Gene Diversity Patterns at 10 X-Chromosomal Loci in Humans and Chimpanzees

Takashi Kitano,[1] *Carsten Schwarz, Birgit Nickel, and Svante Pääbo*

Department of Evolutionary Genetics, Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany

We have investigated the pattern and extent of nucleotide diversity in 10 X-chromosomal genes where mutations are known to cause mental retardation in humans. For each gene, we sequenced the entire coding region from cDNA in humans, chimpanzees, and orangutans, as well as about 3 kb of genomic DNA in 20 humans sampled worldwide and in 10 chimpanzees representing two ''subspecies.'' Overall nucleotide diversity in these genes is about twofold lower in humans than in chimpanzees, and nucleotide diversity within and between species is low, suggesting that a high level of functional constraint acts on these genes. Strikingly, we find that a summary of the allele frequency spectrum is significantly correlated in humans and chimpanzees, perhaps reflecting very similar levels of constraint at these genes in the two species. A possible exception is *FMR2*, which shows a higher number of nonsynonymous than synonymous substitutions on the human lineage, suggesting the action of positive selection.

Introduction

Although humans and chimpanzees are closely related, they differ in population structure as well as in current and past population size. They also differ in a number of cognitive capabilities, some of which may be of fundamental importance for the differences in the culture and in the history of the two species. The compounded effect of these differences is reflected in the contemporary genetic diversity of the species, which in the great apes is significantly higher than in humans (Crouau-Roy et al. 1996; Gagneux et al. 1999; Kaessmann, Wiebe, and Pääbo 1999; Kaessmann et al. 2001). However, comparisons of the extent and distribution of genetic diversity between humans and chimpanzees are still hampered by the fact that very few loci are studied in chimpanzees.

To gain a better overview of genetic diversity in chimpanzees, patterns of variation at genes of functional importance for interesting aspects of human and ape phenotypes need to be studied in a comparable way in the two species. One class of such genes in humans is those in which mutations are known to cause mental retardation. These genes are of interest because they are required for normal cognitive development, and because they could have been the direct targets for selection when human cognitive abilities evolved.

Interestingly, a particularly high density of such genes are found on the X chromosome (reviewed in Chelly 1999; Toniolo and D'Adamo 2000; Castellvi-Bel and Mila 2001). For example, in the ''Online Mendelian Inheritance in Man'' (OMIM) database, of 530 entries of genes that contain the term ''mental retardation'' and that are mapped to a chromosome, 156 (29%) are located on the X chromosome (May 2002). By contrast, only 6% (559/8704) of all mapped genes are on the X chromosome. Although some of this effect may be due to ascertainment bias, this is unlikely to account for the entire overrepresentation (Zechner et al. 2001).

To take a first step toward studying such genes, we have investigated the pattern and extent of nucleotide diversity in 10 well-characterized X-linked genes (table 1) in humans and chimpanzees. In seven of these genes, mutations in humans are known to cause mental retardation in the absence of other symptoms (*FMR2*, *GDI1*, *IL1RAPL1*, *OPHN1*, *PAK3*, *RPS6KA3*, and *TM4SF2*). In the other three genes, mutations cause syndromes that include mental retardation. These genes are *ATRX*, which is involved in alpha thalassemia and mental retardation; *L1CAM*, which is involved in the MASA syndrome (mental retardation, aphasia, shuffling gait, and adducted thumbs); and *TNRC11*, suggested to be involved in X-linked nonspecific mental retardation (Philibert et al. 1998, but see also Friez et al. 2000).

Materials and Methods

Samples

We used genomic DNAs for 20 male humans, including nine Asians (Aboriginal Australian, Warao South American Indian, Chinese, Japanese, Thai, Papua New Guinean Lowlander, Papua New Guinean Highlander, Nasioi, and Iranian), four Europeans (French, German, English, and Italian), and seven Africans (Mbuti Pygmy, two Biaka Pygmy, Ibo, Yoruba, Effik, and Hausa). These samples include individuals from nine language phyla (Australian, Amerind, Sino-Tibetan, Altaic, Austric, Indo-Pacific, Indo-Hittite, Niger-Kordofanian, and Afro-Asiatic). We also used genomic DNAs for 10 male chimpanzees, including three western chimpanzees (*Pan troglodytes verus*) and seven central chimpanzees (*Pan troglodytes troglodytes*).

For determination of cDNA sequences, mRNAs from the brains of single male humans, chimpanzees, and orangutans, respectively, were used.

PCR Amplification and DNA Sequencing

Polymerase chain reactions (PCRs) (6.7–23.8 kb) were performed using the Expand 20 kb^{Plus} PCR System (Roche). The PCR products were then used as template for second PCRs using AmpliTaq Gold (PerkinElmer). The PCR products were purified with a QIAquick PCR Purification Kit (QIAGEN), after which they were

[1] Present address: Division of Population Genetics, National Institute of Genetics, Mishima, Japan.

Key words: chimpanzees, humans, nucleotide diversity, selection.

E-mail: paabo@eva.mpg.de.

Mol. Biol. Evol. 20(8):1281–1289. 2003
DOI: 10.1093/molbev/msg134
Molecular Biology and Evolution, Vol. 20, No. 8,

Table 1
List of X-linked Mental Retardation Related Genes

Gene	Name	Location	cM/Mb[a]	Disease	Reference
ATRX	Alpha thalassemia/mental retardation syndrome X-linked	Xq13.1-q21.1	2.22 (0.86–3.57)	ATR-X[b]	Gibbons et al. (1995)
FMR2	Fragile X mental retardation 2	Xq28	3.92	NSMR[c]	Gu et al. (1996); Gecz et al. (1996)
GDI1	GDP dissociation inhibitor 1	Xq28	3.92	NSMR	D'Adamo et al. (1998)
IL1RAPL1	Interleukin 1 receptor accessory protein-like 1	Xp22.1-p21.3	0.81 (0.45–1.17)	NSMR	Carrie et al. (1999)
L1CAM	L1 cell adhesion molecule	Xq28	3.92	MASA[d]	Jouet et al. (1994); Vits et al. (1994)
OPHN1	Oligophrenin 1 (RhoGTPase activating protein)	Xq12	1.21 (0.85–1.57)	NSMR	Billuart et al. (1998)
PAK3	p21 (CDKN1A)-activated kinase 3	Xq21.3-q24	3.64 (3.38–3.90)	NSMR	Allen et al. (1998)
RPS6KA3	Ribosomal protein S6 kinase, 90 kD, polypeptide 3	Xp22.2-p22.1	3.57	NSMR	Merienne et al. (1999)
TM4SF2	Transmembrane 4 superfamily member 2	Xq11	1.21 (0.85–1.57)	NSMR	Zemni et al. (2000)
TNRC11	Trinucleotide repeat containing gene 11	Xq13	0.81 (0.45–1.17)	NSMR	Philibert et al. (1998)

[a] Where more than one marker pair exists, the highest and lowest values are given in parentheses. cM/Mb indicates centiMorgans (genetic) per Megabase (physical).
[b] ATR-X (alpha thalassemia/mental retardation X-linked) syndrome.
[c] Nonsyndromic mental retardation.
[d] MASA (mental retardation, aphasia, shuffling gait, and adducted thumbs) syndrome.

sequenced with the BigDye Terminator Cycle Sequencing Kit and ABI PRISM 3700 DNA sequencers (PE Biosystems). For each locus, we sequenced a region of about 3 kb. When a disease mutation was known, the region around it was included in the sequenced region. A list of primers used is available from T.K.

Reverse transcription (RT) of mRNAs was done by SuperScript II RT Kit (GIBCO BRL). Subsequently, PCRs and sequencing were performed essentially as described above. Complete gene coding regions for eight loci were sequenced. The other two loci (*L1CAM* and *RPS6KA3*) were not sequenced completely, because their 5′ noncoding regions were not available. *ATRX*, *PAK3*, and *TNRC11* contained alternative splicing products which made direct sequencing impossible. In these cases, the PCR products were cloned with the TOPO TA-cloning kit (Invitrogen), and at least three clones representing the longest products were sequenced in each case.

The sequence data presented in this study have been submitted to the DDBJ/EMBL/GenBank International Nucleotide Sequence Database under accession numbers AB101681–AB102670.

Data Analysis

ClustalW version1.8 (Thompson, Gibson, and Higgins 1994) was used for multiple alignments. Two measures of nucleotide variability, π (Nei 1987) and θ_W (Watterson 1975), were calculated for each locus. Under the standard neutral model of a random-mating population of constant size, and assuming an infinite-site mutation model, both π and θ_W estimate the neutral parameter $3N_e\mu$ for X-chromosomal loci, where N_e is the effective population size and μ is the neutral mutation rate. Tajima's *D* (1989), Fu and Li's *D* (1993), and Fay and Wu's *H* (2000) were calculated to test for deviations from the standard neutral model using the allele frequency spectrum. Tajima's (1989) test examines whether the average number of pairwise nucleotide differences between sequences (π) is larger than expected from the observed number of polymorphic sites (θ_w). The expected difference (*D*) between π and θ_w is roughly zero under the standard neutral model. A positive value of *D* indicates possible balancing selection or population subdivision. A negative value suggests recent directional selection, a population bottleneck, or purifying selection on slightly deleterious alleles (Tajima 1989). The Fu and Li (1993) test is based on the principle of comparing the number of mutations on internal branches with those on external branches. Compared with a neutral model of evolution, directional selection would result in an excess of external mutations, and balancing selection would result in an excess of internal mutations. Fay and Wu's (2000) test compares the difference (*H*) between π (which is influenced most by variants at intermediate frequencies) and θ_H (which is influenced most by high-frequency phylogenetically derived variants). A negative value reflects a relative excess of high-frequency derived alleles, as expected immediately after a selective sweep. Ratios of polymorphism to divergence were compared with the expectations under a neutral model using a 10-locus Hudson, Kreitman, and Aguadé (HKA) test (Hudson, Kreitman, and Aguadé 1987). The coalescence time (*T*) was estimated by the GeneTree package (Griffiths and Tavare 1995), assuming a global panmictic population, no recombination, and using the maximum likelihood estimate for θ. Recombination rates (cM/Mb in table 1) were estimated from GB4 map by identifying the marker closest to the locus of interest and looking up the recombination rate for that marker as given by Payseur and Nachman (2000). Most of the statistical analyses were performed with the DnaSP version 3.50 program (Rozas and Rozas 1999). The program HKA (kindly distributed by J. Hey) was used for HKA tests. A program for performing the *H*-test (Fay and Wu 2000), kindly provided by J. C. Fay, was used. No back-mutations were assumed when calculating *P* values for the *H*-test.

The ancestral protein-coding DNA sequences to humans and chimpanzees were reconstructed by maximum parsimony and maximum likelihood approaches using the programs ''pamp'' and ''baseml'' in the PAML 3.0d package (Yang 1997), with the orangutan sequence as an outgroup. In all cases, the two approaches resulted in identical ancestral sequences. The program ''dists'' in the ODEN package (Ina 1994) was used for estimation of numbers of synonymous substitutions per site (d_S) and

Table 2
Polymorphisms in Humans and in Chimpanzees at 10 X-linked Loci with *Xq13′*

	Human							Chimpanzee						
	Length	*S*	π (%)	θ_w (%)	D_{Tajima}	$D_{Fu\&Li}$	$H_{Fay\&Wu}$	Length	*S*	π (%)	θ_w (%)	D_{Tajima}	$D_{Fu\&Li}$	$H_{Fay\&Wu}$
ATRX	2,947	0	0	0	—	—	—	2,950	9	0.096	0.108	−0.464	−0.911	1.689
	(2,837)	(0)	(0)	(0)	(—)	(—)	(—)	(2,840)	(9)	(0.100)	(0.112)	(−0.464)	(−0.911)	(1.689)
FMR2	2,994	3	0.016	0.028	−1.191	−1.382	−1.063	2,994	7	0.073	0.083	−0.508	−0.924	1.244
	(1,879)	(2)	(0.020)	(0.030)	(−0.812)	(−0.661)	(−1.158*)	(1,879)	(5)	(0.076)	(0.094)	(−0.783)	(−0.951)	(1.067)
GDI1	3,470	11	0.056	0.089	−1.313	−2.992***	0.853	3,468	8	0.058	0.082	−1.229	−1.738	1.600
	(2,495)	(10)	(0.057)	(0.113)	(−1.727*)	(−3.361***)	(0.853)	(2,493)	(8)	(0.081)	(0.113)	(−1.229)	(−1.738)	(1.600)
IL1RAPL1	3,329	6	0.030	0.051	−1.293	−2.562**	0.221	3,330	12	0.085	0.127	−1.512	−2.297**	2.311
	(2,492)	(5)	(0.036)	(0.057)	(−1.094)	(−2.26**)	(0.126)	(2,493)	(11)	(0.105)	(0.156)	(−1.458)	(−2.187**)	(2.133)
L1CAM	3,402	7	0.033	0.058	−1.429	−1.415	0.968	3,409	3	0.028	0.031	−0.356	0.066	0.356
	(2,025)	(7)	(0.055)	(0.097)	(−1.429)	(−1.415)	(0.968)	(2,032)	(3)	(0.047)	(0.052)	(−0.356)	(0.066)	(0.356)
OPHN1	4,013	5	0.041	0.035	0.551	−0.515	0.726	4,050	13	0.109	0.113	−0.171	0.096	1.956
	(3,853)	(5)	(0.043)	(0.037)	(0.551)	(−0.515)	(0.726)	(3,890)	(13)	(0.114)	(0.118)	(−0.171)	(0.096)	(1.956)
PAK3	2,873	3	0.014	0.029	−1.441	−1.382	0.358	2,875	15	0.145	0.184	−0.977	−1.893*	2.578
	(2,438)	(3)	(0.016)	(0.035)	(−1.441)	(−1.382)	(0.358)	(2,440)	(14)	(0.163)	(0.203)	(−0.899)	(−1.777)	(2.400)
RPS6KA3	3,280	1	0.003	0.009	−1.164	−1.596	0.095	3,279	2	0.020	0.022	−0.184	−0.410	0.444
	(3,038)	(1)	(0.003)	(0.009)	(−1.164)	(−1.596)	(0.095)	(3,037)	(2)	(0.022)	(0.023)	(−0.184)	(−0.410)	(0.444)
TM4SF2	3,172	8	0.061	0.071	−0.460	0.162	0.316	3,167	6	0.071	0.067	0.241	−0.596	0.711
	(2,674)	(7)	(0.069)	(0.074)	(−0.210)	(0.679)	(0.221)	(2,669)	(5)	(0.063)	(0.066)	(−0.178)	(−0.951)	(0.711)
TNRC11	3,049	4	0.034	0.037	−0.246	0.128	0.800	3,048	4	0.057	0.046	0.867	1.272	−0.533
	(1,999)	(4)	(0.052)	(0.056)	(−0.246)	(0.128)	(0.800)	(1,998)	(3)	(0.059)	(0.053)	(0.398)	(1.153)	(−0.533)
Xq13′	10,163	20	0.035	0.055	−1.387	−1.395	—[a]	10,151	60	0.180	0.209	−0.676	−1.378	—[a]

NOTE.—Length: total number of compared nucleotide sites (including exons and introns); *S*: number of segregating sites; π: average number of nucleotide differences per site between two sequences; θ_w: nucleotide diversity based on the proportion of segregating sites in a sample; D_{Tajima}: Tajima's *D* (Tajima 1989); $D_{Fu\&Li}$: Fu and Li's *D* (Fu and Li 1993); $H_{Fay\&Wu}$: Fay and Wu's *H* (Fay and Wu 2000). Values in parentheses are estimated from noncoding regions. Asterisks indicate the significance level (*10%, **5%, ***2%).

[a] Because determination of ancestral sites is complex, estimation was not done.

numbers of nonsynonymous substitutions per site (d_N) (Nei and Gojobori 1986). A likelihood ratio test was applied with the ''codeml'' program in the PAML 3.0d package (Yang 1997).

For comparisons to the 10 genes studied here, human, chimpanzee, and orangutan coding regions longer than 100 bp were retrieved from GenBank. After exclusion of members of multigene families and recently duplicated genes, 113 genes remained.

Xq13 Subset Data

To compare our sequence data to a previously sequenced data set of a nontranscribed region at Xq13.3 (*Xq13′* in this article), we chose 20 of the 70 human sequences in Kaessmann et al. (1999), such that the nine language phyla sampled for the X-chromosomal loci were all represented. The individuals sampled were an Aboriginal Australian (AJ241023), a Warao South American Indian (AJ241059), a Chinese (AJ241061), a Japanese (AJ241067), a Thai (AJ241069), a Papua New Guinean Lowlander (AJ241073), a Papua New Guinean Highlander (AJ241084), a Nasioi (AJ241081), an Iranian (AJ241036), a Frenchman (AJ241041), a German (AJ241083), a Briton (AJ241044), an Italian (AJ241042), a Mbuti Pygmy (AJ241085), a Biaka Pygmy (AJ241027), a second Biaka Pygmy (AJ241090), an Ibo (AJ241028), a Yoruba (AJ241091), an Effik (AJ241032), and a Hausa (AJ241045). For western chimpanzees, we chose the *Xq13′* sequenced from the same three individuals (AJ270077, AJ270079, AJ270087) sequenced in this study. For central chimpanzees, we chose same three individuals (AJ270061, AJ270065, AJ270072) sequenced here, as well as four randomly chosen additional individuals (AJ270062, AJ270068, AJ270069, AJ270071).

Results

Levels of polymorphism for each of the 10 X-chromosomal genes studied are summarized in table 2. The average nucleotide diversity (π) in humans and chimpanzees was 0.029% and 0.074%, respectively. Nucleotide diversity in humans ranged from zero at *ATRX* to 0.061% at *TM4SF2*, whereas in chimpanzees it ranged from 0.020% at *RPS6KA3* to 0.145% at *PAK3*. Thus, overall, nucleotide diversity is lower in humans than in chimpanzees, and this is reflected both in coding and noncoding parts of the genes (see also table 3). Interestingly, the nucleotide diversity across genes is not correlated in humans and chimpanzees either when measured as π or when measured as θ_w (fig. 1*A* and *B*).

The average level of nucleotide divergence between humans and chimpanzees was 0.588% (table 3). It ranged from 0.319% at *RPS6KA3* to 0.913% at *FMR2*. The observed divergence between humans and chimpanzees was used to estimate the neutral mutation rate (μ) for each gene. Assuming a species divergence time of 5 MYA and a generation time of 20 years, the average value was 1.18×10^{-8} per nucleotide per generation (table 4). The estimates of the effective population size (N_e) vary between 1,600 and 15,700 for humans and 10,400 and 63,100 for chimpanzees, and the estimates of coalescence times (T_{mode}) vary between 68,000 and 682,000 years for humans and 433,000 and 2,095,000 years for chimpanzees.

Table 3
Sequence Divergence Between Humans and Chimpanzees at 10 X-linked Mental Retardation Loci with *Xq13′*

Locus	Length	*FD*	D_{HC} (%)	*Rπ*	P_H	P_C
ATRX	2,947	9	0.366	—	—	0.262
	(2,837)	(9)	(0.380)	(—)	(—)	(0.262)
FMR2	2,994	25	0.913	4.56	0.018	0.080
	(1,879)	(20)	(1.160)	(3.80)	(0.017)	(0.066)
GDI1	3,468	18	0.593	1.04	0.094	0.098
	(2,493)	(17)	(0.760)	(1.42)	(0.075)	(0.106)
IL1RAPL1	3,325	26	0.851	2.83	0.035	0.100
	(2,488)	(25)	(1.090)	(2.92)	(0.033)	(0.096)
L1CAM	3,401	12	0.390	0.85	0.085	0.072
	(2,024)	(9)	(0.510)	(0.85)	(0.108)	(0.093)
OPHN1	4,010	26	0.751	2.66	0.055	0.145
	(3,850)	(24)	(0.730)	(2.65)	(0.059)	(0.156)
PAK3	2,871	8	0.383	10.36	0.037	0.379
	(2,436)	(7)	(0.410)	(10.19)	(0.039)	(0.401)
RPS6KA3	3,278	10	0.319	6.67	0.009	0.063
	(3,036)	(10)	(0.340)	(7.33)	(0.009)	(0.064)
TM4SF2	3,167	20	0.739	1.16	0.083	0.096
	(2,669)	(20)	(0.860)	(0.91)	(0.081)	(0.074)
TNRC11	3,048	15	0.571	1.68	0.060	0.100
	(1,998)	(13)	(0.750)	(1.13)	(0.070)	(0.079)
Xq13′	10,140	78	0.929	5.14	0.038	0.194

NOTE.—Length: total number of compared nucleotide sites; *FD*: fixed nucleotide difference between humans and chimpanzees; D_{HC}: average number of nucleotide substitutions per site between humans and chimpanzees; *Rπ*: ratio of π (chimpanzee/human); P_H: proportion of human π/D_{HC}; P_C: proportion of chimpanzee π/D_{HC}. Values in parentheses are estimated for noncoding regions.

Tajima's *D* (table 2) is negative for all nine loci where it can be calculated in humans (*ATRX* lacked variable positions in the humans sampled) and for eight of 10 loci in chimpanzees. However, none of the *D* values differ significantly from the neutral expectation of zero. For two loci (*GDI1* and *IL1RAPL1*) in humans and two loci (*IL1RAPL1* and *PAK3*) in chimpanzees, the *P* values for Fu and Li's *D* were less than 0.05 (uncorrected for multiple tests).

When the Tajima's *D* values for the human genes (except *ATRX*) and *Xq13′* are compared to those for the chimpanzee genes (fig. 1*C*), they tend to be positively correlated (Pearson's $R^2 = 0.38$, $P = 0.06$, two-tailed), although not statistically significantly so. Furthermore, Fu and Li's *D* is significantly positively correlated between humans and chimpanzees (fig. 1*D*, $R^2 = 0.54$, $P = 0.02$).

We performed a single, 10-locus HKA test of the neutral expectation of equal ratios of polymorphism to divergence among genes (Hudson, Kreitman, and Aguadé 1987). Even though the polymorphism and divergence levels varied among the loci, the HKA test did not reject the null hypothesis of equal ratios across loci ($\chi^2 = 9.728$, df = 9, $P = 0.37$ for humans; $\chi^2 = 9.487$, df = 9, $P = 0.39$ for chimpanzees). Table 5 shows the numbers of nonsynonymous and synonymous substitutions, as well as the ratio of the nonsynonymous and synonymous rates per position (d_N/d_S), for the evolutionary lineages leading to humans, chimpanzees, and orangutans, respectively. The overall d_N/d_S ratios are very similar on the human and chimpanzee lineages. The fact that the humans and chimpanzees show similar extents of functional constraint is consistent with the correlation of *D* values in the two species. Interestingly, d_N/d_S ratios are lower on the orangutan lineage than on the human and chimpanzee lineages ($P < 0.05$, Fisher's exact test). This may indicate that the overall level of selective constraint is similar in humans and chimpanzees, whereas it is higher in orangutans. This could be the result of a higher effective population size and the absence of population growth in orangutans (Kaessmann et al. 2001). Alternatively, it could indicate positive selection on genes involved in cognitive functions in humans and chimpanzees. When the 113 genes for which human, chimpanzee, and orangutan sequences are available in databases are analyzed in a similar way, d_N/d_S ratios on the human, chimpanzee, and orangutan lineages are 0.31, 0.33, and 0.29, respectively. Thus, genes involved in cognitive functions seem to be more constrained than average genes in the genomes in all three lineages. Furthermore, because d_N/d_S ratios for average genes do not differ between the orangutan lineage and the human and chimpanzee lineages, positive selection may have affected some portion of the genes involved in cognitive functions in human and chimpanzees. Alternatively, such genes may be under more constraints in orangutans than in humans and chimpanzees. However, caution is warranted with respect to this observation because it is based on few genes.

Discussion

Nucleotide Diversity

The nucleotide diversity determined by the SNP (Single Nucleotide Polymorphisms) Consortium (Sachidanandam et al. 2001) for X-chromosomal loci is 0.047%, and the corresponding value for the 10 X-chromosomal genes reported here is 0.029% (range: 0%–0.061%; table 2).

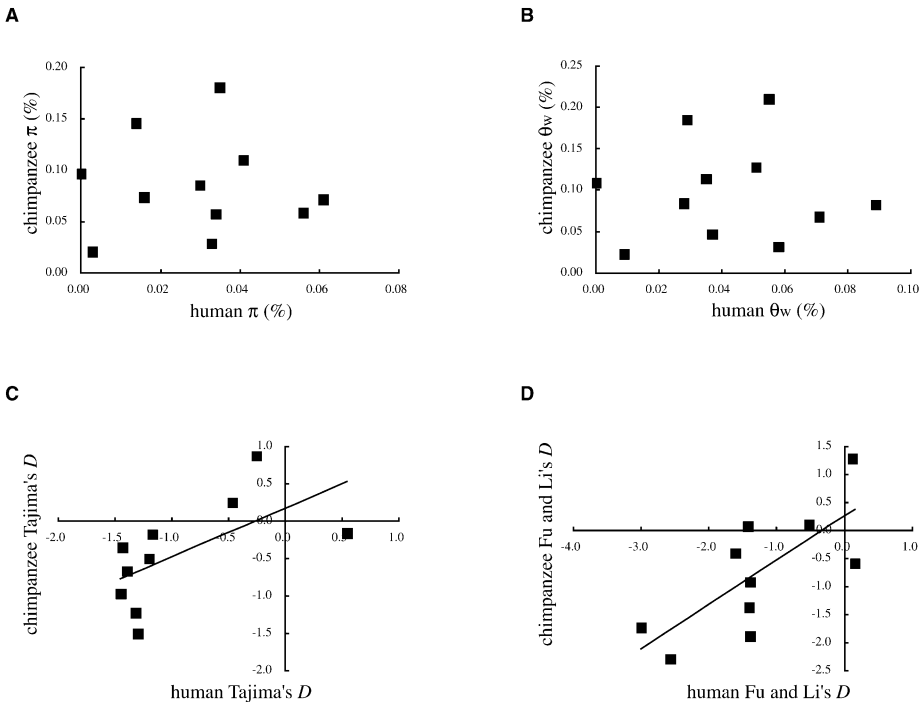

FIG. 1.—Scatterplots of π (*A*), θ_w (*B*), Tajima's *D* (C), and Fu and Li's *D* (D) of 11 X-linked regions for humans and chimpanzees. Ten X-linked mental retardation genes (in cases of Tajima's *D* and Fu and Li's *D*, without *ATRX*) and *Xq13′* are plotted. For Tajima's *D* (*C*) and Fu and Li's *D* (*D*); linear regressions are indicated by solid lines.

Furthermore, the average nucleotide diversity (π) for the noncoding parts of 12 X-chromosomal genes studied in humans by Nachman (2001) is 0.057% (range: 0%–0.19%), and the corresponding value for the 10 genes studied here is 0.047% (ranging from 0% to 0.092%). Thus, overall, nucleotide diversity in these genes tends to be low.

In chimpanzees, the average level of nucleotide diversity (table 2) for the 10 genes also seems low (0.074%) when compared to *Xq13′*, the only locus for which comparable data are available, where it is 0.18%. Because mutations in the genes analyzed here have been shown to cause mental retardation in humans, it is reasonable to assume that they play important roles in neural activities. Hence, polymorphism levels might be lowered by selective constraints. Consistent with this, the average level of sequence divergence between humans and chimpanzees for the 10 genes is low (0.59%) when compared to 12 loci on the X chromosome (summarized by Nachman 2001) (mean 0.87%, range 0.26%–1.63%).

We also estimated nucleotide diversity (π) for three human groups (Asian, European, and African) and two chimpanzee ''subspecies'' (western and central) using the total concatenated sequences (32,509 bp). Among the humans, the Africans showed higher diversity ($\pi = 0.033\%$) than the Asians ($\pi = 0.023\%$) and Europeans ($\pi = 0.025\%$), and the central chimpanzee samples showed higher diversity ($\pi = 0.073\%$) than the western chimpanzee samples ($\pi = 0.025\%$). Further, the diversity in the western chimpanzees is comparable with that of the humans. These results are compatible with previous studies in humans (Kaessmann et al. 1999; Nachman and Crowell 2000; Zhao et al. 2000; Friez et al. 2001; Yu et al. 2001), and chimpanzees (Kaessmann, Wiebe, and Pääbo 1999; Kaessmann et al. 2001).

We used the mutation rates estimated from divergence data and the nucleotide diversity estimated from polymorphism data to estimate the effective population size (N_e) and coalescence time (*T*) for each gene, assuming no recombination (table 4). For humans, the average mutation rate (μ) is 1.18×10^{-8}, and the average effective population size is 8,800, similar to the estimates reported by Nachman et al. (1998) for seven X-chromosomal human genes. When estimated from autosomal loci, the effective population size of humans is around 10,000 (Takahata 1993; Zhao et al. 2000; Yu et al. 2001). Because the number of X chromosomes in a population is three-fourths that of autosomes, the effective population size estimated from X-chromosomal loci is expected to be lower, provided that women and men have similar reproductive behavior. The observation that the effective size estimated from the X-chromosomal data is lower than that for autosomal loci, but higher than the theoretical expectation of three quarters, suggests that male reproductive behavior in the past has tended to reduce the male effective population size.

The average effective population size of chimpanzees is 23,200, about three times higher than that of humans.

Nucleotide Diversity and Recombination Rate

Nucleotide diversity and recombination rate have been found to be positively correlated in humans (Nachman 2001; Lercher and Hurst 2002; Waterston, Lindblad-Toh, and Birney 2002). However, in our data, we do not

Table 4
Mutation Rate, Effective Population Size, and Coalescence Time Estimated from Each of 10 X-linked Mental Retardation Loci

	Human				Chimpanzee		
Locus	μ	N_e	T_{mean}	T_{mode}	N_e	T_{mean}	T_{mode}
ATRX	7.32×10^{-9}	—	—	—	43,700	1,843,000	1,574,000
FMR2	1.83×10^{-8}	2,900	199,000	147,000	13,300	554,000	464,000
GDI1	1.19×10^{-8}	15,700	663,000	548,000	16,300	619,000	496,000
IL1RAPL1	1.70×10^{-8}	5,900	355,000	287,000	16,600	540,000	473,000
L1CAM	7.80×10^{-9}	14,100	632,000	468,000	12,000	768,000	582,000
OPHN1	1.50×10^{-8}	9,100	606,000	528,000	24,200	960,000	910,000
PAK3	7.66×10^{-9}	6,100	327,000	230,000	63,100	2,311,000	2,095,000
RPS6KA3	6.38×10^{-9}	1,600	105,000	68,000	10,400	627,000	433,000
TM4SF2	1.48×10^{-8}	13,800	824,000	682,000	16,000	948,000	807,000
TNRC11	1.14×10^{-8}	9,900	551,000	409,000	16,600	940,000	785,000
Mean	1.18×10^{-8}	8,800	473,000	374,000	23,200	1,011,000	862,000

NOTE.—Mutation rate (μ) per nucleotide per generation was estimated from the divergence data between humans and chimpanzees, assuming a divergence time of 5 MYA and a generation time of 20 years. Effective population size (N_e) was estimated from the neutral expectation for these genes, $\pi = 3\ N_e\ \mu$.

observe a significant positive correlation between nucleotide diversity (π) and the recombination rate (Pearson's product moment correlation coefficient, two-tailed *t*-test, $R^2 = 0.10$, $P = 0.37$), or between θ_w and the recombination rate ($R^2 = 0.01$, $P = 0.85$) (fig. 2*A* and *B*). This is likely to occur because the range of recombination rates of the regions studied here is quite narrow, and because the number of chromosomes analyzed is relatively small.

Interspecies Correlation of *D* Values

Of nine genes in humans, eight and seven showed negative Tajima's *D* and Fu and Li's *D* values, respectively. Similarly, of 10 genes in chimpanzees, eight and seven showed negative Tajima's *D* and Fu and Li's *D* values, respectively. That *D* values tend to be negative in both species may indicate that chimpanzee and human demographic history is characterized by population growth. It will be interesting to see if that is also true for gorillas and orangutans, where *D* values from the single region studied to date (*Xq13′*) are less negative than in humans and chimpanzees (Kaessmann et al. 2001).

We find that Tajima's *D* values among the 10 loci are positively correlated in chimpanzees and humans (fig. 1*C*). This is surprising in view of the fact that humans and chimpanzees are unlikely to have exactly the same demographic history. To see if this is the case in another pair of closely related organisms, we plotted Tajima's *D* values for eight noncoding loci of three autosomal and five X-chromosomal loci from *Drosophila melanogaster* and *Drosophila simulans* (Moriyama and Powell 1996). In this case, the values were not correlated ($R^2 < 0.01$, $P = 0.97$). Taken at face value, the correlation in *D* values in chimpanzees and humans may be explained by the occurrence of very frequent selective sweeps in the same recombinational environment in the two species. Under this scenario, however, we would expect diversity levels to be correlated in the two species (Braverman et al. 1995), which is not the case (fig. 1*A* and *B*). Another possibility is that among the alleles that segregate in the human and chimpanzee populations, a large fraction are weakly deleterious. If such slightly deleterious alleles have very similar distributions of scaled selection coefficients in the two species, this might account for the correlation in *D* values.

Table 5
Evolutionary Rates of Gene Coding Regions at 10 X-linked Mental Retardation Loci

Gene	Compared Sites (bp)	*H* Branch	*C* Branch	*O* Branch
ATRX	7,476	2/8 (0.066)	4/7 (0.150)	13/34 (0.100)
FMR2	3816	6/3 (0.598)	2/3 (0.199)	8/14 (0.169)
GDI1	1,341	0/2 (0)	0/2 (0)	1/8 (0.038)
IL1RAPL1	2,088	0/4 (0)	0/1 (0)	1/8 (0.037)
L1CAM	3,765	2/3 (0.210)	1/3 (0.105)	3/42 (0.022)
OPHN1	2,406	1/1 (0.289)	0/2 (0)	0/12 (0)
PAK3	1,677	0/4 (0)	0/1 (0)	0/8 (0)
RPS6KA3	2,178	0/2 (0)	0/0 (—)	0/11 (0)
TM4SF2	732	0/0 (—)	0/0 (—)	4/1 (1.296)
TNRC11	6,081	0/6 (0)	3/11 (0.085)	0/40 (0)
Total	31,560	11/33 (0.109)	10/30 (0.095)	30/178 (0.048)

NOTE.—Numbers of nonsynonymous/synonymous (ratio) substitutions for each branch of 10 X-linked mental retardation related gene coding regions. Values for total were estimated using total concatenated sequences. Ratios (shown in parentheses) are estimated from d_N/d_S by Nei and Gojobori's (1986) methods. *H*, *C*, and *O* branches mean branches connecting human, chimpanzee, and orangutan, respectively.

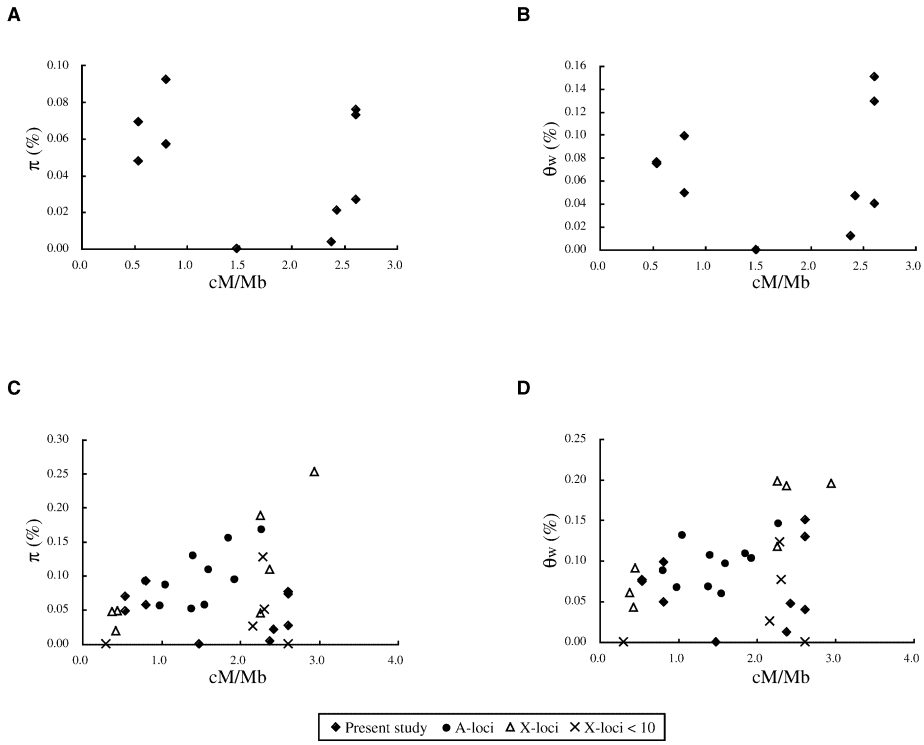

FIG. 2.—Scatterplots of nucleotide diversity versus recombination rate in humans. Panels *A* and *B* show nucleotide diversity expressed as π and as θ_w, respectively, versus recombination rate for the 10 genes in the present study. Panels *C* and *D* show π and θ_w, respectively, versus. recombination rate for these 10 genes, as well as genes studied by Nachman (2001). These are: 10 autosomal genes (A-loci), seven X-chromosomal genes (X-loci), and five X-chromosomal genes excluded by Nachman because of small sample size (X-loci < 10). For all genes, π and θ_w values were estimated from noncoding regions. For the present study, these were: *ATRX* (2,837 bp, $\pi = 0$, $\theta_w = 0$), *FMR2* (1,879 bp, $\pi = 0.020$, $\theta_w = 0.030$), *GDI1* (2,495 bp, $\pi = 0.057$, $\theta_w = 0.113$), *IL1RAPL1* (2,633 bp, $\pi = 0.038$, $\theta_w = 0.064$), *L1CAM* (2,025 bp, $\pi = 0.055$, $\theta_w = 0.097$), *OPHN1* (3,853 bp, $\pi = 0.043$, $\theta_w = 0.037$), *PAK3* (2,415 bp, $\pi = 0.016$, $\theta_w = 0.035$), *RPS6KA3* (3,038 bp, $\pi = 0.003$, $\theta_w = 0.009$), *TM4SF2* (2,674 bp, $\pi = 0.069$, $\theta_w = 0.074$), *TNRC11* (1,999 bp, $\pi = 0.052$, $\theta_w = 0.056$). Then, values for recombination rates and nucleotide diversities were corrected following Nachman (2001).

It should be noted that our results must be interpreted with caution as they are based on a small number of genes. However, if further work confirms that an interspecies correlation in *D* values exists, then the latter scenario would seem to be the most likely explanation.

Coding Regions

The genes *FMR2* and *TM4SF2* showed higher numbers of nonsynonymous changes than of synonymous ones on the human and orangutan evolutionary lineages, respectively. In the case of *TM4SF2*, a partial sequence (AB047628) from *Macaca fascicularis* can be used to show that all four nonsynonymous substitutions occurred on the orangutan lineage. Moreover, these four nonsynonymous substitutions all occurred within an eight–amino acid segment in the protein, suggesting that this region of the protein has changed its function during orangutan evolution.

In the case of the *FMR2* gene, a higher number of nonsynonymous substitutions than of synonymous substitutions was found on the human lineage but not on the chimpanzee and orangutan lineages. Although the difference in d_N/d_S ratios on the human and the chimpanzee and orangutan lineages is not significant ($P = 0.12$) in a likelihood ratio test (Yang 1998), it is noteworthy that four of six human nonsynonymous changes are located around two nuclear localization signals (Gecz et al. 1997). Furthermore, the *FMR2* gene showed relatively low polymorphism in humans but not in chimpanzees. Thus, *FMR2* is a candidate for investigation of possible changes in function on the human lineage, and *TM4SF2* is a candidate for such an event on the orangutan lineage.

Conclusions

Mutations in the 10 genes analyzed here cause mental retardation in humans either as a single clinical feature or in combination with other symptoms. The genes are thus all involved in cognitive functions. It may therefore be of interest to investigate whether any indications of a changed function or positive selection can be detected in these genes, particularly on the human evolutionary line. With the possible exception of *FMR2*, no candidate can be identified by the approaches taken here. However, it seems clear that these genes are more conserved than average genes, a fact that may in principle facilitate the identification of positive selection when larger genomic regions and more chromosomes are investigated. We anticipate that in the near future the availability of a chimpanzee genome sequence and a human haplotype map will allow a genome-wide search for evidence of selection in genes such as these.

Acknowledgments

We thank H. Kaessmann, J. Kidd, P. A. Morin, M. Stoneking, R. Ward, and J. Wickings for preparation of genomic DNAs; P. Khaitovich for preparation of mRNAs; D. Serre for help in cloning; and especially M. Przeworski for suggestions and help with analyses. This work was funded by the Bundesministerium für Bildung und Forschung and the Max Planck Society, Leipzig, Germany.

Literature Cited

Allen, K. M., J. G. Gleeson, S. Bagrodia, M. W. Partington, J. C. MacMillan, R. A. Cerione, J. C. Mulley, and C. A. Walsh. 1998. PAK3 mutation in nonsyndromic X-linked mental retardation. Nat. Genet. **20**:25–30.

Billuart, P., T. Bienvenu, N. Ronce et al. (15 co-authors). 1998. Oligophrenin-1 encodes a rhoGAP protein involved in X-linked mental retardation. Nature **392**:823–826.

Braverman, J. M., R. R. Hudson, N. L. Kaplan, C. H. Langley, and W. Stephan. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics **140**:783–796.

Carrie, A., L. Jun, T. Bienvenu et al. (19 co-authors). 1999. A new member of the IL-1 receptor family highly expressed in hippocampus and involved in *X*-linked mental retardation. Nat. Genet. **23**:25–31.

Castellvi-Bel, S., and M. Mila. 2001. Genes responsible for nonspecific mental retardation. Mol. Genet. Metab. **72**:104–108.

Chelly, J. 1999. Breakthroughs in molecular and cellular mechanisms underlying X-linked mental retardation. Hum. Mol. Genet. **8**:1833–1838.

Crouau-Roy, B., S. Service, M. Slatkin, and N. Freimer. 1996. A fine-scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium and sequence analysis. Hum. Mol. Genet. **5**:1131–1137.

D'Adamo, P., A. Menegon, and C. Lo Nigro et al. (12 co-authors). 1998. Mutations in GDI1 are responsible for X-linked non-specific mental retardation. Nat. Genet. **19**:134–139.

Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics **155**:1405–1413.

Friez, M. J., F. B. Essop, A. Krause et al. (12 co-authors). 2000. Evidence that a dodecamer duplication in the gene HOPA in Xq13 is not associated with mental retardation. Hum. Genet. **106**:36–39.

Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693–709.

Gagneux, P., C. Wills, U. Gerloff, D. Tautz, P. A. Morin, C. Boesch, B. Fruth, G. Hohmann, O. A. Ryder, and D. S. Woodruff. 1999. Mitochondrial sequences show diverse evolutionary histories of African hominoids. Proc. Natl. Acad. Sci. USA **96**:5077–5082.

Gecz, J., S. Bielby, G. R. Sutherland, and J. C. Mulley. 1997. Gene structure and subcellular localization of FMR2, a member of a new family of putative transcription activators. Genomics **44**:201–213.

Gecz, J., A. K. Gedeon, G. R. Sutherland, and J. C. Mulley. 1996. Identification of the gene FMR2, associated with FRAXE mental retardation. Nat. Genet. **13**:105–108.

Gibbons, R. J., D. J. Picketts, L. Villard, and D. R. Higgs. 1995. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell **80**:837–845.

Griffiths, R. C., and S. Tavare. 1995. Unrooted genealogical tree probabilities in the infinitely-many-sites model. Math. Biosci. **127**:77–98.

Gu, Y., Y. Shen, R. A. Gibbs, and D. L. Nelson. 1996. Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. Nat. Genet. **13**:109–113.

Hudson, R. R., M. Kreitman, and M. Aguadé. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

Ina, Y. 1994. ODEN: a program package for molecular evolutionary analysis and database search of DNA and amino acid sequences. Comput. Appl. Biosci. **10**:11–12.

Jouet, M., A. Rosenthal, G. Armstrong, J. MacFarlane, R. Stevenson, J. Paterson, A. Metzenberg, V. Ionasescu, K. Temple, and S. Kenwrick. 1994. X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked hydrocephalus result from mutations in the L1 gene. Nat. Genet. **7**:402–407.

Kaessmann, H., F. Heissig, A. von Haeseler, and S. Pääbo. 1999. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. Nat. Genet. **22**:78–81.

Kaessmann, H., V. Wiebe, and S. Pääbo. 1999. Extensive nuclear DNA sequence diversity among chimpanzees. Science **286**:1159–1162.

Kaessmann, H., V. Wiebe, G. Weiss, and S. Pääbo. 2001. Great ape DNA sequences reveal a reduced diversity and an expansion in humans. Nat. Genet. **27**:155–156.

Lercher M. J., and L. D. Hurst. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. Trends Genet. **18**:337–340.

Merienne, K., S. Jacquot, S. Pannetier, M. Zeniou, A. Bankier, J. Gecz, J. L. Mandel, J. Mulley, P. Sassone-Corsi, and A. Hanauer. 1999. A missense mutation in RPS6KA3 (RSK2) responsible for non-specific mental retardation. Nat. Genet. **22**:13–14.

Moriyama, E. N., and J. R. Powell. 1996. Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13**:261–277.

Nachman, M. W. 2001. Single nucleotide polymorphisms and recombination rate in humans. Trends. Genet. **17**:481–485.

Nachman, M. W., V. L. Bauer, S. L. Crowell, and C. F. Aquadro. 1998. DNA variability and recombination rates at X-linked loci in humans. Genetics **150**:1133–1141.

Nachman, M. W., and S. L. Crowell. 2000. Contrasting evolutionary histories of two introns of the Duchenne muscular dystrophy gene, DMD, in humans. Genetics **155**:1855–1864.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

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

Payseur, B. A., and M. W. Nachman. 2000. Microsatellite variation and recombination rate in the human genome. Genetics **156**:1285–1298.

Philibert, R. A., B. H. King, S. Winfield et al. (12 co-authors). 1998. Association of an X-chromosome dodecamer insertional variant allele with mental retardation. Mol. Psychiatry **3**:303–309.

Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15**:174–175.

Sachidanandam, R., D. Weissman, S. C. Schmidt et al. (38 co-authors). 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409**:928–933.

Tajima, F. 1989. Statistical method for testing the neutral

mutation hypothesis by DNA polymorphism. Genetics **123**:585–595.
Takahata, N. 1993. Allelic genealogy and human evolution. Mol. Biol. Evol. **10**:2–22.
Thompson, J. D., T. J. Gibson, and D. G. Higgins. 1994. ClsutalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids. Res. **22**:4673–4680.
Toniolo, D., and P. D'Adamo. 2000. X-linked non-specific mental retardation. Curr. Opin. Genet. Dev. **10**:280–285.
Vits, L., G. Van Camp, P. Coucke et al. (13 co-authors). 1994. MASA syndrome is due to mutations in the neural cell adhesion gene L1CAM. Nat. Genet. **7**:408–413.
Waterston, R. H., K. Lindblad-Toh, E. Birney et al. (222 co-authors). 2002. Initial sequencing and comparative analysis of the mouse genome. Nature **420**:520–562.
Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7**:256–276.
Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **13**:555–556.
Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.
Yu, N., Z. Zhao, Y. X. Fu et al. (8 co-authors). 2001. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. Mol. Biol. Evol. **18**:214–222.
Zechner, U., M. Wilda, H. Kehrer-Sawatzki, W. Vogel, R. Fundele, and H. Hameister. 2001. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? Trends Genet. **17**:697–701.
Zemni, R., T. Bienvenu, M. C. Vinet et al. (23 co-authors). 2000. A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation. Nat. Genet. **24**:167–170.
Zhao, Z., L. Jin, Y. X. Fu et al. (13 co-authors). 2000. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. USA **97**:11354–11358.

Edward Holmes, Associate Editor

Accepted March 25, 2003