

Positive and Negative Selection on Mammalian Y Chromosomes

Dave T. Gerrard and Dmitry A. Filatov

School of Biosciences, The University of Birmingham, Edgbaston, Birmingham, United Kingdom

Y chromosomes are genetically degenerate in most organisms studied. The loss of genes from Y chromosomes is thought to be due to the inefficiency of purifying selection in nonrecombining regions, which leads to the accumulation of deleterious mutations via the processes of hitchhiking, background selection, and Muller's ratchet. As the severity of these processes depends on the number of functional genes linked together on the nonrecombining Y, it is not clear whether these processes are still at work on the old, gene-poor mammalian Y chromosomes. If purifying selection is indeed less efficient in the Y-linked, compared to the X-linked genes, deleterious nonsynonymous substitutions are expected to accumulate faster on the Y chromosome. However, positive selection on Y-linked genes could also increase the rate of amino acid-changing substitutions. Thus, the previous reports of an elevated nonsynonymous substitution rate in Y-linked genes are still open to interpretation. Here, we report evidence for positive selection in two out of three studied mammalian Y-linked genes, suggesting that adaptive Darwinian evolution may be common on mammalian Y chromosomes. Taking positive selection into account, we demonstrate that purifying selection is less efficient in mammalian Y-linked genes compared to their X-linked homologues, suggesting that these genes continue to degenerate.

Introduction

Old Y chromosomes, such as those in mammals, birds, and in *Drosophila* are genetically degenerate (Bull 1983), containing few functional genes (e.g., Skaletsky et al. 2003). It is thought that Y-linked genes accumulate deleterious mutations due to the reduced efficacy of purifying selection in the nonrecombining regions (reviewed in B. Charlesworth and D. Charlesworth 2000). Deleterious mutations may be carried to fixation by linked advantageous mutations ("selective sweeps") (Rice 1987). Additionally, the selective elimination of deleterious mutations, causing "background selection" (Charlesworth, Morgan, and Charlesworth 1993) could accelerate the stochastic fixation of mildly detrimental mutations (B. Charlesworth and D. Charlesworth 2000). Furthermore, selective sweeps and background selection reduce the effective population size (and therefore variability) of genes in evolving Y chromosomes, allowing the operation of "Muller's ratchet" (the stochastic loss of chromosomes with the fewest mutations) (B. Charlesworth and D. Charlesworth 2000; Gordo and Charlesworth 2000). Both reduced genetic diversity and the accumulation of deleterious mutations were indeed reported for the young (10–20 Myr old) Y chromosomes of the white campion *Silene latifolia* (Guttman and Charlesworth 1998; Filatov et al. 2000; Filatov et al. 2001; Filatov and Charlesworth 2002; Matsunaga et al. 2003; Filatov 2005) and for the neo-Y chromosomes of *Drosophila miranda* (Bachtrog and Charlesworth 2002; Bachtrog 2003), supporting these hypotheses.

With time, the accumulation of deleterious mutations, gene loss, and inability of Y-linked genes to adapt to a changing environment may lead to genetically degenerate Y chromosomes, similar to the mammalian Y. But what is the further fate of the gene-poor mammalian Y chromosome? Could the processes of Y degeneration continue indefinitely until there is nothing left? Only a few mammalian species exist without Y chromosomes (e.g., *Ellobius lutescens* [Fredga 1988]), and the cause of the loss is

unknown but may well represent the final stage of Y degeneration. It has been claimed that the loss of human Y-linked genes is inexorable and that it may lead to the extinction of the entire species because of the active role this chromosome plays in sex determination (Aitken and Marshall Graves 2002; Sykes 2003). These claims seem unlikely to be true because the efficacy of each of the processes causing genetic degeneration depends on the number of functional genes linked together on the nonrecombining Y chromosomes (Charlesworth, Morgan, and Charlesworth 1993; Peck 1994; Orr and Kim 1998; B. Charlesworth and D. Charlesworth 2000; McVean and Charlesworth 2000). With few functional genes surviving on the modern mammalian Y chromosomes (e.g., Skaletsky et al. 2003), the per-chromosome deleterious mutation rate may be too low for background selection to operate (Charlesworth, Morgan, and Charlesworth 1993), and adaptive mutations may also be too rare for selective sweeps to play a major role in the accumulation of deleterious mutations. In contrast to the drastic reduction of genetic diversity on the young *S. latifolia* Y chromosomes (Filatov et al. 2000; Filatov et al. 2001; Matsunaga et al. 2003) and *D. miranda* neo-Y (Bachtrog and Charlesworth 2002; Bachtrog 2003), DNA diversity is only slightly lower on the human Y than on the X (Nachman 1998; Shen et al. 2000). This is consistent with the hypothesis that, after 240–320 Myr of degeneration (Lahn and Page 1999), the human Y may contain too few genes for the degeneration processes to actively proceed further. To address the question of whether the purifying selection is relaxed on the old mammalian Y chromosomes and whether degeneration may still continue, we analyzed the amino acid and synonymous substitution rates in three pairs of mammalian Y- and X-linked genes.

Comparing homologous coding DNA sequences from several species, it is possible to infer selection from the ratio of divergence rates at nonsynonymous (K_a) and silent (K_s) sites. If purifying selection is eliminating most amino acid replacements before they fix in the population and become substitutions, then K_a/K_s will be much less than one, whilst positive selection at many codons should cause this ratio to exceed unity. If the efficacy of purifying selection on the Y chromosome is reduced, as suggested by

Key words: adaptive evolution, mammalian sex chromosomes.

E-mail: d.filatov@bham.ac.uk.

Mol. Biol. Evol. 22(6):1423–1432. 2005

doi:10.1093/molbev/msi128

Advance Access publication March 9, 2005

Table 1
PCR and Sequencing Primers

Gene	Primer Name	Primer Sequence (5'-3')
<i>SMCX</i>	SMCX_+5	CACTTGAGGCCATAATCCGTGAAG
<i>SMCX</i>	SMCX_+1	ATGGTGACCACTACCCCTGC
<i>SMCX</i>	SMCX_-6	TTTGTACAACCCAGCTCCTTCTC
<i>SMCX</i>	SMCX_+12	GTCCACTGGCATCATCGAGC
<i>SMCX</i>	SMCX_-3	GGTAGTTAGGAGGCTCCTCAGGTC
<i>SMCX</i>	SMCX_+7	GGCGGAGGGTCTCAAGTTTG
<i>SMCX</i>	SMCX_-2	TCTGGTTCTCCTGGGTGCTG
<i>SMCY</i>	SMCY_+7	ACCCCTGTCTAGATGACTTGGAG
<i>SMCY</i>	SMCY_-12	TCTCTGAGGTCCTGTGCAGACAG
<i>SMCY</i>	SMCY_+16	AGACAATCCTAGCCTTGCTGG
<i>SMCY</i>	SMCY_-8	AGACACTGAAGGCCTCACC
<i>SMCY</i>	SMCY_+13	AGGTGAGCTTTCAGGCCAG
<i>SMCY</i>	SMCY_-6	TGCGGCAACAGTGAGGACAG
<i>SMCY</i>	SMCY_-2	AGCCTGAATCCTTGTGTGTG
<i>USP9X</i>	USP9X_+1	TGATACCGTAAAGCGCTTGC
<i>USP9X</i>	USP9X_+3	CGTATGTCATACAGGCGCCA
<i>USP9X</i>	USP9X_-4	GGAGGGTTAAGTAAACGCCAT
<i>USP9X</i>	USP9X_-2	TGGCAGAGCCACGGACTACT
<i>USP9Y</i>	USP9Y_+1	TTGCCCTCGGGTCTTGTCTAT
<i>USP9Y</i>	USP9Y_+3	TGATGATGGAGATGTAACAGAATGC
<i>USP9Y</i>	USP9Y_-4	CTGATGGGTCTTGGCAATGTTA
<i>USP9Y</i>	USP9Y_-2	TGGCAGGACCACGAACTATT
<i>UTX</i>	UTX_+1	GGTCAGAGTTCACATTCGGC
<i>UTX</i>	UTX_-2	AGAACTCTGCTGAGCTGGG
<i>UTY</i>	UTY_+1	CAGAGTTCATGTTTGTGTCAGGACC
<i>UTY</i>	UTY_-2	GCATGCTTTCAGAACTTCTGTTG

theory (Felsenstein 1974; Rice 1996), then the K_a/K_s ratio should be higher for the Y-linked genes, compared to the X-linked homologues. Indeed, several studies have already reported an elevated number of amino acid substitutions in Y-linked (or W-linked) versus X-linked (or Z-linked) genes (Agulnik et al. 1998; Fridolfsson and Ellegren 2000; Wyckoff, Li, and Wu 2002; Bachtrog 2003). However, there remains the possibility that positive selection was partly responsible for the increased K_a/K_s on the Y relative to the X. Adaptive nonsynonymous changes at a small number of sites would increase K_a/K_s , but the gene-wide average ratio would be kept below unity by purifying selection operating on the majority of codons. This possibility has not been addressed in the previous comparisons of the K_a/K_s ratios among the X- and Y-linked genes (Agulnik et al. 1998; Fridolfsson and Ellegren 2000; Wyckoff, Li, and Wu 2002; Bachtrog 2003).

To estimate the relative impacts of adaptive and purifying selection on the homologous mammalian X- and Y-linked genes, we sequenced partial coding regions of three pairs of sex-linked genes, *SMCX/SMCY*, *USP9X/USP9Y*, and *UTX/UTY*, from up to 12 mammalian species. These sequences were analyzed using maximum likelihood models allowing for variable selective pressures across codons (Yang et al. 2000). Unexpectedly, we detected positive selection in two out of three studied mammalian Y-linked genes, *USP9Y* and *UTY*. This suggests that positive selection on Y-linked genes is not uncommon and should be accounted for when inferring patterns of selection from substitution rates in X- and Y-linked genes. Furthermore, after correcting for positive selection, we demonstrate that purifying selection is still less efficient on mammalian Y-linked genes than their X-linked homologues—a sign that our Y may still be degenerating.

Materials and Methods

DNA Samples and Extraction

Preextracted DNA was obtained from the Corriel Cell Repository for the following species (genes successfully amplified and sequenced in parentheses): *Lemur catta*, ring-tailed lemur (*SMCX*, *UTX*, *USP9X*); *Gorilla gorilla*, gorilla (*SMCX*, *UTX*, *USP9X*); *Ateles geoffroyi*, black spider monkey (*SMCX*, *SMCY*, *UTX*, *UTY*, *USP9X*, *USP9Y*); *Macaca fascicularis*, crab-eating macaque (*SMCX*, *SMCY*, *UTX*, *UTY*, *USP9X*, *USP9Y*); *Pongo pygmaeus abelii*, orangutan (Sumatran) (*UTX*, *UTY*, *USP9X*, *USP9Y*); *Pan troglodytes*, chimpanzee (*UTY*); and *Homo sapiens*, human (*UTX*, *UTY*). In addition, DNA was extracted from tissue samples obtained from the Institute of Zoology, London, for the following species: *Eulemur fulvus*, brown lemur (*SMCY*, *UTY*, *USP9X*, *USP9Y*); *Cheirogaleus medius*, fat-tailed dwarf lemur (*UTX*); *Leontopithecus rosalia*, golden lion tamarin (*UTX*, *USP9X*); *Presbytis entellus*, hanuman langur (*UTX*); *Colobus* sp., black and white colobus monkey (*UTX*); *Hylobates lar*, lar gibbon (*UTX*, *UTY*, *USP9X*, *USP9Y*, *SMCY*); *G. gorilla*, gorilla (western lowland) (*UTY*, *USP9Y*, *SMCY*). Human, chimpanzee, and mouse sequences were obtained from GenBank unless specified otherwise above. Total genomic DNA was extracted using DNAzol reagent (Invitrogen).

Polymerase Chain Reaction, Cloning, and Sequencing

Polymerase chain reaction (PCR) was carried out in 25- μ l reactions using the following general recipe in which “ T_1 ” represents the primer-specific annealing temperature and T_2 is $T_1 + 2^\circ\text{C}$: 95°C for 2 min 30 s; $T_2^\circ\text{C}$ for 30 s; 72°C for 3 min 30 s; 92°C for 20 s; $T_1^\circ\text{C}$ for 30 s; 72°C for 3 min; 33 cycles of steps 4–6; 72°C for 5 min; 7°C hold. Table 1 lists the primers used for PCR in each gene and also the internal primers where they were necessary. All primers were designed on GenBank human sequence using Vector NTI. Where PCR products were weak, or direct sequencing indicated multiple products, they were cloned into *pCR plasmids* using the TA cloning kit (Invitrogen, Carlsbad, Calif.). *EcoRI* digestion was used to confirm the presence/absence of cloned inserts. Sequencing was carried out in 5- μ l reactions using 2 μ l of gel-extracted DNA for direct sequencing or 0.5 μ l of plasmid from cloned DNA. The sequencing used the BigDye v3.1 kit (ABI, Foster City, Calif.) on an ABI3700 capillary sequencer. All sequences were covered in forward and reverse directions. For each gene, a single, multispecies “contig” was generated in Gap4 (Staden, Beal, and Bonfield 2000) using PreGap4 to feed in the raw ABI sequence files. Assembling one contig for all species automatically aligned the sequences, and divergent indels and site substitutions were revealed early on and could be checked either by simply referring back to the trace or by resequencing a piece of sequence if necessary. Novel sequences have been submitted to GenBank under accession numbers AY591386–AY591426 and AY699605–AY699606.

Sequence Analysis

Species consensus sequences were extracted from Gap4 with alignment gaps (indels) intact. In Proseq 2.9

(Filatov 2002), coding sequences were assigned based on the human intron-exon structure given in GenBank. The *USP9Y* sequence from black spider monkey (*A. Geoffroyi*) contained both a stop codon and, further downstream, a frameshift mutation. As the region sequenced for this study began in exon 36 of 46 and 70% of the coding sequence lies before this point (based on human sequence), we cannot be sure whether *USP9Y* is nonfunctional or just curtailed in this species. These were the only stop codons and frameshifts detected in the study, and this sequence was excluded from any analyses.

Maximum Likelihood Estimation and Likelihood Ratio Tests

The maximum likelihood estimates of various evolutionary models were calculated using the program *codeml*, a constituent of the PAML package (Version 3.14 beta, Yang 1997), and its notation is followed here: the K_a/K_s ratio is termed ω , and model names are as in Yang et al (2000) and Swanson, Nielsen, and Yang (2003). In what follows, when we refer to different codons within a gene, we mean different positions along the protein sequence and not every occurrence of a certain coding triplet.

Most codons in most genes are probably under purifying selection and have $\omega < 1$, while some genes may have few codons under positive selection (with $\omega > 1$). Using models allowing for several classes of codons with separate ω ratios (Yang et al. 2000; Swanson, Nielsen, and Yang 2003), it is possible to test for the variation of selective pressures across the codons and for the presence of codons under positive selection. The simplest model, M0, assumes a single ω ratio for all codons and all the branches of the phylogeny. Fitting a single ω ratio assumes that all the codons are under the same selective pressure and the estimated ω averages across the codons. As only few codons in a gene may be under positive selection, while the rest of the gene is under purifying selection, the average ω is likely to be below 1 despite the action of positive selection on the gene. Thus, estimating a single ω ratio is an extremely conservative approach to test for positive selection. Adding another class of codons with a separate ω_1 may better describe the distribution of ω across codons. Model M3 allows for two or more classes of codons with separate ω ratios (Yang et al. 2000). The program *codeml* was used to calculate the likelihood of each model separately, for example, $L(M0)$ and $L(M3)$ and a likelihood ratio test (LRT) were used to test if the more general model M3 fits the data better than M0. Under the null hypothesis, that M3 does not improve the fit of the model to the data, compared to model M0, the log-likelihood ratio, $2\Delta L = 2[\ln L(M0) - \ln L(M3)]$ is χ^2 distributed with degrees of freedom equal to the difference in number of parameters between the two models (Yang et al. 2000).

The significantly better fit of M3, compared to M0 indicates heterogeneity of ω among codons within the coding region of a gene. To test for the presence of a class of codons under positive selection ($\omega > 1$), we used models M14 and M15 (Swanson, Nielsen, and Yang 2003). M14 fits a beta distribution, which may be unimodal, bimodal, or flat, in the interval $0 \leq \omega \leq 1$ but does not allow any class

of codons to have $\omega > 1$. M15 features an equally limited distribution but allows for an additional class of codons with $\omega > 1$. For the M14/M15 test, the P value of the LRT is half that obtained from the normal χ^2 statistic with one degree of freedom (Swanson, Nielsen, and Yang 2003).

The models above were devised to detect positive selection but do not allow testing of whether purifying selection differs significantly among the lineages. To test whether the Y-linked genes are less constrained by purifying selection, compared to the X-linked homologues, we compared a null model with a single ω ratio across the tree to the model with separate ω_x and ω_y ratios for the X- and Y-linked branches, respectively. The two models were compared using a LRT, assuming that under the null hypothesis of no difference between the two nested models the test statistic, $2\Delta L = 2[\ln L(\text{model}_1) - \ln L(\text{model}_2)]$ is χ^2 distributed with degrees of freedom equal to the difference in number of parameters between the two models (Yang 1998). For this analysis we used a combined tree for all the X- and Y-linked sequences for every gene (e.g., as shown in fig. 2) to calculate the likelihood under the two models. This test assumes a single ω ratio across codons, hence its results are equivalent to the previous comparisons of the K_a/K_s ratios among the X- and Y-linked genes (Agulnik et al. 1998; Fridolfsson and Ellegren 2000; Wyckoff, Li, and Wu 2002; Bachtrog 2003) and suffer from the same problem—averaging across the codons may bias the ω ratio upwards if some codons evolve under positive selection. Thus, a model with variable ω ratios across codons has to be used to account for the possibility of positive selection.

To test whether purifying selection is relaxed in the Y-linked, compared to the X-linked gene, taking into account variable selective pressures across the codons, one needs to use a model which allows for several classes of codons and lineage-specific ω ratios (e.g., ω_{0X} , ω_{0Y} , ω_{1X} , and ω_{1Y}), the so called “branch-site model.” Two branch-site models have been implemented in *codeml* (Yang and Nielsen 2002); however, the tests using these models appear to be too liberal and unreliable (Zhang 2004). Instead, we followed a parametric bootstrap approach to generate a distribution for the ω_{0X} , ω_{0Y} , and the proportion of codons falling into the ω_0 class under the M3 model with two site classes. The bootstrapping was conducted using a Perl script, which generated 1,000 pseudoreplicates bootstrapping across the codons. Each “column” of the data set, containing the same codon position from all the sequences in the alignment was stored as a separate entity in an array. Each pseudoreplicate was generated by randomly selecting the “columns” of codons from this array until a new alignment was created matching the length of the original (Felsenstein 1988). As the bootstrap pseudoreplicates are nonindependent, standard tests of significance (e.g., t -test) are not applicable, hence, the bootstrap distribution was represented graphically (fig. 3).

All the maximum likelihood analyses described above are conditional on a topology of a phylogenetic tree. Initially topologies were calculated for each gene using both the neighbor-joining (Tamura-Nei distance) and maximum parsimony methods in MEGA2 (Kumar et al. 2001). Phylogenetic trees using combined X and Y sequence

Table 2
Sequences Obtained for Each Gene

Gene	<i>N</i> ^a	Exons ^b	Codons ^c	cDNA ^d (%)
X linked				
<i>SMCX</i>	6	21–27	457	30
<i>USP9X</i>	10	36–37	235	9
<i>UTX</i>	12	17	223	19
Y linked				
<i>SMCY</i>	9	21–27	447	29
<i>USP9Y</i>	7	36–37	242	9
<i>UTY</i>	9	17	211	18

^a The number of species studied for each gene. See experimental procedures for the list of species.

^b Based on the exons annotated on human sequence in GenBank.

^c The number of codons is one-third the length of coding DNA sequence.

^d Based on length of human coding sequence annotated in GenBank.

alignments were constructed using the neighbor-joining method to test the phylogenetic relationships of the sequences and specifically to check for monophyletic origins of all X and all Y sequences. This may not have been the case (1) if the cessation of recombination between X and Y had occurred independently in different lineages, as it was demonstrated for bird sex chromosomes (Ellegren and Carmichael 2001) or (2) if intraspecific gene conversion had overwritten X with Y or vice versa as has happened at least twice to the *ZFY* gene during the radiation of cats (Pecon Slattery, Sanner-Wachter, and O'Brien 2000). The first possibility seems unlikely for rodents and primates as the studied genes are thought to have become sex linked before the rodent-primate split (Lahn and Page 1999). Additionally, to test the sensitivity of the results of the LRTs to changes in the tree topology, we reran the significant M14/M15 results using up to 47 alternative best trees generated in PAUP* (Swofford 1998).

Pairwise Human-Mouse X-Linked Measurements

To place the substitution rates in the *UTX/UTY*, *SMCX/SMCY*, and *USP9X/USP9Y* genes in context, we compared the pairwise ω estimates (between human and mouse) of nine ubiquitously expressed X-linked genes possessing Y homologues with 121 other X-linked genes. To do this, the map of human-mouse orthologous loci was cross-referenced with the list of mouse and human homologous sequence pairs downloaded from LocusLink (<ftp://ftp.ncbi.nih.gov/refseq/LocusLink/>) using locus ID numbers common to both. This resulted in a list of GenBank messenger RNA (mRNA) accessions of human X-linked loci for which reciprocal best matches had been made with known and available rodent (mouse and rat) mRNAs. For each pair of mRNA accessions, the full GenBank entries were downloaded from the NCBI website. The following EMBOSS (Rice, Longden, and Bleasby 2000) programs were used: Coderet—to extract and translate the coding nucleotide sequence; Needle—to align each pair of protein sequences; Tralign—to align the nucleotide sequences based on the protein alignment. The PAML program *codeml* (Yang 1997) was then used to estimate the pairwise ω for each nucleotide alignment using the method of Goldman and Yang (Goldman and Yang 1994). The

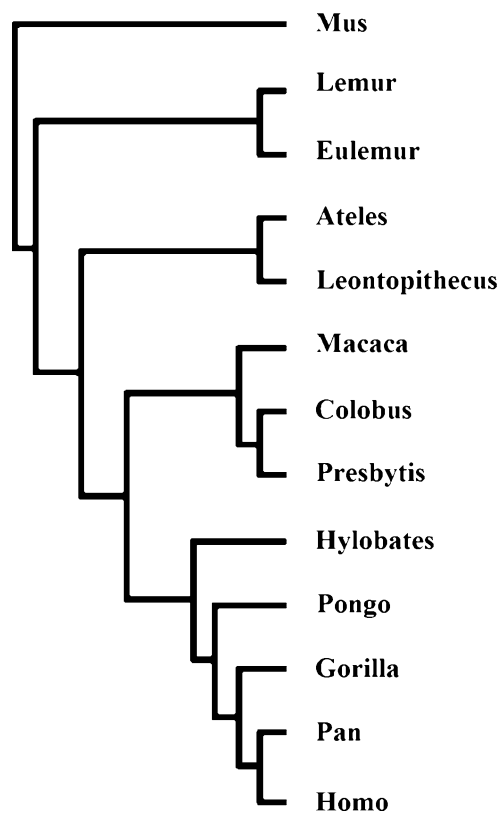


FIG. 1.—Phylogenetic relationship of the genera used in this study. Based on Purvis (1995).

sequence downloads, alignments, and ω estimations were automated using a Perl script running on a LINUX platform. Handling and conversion of sequences and alignments were made possible with BioPerl extension modules (www.bioperl.org).

Results

To estimate the relative impacts of adaptive and purifying selection on the homologous X- and Y-linked genes, we sequenced partial coding regions of three pairs of sex-linked genes, *SMCX/SMCY*, *USP9X/USP9Y*, and *UTX/UTY* from up to 12 mammalian species (table 2, and *Materials and Methods*).

Phylogeny of New X and Y Sequences

Before carrying out any tests of evolutionary models, we examined the phylogenetic relationships of the sequences. The trees from *SMCX/SMCY* followed the standard primate phylogeny shown in figure 1 (Purvis 1995) irrespective of the method used. However, whilst they still separated into X and Y clades, the trees from *USP9X/USP9Y* and *UTX/UTY* were not entirely consistent with the same traditional phylogeny and also differed depending on the tree-building method used (neighbor-joining with nucleotide or amino acid sequences and maximum parsimony with nucleotide sequences). Figure 2 shows the nucleotide neighbor-joining tree using all sites in the *USP9X/USP9Y* alignment. The positions of the *Pongo* and *Hylobates*

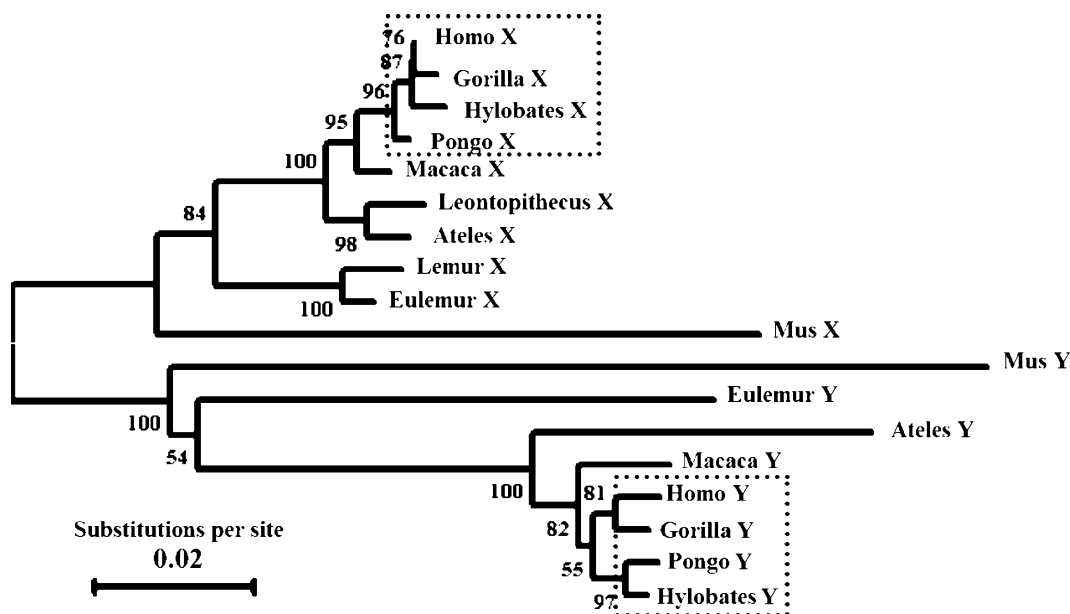


FIG. 2.—The neighbor-joining tree (Tamura-Nei distance) of nucleotide sequences for *USP9X* and *USP9Y* sequences combined. The topologies of the ape clades (boxed) in both X and Y halves of the tree do not agree with the traditional phylogeny.

sequences within both X and Y halves of the tree were not as expected: *Pongo* clustered with *Hylobates* for the *USP9Y*, while *Hylobates* was closer to *Homo* and *Gorilla* than *Pongo* using the *USP9X* gene. The *SMCX/SMCY* sequences were substantially longer than those of the other genes (see table 2), and because we reasoned that the Y-linked genes at least, must have had the same phylogeny, we accepted the traditional primate phylogeny over any other to use in our analyses. Using the alternative phylogenies made no difference to the results of the LRTs (see below).

Testing for Heterogeneity in K_a/K_s (ω) Among Codons

Data sets for each of the six genes were analyzed separately using maximum likelihood models allowing for variable selective pressures across codons (Yang et al. 2000). The models allow the estimation of both the K_a/K_s ratio (ω , using the notation of the program PAML

[Yang, 1997]) for several discrete classes of codons and the proportion of codons within each class. The nested models with increasing numbers of parameters can be compared by the LRT, providing a test of whether a more parameter rich model fits the data significantly better (see Methods).

The analysis using model M0 (Yang et al. 2000) is similar to that from the pairwise comparisons conducted by others (Agulnik et al. 1998; Fridolfsson and Ellegren 2000; Wyckoff, Li, and Wu 2002; Bachtrog 2003) as it allowed for just a single class of codons. Consistent with previous results, the estimated ω ratio for the Y-linked genes is greater than that for their X-linked homologues (table 3). The significance of this difference can be shown on a combined tree of X and Y sequences using a model allowing for separate ω ratios for X- and Y-linked genes and by testing this model against the model with a single ω ratio for all branches. For all three pairs of genes, the LRT showed that separate ω for X and Y branches results in a

Table 3
Likelihood Ratio Tests of Models M0 versus M3

Gene	Model M0		Model M3 With Two Classes of Codons				M3 Versus M0	
	Ln L ^a	ω^b	Ln L	ω_0	ω_1	% With ω_1	2 ΔL^c	P Value
X linked								
<i>SMCX</i>	-3,240.7	0.149	-3206.0	0.000	1.060	19.3	69.385	< 0.001
<i>USP9X</i>	-1,590.0	0.089	-1588.4	0.076	1.582	1.0	3.117	0.210
<i>UTX</i>	-1,237.3	0.270	-1232.4	0.014	1.580	16.5	9.897	0.042
Y linked								
<i>SMCY</i>	-4,389.2	0.286	-4357.0	0.140	1.201	19.0	64.361	< 0.001
<i>USP9Y</i>	-1,750.5	0.207	-1732.6	0.090	1.387	15.0	35.810	< 0.001
<i>UTY</i>	-1,992.8	0.962	-1987.4	0.512	2.120	35.8	10.744	0.030

NOTE.—Values in bold show significant values ($p < 0.05$).

^a Ln L represents the log likelihood.

^b ω Represents the codon-specific K_a/K_s estimate from PAML.

^c 2 ΔL represents double the difference in log likelihoods between the two models.

Table 4
Likelihood Ratio Tests of X and Y Branch-Specific Models on a Combined Tree

Genes (number of X/Y)	Number of Codons	Separate Rates	Ln L	Branch ω s		
				ω_X	ω_0	ω_Y
<i>UTX/UTY</i> (12/9)	207	ω_0	-2,537.246		0.743	
		ω_Y, ω_0	-2,529.545**		0.383	1.063
		ω_X, ω_0	-2,527.548**	0.243		0.965
<i>USP9X/USP9Y</i> (10/7)	233	ω_0	-2,699.612		0.167	
		ω_Y, ω_0	-2,687.061**		0.078	0.272
		ω_X, ω_0	-2,685.333**	0.065		0.261
<i>SMCX/SMCY</i> (6/9)	435	ω_0	-5,811.661		0.228	
		ω_Y, ω_0	-5,802.744**		0.146	0.287
		ω_X, ω_0	-5,800.538**	0.134		0.289

NOTE.—For each combined X and Y tree, three models were fitted to the data. The first model gave one ω ratio to all branches on the phylogeny (equivalent to model 0 in previous models). The second and third models allowed two separate ω ratios: one for the Y or the X-specific branches, respectively, with the rest of the tree having ω_0 . In all cases a model with two different ω ratios within the tree fit the data better than with one overall ω_0 . Values in bold show significant values ($p < 0.05$).

** LRT between one ω for all branches and two ω s significant at $P < 0.0001$.

significantly better fit of the model to the data ($P < 0.0001$, table 4).

Model M3 (Yang et al. 2000) takes into account the variable selective pressures across the codons by allowing for several classes of codons, with ω ratios estimated separately for each class. Model M3 with two classes of codons fits the data significantly better than the model M0 for all the genes except *USP9X* (table 3), demonstrating that selective pressure significantly varies across the codons in five out of six genes studied. The addition of extra classes of codons to model M3 did not improve the fit of the model to the data for any of the genes (data not shown), suggesting that two classes of codons are sufficient to accommodate the variation in selective regimes among the codons.

Under model M3 with two classes of codons, the majority of the codons fall into the class with the lower ω_0 ratio (table 3), reflecting the fact that most codons in these genes are under purifying selection. Interestingly, the proportion of codons falling into the class with ω_0 is substantially lower in *USP9Y* (85.0%) and *UTY* (64.2%) compared to *USP9X* (99.0%) and *UTX* (83.5%). This reduction in the number of codons under purifying selection in *USP9Y* and *UTY* compared to the X-linked homologues can be either due to reduced efficacy of purifying selection on the nonrecombining Y-linked genes or due to positive selection at some codons in the *USP9Y* and *UTY* genes. The estimated ω_0 for this class of codons is higher for all the Y-linked genes compared to the X-linked homologues (table 3). This is consistent with the results of the previous studies which reported an elevated number of amino acid substitutions in Y-linked (or W-linked) versus X-linked (or Z-linked) genes (Agulnik et al. 1998; Fridolfsson and Ellegren 2000; Wyckoff, Li, and Wu 2002; Bachtrog 2003) and with the theoretical prediction of a reduced efficacy of purifying selection on the nonrecombining Y chromosomes (Charlesworth and Charlesworth 2000), suggesting that mammalian Y-linked genes continue to accumulate deleterious mutations at a greater rate than those on the X.

To test whether the ω_0 values for the Y-linked genes are significantly higher than for the X-linked homologues,

we followed a parametric bootstrap approach. For each gene we bootstrapped across codons to generate distributions of ω_0 (under model M3) and the proportion of codons falling into the ω_0 class, p_0 (fig. 3). Assuming independence of bootstrap replicates, t -tests showed that ω_{0Y} is significantly greater than ω_{0X} for all three pairs of genes ($P < 0.001$). However, the nonindependence of the pseudoreplicates may make such a test too liberal. For *SMCX/SMCY* (fig. 3a) and *UTX/UTY* (fig. 3b) the “clouds” of values for the X- and Y-linked genes are clearly distinct, suggesting that the increase in ω_{0Y} over ω_{0X} is significant for these genes. On the other hand, for *USP9X/USP9Y* there is considerable overlap of ω_{0Y} and ω_{0X} distributions (fig. 3c), suggesting that ω_{0Y} is not significantly greater than ω_{0X} for this pair of genes.

Testing for Positive Selection at Some Codons

For the second class of codons, the ω_1 ratio exceeds unity for all the X- and Y-linked genes (table 3), suggesting that some codons in these genes may be under positive selection. To test whether adaptive selection is indeed acting in all or some of these genes, we have to demonstrate that the ω_1 ratios are significantly greater than unity (i.e., are nonneutral) for this class of codons. This can be done by comparing two nested models, one allowing for a class of codons free to take any $\omega > 1$ and the other with this class having fixed $\omega = 1$. Models M14 and M15 (Swanson, Nielsen, and Yang 2003) implemented in PAML allow the ω ratios below 1 to be distributed according to a beta distribution and thus may better account for the variation in purifying selection across the codons compared to models M0 and M3. To accommodate the codons with ω ratios above 1.0, these models have a further class (ω_1), that, in M15 is free to take any K_a/K_s value above 1.0, and is fixed at $K_a/K_s = 1$ for model M14 (fig. 3). The comparison of M14 and M15 by the LRTs demonstrated that model M15 fits the data significantly ($P < 0.05$) better than M14 for *UTY* and *USP9Y* but not for *SMCY* or any of the three X-linked genes (table 5), suggesting the presence of codons under positive selection in *UTY* and *USP9Y* genes.

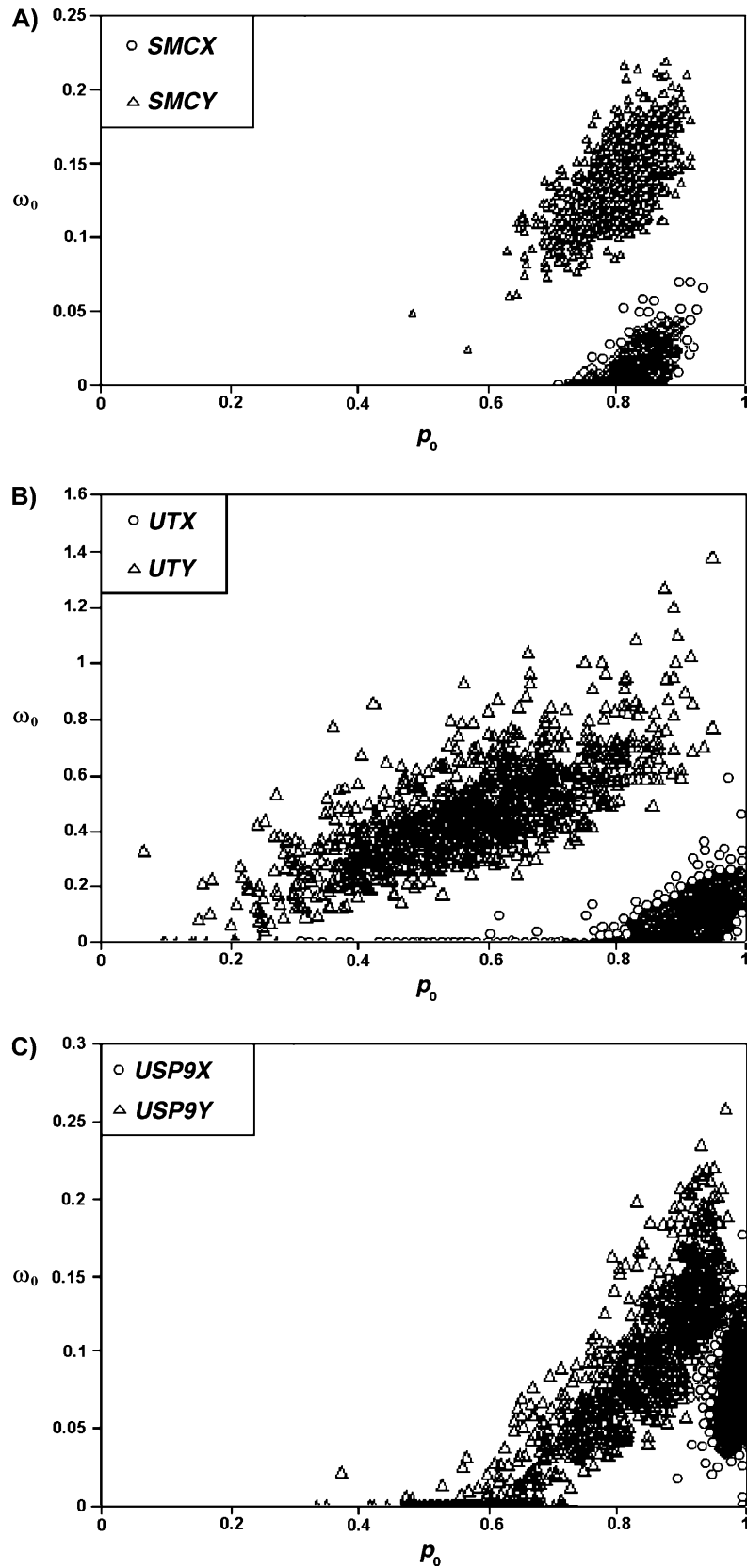


FIG. 3.—Plots of ω_0 against the proportion of codons falling into the ω_0 class, p_0 , from the 1,000 bootstrap replicates under model M3 for each of (a) *SMCX/SMCY*, (b) *UTX/UTY*, and (c) *USP9X/USP9Y*.

Table 5
Likelihood Ratio Tests of Models M14 Versus M15

Gene	M14		M15		LRT		P Value
	Ln L	% With $\omega = 1.0$	Ln L	ω_1	% With ω_1	$2\Delta L$ M15 versus M14	
X linked							
<i>SMCX</i>	-3206.1	19.7	-3206.0	1.060	19.3	0.095	0.379
<i>USP9X</i>	-1595.3	16.4	-1595.3	1.000	16.4	0.000	0.500
<i>UTX</i>	-1232.7	24.0	-1232.4	1.646	14.9	0.609	0.218
Y linked							
<i>SMCY</i>	-4357.4	22.8	-4357.0	1.211	18.7	0.7788	0.189
<i>USP9Y</i>	-1732.7	18.7	-1730.7	4.089	3.7	3.859	0.025
<i>UTY</i>	-1989.8	81.2	-1987.4	1.988	38.9	4.777	0.014

NOTE.—Values in bold show significant values ($p < 0.05$).

To assess whether the topology of the tree that we chose affected the results of the LRTs, we ran the M14/M15 tests for *UTY* and *USP9Y* using different tree topologies. We used PAUP 4.0 (Swofford 1998) to generate alternative trees close to the best tree topology. The trees were allowed to contain multifurcating nodes but were all rooted using the mouse sequence as an out-group. For *UTY*, 45 alternative trees were used. Twelve of these trees were no worse than the best tree ($P < 0.05$, KH test [Kishino and Hasegawa, 1989]). None of the 45 trees failed to show the significantly better fit of model 15 compared to model 14 in the log-likelihood ratio test after these models were run for each tree. For *USP9Y*, 47 similar trees were used with 25 as good as the best tree (KH test, $P < 0.05$). Again, none of the tree topologies affected the result of the M14/M15 ratio test, and the model including adaptive selection better explained the data for all trees.

Discussion

The accuracy and power of the phylogenetic maximum likelihood analysis to detect positive selection has been extensively tested (Anisimova, Bielawski, and Yang 2001; Zhang 2004). Zhang (2004) demonstrated that branch-site likelihood methods (Yang and Nielsen 2002) are unreliable as they detect positive selection more often than 5% when applied to data simulated with an absence of selection. However, this does not apply to the codon-based likelihood analyses used in our paper. The site-branch methods are designed to infer selection at some of the sites along some of the branches in the phylogeny, and the authors are very cautious about these methods (Yang and Nielsen 2002). On the contrary, the method used in our paper tests for selection on individual codon sites over an entire phylogenetic tree (Yang et al. 2000). It was demonstrated that this approach is fairly conservative and reliable (Anisimova, Bielawski, and Yang 2001). Models M14 and M15 (which are modifications of model M8) are much more stringent compared to M0/M3 (Anisimova, Bielawski, and Yang 2001). Indeed, using alternative tree topologies had no effect on the result of the M14/M15 ratio tests. Thus, the rejection of model M14 for *UTY* and *USP9Y* suggests the presence of codons under positive selection in these genes.

Our results provide evidence for positive selection on single-copy mammalian Y-linked genes. Although we used

three X-linked and three Y-linked genes in our study, the P values reported are without adjustment for multiple testing because the hypothesis of interest is a comparison of models for each gene, not for all of the genes. While we concur with Perneger (1998) that the outcome of the test for positive selection within *SMCY* is irrelevant to the outcome of the test for *UTY*, we have provided the means for investigators to examine the data and if desired perform corrections themselves. If the most stringent correction, Bonferroni, is used, one of the three Y-linked genes, *UTY*, still shows evidence for positive selection. We also applied the false discovery rate correction method (Verhoeven, Simonsen, and McIntyre 2005), which maintains the type I error rate but which also better maintains the type II error rate (false acceptance of null). In this case, the null hypothesis was rejected for both *USP9Y* and *UTY* but not *SMCY*.

The detection of positive selection for two of the three Y-linked genes studied suggests that adaptive selection may not be a rare phenomenon on the mammalian Y chromosome. Indeed, an analysis of the mammalian Y-linked *DAZ* gene family demonstrated that the high ω ratio in this gene is caused not by the absence of purifying selection, as thought before (Agulnik et al. 1998), but by adaptive selection at some of the codons, while most codons are under purifying selection (Bielawski and Yang 2001). The *DAZ* gene family belongs to class II of Y-linked genes according to the classification proposed by Lahn, Pearson, and Jegalian (2001), i.e., it is a Y-specific gene family with a male-specific function, which evolved on the Y chromosome. Such genes are probably more free to evolve new functions than the genes in this study, which match their X homologues in the breadth of expression and, to an unknown extent, function.

The causes of the positive selection on *UTY* and *USP9Y* are not completely clear. *UTX/UTY* and *USP9X/USP9Y* belong to a small group of ubiquitously expressed housekeeping genes with active copies resident singly on X and Y. The genes in this group are under fairly strong purifying selection: the pairwise ω for the mouse-human divergence for nine X-linked genes (*USP9X*, *SMCX*, *UTX*, *UBE1X*, *ZFX*, *SOX3*, *DBX*, *RBMX*, *RPS4X*) which retained expressed Y-linked homologues are significantly lower compared to 121 X-linked genes without active human Y homologues ($P = 0.022$, Kruskal-Wallis test). Thus, it is possible that these genes retained active Y-linked

homologues due to more stringent purifying selection compared to X-linked genes that lost Y homologues. Positive selection on the *UTY* and *USP9Y* but not on the X-linked copies of these genes may suggest that the X- and Y-linked copies are somewhat diverged in function. However, the X copies of these genes are not dosage compensated (Lahn, Pearson, and Jegalian 2001), suggesting that both X and Y copies of these genes perform similar functions. Alternatively, the positive selection on the *UTY* and *USP9Y* genes may be for compensatory mutations, which maintain the function of the Y-linked genes despite the accumulation of deleterious mutations in the nonrecombining genes.

Our finding that two out of three Y-linked genes studied undergo positive Darwinian selection indicates that it may be fairly common on the Y chromosomes. Previously, the elevated ω ratios in the mammalian (or avian) single-copy Y-linked (or W-linked) genes compared to the X-linked (or Z-linked) homologues were interpreted as evidence for relaxation of purifying constraint on the Y chromosome (Agulnik et al. 1998; Fridolfsson and Ellegren 2000; Wyckoff, Li, and Wu 2002; Bachtrog 2003). Our results demonstrate that such interpretations have to be taken with caution because the situation may be complicated by the presence of adaptive selection on the Y-linked genes.

Acknowledgments

We thank Lauren McIntyre for helpful suggestions and for sending us a prepublished manuscript. This work was supported by a grant to D.A.F. from the BBSRC. D.T.G. was supported by a PhD studentship from the BBSRC.

Literature Cited

- Agulnik, A. I., A. Zharkikh, H. Boettger-Tong, T. Bourgeron, K. McElreavey, and C. E. Bishop. 1998. Evolution of the DAZ gene family suggests that Y-linked DAZ plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. *Hum. Mol. Genet.* **7**:1371–1377.
- Aitken, R. J., and J. A. Marshall Graves. 2002. The future of sex. *Nature* **415**:963.
- Anisimova, M., J. P. Bielawski, and Z. Yang. 2001. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Mol. Biol. Evol.* **18**:1585–1592.
- Bachtrog, D. 2003. Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in *Drosophila*. *Nat. Genet.* **34**:215–219.
- Bachtrog, D., and B. Charlesworth. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**:323–326.
- Bielawski, J. P., and Z. Yang. 2001. Positive and negative selection in the DAZ gene family. *Mol. Biol. Evol.* **18**:523–529.
- Bull, J. J. 1983. *Evolution of sex determining mechanisms*. The Benjamin/Cummings Publishing Company, Menlo Park, Calif.
- Charlesworth, B., and D. Charlesworth. 2000. The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**:1563–1572.
- Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**:1289–1303.
- Ellegren, H., and A. Carmichael. 2001. Multiple and independent cessation of recombination between avian sex chromosomes. *Genetics* **158**:325–331.
- Felsenstein, J. 1974. The evolutionary advantage of recombination. *Genetics* **78**:737–756.
- . 1988. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**:521–565.
- . 2002. ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. *Mol. Ecol. Notes* **2**:621–624.
- Filatov, D. A. 2005. Substitution rates in a new *Silene latifolia* sex-linked gene, *SlssX/Y*. *Mol. Biol. Evol.* **22**:1–7.
- Filatov, D. A., and D. Charlesworth. 2002. Substitution rates in the X- and Y-Linked genes of the plants, *silene latifolia* and *S. dioica*. *Mol. Biol. Evol.* **19**:898–907.
- Filatov, D. A., V. Laporte, C. Vitte, and D. Charlesworth. 2001. DNA diversity in sex-linked and autosomal genes of the plant species *silene latifolia* and *silene dioica*. *Mol. Biol. Evol.* **18**:1442–1454.
- Filatov, D. A., F. Moneger, I. Negrutiu, and D. Charlesworth. 2000. Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**:388–390.
- Fredga, K. 1988. Aberrant chromosomal sex-determining mechanisms in mammals, with special reference to species with XY females. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **322**:83–95.
- Fridolfsson, A. K., and H. Ellegren. 2000. Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes. *Genetics* **155**:1903–1912.
- Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**:725–736.
- Gordo, I., and B. Charlesworth. 2000. On the speed of Muller's ratchet. *Genetics* **156**:2137–2140.
- Gutman, D. S., and D. Charlesworth. 1998. An X-linked gene with a degenerate Y-linked homologue in a dioecious plant. *Nature* **393**:263–266.
- Kishino, H., and M. Hasegawa. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* **29**:170–179.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**:1244–1245.
- Lahn, B. T., and D. C. Page. 1999. Four evolutionary strata on the human X chromosome. *Science* **286**:964–967.
- Lahn, B. T., N. M. Pearson, and K. Jegalian. 2001. The human Y chromosome, in the light of evolution. *Nat. Rev. Genet.* **2**:207–216.
- Matsunaga, S., E. Isono, E. Kejnovsky, B. Vyskot, J. Dolezel, S. Kawano, and D. Charlesworth. 2003. Duplicative transfer of a MADS box gene to a plant Y chromosome. *Mol. Biol. Evol.* **20**:1062–1069.
- McVean, G. A., and B. Charlesworth. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**:929–944.
- Nachman, M. W. 1998. Y chromosome variation of mice and men. *Mol. Biol. Evol.* **15**:1744–1750.
- Orr, H. A., and Y. Kim. 1998. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics* **150**:1693–1698.
- Peck, J. R. 1994. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* **137**:597–606.
- Pecon Slattery, J., L. Sanner-Wachter, and S. J. O'Brien. 2000. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in *Felidae* (*Mammalia*). *Proc. Natl. Acad. Sci. USA* **97**:5307–5312.
- Perneger, T. V. 1998. What's wrong with Bonferroni adjustments. *BMJ* **316**:1236–1238.

- Purvis, A. 1995. A composite estimate of primate phylogeny. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **348**:405–421.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**:276–277.
- Rice, W. R. 1996. Evolution of the Y sex chromosome in animals. *Bioscience* **46**:331–343.
- . 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**:161–167.
- Shen, P., F. Wang, P. A. Underhill et al. (13 co-authors). 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* **97**:7354–7359.
- Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx et al. (37 co-authors) 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**:825–837.
- Staden, R., K. F. Beal, and J. K. Bonfield. 2000. The Staden package, 1998. *Methods Mol. Biol.* **132**:115–130.
- Swanson, W. J., R. Nielsen, and Q. Yang. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**:18–20.
- Swofford, D. L. 1998. PAUP*. phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.
- Sykes, B. 2003. Adam's curse—a future without men. Bantam Press, London.
- Verhoeven, K. J. F., K. L. Simonsen, and L. M. McIntyre. 2005. Implementing false discovery rate control: increasing your power. *Oikos* **108**:643–647.
- Wyckoff, G. J., J. Li, and C. I. Wu. 2002. Molecular evolution of functional genes on the mammalian Y chromosome. *Mol. Biol. Evol.* **19**:1633–1636.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
- . 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **15**:568–573.
- Yang, Z., and R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**:908–917.
- 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. 2004. Frequent false detection of positive selection by the likelihood method with branch-site models. *Mol. Biol. Evol.* **21**:1332–1339.

Lauren McIntyre, Associate Editor

Accepted March 1, 2005