........R. G.N.TV....
Talapoin              ......R... .R.R.....- ------RRAK KDS....... .......L.S .....K.... .........Y .......... ........R. ....A..R.. .K......S. .......... ........R. G...TV....
Douc langur           .......R.. .R.......- ------RRAK KDS.....R. .......LK. .....K.... ...E....EY .......... ........R. ....VQ.... .R......S. .....T.... .V......RS GK...V....
Marmoset              .....KD... I...I.NET- ---------- ---..DNL.. .......... ...I.KP... .RRRS.P... ..V....... ........R. ....A..L.. ..I....FS. V......... ....K..G.S G..SFVDQQ.
Tamarin               .....KD... I...I.NET- ---------- ---..DNL.. .......... ...I.KP... .RRRS.P... ..V....... ........R. ....A..L.. ..I....FS. V......... ....K..G.S G..SFVDQQ.
Owl monkey            .....ED... IE---KNKT- ---------- ---..DNLR. .....MK... ..GIDKP... .QRL..PRV. ..I....... R.......R. ....A..... .......FS. V.T....... ....K..L.S G..SFVDQEM
Squirrel monkey       .....EDA.. I...M.NE.- ---------- ---..VNL.. .....M.... ...IDKP... SGCR..A.-- ..V....... K.......R. ....A..L.. A......FS. .....V.... ....G..RKS G.----DQEM
Woolly monkey         .....ED... IEN.M.NER- ---------- ---..DNL.. ........K. ...IDKPR.. .RR...P... ..V....... ........R. ....V..... ........S. V...T..... ....E..L.S G..SFVDQEM

Human                 GKDAHATHLQ STEASVPAKS GPSGPDNVQS LPLWPLPKGQ MSREKQPDPE SAPSQKLTGI AQPKKKVKVS -VTSPSSPEL VSP------E EHADFSSFLL LVDAAVQRAA ELELEKKQEP NP
Chimpanzee            .......... .......... .......... .......... .......... .......... .......... -I........ ...------. .Y........ .......... .......... ..
Gorilla               .......... ..K....... .......... .......... ..G....... ........R. .......... -I.......P ...------. .Y........ .......... .......... ..
Orangutan             ..N....... ..D.....S. .......... .......... ..G..L.... .........K ...E....F. -I...F...P ...------. .Y......Q. .......... ........D. ..
Gibbon                .......... ..D....... .....E.... ..V....... ..G..L.... .....NP.V. .......... -........P .P.------. .YP.....Q. .......... .......... ..
Rhesus monkey         .........R ..D....... ..R.S..... ...RSS.... ..G..I.E.G ........M. .......... NI..L....P ..T------. .Y......Q. .......... .........S ..
Crab-eating monkey    .........R ..D....... ..R.S..... ...RSS.... ..G..I.E.G ........V. .......... NI..L....R ..T------. .Y......Q. .......... .........S ..
Baboon                ....N..... ..D....... ..R.S..... ...RSS.... ..G..I.E.G ........M. .......... NI..S....P ..T------. .Y......Q. .......... .........S ..
African green monkey  .......... ..N....... ..R.S..... ...RSS.... ..G..I.E.G ........M. .......... NI.AS....P ..T------. .Y....N.Q. .......... .........S ..
Talapoin              .......... ..D....... ..R.S..... ...RSS.... ..G..I.E.G ........M. .......... NI..S....P ..T------. .Y......Q. .......... ......N..S ..
Douc langur           S......... ..D....... ..R.S..A.. ...RS..... ..G..I...G ........M. .......... NI..S....P ..SPEPVSP. ......G.Q. .......... ......Q..S ..
Marmoset              ..-DNDN... G.DDF.S... R.RD..Q... .--..V.M.. ..GK.L...G W....EVAV. ......L... TN.TR...KP .P.------. .YP...N..I ..EV.A.... .........S .-
Tamarin               ..-DNDN... G.DDF.S... R.RD..Q... .--..V.M.. ..GK.L...G W....EVAV. ......L... TN.TR...KP .P.------. .YP...N..I ..EV.A.... .........S .-
Owl monkey            ...--D.... G.DD..FP.. ..RYLGK... I--..V.M.. .....L...R ..C.KE.AVK ......L... TN.T.....P ...------K .YP..TR.HI ..EV.A...V .........- --
Squirrel monkey       R..DND.N.. D.DDF.S..L R.RH..K... .--..V.M.. ..G.NL...G L...RE.AV. ......LQ.. TNAT.....P ...------. .YP.LT..QI ..EV...... .........- --
Woolly monkey         ...DDD...R G.DEF.S... RRRD..K... .--..V.MC. ..G.QL...G .....E.AV. ......F.I. TN.T.....P ...------. .YP.....QI ..EV...... .........S .-
```

Figure 2.—Alignment of TGIFLX sequences of 16 primates. A dot indicates identity to the human sequence and a dash indicates a gap. The first 7 and last 10 amino acid positions are primer encoded in various sequences and are not used in subsequent sequence analysis. The four positively selected sites with posterior probabilities >90% (see text) are in boldface type.

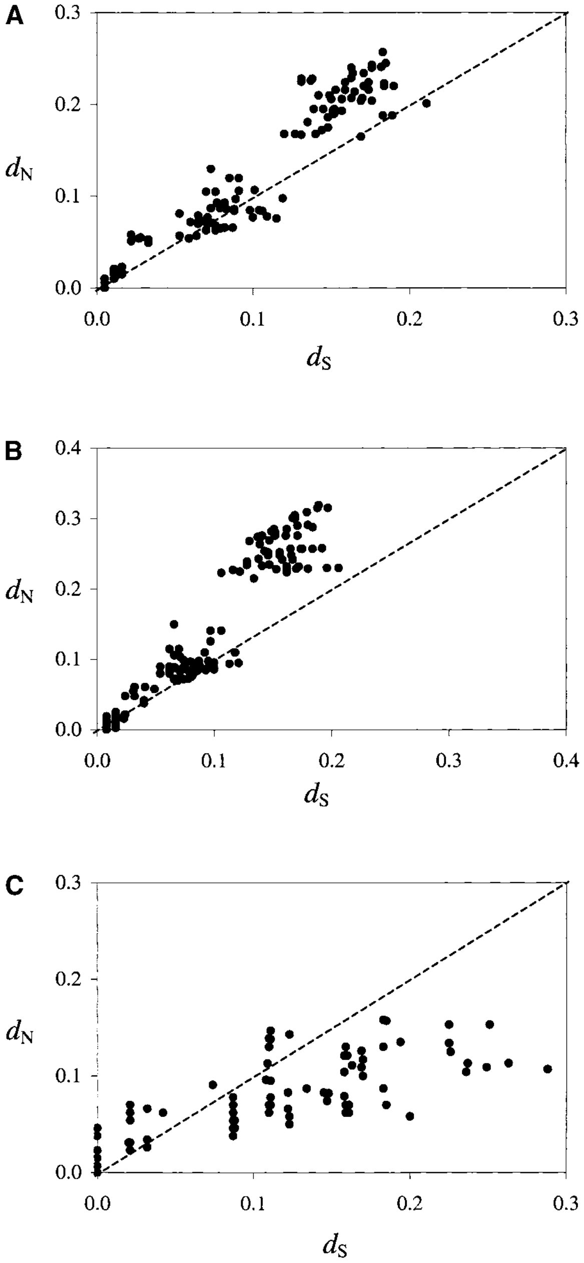

FIGURE 3.—Pairwise comparisons of d_S and d_N among 16 primate *TGIFLX* sequences for (A) the entire sequence, (B) nonhomeodomain regions, and (C) the homeodomain.

class of sites to model 7. We found that model 8 fits the data significantly better than model 7 (χ^2 = 15.2, d.f. = 2, $P < 0.001$), with an additional class of sites of d_N/d_S = 2.42. Four codons were identified to be under positive selection with posterior probabilities >90%, and they are marked on the sequences shown in Figure 2. Similar results were obtained when models 1 and 2 were compared (see YANG *et al.* 2000 for details of the model description). Because the likelihood method has been shown to generate false-positive results occasionally (SUZUKI and NEI 2002), we examined the evidence for selection at the four codons by a more conservative parsimony-based method (SUZUKI and GOJOBORI 1999). None of the four codons show significant results of positive selection when they are tested individually (P = 0.19–0.59). When they are tested together, however, significant evidence for positive selection is found (average d_N/d_S = 5.10, P = 0.021), suggesting that one or more of the four codons are under positive selection. It is interesting to note that two of the four codons are located within the homeodomain while the other two are adjacent to the 3′ end of the homeodomain, suggesting that the homeodomain may indeed be under positive selection (Figure 2). The two residues within the homeodomain are not among the completely conserved residues of all homeodomains, indicating that substitutions at these sites are unlikely to disrupt the basic structure and function of homeodomains. Furthermore, crystal structures of homeodomains show that the first of the two residues is involved in DNA-protein binding and that it contributes significantly to the functional specificity of homeodomains (GEHRING *et al.* 1994b). The second of the two residues belongs to helix I of the homeodomain, and it may also be involved in DNA-protein binding, although a more specific molecular function has yet to be defined.

Selection promotes the diversity of charge profile: To investigate what types of nonsynonymous substitutions are favored by selection, we counted the numbers of conservative and radical nonsynonymous substitutions on each branch of the tree in Figure 4. Conservative nonsynonymous substitutions are those that do not alter the charge of the encoded amino acids and radical substitutions are those that alter the charge of the amino acids. We found a total number of r = 91.5 radical substitutions and c = 104 conservative substitutions in the tree for the nonhomeodomain regions. The potential numbers of radical and conservative sites are R = 128 and C = 195, respectively. The radical substitution rate (r/R = 0.715) is significantly greater than the conservative substitution rate (c/C = 0.533) at P = 0.027 (binomial test). This is in sharp contrast to the situation in most mammalian genes where the radical substitution rate is below the conservative rate (ZHANG 2000). This result suggests that selection may favor alternations of amino acid charge in TGIFLX evolution. We also tested the hypothesis that selection may favor an alternation of amino acid polarity, but obtained no supporting evidence. For the homeodomain, there is no evidence for selection promoting the diversity of either amino acid polarity or charge.

In the above, we compared the number of radical substitutions per radical site (r/R) with the number of conservative substitutions per conservative site (c/C). This comparison provides information on differential selections at radical *vs.* conservative sites, as long as the four parameters (r, c, R, and C) are correctly estimated (SMITH 2003). In contrast, comparisons between r and c can be misleading, because the potential numbers of

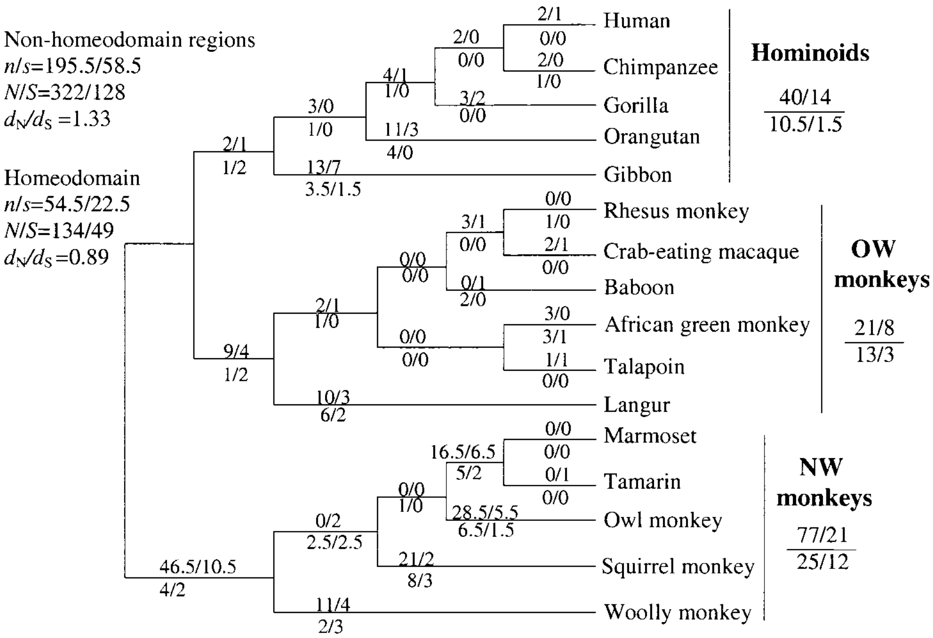

Figure 4.—Numbers of synonymous (s) and nonsynonymous (n) substitutions in the evolution of primate *TGIFLX* genes. Shown above each branch is n/s for the nonhomeodomain regions and below each branch is n/s for the homeodomain. N and S are the potential numbers of nonsynonymous and synonymous sites, respectively (see text).

radical (R) and conservative (C) sites in a gene sequence are usually different and they are affected by many factors unrelated to selection (Dagan *et al.* 2002).

Rapid evolution of mammalian X-linked testis-expressed homeobox genes: As mentioned, two other homeobox genes, *Pem* and *OdsH*, were reported to evolve rapidly (Sutton and Wilkinson 1997; Ting *et al.* 1998). The d_N/d_S ratio of *Pem* ranges from 0.65 to 1.56 for the homeodomain between *Mus musculus* and several related rodents (Sutton and Wilkinson 1997). We reanalyzed the *OdsH* homeodomain sequences from *D. simulans* and *D. mauritiana* (Ting *et al.* 1998) and obtained a d_N/d_S ratio of 1.55. Interestingly, *TGIFLX*, *Pem*, and *OdsH* are all located in X chromosomes and are all testis expressed. This observation prompted us to wonder whether it is a general pattern for X-linked testis-expressed homeobox genes to evolve rapidly. To test this hypothesis, we searched for orthologous homeobox genes from the human and mouse genome sequences. Our search was not exhaustive, but random. Of the 64 genes found, 4 are X-linked and testis expressed, 3 are X-linked and non-testis expressed, 13 are autosomal and testis expressed, and 44 are autosomal and non-testis expressed. Note that there appear to be only 7 X-linked homeobox genes, as a further exhaustive search did not find additional genes. Here "testis expression" simply means that the gene is expressed in testis, regardless of its expression in other tissues. We aligned the sequences and computed the amino acid p-distance for each orthologous pair. As shown in Table 1 and Figure 5A, when the entire protein is considered, autosomal homeobox genes (regardless of the expression pattern) and X-chromosomal non-testis-expressed homeobox genes have similar amino acid p-distances on average, which are an order of magnitude lower than those of X-linked testis-expressed homeobox genes, and their difference is statistically significant ($P < 0.0001$, permutation test). The same pattern is observed when only the homeodomain or nonhomeodomain regions are considered (Table 1; Figure 5, B and C). These results suggest that it is a general pattern for mammalian X-linked testis-expressed homeobox genes to evolve rapidly. In addition to *TGIFLX*, the other X-linked testis-expressed homeobox genes are *ESX1L*, *OTEX*, and *PEPP-2*. While the mouse ortholog of human *ESX1L* is clearly defined by a phylogenetic analysis (data not shown) and chromosomal locations, the orthologs of human *OTEX* and *PEPP-2* are not uniquely defined, probably because of independent gene duplications in rodents and primates after their separation (Wayne *et al.* 2002). From the mouse genome sequence, we identified a total of 15 homologs of the human *OTEX* and *PEPP-2* genes and conducted a phylogenetic analysis of these genes. The phylogeny is not well resolved and has low bootstrap supports (see supplementary Figure 1 at http://www.genetics.org/supplemental/). To be conservative, we computed protein p-distances for the human *OTEX* with each of the 15 mouse genes and presented the smallest distance in Table 1. We also did the same for the human *PEPP-2* gene. Considering possible nonindependent comparisons involved, we also repeated all the statistical tests when only one of the *OTEX* and *PEPP-2* genes was used. We found that the statistical results remain unchanged.

DISCUSSION

In this report, we provide evidence that *TGIFLX* evolves rapidly under positive selection in primates and that the selection favors diversity in charge profile. Al-

TABLE 1

Protein *p*-distances between orthologous human and mouse homeobox genes

		Protein *p*-distance		
Gene name	Protein length (amino acids)	Entire protein	Homeodomain	Nonhomeodomain regions
X-linked, testis expressed				
TGIFLX	222	0.550	0.456	0.582
ESXIL	310	0.565	0.333	0.620
OTEX	176	0.625	0.544	0.664
PEPP-2	208	0.606	0.526	0.636
Mean ± standard error of the mean		0.587 ± 0.018	0.465 ± 0.048	0.626 ± 0.017
X-linked, non-testis expressed				
ARX	560	0.036	0.000	0.040
CDX4	282	0.167	0.017	0.207
POU3F4	361	0.011	0.000	0.013
Mean ± standard error of the mean		0.071 ± 0.048	0.006 ± 0.006	0.087 ± 0.061
Autosomal, testis expressed				
IRX2	471	0.104	0.000	0.119
LHX2	389	0.010	0.000	0.012
LHX9	321	0.006	0.000	0.007
NKX3.1	230	0.322	0.000	0.435
NKX6-2	277	0.029	0.000	0.036
PBX2	430	0.021	0.000	0.024
PKNOX2	305	0.011	0.000	0.012
TIX1[a]	949	0.144	0.037	0.175
ZHX3[a]	522	0.123	0.030	0.154
SIX1	273	0.015	0.000	0.019
TGIF	272	0.103	0.000	0.134
TGIF2	237	0.063	0.000	0.084
ZFHX1B	1214	0.034	0.017	0.035
Mean ± standard error of the mean		0.076 ± 0.024	0.006 ± 0.004	0.096 ± 0.033
Autosomal, non-testis expressed				
ALX3	343	0.085	0.000	0.102
ALX4	397	0.111	0.000	0.129
BAPX1	333	0.153	0.000	0.187
BARX2	254	0.130	0.000	0.162
CRX	299	0.033	0.000	0.042
DLX4	168	0.274	0.017	0.417
GHS-2	303	0.092	0.000	0.114
HHEX	303	0.070	0.018	0.085
IPF1	283	0.120	0.000	0.150
IRX3	501	0.102	0.000	0.116
IRX4	512	0.158	0.000	0.180
IRX5	417	0.113	0.000	0.132
IRX6	438	0.233	0.048	0.263
LHX1	406	0.005	0.000	0.006
LHX3	398	0.101	0.000	0.117
LHX4	367	0.008	0.000	0.010
LHX5	402	0.012	0.000	0.014
LHX6	340	0.168	0.000	0.201
LMX1A	382	0.029	0.000	0.034
LMX1B	372	0.003	0.000	0.003
OTX1	354	0.025	0.000	0.030
PHOX2A	280	0.021	0.000	0.027
PHOX2B	314	0.000	0.000	0.000
PITX1	314	0.035	0.000	0.043
PITX2	317	0.013	0.000	0.015
PITX3	302	0.017	0.000	0.020

(*continued*)

TABLE 1

(Continued)

		Protein *p*-distance		
Gene name	Protein length (amino acids)	Entire protein	Homeodomain	Nonhomeodomain regions
Autosomal, non-testis expressed				
PKNOX1	314	0.039	0.000	0.045
PROP1	223	0.265	0.070	0.331
PROX1	736	0.023	0.000	0.025
PRX2	246	0.077	0.000	0.102
RAX	342	0.140	0.000	0.170
SHOX2	330	0.015	0.000	0.019
SIX2	436	0.014	0.023	0.012
SIX3	332	0.024	0.000	0.029
SIX4	753	0.089	0.000	0.103
SIX5	657	0.139	0.000	0.150
SIX6	246	0.024	0.017	0.027
TLX1	330	0.027	0.000	0.033
TLX2	284	0.070	0.000	0.088
TLX3	291	0.010	0.000	0.013
VAX1	279	0.029	0.000	0.036
VAX2	290	0.121	0.000	0.150
VSX1	354	0.229	0.040	0.260
ZFH4	3525	0.082	0.009	0.087
Mean ± standard error of the mean		0.080 ± 0.011	0.006 ± 0.002	0.097 ± 0.014

[a] The mouse sequence is not available. Instead, the rat sequence is analyzed here.

though positive selection acts mainly in the nonhomeodomain regions of the protein, it may also operate at a few sites in the homeodomain. The homeodomain is used in binding DNA sequences in transcription regulation, while the nonhomeodomain regions in TGIFLX might be used in protein-protein interaction as in the case of TGIF and TGIF2 (Bertolino *et al.* 1995; Melhuish and Wotton 2000; Melhuish *et al.* 2001). Rapid evolution at these sites thus may alter the DNA- and protein-binding properties of TGIFLX. In mouse, the *TGIFLX* ortholog *Tex1* is exclusively expressed in the germ cells at the spermatid stage (Lai *et al.* 2002) and apparently escapes the inactivation that most X-linked genes are supposed to experience in spermatogenesis (Lifschytz and Lindsley 1972). Although the physiological function of *TGIFLX* is unknown, the restricted temporal and spatial expression pattern suggests a role of this gene in spermatogenesis and the detected positive selection on *TGIFLX* may be related to spermatogenesis as well.

Our analysis of homeobox genes of humans and mice revealed a general pattern of rapid evolution of X-linked, testis-expressed homeobox genes, although the number of such genes is relatively small. It is interesting to note that among autosomal homeobox genes, testis-expressed genes and non-testis-expressed genes show similar rates of amino acid substitution (Figure 5). Thus, testis expression alone does not explain high rates of protein evolution. Among non-testis-expressed homeobox genes, there is also no significant difference in substitution rate between autosomal genes and X-linked genes, suggesting that chromosomal location alone also does not explain the difference in amino acid substitution rate. We noted in collecting the expression pattern data that 3 of the 4 X-linked testis-expressed genes (*TGIFLX*, *OTEX*, and *PEPP-2*), but only 1 (*NKX3.1*) of the 13 autosomal testis-expressed genes, have exclusive or highly selective expressions in testis. This difference suggests that the majority of the autosomal testis-expressed genes may be under greater functional constraints due to their multifaceted roles in many tissues and developmental processes and thus evolve more slowly. Indeed, *NKX3.1*, which is expressed only in testis, has the highest substitution rate among the 13 autosomal testis-expressed genes (Table 1). On the contrary, most of the X-linked testis-expressed homeobox genes are expressed exclusively or highly in testis and may thus be specifically involved in male reproduction. Many authors showed that genes involved in male reproduction evolve rapidly under positive selection (*e.g.*, Lee *et al.* 1995; Swanson and Vacquier 1995; Metz and Palumbi 1996; Tsaur and Wu 1997; Rooney and Zhang 1999; Wyckoff *et al.* 2000; Swanson and Vacquier 2002; Podlaha and Zhang 2003). In particular, Torgerson and Singh (2003) recently showed that mammalian X-linked sperm proteins evolve faster than

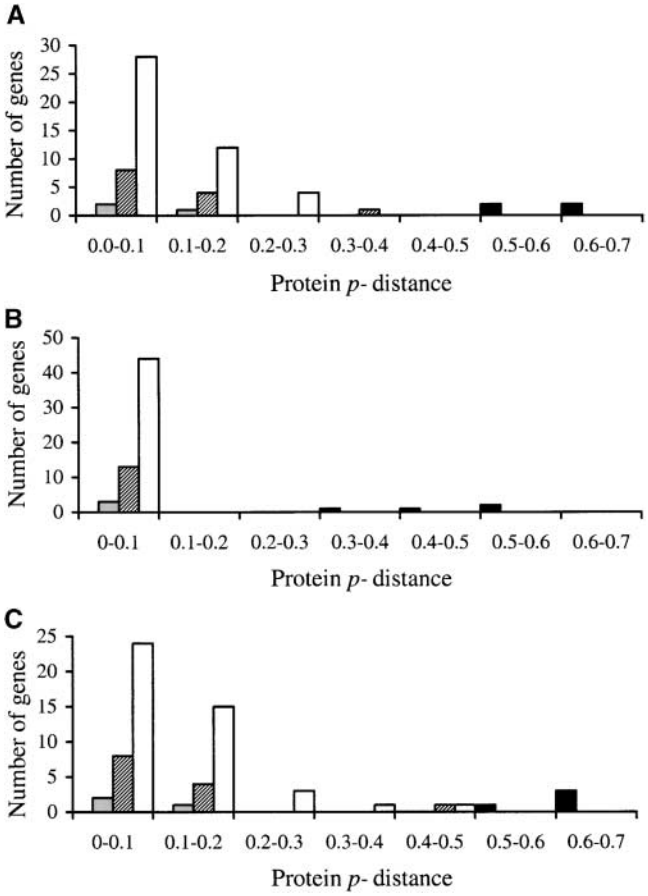

Figure 5.—Distribution of the evolutionary rate of 64 mammalian homeobox genes. The evolutionary rate is measured by protein *p*-distance between the human and mouse orthologous genes for (A) the entire sequence, (B) the homeodomain, and (C) nonhomeodomain regions. Solid bars, X-linked testis-expressed genes; shaded bars, X-linked non-testis expressed; hatched bars, autosomal testis expressed; open bars, autosomal non-testis expressed.

autosomal ones. Our finding of rapid evolution of mammalian X-linked testis-expressed homeobox genes is thus consistent with these previous observations.

Wang *et al.* (2001) reported that the mammalian X chromosome harbors disproportionately more spermatogonia-expressed genes than autosomes. Spermatogonia are the mitotic germ cells of the testis from which sperm arise by spermatogenesis. Spermatogonia-expressed genes are probably involved in male reproduction. In our random sample of 64 homeobox genes, 57% of the 7 X-linked genes and 23% of the 57 autosomal genes are testis expressed. Thus, even for homeobox genes, the X chromosome appears to harbor a higher proportion of testis-expressed genes than autosomes (P = 0.074). If only those genes that are exclusively (or highly selectively) expressed in testis are considered, the X chromosome harbors an even higher percentage of such genes (3/7 = 43%) than autosomes (1/57 = 2%), and their difference is significant (P = 0.003). Sex-chromosome meiotic drive and/or sexual antagonism have been invoked as possible explanations for a higher proportion of X-linked genes to function in male reproduction, and these hypotheses have been discussed extensively in Wang *et al.* (2001).

It has also been proposed that X-linked genes evolve more rapidly than autosomal genes (Charlesworth *et al.* 1987). This is particularly so when the X-linked genes are expressed only in males, because all newly arising advantageous alleles, dominant or recessive, are exposed to positive Darwinian selection. In contrast, recessive advantageous alleles at autosomal loci are effectively neutral when the allele frequencies are very low. This might explain the effectiveness of positive selection on X-linked testis-expressed genes.

The X chromosome has been shown to be of special importance in hybrid sterility between closely related species (reviewed in Coyne 1992). The importance of homeobox genes in hybrid sterility, however, is not well recognized, probably because most homeobox genes are evolutionarily conserved. It was thus a surprise to identify the rapidly evolving *OdsH*, an X-linked testis-expressed homeobox gene that is in part responsible for the hybrid male sterility between *D. simulans* and *D. mauritiana* (Ting *et al.* 1998). This study showed that it is a general pattern for mammalian X-linked testis-expressed homeobox genes to evolve rapidly. This suggests the intriguing possibility that it is a rule rather than an exception that homeobox genes such as *OdsH* play important roles in reproductive isolation. In the future, it will be of great interest to work out the developmental pathways in which these homeobox genes function and the biological significance of their rapid pace of evolution.

We thank Lizhi Gao, Ondrej Podlaha, David Webb, and two anonymous reviewers for valuable comments. This work was supported by a startup fund from the University of Michigan and a National Institutes of Health grant (GM67030) to J.Z.

LITERATURE CITED

Banerjee-Basu, S., and A. D. Baxevanis, 2001 Molecular evolution of the homeodomain family of transcription factors. Nucleic Acids Res. **29:** 3258–3269.

Bertolino, E., B. Reimund, D. Wildt-Perinic and R. G. Clerc, 1995 A novel homeobox protein which recognizes a TGT core and functionally interferes with a retinoid-responsive motif. J. Biol. Chem. **270:** 31178–31188.

Bharathan, G., B. J. Janssen, E. A. Kellogg and N. Sinha, 1997 Did homeodomain proteins duplicate before the origin of angiosperms, fungi, and metazoa? Proc. Natl. Acad. Sci. USA **94:** 13749–13753.

Blanco-Arias, P., C. A. Sargent and N. A. Affara, 2002 The human-specific Yp11.2/Xq21.3 homology block encodes a potentially functional testis-specific TGIF-like retroposon. Mamm. Genome **13:** 463–468.

Burglin, T. R., 1997 Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. Nucleic Acids Res. **25:** 4173–4180.

Carroll, S. B., J. K. Grenier and S. D. Weatherbee, 2001 *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design.* Blackwell Scientific, Malden, MA.

Charlesworth, B., J. A. Coyne and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130:** 113–146.

Coyne, J. A., 1992 Genetics and speciation. Nature **355:** 511–515.

Dagan, T., Y. Talmor and D. Graur, 2002 Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection. Mol. Biol. Evol. **19:** 1022–1025.

Garcia-Fernandez, J., and P. W. Holland, 1994 Archetypal organization of the amphioxus Hox gene cluster. Nature **370:** 563–596.

Gehring, W. J., M. Affolter and T. Burglin, 1994a Homeodomain proteins. Annu. Rev. Biochem. **63:** 487–526.

Gehring, W. J., Y. Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A. F. Schier *et al.*, 1994b Homeodomain-DNA recognition. Cell **78:** 211–223.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider *et al.*, 1998 Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. **9:** 585–598.

Kappen, C., 2000 The homeodomain: an ancient evolutionary motif in animals and plants. Comput. Chem. **24:** 95–103.

Kappen, C., K. Schughart and F. H. Ruddle, 1993 Early evolutionary origin of major homeodomain sequence classes. Genomics **18:** 54–70.

Kumar, S., K. Tamura, I. Jakobsen and M. Nei, 2001 MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:** 1244–1245.

Lai, Y. L., H. Li, H. S. Chiang and H. M. Hsieh-Li, 2002 Expression of a novel TGIF subclass homeobox gene, Tex1, in the spermatids of mouse testis during spermatogenesis. Mech. Dev. **113:** 185–187.

Lee, Y. H., T. Ota and V. D. Vacquier, 1995 Positive selection is a general phenomenon in the evolution of abalone sperm lysin. Mol. Biol. Evol. **12:** 231–238.

Li, W. H., 1997 *Molecular Evolution.* Sinauer Associates, Sunderland, MA.

Lifschytz, E., and D. L. Lindsley, 1972 The role of X-chromosome inactivation during spermatogenesis. Proc. Natl. Acad. Sci. USA **69:** 182–186.

Long, M., 2001 Evolution of novel genes. Curr. Opin. Genet. Dev. **11:** 673–680.

Makalowski, W., and M. S. Boguski, 1998 Evolutionary parameters of the transcribed mammalian genome: an analysis of 2,820 orthologous rodent and human sequences. Proc. Natl. Acad. Sci. USA **95:** 9407–9412.

McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa and W. J. Gehring, 1984 A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. Cell **37:** 403–408.

Melhuish, T. A., and D. Wotton, 2000 The interaction of the carboxyl terminus-binding protein with the Smad corepressor TGIF is disrupted by a holoprosencephaly mutation in TGIF. J. Biol. Chem. **275:** 39762–39766.

Melhuish, T. A., C. M. Gallo and D. Wotton, 2001 TGIF2 interacts with histone deacetylase 1 and represses transcription. J. Biol. Chem. **276:** 32109–32114.

Metz, E. C., and S. R. Palumbi, 1996 Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. Mol. Biol. Evol. **13:** 397–406.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Nei, M., and S. Kumar, 2000 *Molecular Evolution and Phylogenetics.* Oxford University Press, New York.

Page, S. L., and M. Goodman, 2001 Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol. Phylogenet. Evol. **18:** 14–25.

Podlaha, O., and J. Zhang, 2003 Positive selection on protein-length in the evolution of a primate sperm ion channel. Proc. Natl. Acad. Sci. USA **100:** 12241–12246.

Rooney, A. P., and J. Zhang, 1999 Rapid evolution of a primate sperm protein: Relaxation of functional constraint or positive Darwinian selection? Mol. Biol. Evol. **16:** 706–710.

Scott, M. P., and A. J. Weiner, 1984 Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. Proc. Natl. Acad. Sci. USA **81:** 4115–4119.

Shepherd, J. C., W. McGinnis, A. E. Carrasco, E. M. De Robertis and W. J. Gehring, 1984 Fly and frog homeodomains show homologies with yeast mating type regulatory proteins. Nature **310:** 70–71.

Singer, S. S., J. Schmitz, C. Schwiegk and H. Zischler, 2003 Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol. Phylogenet. Evol. **26:** 490–501.

Smith, N. G., 2003 Are radical and conservative substitution rates useful statistics in molecular evolution? J. Mol. Evol. **57:** 467–478.

Steiper, M. E., and M. Ruvolo, 2003 New World monkey phylogeny based on X-linked G6PD DNA sequences. Mol. Phylogenet. Evol. **27:** 121–130.

Sutton, K. A., and M. F. Wilkinson, 1997 Rapid evolution of a homeodomain: evidence for positive selection. J. Mol. Evol. **45:** 579–588.

Suzuki, Y., and T. Gojobori, 1999 A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **16:** 1315–1328.

Suzuki, Y., and M. Nei, 2002 Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **19:** 1865–1869.

Swanson, W. J., and V. D. Vacquier, 1995 Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. Proc. Natl. Acad. Sci. USA **92:** 4957–4961.

Swanson, W. J., and V. D. Vacquier, 2002 The rapid evolution of reproductive proteins. Nat. Rev. Genet. **3:** 137–144.

Ting, C., S. C. Tsaur, M. L. Wu and C.-I Wu, 1998 A rapidly evolving homeobox at the site of a hybrid sterility gene. Science **282:** 1501–1504.

Torgerson, D. G., and R. S. Singh, 2003 Sex-linked mammalian sperm proteins evolve faster than autosomal ones. Mol. Biol. Evol. **20:** 1705–1709.

Tsaur, S. C., and C.-I Wu, 1997 Positive selection and the molecular evolution of a gene of male reproduction, Acp26Aa of Drosophila. Mol. Biol. Evol. **14:** 544–549.

Wang, P. J., J. R. McCarrey, F. Yang and D. C. Page, 2001 An abundance of X-linked genes expressed in spermatogonia. Nat. Genet. **27:** 422–426.

Wayne, C. M., J. A. MacLean, G. Cornwall and M. F. Wilkinson, 2002 Two novel human X-linked homeobox genes, hPEPP1 and hPEPP2, selectively expressed in the testis. Gene **301:** 1–11.

Wyckoff, G. J., W. Wang and C.-I Wu, 2000 Rapid evolution of male reproductive genes in the descent of man. Nature **403:** 304–309.

Xia, X., and Z. Xie, 2001 DAMBE: software package for data analysis in molecular biology and evolution. J. Hered. **92:** 371–373.

Yang, Z., R. Nielsen, N. Goldman and A. M. Pedersen, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:** 431–449.

Zhang, J., 2000 Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. J. Mol. Evol. **50:** 56–68.

Zhang, J., 2003 Evolution by gene duplication: an update. Trends Ecol. Evol. **18:** 292–298.

Zhang, J., and M. Nei, 1996 Evolution of Antennapedia-class homeobox genes. Genetics **142:** 295–303.

Zhang, J., and M. Nei, 1997 Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. **44** (Suppl 1): S139–S146.

Zhang, J., and H. F. Rosenberg, 2002 Diversifying selection of the tumor-growth promoter angiogenin in primate evolution. Mol. Biol. Evol. **19:** 438–445.

Zhang, J., S. Kumar and M. Nei, 1997 Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol. Biol. Evol. **14:** 1335–1338.

Zhang, J., H. F. Rosenberg and M. Nei, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc. Natl. Acad. Sci. USA **95:** 3708–3713.

Communicating editor: Y. X. Fu