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

Takashi Kitano, 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 (Zechnner 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, ILIRAPLI, 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; LICAM, 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, Nasion, 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 (Australoid, Amerind, Sino-Tibetan, Altaic, Austro-Asiatic, Indo-Pacific, Indo-Hittite, Nger-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 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...
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 as described above. Complete gene coding regions for eight loci were sequenced. The other two loci (LICAM and RP56K3) were not sequenced completely, because their 5′ non-coding 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 version 1.8 (Thompson, Gibson, and Higgins 1994) was used for multiple alignments. Two measures of nucleotide variability, \( \pi \) (Nei 1987) and \( \theta_w \) (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 \( \pi \) and \( \theta_w \) estimate the neutral parameter 3N_eμ for X-chromosomal loci, where \( N_e \) is the effective population size and \( \mu \) 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 (\( \pi \)) is larger than expected from the observed number of polymorphic sites (\( \theta_w \)). The expected difference (\( D \)) between \( \pi \) and \( \theta_w \) 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 \( \pi \) (which is influenced most by variants at intermediate frequencies) and \( \theta_w \) (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 \( \theta_w \). 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 1

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Location</th>
<th>cM/Mb</th>
<th>Disease</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATRX</td>
<td>Xq13.1-q21.1</td>
<td>2.22 (0.86–3.57)</td>
<td>NSMR</td>
<td>Gibbons et al. (1995)</td>
</tr>
<tr>
<td>FMR2</td>
<td>Xq28</td>
<td>3.92</td>
<td>NSMR</td>
<td>Gu et al. (1996); Geccz et al. (1996)</td>
</tr>
<tr>
<td>GD1</td>
<td>Xq28</td>
<td>3.92</td>
<td>NSMR</td>
<td>D’Adamo et al. (1998)</td>
</tr>
<tr>
<td>IL1RAPL1</td>
<td>Xp22.1-p21.3</td>
<td>0.81 (0.45–1.17)</td>
<td>NSMR</td>
<td>Carrie et al. (1999)</td>
</tr>
<tr>
<td>LICAM</td>
<td>Xq28</td>
<td>3.92</td>
<td>MASA</td>
<td>Jouet et al. (1994); Vits et al. (1994)</td>
</tr>
<tr>
<td>OP1N1</td>
<td>Xq12</td>
<td>1.21 (0.85–1.57)</td>
<td>NSMR</td>
<td>Billuart et al. (1998)</td>
</tr>
<tr>
<td>PAK3</td>
<td>Xq21.3-q24</td>
<td>3.64 (3.38–3.90)</td>
<td>NSMR</td>
<td>Allen et al. (1998)</td>
</tr>
<tr>
<td>RP56K3</td>
<td>Xp22.2-p22.1</td>
<td>3.57</td>
<td>NSMR</td>
<td>Merienne et al. (1999)</td>
</tr>
<tr>
<td>TM4SF2</td>
<td>Xq11</td>
<td>1.21 (0.85–1.57)</td>
<td>NSMR</td>
<td>Zemni et al. (2000)</td>
</tr>
<tr>
<td>TNRC11</td>
<td>Xq13</td>
<td>0.81 (0.45–1.17)</td>
<td>NSMR</td>
<td>Philibert et al. (1998)</td>
</tr>
</tbody>
</table>

* Where more than one marker pair exists, the highest and lowest values are given in parentheses. cM/Mb indicates centiMorgans (genetic) per Megabase (physical).

** ATR-X (alpha thalassemia/mental retardation X-linked) syndrome.

β Nonsyndromic mental retardation.

γ MASA (mental retardation, aphasia, shuffling gait, and adducted thumbs) syndrome.
numbers of nonsynonymous substitutions per site (dN) (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 (AJ241069), a Thai (AJ241069), a Papua New Guinea Lowlander (AJ241073), a Papua New Guinea Highlander (AJ241084), a Niasioi (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 Efik (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 d_m (fig. 1A 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′

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

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; R_{π}: ratio of π (chimpanzee/human); P_{HC}: proportion of human π/D_{HC}; P_{PC}: 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. 1C), they tend to be positively correlated (Pearson’s R squared = 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. 1D, R squared = 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 non-synonymous and synonymous substitutions, as well as the ratio of the non-synonymous and synonymous rates per position ($d_{ns}/d_{s}$), for the evolutionary lineages leading to humans, chimpanzees, and orangutans, respectively. The overall $d_{ns}/d_{s}$ 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 ratio of D values in the two species. Interestingly, $d_{ns}/d_{s}$ 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, $d_{ns}/d_{s}$ 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 $d_{ns}/d_{s}$ 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).
Furthermore, the average nucleotide diversity ($\pi$) 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 ($\pi$) 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 ($\mu$) is $1.18 \times 10^{-8}$, 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 data 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...
observed a significant positive correlation between nucleotide diversity ($\pi$) and the recombination rate (Pearson’s product moment correlation coefficient, two-tailed t-test, $R^2 = 0.10, P = 0.37$), or between $\theta_w$ and the recombination rate ($R^2 = 0.01, P = 0.85$) (fig. 2A 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, 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. 1C). 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. 1A 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 4**

<table>
<thead>
<tr>
<th>Locus</th>
<th>$\mu$</th>
<th>$N_e$</th>
<th>$T_{mean}$</th>
<th>$T_{mode}$</th>
<th>$\pi$</th>
<th>$T_{mean}$</th>
<th>$T_{mode}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATRX</td>
<td>$7.32 \times 10^{-9}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>43,700</td>
<td>1,843,000</td>
</tr>
<tr>
<td>FMR2</td>
<td>$1.83 \times 10^{-8}$</td>
<td>2.900</td>
<td>199,000</td>
<td>147,000</td>
<td></td>
<td>13,300</td>
<td>554,000</td>
</tr>
<tr>
<td>GDH1</td>
<td>$1.19 \times 10^{-8}$</td>
<td>15.700</td>
<td>663,000</td>
<td>548,000</td>
<td></td>
<td>16,300</td>
<td>619,000</td>
</tr>
<tr>
<td>ILIRAP1</td>
<td>$1.70 \times 10^{-9}$</td>
<td>5.900</td>
<td>355,000</td>
<td>287,000</td>
<td></td>
<td>16,600</td>
<td>540,000</td>
</tr>
<tr>
<td>L1CAM</td>
<td>$7.80 \times 10^{-9}$</td>
<td>14.100</td>
<td>632,000</td>
<td>468,000</td>
<td></td>
<td>12,000</td>
<td>768,000</td>
</tr>
<tr>
<td>OPHN1</td>
<td>$1.50 \times 10^{-8}$</td>
<td>9.100</td>
<td>606,000</td>
<td>528,000</td>
<td></td>
<td>24,200</td>
<td>960,000</td>
</tr>
<tr>
<td>PAK3</td>
<td>$7.66 \times 10^{-9}$</td>
<td>6.100</td>
<td>327,000</td>
<td>230,000</td>
<td></td>
<td>63,100</td>
<td>2,311,000</td>
</tr>
<tr>
<td>RPS6KA3</td>
<td>$6.38 \times 10^{-9}$</td>
<td>1.600</td>
<td>105,000</td>
<td>68,000</td>
<td></td>
<td>10,400</td>
<td>627,000</td>
</tr>
<tr>
<td>TM4SF2</td>
<td>$1.48 \times 10^{-8}$</td>
<td>13.800</td>
<td>824,000</td>
<td>682,000</td>
<td></td>
<td>16,000</td>
<td>948,000</td>
</tr>
<tr>
<td>TNRC11</td>
<td>$1.14 \times 10^{-8}$</td>
<td>9.900</td>
<td>551,000</td>
<td>409,000</td>
<td></td>
<td>16,600</td>
<td>940,000</td>
</tr>
<tr>
<td>Mean</td>
<td>$1.18 \times 10^{-8}$</td>
<td>8.800</td>
<td>473,000</td>
<td>374,000</td>
<td></td>
<td>23,200</td>
<td>1,011,000</td>
</tr>
</tbody>
</table>

*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.
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


Tajima, F. 1989. Statistical method for testing the neutral
Edward Holmes, Associate Editor
Accepted March 25, 2003