



Four Evolutionary Strata on the Human X Chromosome

Bruce T. Lahn and David C. Page

Science **286**, 964 (1999);

DOI: 10.1126/science.286.5441.964

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

The following resources related to this article are available online at www.sciencemag.org (this information is current as of May 10, 2012):

A correction has been published for this article at:

<http://www.sciencemag.org/content/286/5448/2269.6.full.html>

Updated information and services, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/286/5441/964.full.html>

This article **cites 24 articles**, 9 of which can be accessed free:

<http://www.sciencemag.org/content/286/5441/964.full.html#ref-list-1>

This article has been **cited by** 281 article(s) on the ISI Web of Science

This article has been **cited by** 100 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/286/5441/964.full.html#related-urls>

- mann, J. R. Ecker, *Cell* **72**, 427 (1993)] (23). The largest of 16 independent *NPH3* cDNAs was sequenced (24) completely (GenBank accession number AF180390).
11. GenBank searches were accomplished with the gapped BLAST program [S. F. Altschul et al., *Nucleic Acid Res.* **25**, 3389 (1997)].
 12. The data are available at www.sciencemag.org/feature/data/1042358.shl.
 13. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
 14. T. Patschinsky, T. Hunter, F. S. Esch, J. A. Cooper, B. M. Sefton, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 973 (1982).
 15. The BTB/POZ domain was identified with SMART [J. Schultz, F. Milpetz, P. Bork, C. P. Ponting, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857 (1998)]. The coiled-coil structure was identified with COILS [A. Lupas, M. Van Dyke, J. Stock, *Science* **252**, 1162 (1991)].
 16. O. Albagli, P. Dhordain, C. DeWeindt, G. LeCocq, D. LePince, *Cell Growth Differ.* **6**, 1193 (1995); L. Aravind and E. V. Koonin, *J. Mol. Biol.* **285**, 1353 (1999).
 17. C. Cohen and D. A. D. Parry, *Proteins* **7**, 1 (1990); A. Lupas, *Trends Biochem. Sci.* **21**, 375 (1996).
 18. Structural analyses were performed with the Protean program (DNASTAR, Madison, WI).
 19. S. Fields, *Methods* **5**, 116 (1993); S. Fields and R. Sternglanz, *Trends Genet.* **10**, 286 (1994).
 20. M. Nagao and K. Tanaka, *J. Biol. Chem.* **267**, 17925 (1992).
 21. M. C. Faux and J. D. Scott, *Cell* **85**, 9 (1996); T. Pawson and J. D. Scott, *Science* **278**, 2075 (1997); E. A. Elion, *Science* **281**, 1625 (1998).
 22. S. D. Choi, R. Creelman, J. Mullet, R. A. Wing, *Weeds World* **2**, 17 (1995), <http://genome-www.stanford.edu/Arabidopsis/www/home.html>.
 23. J. Sambrook, E. F. Fritsch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Plainview, NY, 1989).
 24. Sequencing templates were prepared by polymerase chain reaction and sequenced with an ABI377 automated sequencer (Perkin-Elmer, Norwalk, CT).
 25. Phototropism and hypocotyl growth was assayed as described previously [E. L. Stowe-Evans, R. M. Harper, A. V. Motchoulski, E. Liscum, *Plant Physiol.* **118**, 1265 (1998)].
 26. E. Liscum and R. P. Hangarter, *Plant Cell* **3**, 685 (1991).
 27. J. W. Reed, P. Nagpal, D. S. Poole, M. Furuya, J. Chory, *Plant Cell* **5**, 147 (1993).
 28. C. Bell and J. R. Ecker, *Genomics* **19**, 137 (1994).
 29. H.-G. Nam et al., *Plant Cell* **1**, 699 (1989).
 30. Information about markers AM40 and AM80 is available at <http://www.kazusa.or.jp/arabi/chr5/map/24-26Mb.html>.
 31. Y. Nakamura et al., *DNA Res.* **4**, 401 (1997); <http://www.kazusa.or.jp/arabi/chr5/map/24-26Mb.html>.
 32. Soluble and total microsomal membrane fractions

- were separated by ultracentrifugation, followed by two-phase partitioning to enrich for plasma membranes, as described previously [T. W. Short, P. Raymond, W. R. Briggs, *Plant Physiol.* **101**, 647 (1993)].
33. Antibodies against NPH1 were previously described (7). Rabbit polyclonal antisera were raised (22) against a COOH-terminal NPH3 fusion protein [CBD-NPH3C2 (see Fig. 3A)]. CBD-NPH3 protein was expressed from pET34-Ek/LIC in *Escherichia coli* and purified according to manufacturer's instructions (Novagen, Madison, WI).
 34. NPH1-NPH3 interaction was examined in yeast with the Matchmaker Gal4 II System (Clontech, Palo Alto, CA). Expression of fusion peptides was verified by immunoblot analysis (9, 22) with monoclonal antibodies raised against the Gal4 DNA binding domain (GBD) and Gal4 activation domain (GAD) (Clontech).
 35. J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Plainview, NY, 1972).
 36. We thank R. Harper for data in Fig. 1; J. M. Christie and W. R. Briggs for GBD-NPH1 constructs and NPH1 antisera; D. Randall for production of NPH3 antisera; the Arabidopsis Biological Resource Center in Columbus, Ohio, for BAC clones and cDNA libraries; and members of our laboratory for helpful comments on the manuscript. This work was funded by USDA National Research Initiative grant 96-35304-3709, NSF grant MCB-9723124, and University of Missouri Research Board grant RB96-055.

3 June 1999; accepted 17 September 1999

Four Evolutionary Strata on the Human X Chromosome

Bruce T. Lahn* and David C. Page†

Human sex chromosomes evolved from autosomes. Nineteen ancestral autosomal genes persist as differentiated homologs on the X and Y chromosomes. The ages of individual X-Y gene pairs (measured by nucleotide divergence) and the locations of their X members on the X chromosome were found to be highly correlated. Age decreased in stepwise fashion from the distal long arm to the distal short arm in at least four "evolutionary strata." Human sex chromosome evolution was probably punctuated by at least four events, each suppressing X-Y recombination in one stratum, without disturbing gene order on the X chromosome. The first event, which marked the beginnings of X-Y differentiation, occurred about 240 to 320 million years ago, shortly after divergence of the mammalian and avian lineages.

The human X and Y chromosomes, like those of other animals, are thought to have evolved from an ordinary pair of autosomes (1). The pseudoautosomal regions at the termini of the X and Y chromosomes still recombine during male meiosis, ensuring X-Y nucleotide sequence identity there. Elsewhere on the X and Y chromosomes, however, X-Y recombination has been suppressed. These nonrecombining regions of the X and Y chromosomes have become highly differentiated during evolution, and only a few X-Y sequence similarities per-

sist within them. These modern X-Y gene pairs are the remaining "fossils" where extensive sequence identity between ancestral X and Y chromosomes once existed. The recent discovery of many X-Y genes has made it possible to examine the entire group to search for patterns of human sex chromosome evolution. Thus far, the human sex chromosomes—the best characterized mammalian sex chromosomes—have been found to contain 19 X-Y gene pairs (2).

We first compared the locations of all 19 pairs of genes on the human X and Y chromosomes (Fig. 1). We determined the relative positions of the X-linked genes through radiation hybrid analysis, in many cases confirming previously published localizations (3). Map positions of the Y-linked homologs were obtained principally from the literature (4–6). On the X chromosome, most of the X-Y genes map to the short arm, where they are concentrated toward the distal end. By contrast, the X-Y genes are

found as singletons or small clusters throughout the euchromatic portion of the Y chromosome. In general, the map order of the X-linked genes corresponds poorly to that of the Y-linked homologs. Local exceptions to this rule are provided by three small gene clusters that are present on both X and Y chromosomes (Fig. 1).

We next measured, for each of the 19 X-Y gene pairs, synonymous nucleotide divergence between the X-linked and Y-linked coding regions (7). Because synonymous substitutions do not alter the encoded protein, they are generally assumed to be nearly neutral with respect to selection. The statistic K_S (the estimated mean number of synonymous substitutions per synonymous site) is often used to gauge evolutionary time (8). In the present context, K_S values provide a measure of the evolutionary time that has elapsed since the gene pairs started differentiating into distinct X and Y forms. The calculated K_S values are given in Table 1, where gene pairs are listed according to map order on the X chromosome.

We noted that the 19 K_S values appeared to cluster into approximately four groups (Fig. 2): 0.94 to 1.25 (group 1), 0.52 to 0.58 (group 2), 0.23 to 0.36 (group 3), and 0.05 to 0.12 (group 4). Each X-Y gene pair's K_S value differed significantly from those of all gene pairs in other groups ($P \leq 0.02$). The most striking observation was that, on the X chromosome, the four K_S -defined groups of genes are arranged in an orderly sequence (Fig. 2). X-Y genes are stratified by age along the length of the X chromosome. By contrast, on the Y chromosome, the K_S -defined groups appear to be scrambled (compare Table 1 and Fig. 1).

What might account for the orderly stratification of X-Y genes by age on the human X chromosome? We hypothesize that, during evo-

Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, MA 02142, USA.

*Present address: Department of Human Genetics, University of Chicago, 924 East 57th Street, Chicago, IL 60637, USA.

†To whom correspondence should be addressed. E-mail: dcpage@wi.mit.edu

lution, differentiation of the X from the Y chromosome was initiated one region, or stratum, at a time. Regions were recruited in the order of their physical position, with stratum 1 (containing the genes of group 1) having been the first to embark on X-Y differentiation, and stratum 4 having been the most recent. Genes in the same stratum began differentiating into X and Y homologs at about the same time, accounting for their similar K_S values.

X-Y differentiation would have occurred only after X-Y recombination ceased (9). Our findings suggest that during evolution, X-Y recombination was suppressed regionally, beginning with stratum 1 and subsequently expanding in discrete steps to include strata 2, 3, and 4. Chromosomal inversions, which are known to be capable of suppressing recombination across broad regions in mammals (10), would appear to be the most likely mechanism. These inversions must have occurred on the evolving Y chromosome, where the strata have been scrambled, but not on the X chromosome, where the order of strata apparently has been preserved (Figs. 1 and 2). [Had the strata on the human X chromosome been extensively shuffled during evolution—as may have occurred on the mouse X chromosome after divergence of the human and murine lineages (11)—we would have observed no correlation between the age of X-Y gene pairs and the map positions of their X-chromosomal members.] In the modern human sex chromosomes, the proximal boundary of the pseudoautosomal region is spanned by a gene that is intact on the X chromosome, but grossly interrupted on the Y chromosome (12), consistent with disruption of

an ancient pseudoautosomal region by a Y-chromosomal inversion. We speculate that this particular event was the most recent in a series of inversions, each of which enabled X-Y differentiation to begin in one stratum.

This model of staged, region-by-region initiation of X-Y differentiation also accounts for two global features of the X chromosome's gene content: (i) the concentration in strata 3 and 4 of genes with detectable Y homologs (Fig. 1) and (ii) the concentration on the short arm (strata 2, 3, and 4) of genes that escape X inactivation, some with and some without Y homologs (13). Evolutionary theory predicts that once X-Y recombination ceased within a stratum, the genes on the affected portion of the Y chromosome began to decay, with most of the Y-linked genes ultimately being obliterated (1). As an adaptive response, homologous genes on the X chromosome were up-regulated, and subsequently became subject to X inactivation, processes thought to have spread during evolution on a gene-by-gene or cluster-by-cluster basis (14). If decay of Y-linked genes and adaptation of X-linked homologs were gradual evolutionary processes, then one would expect the youngest X strata to exhibit the highest densities of (i) genes with detectable Y homologs and (ii) genes that escape inactivation. Both predictions are met (Fig. 1) (13).

A comparison of the youngest (group 4) gene pairs with the older (groups 1 through 3) gene pairs illustrates certain temporal features of X-Y differentiation. We measured both synonymous and nonsynonymous substitutions for each gene pair (Table 1). Nonsynonymous sub-

stitutions alter the encoded protein and are constrained by selection. Thus, their frequency (K_A , the estimated mean number of nonsynonymous substitutions per nonsynonymous site) is a function of both evolutionary time and selective constraints on the encoded proteins. The degree

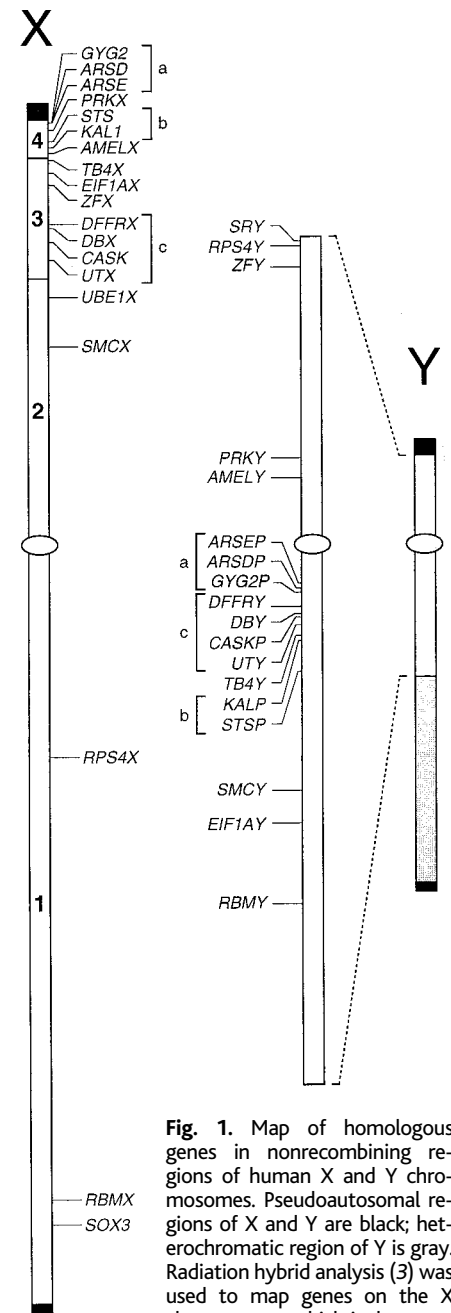


Fig. 1. Map of homologous genes in nonrecombining regions of human X and Y chromosomes. Pseudoautosomal regions of X and Y are black; heterochromatic region of Y is gray. Radiation hybrid analysis (3) was used to map genes on the X chromosome, which is drawn on a centiRay scale. K_S -defined strata on the X chromosome are indicated. The boundary between strata 2 and 1 is somewhere between *SMCX* and *RPS4X*; here, it is arbitrarily shown at the centromere (white oval). Genes and pseudogenes on the Y chromosome were ordered previously by analysis of naturally occurring deletions (4, 5). *UBE1X* has a homolog on the squirrel monkey Y chromosome but not on the human Y chromosome (29). Brackets denote three small gene clusters (labeled a, b, c) that are present on both X and Y chromosomes.

Table 1. Sequence divergence between homologous X- and Y-linked genes.

Gene pair	K_S	K_A	K_S/K_A	DNA divergence (%)	Protein divergence (%)	Sequence compared (nucleotides)
Group 4						
<i>GYG2/GYG2P*</i>	0.11	0.06	1.8	7	12	525
<i>ARSD/ARSDP*</i>	0.09	0.07	1.3	7	13	846
<i>ARSE/ARSEP*</i>	0.05	0.04	1.2	4	9	615
<i>PRKX/Y</i>	0.07	0.03	2.3	5	8	1020
<i>STS/STSP*</i>	0.12	0.10	1.2	11	18	852
<i>KAL1/KALP*</i>	0.07	0.06	1.2	6	12	1302
<i>AMELX/Y</i>	0.07	0.07	1.0	7	12	576
Group 3						
<i>TB4X/Y</i>	0.29	0.04	7.3	7	7	135
<i>EIF1AX/Y</i>	0.32	0.01	32	9	2	432
<i>ZFX/Y</i>	0.23	0.04	5.8	7	7	2394
<i>DFFRX/Y</i>	0.33	0.05	6.6	11	9	7671
<i>DBX/Y</i>	0.36	0.04	9.0	12	9	1932
<i>CASK/CASKP*</i>	0.24	0.22	1.1	15	32	156
<i>UTX/Y</i>	0.26	0.08	3.3	12	15	4068
Group 2						
<i>UBE1X/Y</i>	0.58	0.07	8.3	16	13	693
<i>SMCX/Y</i>	0.52	0.08	6.5	17	15	4623
Group 1						
<i>RPS4X/Y</i>	0.97	0.05	19	18	18	792
<i>RBMX/Y</i>	0.94	0.25	3.8	29	38	1188
<i>SOX3/SRY</i>	1.25	0.19	6.6	28	29	264

*Y copy is pseudogene. DNA and protein divergence refer to uncorrected nucleotide (coding region) and amino acid divergence (nonidentity).

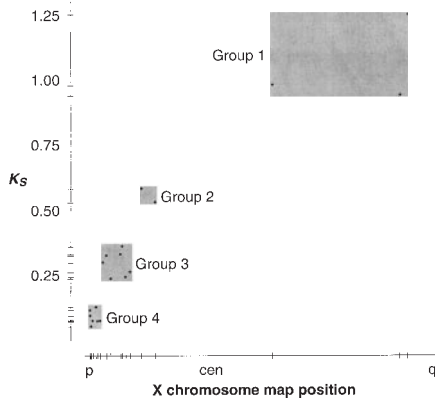


Fig. 2. Plot of K_S (Table 1) versus X-chromosome map position (Fig. 1) for 19 X-Y gene pairs.

of constraint can be reflected in the ratio K_S/K_A ; values greater than one indicate the presence of constraints on both homologs, and values in the vicinity of one are consistent with lack of constraint on at least one homolog (8, 15). In groups 1 through 3, 10 of 11 gene pairs exhibit K_S/K_A ratios of 3 or higher (Table 1), suggesting that natural selection has preserved the Y copies of these genes. Without such selection, these X-Y homologies (especially those in groups 1 and 2) would no longer be visible. By contrast, the seven gene pairs in group 4 show K_S/K_A ratios of 1 to 2, and in five of these pairs, the Y copy is known to be a pseudogene. Among the group 4 pairs, X-Y homology is readily apparent even in the absence of selective constraint, because there has been little time for erosion of sequence similarity. Thus, the Y-chromosomal genes of the older groups, and especially those of groups

1 and 2, are survivors of an early winnowing process that is still ongoing in group 4.

To determine the age of the K_S -defined strata, we used two methods. First, we considered published information on homologs of representative genes in diverse mammals. The maximum age of stratum 4, for example, was suggested by the prior observation that homologs of *STS* and *KAL1* are pseudoautosomal or autosomal in prosimians (16–18). Assuming that suppression of X-Y recombination is an irreversible evolutionary step (14), this implies that X-Y differentiation in stratum 4 began less than 50 million years ago (Ma), when the simian and prosimian lineages diverged (19). Minimum ages of the strata could also be inferred. For example, *STS* and *KAL1* have been shown to have X- and Y-specific homologs in both New and Old World monkeys (16, 17), suggesting that X-Y differentiation in stratum 4 began at least 30 Ma, when the New and Old World monkey lineages diverged (19, 20). Using similar logic, we inferred the ages of stratum 3 (80 to 130 million years), stratum 2 (130 to 170 million years), and stratum 1 (130 to 350 million years) from prior data on gene homologs in more-distantly related species, including nonprimate mammals, marsupials, monotremes, and birds (21).

These cross-species comparisons yielded reasonably precise estimates of age for strata 2, 3, and 4—the younger strata—but only crude estimates of age for stratum 1. Because this oldest stratum might contain information about the origins of mammalian sex chromosomes, its age is of great interest. Here, we used a second

dating method, based on K_S values for X-Y gene pairs. Theory predicts that among human X-Y gene pairs, K_S values should be roughly proportional to age (8). This expectation is met by the X-Y gene pairs of strata 2, 3, and 4 (Fig. 3). By extrapolation, we estimated that X-Y differentiation began 240 to 320 Ma in stratum 1 (Fig. 3). These findings suggest that X-Y divergence began shortly after the mammalian lineage arose, having diverged from the lineage of birds (with Z-W sex chromosomes) between 300 and 350 Ma (19). [Because the sex chromosomes of birds appear to be completely unrelated to the mammalian sex chromosomes, it is thought that they arose independently, from a different autosomal pair (22).] Interestingly, our K_S findings indicate that *SOX3* and *SRY* (the primary sex-determining gene) are among the oldest known X-Y gene pairs in humans (Table 1). This finding strengthens an hypothesis, by Foster and Graves, which states that an ordinary autosomal pair became sex chromosomes when mutations fashioned one allele of *SOX3*, originally an autosomal gene, into the male-determining factor *SRY* (23). Indeed, formal cluster analysis of the K_S values we report suggests that the X-Y genes of group 1 might actually comprise two distinct strata, with *SRY/SOX3* perhaps being older than the two other X-Y gene pairs of group 1 (*RPS4X/Y* and *RBMY/Y*) (24). Although the difference in K_S values between *SRY/SOX3* and the two other X-Y gene pairs is not statistically significant, the evidence is suggestive.

If future studies establish that the group 1 genes are divisible into two strata, these results

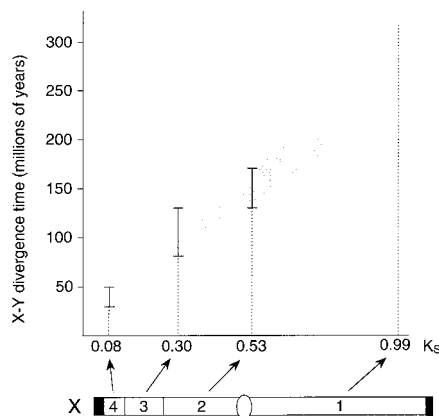


Fig. 3. Plot of X-Y divergence time (age) versus average K_S value for X-Y gene pairs (weight-averaged) in each stratum. The X chromosome schematic is adapted from Fig. 1. Maximum and minimum age estimates for strata 2, 3, and 4 are bracketed; these are not statistical confidence intervals. Theory predicts an approximately linear relationship between age and K_S value (8); the shaded area is calibrated with respect to stratum 2, whose age is 130 to 170 million years (21) and whose average K_S value is 0.53. By extrapolation, the age of stratum 1 is estimated between 240 and 320 million years.

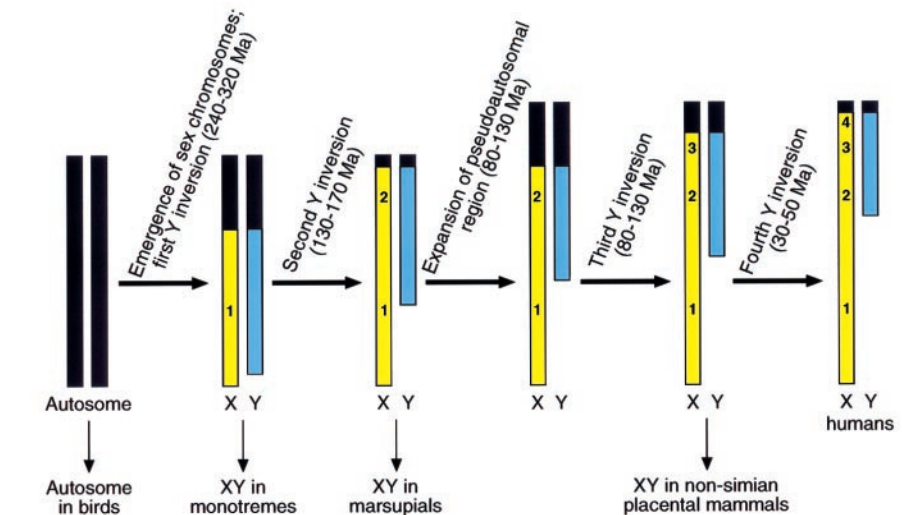


Fig. 4. A proposed sequence of evolutionary events that generated four strata on the human X chromosome. Four inversions on the Y chromosome are postulated. Each inversion reduced the size of the pseudoautosomal (X-Y recombining) region (black; for simplicity, only one pseudoautosomal region is shown for each chromosome) and enlarged the portions of the X (yellow) and Y (blue) chromosomes that did not recombine during male meiosis. Ongoing decay and loss of Y genes offset these periodic expansions of the nonrecombining region of the Y chromosome. Points of divergence from the sex chromosomes of other mammals are indicated. This model does not preclude the occurrence of (i) additional inversions or other rearrangements within the nonrecombining portion of the evolving Y chromosome or (ii) similar rearrangements on the evolving X chromosome, so long as they do not disturb the fundamental order among the four strata.

would also help date the emergence of X inactivation during mammalian sex chromosome evolution. *XIST*, an X-specific gene which plays a pivotal role in X inactivation (25), is located near *RPS4X* and therefore would be in the younger of the two strata—not in the stratum where the nascent X and Y chromosomes first differentiated. This would controvert the hypothesis of Chandra, who speculated that X inactivation emerged contemporaneously with the chromosomal sex-determining mechanism (26).

Consistent with our evolutionary map, Graves and colleagues have postulated that the long arm and proximal short arm of the human X chromosome are at least 170 million years old (27, 28). They have referred to this portion of the X as the “XCR” (X conserved region). Graves’s XCR corresponds approximately to our strata 1 and 2. They have also postulated that the distal short arm of the human X chromosome is younger. This “XAR” (X added region) was attributed to translocation of an autosome to the pseudoautosomal region of both X and Y after divergence of placental mammals from marsupials (27, 28). Our strata 3 and 4 are found within Graves’s XAR.

In conclusion, we postulate that the evolution of human sex chromosomes was punctuated by at least four events, plausibly a series of inversions on the Y chromosome (Fig. 4). Each event suppressed X-Y recombination in one stratum and enabled X-Y differentiation to proceed there. The first of these events, which created stratum 1, was roughly contemporaneous with the birth of the mammalian sex chromosomes and the emergence of *SRY* as the primary sex determinant. This occurred about 240 to 320 Ma, shortly after the mammalian and avian lineages diverged. The pseudoautosomal region was expanded by translocation of autosomal material between the second and third events (which created strata 2 and 3, respectively). The fourth event occurred relatively recently, during primate evolution, creating stratum 4, where X-Y differentiation is still in its earliest stages.

References and Notes

- J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, Menlo Park, CA, 1983); J. A. Graves, *Annu. Rev. Genet.* **30**, 233 (1996); B. Charlesworth, *Curr. Biol.* **6**, 149 (1996); W. R. Rice, *Bioscience* **46**, 331 (1996).
- The 19 X-Y gene pairs studied include the following: *GYG2/GYG2P* [J. Mu, A. V. Skurat, P. J. Roach, *J. Biol. Chem.* **272**, 27589 (1997); (6)], *ARSD/ARSDP*, *ARSE/ARSEP* [G. Meroni et al., *Hum. Mol. Genet.* **5**, 423 (1996)], *PRKX/Y* [A. Klink et al., *Hum. Mol. Genet.* **4**, 869 (1995); K. Schiebel et al., *Hum. Mol. Genet.* **6**, 1985 (1997)], *STS/STSP* (16), *KAL1/KALP* [B. Franco et al., *Nature* **353**, 529 (1991); R. Legouis et al., *Cell* **67**, 423 (1991); (17)], *AMELX/Y* [Y. Nakahori, O. Takenaka, Y. Nakagome, *Genomics* **9**, 264 (1991)], *TB4X/Y* [H. Gondo et al., *J. Immunol.* **139**, 3840 (1987); (5)], *ZFX/Y* [D. C. Page et al., *Cell* **51**, 1091 (1987); A. Schneider-Gädick, P. Beer-Romero, L. G. Brown, R. Nussbaum, D. C. Page, *Cell* **57**, 1247 (1989)], *EIF1AX/Y* [T. E. Dever et al., *J. Biol. Chem.* **269**, 3212 (1994); (5)], *DFFRX/Y* [M. H. Jones et al., *Hum. Mol. Genet.* **5**, 1695 (1996); (5)], *DBX/Y* (5), *CASK/CASKP* [A. R. Cohen et al., *J. Cell Biol.* **142**, 129 (1998); (6)], *UTX/Y* (5), *SMCX/Y* [J. Wu et al., *Hum. Mol. Genet.* **3**, 153 (1994); A. I. Agulnik et al., *Hum. Mol. Genet.* **3**, 879 (1994)], *RPS4X/Y* [E. M. Fisher et al., *Cell* **63**, 1205 (1990)], *RBMX/Y* [M. Souillard et al., *Nucleic Acids Res.* **21**, 4210 (1993); K. Ma et al., *Cell* **75**, 1287 (1993); M. L. Delbridge, P. A. Lingenfelter, C. M. Distech, J. A. Graves, *Nature Genet.* **22**, 223 (1999); S. Mazeyrat, N. Saut, M. G. Mattei, M. J. Mitchell, *Nature Genet.* **22**, 224 (1999)], *SOX3/SRY* [M. Stevanovic, R. Lovell-Badge, J. Collignon, P. N. Goodfellow, *Hum. Mol. Genet.* **2**, 2013 (1993); A. H. Sinclair et al., *Nature* **346**, 240 (1990)]. One interspecies pair was also studied: human *UBE1X* [P. M. Handley, M. Mueckler, N. R. Siegel, A. Ciechanover, A. L. Schwartz, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 258 (1991)] and squirrel monkey *UBE1Y* (29). In humans, *UBE1Y* was deleted from the Y chromosome (29). We used squirrel monkey *UBE1Y* as a substitute.
- Using polymerase chain reaction (PCR), we tested DNAs from the 93 hybrid cell lines of the GeneBridge 4 panel (Research Genetics) [G. Gyapay et al., *Hum. Mol. Genet.* **5**, 339 (1996)] for the presence of each of the X-linked genes. PCR conditions and primer sequences have been deposited at GenBank, where accession numbers are as follows: *GYG2*, G49430; *ARSD*, G42687; *ARSE*, G42688; *PRKX*, G42689; *STS*, G42690; *KAL1*, G42691; *AMELX*, G42692; *TB4X*, G34979; *EIF1AX*, G34989; *ZFX*, G42693; *DFFRX*, G34982; *DBX*, G34988; *CASK*, G49441; *UTX*, G34976; *UBE1X*, G42694; *SMCX*, G42695; *RPS4X*, AF041428; *RBMX*, G42696; and *SOX3*, G42697. Analysis of the results positioned the genes with respect to the radiation hybrid map of the X chromosome constructed at the Whitehead/MIT Center for Genome Research [T. J. Hudson et al., *Science* **270**, 1945 (1995); www-genome.wi.mit.edu/cgi-bin/contig/phys_map].
- D. Vollrath et al., *Science* **258**, 52 (1992).
- B. T. Lahn and D. C. Page, *Science* **278**, 675 (1997).
- C. Sun et al., *Nature Genet.*, in press.
- Homologous X and Y DNA sequences were aligned by means of MegAlign software (DNASTAR, Madison, WI). For each X-Y gene pair, estimates of the mean numbers of synonymous substitutions per synonymous site (K_s), and of nonsynonymous substitutions per nonsynonymous site (K_a)—all corrected for multiple changes—were calculated using published algorithms (8) as implemented in GCG software (Genetics Computer Group, Madison, WI). Insertions and deletions were ignored in these calculations. In the case of *SOX3* and *SRY*, sequence similarity is limited to, and our analysis was restricted to, the HMG box domain. Our analyses of other X-Y gene pairs employed all available coding sequences. Only a partial *UBE1Y* (squirrel monkey) coding sequence was available for comparison with its human X homolog. Sequences for all pseudogenes were extracted from genomic sequences: *GYG2P*, *ARSDP*, and *ARSEP* from BAC (bacterial artificial chromosome) clone 203M13 (GenBank AC002992); *STSP* from BAC clone NH0494J04 (GenBank AC006382); *KALP* from BAC clone NH0292P09 (GenBank AC006370); *CASKP* from BAC clone 47511 (GenBank AC004474). Sequences for all other genes were obtained from published cDNAs, whose GenBank accession numbers are as follows: *GYG2*, U94362; *ARSD*, X83572; *ARSE*, X83573; *PRKX*, X85545; *PRKY*, Y15801; *STS*, M16505; *KAL1*, M97252; *AMELX*, M86932; *AMELY*, M86933; *TB4X*, M17733; *TB4Y*, AF000989; *ZFX*, X59739; *ZFY*, M30607; *EIF1AX*, L18960; *EIF1AY*, AF000987; *DFFRX*, X98296; *DFFRY*, AF000986; *DBX*, AF000982; *DBY*, AF000984; *CASK*, AF032119; *UTX*, AF000992; *UTY*, AF000994; *UBE1X*, M58028; *UBE1Y*, AJ003105; *SMCX*, L25270; *SMCY*, U52191; *RPS4X*, M58458; *RPS4Y*, M58459; *RBMX*, Z23064; *RBMY*, X76059; *SOX3*, X71135; *SRY*, X53772.
- W. H. Li, *J. Mol. Evol.* **36**, 96 (1993); *Molecular Evolution* (Sinauer Associates, Sunderland, MA, 1997).
- B. O. Bengtsson and P. N. Goodfellow, *Ann. Hum. Genet.* **51**, 57 (1987).
- L. M. Silver, *Trends Genet.* **9**, 250 (1993); R. H. Martin et al., *Hum. Genet.* **93**, 135 (1994); M. Jaarola, R. H. Martin, T. Ashley, *Am. J. Hum. Genet.* **63**, 218 (1998).
- H. J. Blair, V. Reed, S. H. Laval, Y. Boyd, *Genomics* **19**, 215 (1994); W. J. Murphy, S. Sun, Z.-Q. Chen, J. Pecon-Slattey, S. J. O'Brien, *Genome Res.*, in press.
- P. A. Weller, R. Critcher, P. N. Goodfellow, J. German, N. A. Ellis, *Hum. Mol. Genet.* **4**, 859 (1995).
- C. J. Brown, L. Carrel, H. F. Willard, *Am. J. Hum. Genet.* **60**, 1333 (1997).
- K. Jegalian and D. C. Page, *Nature* **394**, 776 (1998).
- K_s/K_a ratios can be depressed by positive selection, which accelerates protein divergence (8). However, among the X-Y pairs shown here to have relatively low K_s/K_a ratios, the abundance of Y pseudogenes (Table 1) suggests that absence of selective constraint is the more significant factor.
- P. H. Yen et al., *Cell* **55**, 1123 (1988).
- I. del Castillo, M. Cohen-Salmon, S. Blanchard, G. Lutfalla, C. Petit, *Nature Genet.* **2**, 305 (1992); B. Incerti et al., *Nature Genet.* **2**, 311 (1992).
- R. Toder, G. A. Rappold, K. Schiebel, W. Schempp, *Hum. Genet.* **95**, 22 (1995).
- S. Kumar and S. B. Hedges, *Nature* **392**, 917 (1998); M. J. Benton, *Vertebrate Paleontology* (Chapman & Hall, New York, 1997).
- D. Pilbeam, *Sci. Am.* **250** (no. 3), 84 (1984).
- Stratum 3: Homologs of *ZFX* are autosomal in marsupials (27), which diverged from placental mammals 130 Ma (19). For *ZFX/Y* and *UTX/Y* (and for *UBE1X/Y* and *SMCX/Y*, in stratum 2), we employed sequence-based phylogenetic analysis to determine if differentiation into X and Y forms had begun before or after mouse/human divergence. For each X-Y gene pair, we used GCG software to construct a phylogenetic tree relating human X, human (or monkey) Y, mouse X, and mouse Y homologs. In each of the four cases, the X homologs in human and mouse formed a branch which was distinct from a second branch formed by the Y homologs in human (or monkey) and mouse. These findings suggest that X-Y differentiation of these four gene pairs began before divergence of humans and mice. This is consistent with X-Y divergence having initiated before the placental mammalian radiation that occurred 80 to 100 Ma. Stratum 2: Distinct X- and Y-linked forms of *UBE1* have been found in both placental mammals and marsupials [M. J. Mitchell, D. R. Woods, S. A. Wilcox, J. A. Graves, C. E. Bishop, *Nature* **359**, 528 (1992)], but their homologs are autosomal in monotremes (29), which diverged from placental mammals and marsupials 170 Ma (19). Stratum 1: Distinct X- and Y-linked forms of *RPS4* have been found in placental mammals and marsupials (K. Jegalian and D. C. Page, unpublished results), but their homologs are autosomal in birds, whose lineage diverged from that of mammals 300 to 350 Ma (19). Y-specific *SRY* sequences have been identified in both placental mammals and marsupials [J. W. Foster et al., *Nature* **359**, 531 (1992)].
- A. K. Fridolfsson et al., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8147 (1998).
- J. W. Foster and J. A. Graves, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1927 (1994).
- Dendrograms of the 19 K_s values (Table 1) were constructed using five clustering algorithms (average, centroid, Ward’s, single linkage, and complete linkage) implemented in JMP statistics software (SAS Institute, Cary, NC). The most significant branch classification (using any of the algorithms) had five clusters corresponding to the four groups shown in Table 1 and Fig. 2, but with group 1 divided into subgroups 1A (*SOX3/SRY*) and 1B (*RPS4X/Y* and *RBMX/Y*). However, the difference in K_s value between *SOX3/SRY* (1.25 ± 0.41) and *RPS4X/Y* (0.97 ± 0.16) or *RBMX/Y* (0.94 ± 0.15) was not statistically significant. At present, any distinction between subgroups 1A and 1B is tentative.
- H. F. Willard, *Cell* **86**, 5 (1996).
- H. S. Chandra, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 6947 (1985).
- J. A. Spencer, A. H. Sinclair, J. M. Watson, J. A. Graves, *Genomics* **11**, 339 (1991).
- J. A. Graves, *Philos. Trans. R. Soc. London Ser. B* **350**, 305 (1995); R. Toder and J. A. Graves, *Mamm. Genome* **9**, 373 (1998); J. M. Watson, J. A. Spencer, J. A. Graves, M. L. Snead, E. C. Lau, *Genomics* **14**, 785 (1992); J. A. Spencer, J. M. Watson, J. A. Graves, *Genomics* **9**, 598 (1991).
- M. J. Mitchell et al., *Hum. Mol. Genet.* **7**, 429 (1998).
- We thank F. Lewitter for help with sequence comparisons, H. Skaltsky and N. Kawaguchi for help with database searches and analysis of mapping data, and P. Bain, D. Bartel, A. Bortvin, B. Charlesworth, D. Charlesworth, A. Chess, A. Clark, C. Distech, G. Fink, S. Gilbert, J. Graves, R. Jaenisch, K. Jegalian, E. Lander, D. Menke, W. Rice, S. Rozen, C. Tilford, and J. Wang for discussions and comments on the manuscript. Supported by NIH grant HG00257.

14 June 1999; accepted 17 September 1999