

# Evidence for a Complex Demographic History of Chimpanzees

Anne Fischer, Victor Wiebe, Svante Pääbo, and Molly Przeworski<sup>1</sup>

Max Plank Institute for Evolutionary Anthropology, Leipzig, Germany

To characterize patterns of genomic variation in central chimpanzees (*Pan troglodytes troglodytes*) and gain insight into their evolution, we sequenced nine unlinked, intergenic regions, representing a total of 19,000 base pairs, in 14 individuals. When these DNA sequences are compared with homologous sequences previously collected in humans and in western chimpanzees (*Pan troglodytes verus*), nucleotide diversity is higher in central chimpanzees than in western chimpanzees or in humans. Consistent with a larger effective population size of central chimpanzees, levels of linkage disequilibrium are lower than in humans. Patterns of linkage disequilibrium further suggest that homologous gene conversion may be an important contributor to genetic exchange at short distances, in agreement with a previous study of the same DNA sequences in humans. In central chimpanzees, but not in western chimpanzees, the allele frequency spectrum is significantly skewed towards rare alleles, pointing to population size changes or fine-scale population structure. Strikingly, the extent of genetic differentiation between western and central chimpanzees is much stronger than what is seen between human populations. This suggests that careful attention should be paid to geographic sampling in studies of chimpanzee genetic variation.

## Introduction

Studies of genetic variation in close relatives of humans can help to identify which features of our evolution are unusual and which are shared. Specifically, a comparison of humans and other apes can help to identify targets of natural selection in the human lineage, compare demographic histories, and assess the extent to which recombination and mutation rates are conserved among closely related species (e.g., Kaessmann, Wiebe and Pääbo 1999; Huttley et al. 2000; Nachman and Crowell 2000; Kaessmann et al. 2001; Enard et al. 2002; Yi, Ellsworth, and Li 2002; Wall et al. 2003). In those respects, the study of the closest living relative of humans, the chimpanzee, may be especially enlightening (cf. Olson and Varki [2003]).

Chimpanzees are classified into two species: the common chimpanzee, *Pan troglodytes*, and the pygmy chimpanzee or bonobo, *Pan paniscus*. Among common chimpanzees, three “subspecies” are recognized on the basis of their geographic distribution (Napier and Napier 1967): *Pan troglodytes verus* in western Africa, *P. t. troglodytes* in central Africa, and *P. t. schweinfurthii* in eastern Africa. Recently, a fourth subspecies, *P. t. vellerosus*, has been suggested in western Africa based on mitochondrial DNA divergence (Gonder et al. 1997). Very little is known about the distribution of variation among these subspecies or between species (cf. Morin et al. [1994]).

Indeed, while we now have polymorphism data for extensive regions of the human genome, including coding as well as intergenic regions (Ptak and Przeworski 2002 and references therein), there are few surveys of genetic variation in chimpanzees. Several studies have examined the phylogenetic relationship of humans to other primates (Ruvolo 1997; Gagneux et al. 1999; Chen and Li 2001;

Jensen-Seaman, Deinard, and Kidd 2001) and estimated the sequence divergence (Chen et al. 2001; Ebersberger et al. 2002). However, intraspecific variation among chimpanzees has been studied in few regions: mitochondrial DNA (Ferris et al. 1981; Morin et al. 1994; Wise et al. 1997), the HOXB6 intergenic region (Deinard and Kidd 1999), a nongenic region at Xq13.3 (Kaessmann, Wiebe, and Pääbo 1999; Kaessmann et al. 2001), four genes of the renin-angiotensin system (Dufour et al. 2000), four autosomal and two X chromosomal genes in two individuals (Satta 2001), the male-specific part of the Y chromosome (Stone et al. 2002), and 10 genes on the X chromosome (Kitano et al. 2003). The main conclusions to emerge from these studies are that the mutation rate is similar in humans and chimpanzees, but levels of diversity are twofold to sixfold higher in chimpanzees, reflective of a larger effective population size (e.g., Kaessmann, Wiebe, and Pääbo [1999]).

Many of the regions surveyed in chimpanzees are genic or lie in regions of low or no recombination. Thus, they are more likely affected by natural selection (directly or at linked sites). To study nonselective processes shaping patterns of variation, surveys of intergenic regions in areas of normal to high recombination may be more informative. Recently, short fragments from 50 intergenic regions were sequenced in 17 chimpanzees, including five known to be central and six known to be western (Yu et al. 2003). The authors reported that, in contrast to previous studies, levels of diversity were only 50% higher in chimpanzees than in worldwide samples of humans.

We collected DNA sequence variation from nine intergenic, autosomal regions in 14 central chimpanzees. In each region, we sequenced approximately 1 kb from each end of a 10-kb segment to characterize patterns of linkage disequilibrium as well as of diversity. These regions have previously been sequenced in three human populations (Frisse et al. 2001) and partially in 16 western chimpanzees (Gilad et al. 2003). These data suggest that central chimpanzees are 2-fold to 2.5-fold more diverse than western chimpanzees and worldwide samples of humans. Consistent with this, levels of linkage disequilibrium are lower in central chimpanzees than in humans. Interestingly, there appears to be much stronger genetic

<sup>1</sup>Present address: Department of Ecology and Evolutionary Biology, Brown University, Providence, Rhode Island.

Key words: chimpanzees, chimpanzee subspecies, diversity, population structure, linkage disequilibrium.

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

*Mol. Biol. Evol.* 21(5):799–808. 2004

DOI:10.1093/molbev/msh083

Advance Access publication February 12, 2004

differentiation between western and central subspecies than between human populations. Together with the high proportion of rare alleles observed in central but not western samples, this finding suggests a complex demographic history of chimpanzees.

## Material and Methods

### DNA Samples

We chose 14 unrelated central chimpanzees (*Pan troglodytes troglodytes*) for this study, almost all of which were wild born (in Gabon), and confirmed that they were not close relatives by typing a number of microsatellites (data not shown). One orangutan and one human were sequenced to infer the ancestral alleles and estimate the mutation rate. We also used polymorphism data previously collected by Gilad et al. (2003) for 16 western chimpanzees.

### Genomic Regions and Characteristics

We resequenced nine of 10 genomic regions chosen in a previous study (Frisse et al. 2001) in the 14 central chimpanzees. The 10 regions were selected to have large-scale recombination rates and GC contents close to the genome average (Frisse et al. 2001). One region (region 9) was not amplified in chimpanzees because of a 400-bp complex repeat, which made sequencing difficult. Sequences are available from GenBank under accession numbers AY458900 to AY459151.

The GC content for each region was determined using a Unix script, kindly provided by Ingo Ebersberger. We checked for conserved regions between mouse and human sequences using the Berkeley genome pipeline (<http://pipeline.lbl.gov/>).

To estimate the recombination rates ( $c$ ) of the nine regions studied, we made the assumption that large-scale rates are similar in humans and chimpanzees and used the human recombination rates reported in Kong et al. (2002). This study provides average recombination rates for 3 Mb windows centered on a marker. We assigned each region to the closest marker on this map; these were located 4 kb to 370 kb away from the 10-kb region of interest.

### DNA Extraction

DNA was extracted from 50 ml cell cultures containing between  $5 \times 10^6$  to  $50 \times 10^6$  cells. These cell cultures were obtained from lymphocytes transformed with Epstein-Barr virus. The cell pellets were first resuspended in 5 ml TEN (10 mM Tris/HCl, pH 8.2 and 400 mM NaCl, 2 mM EDTA, pH 8.2). Ten milliliters of TEN, 200 ml 10% SDS (sodium-dodecyl-sulfate) and 100 ml proteinase K were then added. After an overnight incubation at 37°C and cooling for 1 h in the refrigerator, 5 ml of saturated NaCl were added. The solutions were well shaken and centrifuged for 15 min at 4°C (3,500 rpm). The supernatants were transferred in twice their volume of 100% ethanol for DNA precipitation. The DNAs were then washed with 70% ethanol. The dried pellets were redissolved in 1 ml water. This DNA was then treated 1 h with RNase A (10 µg/µl). The RNase and nucleotides were removed by a phenol-chloroform extraction. The

remaining pellets were dissolved in 1 ml water and aliquoted to a concentration of 100 ng/ml for further use.

### PCR Amplification

For all nine regions of lengths ~10 kb, approximately 1 kb was amplified at each end. Polymerase chain reaction (PCR) primers were designed using the human sequence. Amplification reactions were performed in 96-well microtiter plate thermal cycler (Applied Biosystem). The PCR reaction mixture (100 µl) contained a standard buffer (10 mM Tris-HCl, 5 mM MgCl<sub>2</sub>), the four deoxynucleotide triphosphates (0.25 mM each), primers (0.5 pM each), Amplitaq Gold™ (PerkinElmer) and genomic DNA (70 ng). Thermal cycling was performed by touchdown PCR with an initial denaturation step at 95°C for 5 min, followed by 16 cycles of denaturation at 94°C for 45 s, primer annealing for 45 s decreasing by 1°C every cycle from 65°C to 50°C (touchdown PCR), and primer extension at 72°C for 2 min. After 16 cycles, 24 cycles at 50°C were added with the same denaturation and annealing conditions and a final extension was carried out at 72°C for 3 min.

### DNA Sequencing

Sequencing primers were designed to anneal approximately every 400 bp, for complete coverage in both orientations of each locus pair. After DNA amplification, PCR products were purified with the QIAquick PCR Purification Kit (QIAGEN), and the amount of purified DNA was estimated by electrophoresis in a 1% agarose gel and by measurement with a spectrophotometer. Approximately 10 ng of the purified sample (1 to 4 µl) was used as sequencing template. These templates were diluted in water to a volume of 7 µl, and 1 µl primer (5 pM) and 2 µl dye-terminator were added. Cycle sequencing was performed according to the manufacturer's instructions using the BigDye Terminator Cycle Sequencing kit and ABI Prism 3700 (PerkinElmer Biosystems).

Chromatograms were imported to Sequencher version 4.0 (Gene Codes Corporation) for the assembly into contigs and the identification of polymorphic sites. Diploid sequence was determined by inspection of each nucleotide position in high-quality chromatograms.

Hardy-Weinberg equilibrium was tested for each site using Arlequin version 2.0 (<http://lgb.unige.ch/arlequin/>). Overall, there were no more departures than expected by chance ( $P < 0.05$  for 12 variable sites out of 167 across the different regions). Nonetheless, at some sites that appeared out of Hardy-Weinberg equilibrium, there were individuals homozygous for the rare allele, raising the possibility of allelic dropout. We assessed whether these anomalies are caused by allelic dropout by designing new primers, performing a new PCR, and resequencing the product. Eight of the 12 sites identified as out of Hardy-Weinberg turned out to be heterozygotes. This method was also used on 22 additional homozygote sites, to test whether any additional heterozygous sites had been missed. None of them were found to be heterozygotes.

For insertion-deletions that made direct sequencing difficult, we carried out following procedure: After the

PCR was performed, the products were cloned using the TOPO TA-cloning kit (Invitrogen). Once the bacteria had grown on agar plates, 10 white colonies were picked and a new PCR was performed with 40 cycles beginning with a denaturing temperature of 94°C for 1 min, followed by primer annealing at 57°C for 1 min, and an elongation temperature of 72°C for 2 min. Fifteen minutes of additional elongation were added at the end of the program. The PCR products were then sequenced as above for every individual.

### Statistical Analyses

We used the program DNAsp (Rozas and Rozas 1999) to obtain a number of commonly used statistics. To summarize diversity levels, we calculated the average pairwise difference,  $\pi$  (Nei and Li 1979) and  $\theta_w$  (Watterson 1975). Under the standard neutral model of a random-mating population of constant size with no selection, both are estimates of the population mutation rate,  $\theta = 4N_e\mu$  (where  $\mu$  is the mutation rate per generation and  $N_e$  the effective population size) (Tajima 1989). We also calculated two summaries of allele frequencies, Tajima's  $D$  statistic (Tajima 1989) and Fu and Li's  $D$  test (Fu and Li 1993).

We tested the goodness of fit of a standard neutral model of a random-mating population of constant size by asking how often the mean Tajima's  $D$  statistic across the nine loci or the mean Fu and Li's  $D$  test are as low or lower than observed in 10,000 simulations. We also assessed whether there was significant heterogeneity in the ratio of polymorphism to divergence across loci, using a multilocus Hudson-Kreitman-Aguadé (HKA) test (Hudson, Kreitman, and Aguadé 1987). All three tests were implemented using the program HKA, available from Jody Hey's home page (<http://lifesci.rutgers.edu/~heylab/>). The program assumes an infinite-site mutation model and no recombination within locus pairs but free recombination between them.

To estimate the chimpanzee mutation rate, we reconstructed the sequence for the ancestor of humans and chimpanzees, with the orangutan as an outgroup (using PAML [Yang 1997]). This allowed us to estimate the number of substitutions,  $\kappa$ , separately for each lineage. We obtained an estimate of the mutation rate per generation by dividing  $\kappa$  by the number of generations to the most recent common ancestor, assuming 6 Myr to the common ancestor and 20 years per generation (L. Vigilant, personal communication). Given this estimate of the mutation rate, the diploid effective population size  $\hat{N}_e$  of autosomal loci can be estimated by  $\bar{\pi}/(4\hat{\mu})$  or  $\hat{\theta}_w/(4\hat{\mu})$  under the standard neutral model (where  $\bar{\pi}$  and  $\hat{\theta}_w$  are average values across the nine locus pairs).

### Comparison of Central and Western Chimpanzees

Seven single fragments of 670 to 910 bp from the nine locus pairs were previously sequenced in 16 western chimpanzees, totaling 5.4 kb (Gilad et al. 2003). On the basis of these data and polymorphism data collected for the same 5.4 kb in central chimpanzees, we built a neighbor-joining tree, with the orangutan as an outgroup, for each of

the regions (using MEGA version 2.1 and the Kimura two-parameter method [Kimura 1980]). To summarize differences in allele frequencies between the central and western chimpanzee samples, we calculated the ratio of the estimated variance component due to differences between populations over the estimated total variance,  $\Phi_{st}$  (Excoffier, Smouse, and Quattro 1992), using the program Arlequin (<http://lgb.unige.ch/arlequin/>). We also tabulated the number of polymorphisms shared between western and central population samples, found only in one sample, or fixed between samples. Assuming a simple model in which an ancestral population splits into two descendant populations of constant (but possibly unequal) size, with no subsequent migration, these summaries can be used to estimate the split time and the effective population sizes. We did so using a method of moments developed by Wakeley and Hey (1997), using the mutation rate that we estimated for chimpanzees.

### Measures of Linkage Disequilibrium (LD)

Haplotypes were inferred with the program PHASE (Stephens, Smith, and Donnelly 2001), after exclusion of singletons. Pairwise LD was then calculated using the inferred haplotypes (using DNAsp). To summarize levels of pairwise LD, we used  $r^2$ , a commonly used statistic (Hill and Robertson 1968). An alternative way to quantify levels of LD is to estimate the population crossing-over parameter  $\rho = 4Nr$ . The parameter  $\rho$  is a key determinant of LD patterns, with the strength of LD decreasing when  $\rho$  increases. The relation of  $\rho$  to the mean  $r^2$  is known under simple models (McVean 2002). Under the standard neutral model

$$E(r^2) \approx (4\theta + \rho + 10)/(8\theta^2 + 60\rho + \rho^2 + 320 + 13\rho + 22) \quad (1)$$

so long as the allele frequencies are between 0.05 and 0.95 (cf. Hudson [2001a]). Thus, the expected value of  $r^2$  can be predicted using estimates of  $\rho$  and  $\theta$ .

In humans, there exist two approaches to estimating  $\rho$ . In the first,  $\rho$  is estimated as  $\rho_{\text{map}} = 4\hat{N}_e\hat{c}$ , where  $\hat{c}$  is an estimate of the recombination rate per bp, based on a comparison of physical and genetic maps, and  $\hat{N}_e$  is an estimate of the effective population size estimate, based on diversity and divergence data. In the second method,  $\rho$  is estimated from patterns of LD under the standard neutral model. One way to think of this second estimate is as the population recombination rate needed to produce the observed levels of LD under the standard neutral model. Thus, larger  $\rho$  correspond to lower levels of LD. If the underlying assumptions are valid, the two estimates of  $\rho$  should be similar.

We used the approach of Hudson (2001b) to estimate the maximum composite likelihood of  $r$ , referred to as  $\rho_{\text{ci}}$ . Specifically, we used an extension of the method that allows the simultaneous estimation of  $\rho$  and  $f$ , where  $f$  is the ratio of gene conversion to crossing-over events (Frisse et al. 2001). The program to do so was kindly provided by R. Hudson. The method assumes that gene conversion and crossing-over are alternative resolutions of

**Table 1**  
**Summaries of Polymorphism Data from Central Chimpanzees**

Region <sup>a</sup>	Length (bps)	$\pi$ (%) <sup>b</sup>	$\theta_w$ (%) <sup>b</sup>	Tajima's <i>D</i> Statistic	Fu and Li's <i>D</i> Test	Divergence from Orangutan (%) <sup>b</sup>	Divergence from Human(%) <sup>b</sup>
1	1,820	0.13	0.16	-0.45	-0.64	2.80	1.06
2	2,209	0.24	0.27	-0.39	0.87	3.79	1.20
3	2,567	0.16	0.24	-1.20	-1.29	3.30	1.59
4	1,890	0.08	0.14	-1.26	-0.25	3.11	1.49
5	2,237	0.12	0.16	-0.90	-0.65	2.08	0.90
6	2,417	0.18	0.24	-0.99	-0.66	3.14	1.12
7	2,036	0.10	0.23	-1.94	-2.43	3.80	1.08
8	1,826	0.27	0.37	-0.96	-0.17	2.73	1.25
10	1,997	0.25	0.35	-1.23	-0.09	3.08	0.74
Mean	2,111	0.17	0.24	-1.02	-0.59	3.09	1.16

<sup>a</sup> Numbered as in Frisse et al. (2001).<sup>b</sup> Per base pair.

a Holliday junction and that the conversion-tract length is geometrically distributed with mean length  $L$ . Given this model of gene conversion, an effective population recombination parameter  $\rho_e$  can be calculated as follows:  $\rho_e = \rho d + 2fL(1 - \exp(-d/L))$  (Wiuf and Hein 2000). This  $\rho_e$  can then be also used to estimate the  $E(r^2)$  by substituting  $\rho_e$  for  $\rho$  in equation 1. Following Frisse et al. (2001), we assumed that  $\rho$  and  $f$  are fixed across loci and estimated the two parameters from all loci jointly. To make our results comparable to those in Frisse et al. (2001), we further assumed that  $L = 500$  bp; similar results are obtained if  $L = 1,000$  (results not shown). To test the hypothesis that  $f = 0$ , we compared the observed value of  $\lambda = CLik(f=0)/CLik(\hat{f})$  with a distribution of  $\lambda$  values obtained from 100 simulations of the null model, where  $\theta = \theta_w$ ,  $f = 0$ , and  $\rho = \rho_{cl}$  (*CLik* denotes the composite likelihood). A small ratio indicates that the data are more likely under the alternative where  $f$  is not constrained to be 0.

## Results

### Regions and Genomic Characteristics

Frisse et al. (2001) surveyed 10 autosomal regions in 15 humans sampled from three putative populations: Hausa in Africa, Italians in Europe, and Han Chinese in Asia (data available from <http://geneapps.uchicago.edu/labweb/index.html>). We sequenced nine of the same 10 regions in 14 central chimpanzees, as well as one human and one orangutan. Each region, referred to as a "locus pair," following Frisse et al. (2001), consists of a noncoding 10-kb DNA sequence of which approximately 1,000 bp from each end were determined from each individual. The mean sex-averaged recombination rate estimated in humans for the nine regions is 1.54 cM/Mb, slightly higher than the genome-wide average of 1.19 cM/Mb (Kong et al. 2002). The average GC content of the chimpanzee sequences is 40.8%, close to the human genome-wide average of 42% (Lander et al. 2001). For eight regions, the closest known human gene is more than 90 kb away from the locus, whereas for region 6, the closest gene is 21 kb away.

### Divergence, Diversity, and Effective Population Size

A total of 19,000 bases were sequenced per individual. The average divergence (table 1) between chimpanzees and humans is 1.16% (range, 0.74% to 1.59%) and between chimpanzees and orangutan 3.09% (range, 2.08% to 3.80%). This is in line with previous estimates (Chen and Li 2001; Ebersberger et al. 2002) of the average intergenic divergence of chimpanzee to human and to orangutan (e.g., 1.24% and 3.12%, respectively, in Chen and Li [2001]). In total, 167 variable positions are found among the 28 central chimpanzee chromosomes. The average number of pairwise differences ( $\pi$ ) is 0.17% and ranges from 0.08% to 0.27% across the nine locus pairs. A summary of diversity based on the number of polymorphic sites in the sample,  $\theta_w$  (Watterson 1975), is 0.24% when averaged across all regions and ranges from 0.14% to 0.37% (table 1).

We estimated the mutation rate on the chimpanzee and human evolutionary lineages by reconstructing the common ancestor sequence of chimpanzees and humans (using the orangutan as an outgroup), then inferring the substitutions that occurred on each lineage since the common ancestor. The estimate of the mutation rate per site per generation,  $\mu$ , obtained in this way is  $1.72 \times 10^{-8}$  for chimpanzees and  $2.26 \times 10^{-8}$  for humans; these values are not significantly different from one another (results not shown). Given this estimate of the mutation rate in chimpanzees and the observed  $\pi$ , we estimate the diploid effective population size to be roughly 25,000 (see *Materials and Methods*). If instead  $\theta_w$  is used as a measure of diversity, we obtain  $\sim 35,000$ . By comparison, estimates of  $N_e$  for central chimpanzees obtained from other loci are  $\sim 47,000$  for Xq13,  $\sim 20,000$  for NRY, and  $\sim 25,000$  for HOXB6 (using  $\theta_w$  and  $\mu$  estimates reported in Stone et al. [2002] and assuming 20 years per generation) and only  $\sim 18,500$  for the intergenic regions surveyed by Yu et al. (2003) (using their diversity and divergence data, under the same assumptions as for our estimate).

### Selection and Demography

To assess the evidence for differences in selective constraints among the nine regions in the chimpanzees, we

used a multilocus HKA test with the orangutan to estimate divergence (see *Materials and Methods*). By this approach, there is no evidence for heterogeneity among the regions ( $P = 0.964$ , ignoring intralocus recombination). Further, the regions were not unusually conserved in a comparison of human to mouse (data not shown). Thus, there is no reason to assume that selection constraints have influenced the variation observed in these regions.

To assess the fit of allele frequencies to the expectations of the standard neutral model, we used the Tajima's  $D$  (see *Materials and Methods*). This statistic considers the approximately normalized difference between  $\pi$  and  $\theta_w$ . Under the standard neutral model, the expectations of  $\pi$  and  $\theta_w$  are equal, so the mean Tajima's  $D$  statistic is roughly 0. Because rare alleles contribute more to  $\theta_w$  than to  $\pi$ , a negative value of Tajima's  $D$  statistic reflects a relative excess of low-frequency polymorphisms. As table 1 shows, Tajima's  $D$  statistic is negative in all nine regions. Moreover, although only region 7 exhibits a  $D$  statistic value significantly different from 0 at the 5% level, the mean  $D$  statistic value across loci is significantly negative ( $-1.02$ ;  $P < 10^3$  [see *Materials and Methods*]), reflecting an excess of rare variants relative to standard neutral expectations. The skew in the allele frequency spectrum explains the discrepancy between estimates of the effective population size obtained from  $\pi$  and from  $\theta_w$  and suggests that one or more assumptions of the standard neutral model may be invalid. Because natural selection is unlikely to have affected these sequences, a plausible explanation is a demographic departure from model assumptions such as past population expansions, an old population bottleneck, or fine-scale population structure (Slatkin and Hudson 1991; Fay and Wu 1999; Wakeley and Aliacar 2001).

To assess whether the relative excess of rare variants is caused primarily by singletons, we used Fu and Li's  $D$  test, with the orangutan to infer ancestral and derived states (see *Materials and Methods*). Fu and Li's  $D$  test is based on the difference between  $\eta$ , the total number of variable sites, and  $\eta_s$ , the number of derived singletons (i.e., nonancestral mutations appearing only once in the sample). Fu and Li's  $D$  test is negative in eight out of the nine regions but is only significant at the 5% level for region 7 (table 1). The mean Fu and Li's  $D$  test across locus pairs is not significantly different from the neutral expectation 0 at the 5% level, although the  $P$ -value is low ( $P = 0.07$ ). Furthermore, visual inspection of the entire frequency spectrum suggests that there is an overall excess of low-frequency alleles, rather than an excess of singletons alone (results not shown). Because the main effect of a simple growth model is expected to be on the proportion of singletons (Slatkin and Hudson 1991; Hey and Harris 1999) and the use of Fu and Li's  $D$  test as a test statistic should be more powerful than that of Tajima's  $D$  statistic (Fu 1997; Pluzhnikov, Di Rienzo, and Hudson 2002), a more likely explanation for the high proportion of rare alleles observed in central chimpanzees may be an old bottleneck (Simonsen, Churchill, and Aquadro 1995; Fay and Wu 1999) or fine-scale population structure (Wakeley and Aliacar 2001).

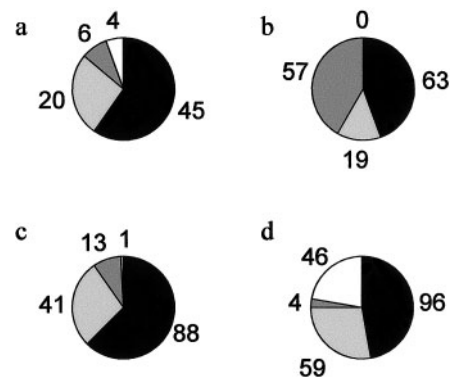


FIG. 1.—Polymorphic sites shared between samples (dark gray), fixed between samples (white), or exclusive to a sample (black and light gray, respectively) from (a) locus pairs in central and western chimpanzees, (b) locus pairs in Hausa and Chinese (Frisse et al. 2001), (c) intergenic segments in central and western chimpanzees (Yu et al. 2003), and (d) intergenic segments in central chimpanzees and bonobos (Yu et al. 2003).

### Chimpanzee Subspecies and Human Populations

Gilad et al. (2003) surveyed 5,420 base pairs from seven of the loci studied here in 16 western chimpanzees. The nucleotide diversity is 1.2-fold to 2-fold higher in central chimpanzees (0.15% versus 0.12% for  $\pi$ , and 0.24% versus 0.12% for  $\theta_w$ ). The difference in  $\theta_w$  value between samples is significant at the 5% level ( $P = 0.018$ , as assessed by a two-tailed Mann-Whitney U test), but the difference in  $\pi$  is not ( $P = 0.337$ ). This is reflected in the Tajima's  $D$  statistic values: whereas the mean Tajima's  $D$  statistic is significantly below its expectation of  $\sim 0$  in central chimpanzees (mean  $D$  value =  $-0.86$ ,  $P = 0.009$ ), it is very close to 0 for the western chimpanzees (mean  $D$  value =  $+0.08$ ).

To examine how DNA sequences in the two groups are evolutionarily related to each other, we built neighbor-joining trees for each locus pair, using human as an outgroup. In these trees, the DNA sequences from the western chimpanzees are always monophyletic, suggesting an old split time between populations (results not shown). Moreover, a large proportion of the genetic variance is caused by differences between the populations ( $\Phi_{st} = 0.62$ ).

Figure 1 presents the polymorphic sites partitioned into variants that are shared between samples, fixed between the samples, or exclusive to one sample. Strikingly, only six out of 71 variable sites are shared between western and central chimpanzee samples, whereas four are fixed between samples. In contrast, in humans, 57 out of 139 polymorphisms are shared between Hausa and Chinese samples and none are fixed (Di Rienzo, personal communication); for other population pairs, the proportion of shared polymorphisms is the same or higher (results not shown). Thus, human populations appear to have much more of their evolutionary history in common than do subspecies of chimpanzees. If we assume a simple split model for the evolution of chimpanzee subspecies and estimate the split time from the locus pair data in figure 1 (see *Materials and Methods*), we obtain an ancestral

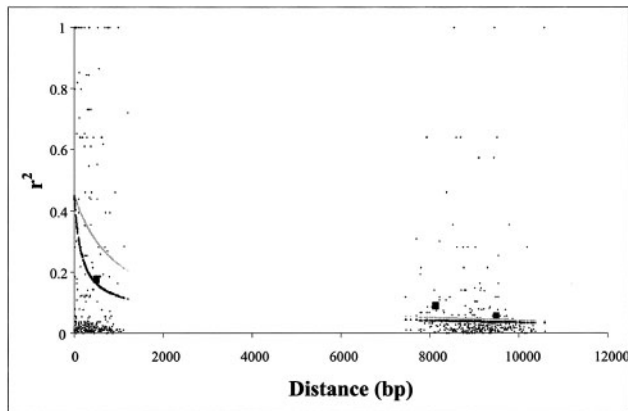


FIG. 2.—Decay of pairwise LD with physical distance. The dark gray dots are the observed  $r^2$  values for each pair of sites with minor allele frequencies greater than 0.05. The black squares are the mean  $r^2$  for sites less than 1.2 kb apart, for sites 7.5 to 9 kb and for sites 9 to 10.5 kb apart. The light gray and black lines are the expected decay of LD under a model of crossing-over only ( $\rho = \rho_{\text{map}}$ ) and under a model with gene conversion (for  $\rho = \rho_{\text{map}}$  and  $f = 2$ ), respectively.

effective population size estimate of  $N_A \approx 51,000$  and a population split time estimate  $\hat{t} = 650,000$  years.

For a comparison, we reanalyzed the data of Yu et al. (2003). Between western and central chimpanzees, 13 out of 142 variable sites are shared and one is fixed between samples, whereas between bonobos and central chimpanzees, the numbers are four and 46, respectively. Using the same approach as above, the estimates are  $N_A \approx 22,000$  and  $\hat{t} = 430,000$  years for western and central chimpanzees and  $N_A \approx 43,000$  and  $\hat{t} = 800,000$  for bonobos and chimpanzees. As evidenced by the difference between estimates based on the intergenic regions of Yu et al. (2003) and on the locus pairs, more data are needed for a precise estimate of these parameters.

Bonobos and chimpanzees occupy ranges that are separated by the Congo-Lualaba river system, thought to have formed approximately 1.5 MYA (Beadle 1981); thus, a sudden split may not be an unreasonable model for the evolution of these two species. Although there are also rivers delimiting the range of chimpanzee subspecies, the historical boundaries are much less clear (Morin et al. 1994; Gagneux 2002), and there may have been ongoing migration between western and central subspecies. If so, application of a sudden split model would tend to lead to an overestimate of the ancestral effective population size and to an underestimate of the split time (cf. Wall [2003]).

### Linkage Disequilibrium in Central Chimpanzees

Levels of linkage disequilibrium (LD) vary across the genome by chance as well as because of local differences in recombination and mutation (cf. Pritchard and Przeworski [2001]). Levels of LD are often summarized by a pairwise summary,  $r^2$ , which measures the correlation between alleles (cf. Hartl and Clark [1997]). If we consider  $r^2$  values for all pairs of single-nucleotide polymorphisms with minor allele frequencies above 0.05 in the central chimpanzee data (fig. 2), there is a clear decay with physical distance, such that the mean  $r^2$  for sites separated

by 1 kb or less is 0.174, whereas it is only 0.066 for sites separated by 8 to 10 kb.

As a first guess at to how LD might be expected to decay, we plot the expected decay of  $r^2$  under the standard neutral model, assuming  $\rho_{\text{map}} = 0.0021$  ( $\rho_{\text{map}} = 4\hat{N}_e\bar{c}$  where  $\hat{N}_e$  is the effective population size estimated from  $\theta_w$  and  $\bar{c}$  is the average recombination rate estimated from human genetic and physical maps). As can be seen in figure 2, whereas the fit for longer distances is adequate,  $r^2$  values over distances less than 1 kb are lower than expected. In other words, there is unexpectedly little LD at short distances. This observation is consistent with a number of previous reports in humans (Frisse et al. 2001; Przeworski and Wall 2001). As pointed out by these authors, one explanation is that the recombination model is lacking an important feature: homologous gene conversion.

Over short distances, homologous gene conversion is likely to make an important contribution to the rate of genetic exchange and, hence, to the decay of LD (Andolfatto and Nordborg 1998). However, over larger distances, the effects of gene conversion will be negligible. Because genetic maps are based on markers 1 Mb or so apart, estimates of the recombination rate based on these markers (such as  $\rho_{\text{map}}$ ) will essentially be estimates of the crossover rate alone (Przeworski and Wall 2001). In contrast, estimates based on patterns of LD over a small scale will be affected by both crossing-over and gene conversion.

To examine whether this difference might account for the poor fit of the predictions to our data, we coestimated  $\rho$  and the ratio of gene conversion to crossing-over events ( $f$ ). Assuming that  $\rho$  and  $f$  are fixed across loci and that the mean gene conversion tract length is 500 bp (see *Materials and Methods*), we obtain  $\rho_{\text{cl}} = 0.0027$  and  $\hat{f} = 2$ , respectively. This estimate of the crossover rate is close to our a priori estimate (i.e.,  $\rho_{\text{map}} = 0.0021$ ). Furthermore, assuming the standard neutral model, we can marginally reject the null hypothesis of no gene conversion ( $\lambda$  is smaller in four out of 100 simulations [see *Materials and Methods*]). The use of the standard neutral model is inappropriate, as there are more rare alleles in the data than expected under this model (and these are less informative about recombination rates). This said, if we try to predict the decay of  $E(r^2)$  with distance from  $\rho_{\text{map}}$ , visual inspection of the data suggests a better fit with the inclusion of gene conversion (this is illustrated in figure 2 for  $f = 2$ ). Given that gene conversion is a general feature of the recombination process across taxa (cf. Pittman and Schimenti [1998]) and that data from the same regions in humans also support the occurrence of gene conversion (Frisse et al. 2001), it seems a plausible explanation for the lower than expected levels of LD in central chimpanzees.

Estimates of recombination parameters obtained from the human and chimpanzee polymorphism data can be compared to assess the evidence for a large change in rates. The analysis of the same nine regions in a Hausa sample leads to  $\rho_{\text{cl}} = 0.0010$  and  $\hat{f} = 5$ . Simulations suggest that the values of  $\hat{f}$  in Hausa and central chimpanzees are not significantly different from one another (Ptak et al. 2004; results not shown). Moreover, the ratio of  $\rho = 4N_e r$

**Table 2**  
**Summary of the Polymorphism Data from Western (Ptv) and Central (Ptt) Chimpanzees**

Source	Chimpanzee Subspecies	Number of Chromosomes	Length	$\pi(\%)^a$	$\theta_w(\%)^a$	Tajima's <i>D</i> Statistic	<i>P</i> -value <sup>b</sup>
HOXB6	Ptt	38	1,000	0.09	0.11	-0.48	NS
	Ptv	116	1,000	0.11	0.11	-0.02	NS
Xq13.3	Ptt	12	10,151	0.18	0.21	-0.75	NS
	Ptv	15	10,151	0.05	0.07	-1.15	NS
Intergenic segments	Ptt	10	23,113	0.13 <sup>c</sup>	0.15 <sup>c</sup>	-0.32 <sup>d</sup>	0.04
	Ptv	12	23,113	0.08 <sup>c</sup>	0.08 <sup>c</sup>	0.15 <sup>d</sup>	NS
Locus pairs	Ptt	28	5,420	0.15 <sup>c</sup>	0.24 <sup>c</sup>	-0.86	<10 <sup>-3</sup>
	Ptv	32	5,420	0.12 <sup>c</sup>	0.12 <sup>c</sup>	0.08	NS

NOTE.—“Locus pairs” refers to the data collected in this study and in Gilad et al. (2003). Additional data are from the HoxB6 intergenic region (Deinard and Kidd 1999), an X-linked intergenic region, Xq13.3 (Kaessmann, Wiebe, and Pääbo 1999), and 50 intergenic segments (Yu et al. 2003).

<sup>a</sup> Per base pair.

<sup>b</sup> Significance assessed at the 5% level (see *Materials and Methods* for details).

<sup>c</sup> Mean value across loci.

<sup>d</sup> Including only loci with at least one polymorphic site (42 in central chimpanzees and 32 in western chimpanzees).

estimates (2.7-fold) is roughly as expected from the ratio of  $N_e$  estimates (2.3-fold). Thus, on the basis of this small data set, there is no evidence that rates of recombination have changed between humans and chimpanzees. It should be noted however that this analysis is limited by the assumption that  $\rho$  is the same across loci. Although these loci were chosen to have similar large-scale recombination rates in humans, local heterogeneity in the recombination rate renders the interpolation of these rates to smaller scales problematic. Unfortunately, there is not enough information in these data to obtain reliable estimates of  $\rho$  for each locus pair separately (Ptak et al. 2004).

## Discussion

### Diversity in Humans and Chimpanzees

The average diversity,  $\theta_w$ , is 0.24% in the central chimpanzees. By comparison, in these same regions,  $\theta_w$  is 0.12% in the Hausa, 0.10% in the Sardinians, and 0.09% in the Han Chinese (Frisse et al. 2001). Estimates of  $\theta_w$  in worldwide samples of humans are  $\sim 0.10\%$  (Przeworski, Hudson, and Di Rienzo 2000; Zhao et al. 2000). Thus, central chimpanzees are approximately 2.4 times more diverse as humans worldwide. For the shorter DNA sequences collected in western chimpanzees (Gilad et al. 2003),  $\theta_w$  is 0.12% compared with 0.24% for the central chimpanzees (using the same DNA sequences). Western chimpanzees, therefore, appear similar to humans in being about half as diverse as central chimpanzees.

As summarized in table 2, higher diversity in central than in western chimpanzees has been reported previously for an approximately 10-kb X-linked region (Kaessmann, Wiebe, and Pääbo 1999) and a number of intergenic regions (Yu et al. 2003). In these respects, available autosomal and X-linked data are consistent (with the exception of a 1-kb region near HOXB6). This said, levels of diversity vary substantially across loci; in particular, our diversity estimates are larger than found by Yu et al. (2003). Because there is a high proportion of rare alleles in central chimpanzees, one explanation may be the smaller sample of the Yu et al. (2003) study (10 chromosomes versus 28). For western chimpanzees, the discrepancy is

harder to explain. One possibility is that it reflects differences in the origin of the western chimpanzees in the two studies; further sampling of western chimpanzees is needed to resolve this point.

### The Population History of Chimpanzees

Allele frequencies in the sample of central chimpanzees are significantly skewed towards rare alleles relative to the expectations of the standard neutral model, as reflected by an average Tajima's *D* of  $-1.02$ . This value is consistent with the findings for other intergenic regions (see table 2). The mean *D* statistic value across the loci surveyed by Yu et al. (2003) is lower than expected under the standard neutral model ( $P = 0.04$ ) and the *D* values for the other two loci are negative as well. In some ways, the high proportion of rare alleles is surprising. In humans, the same observation is often interpreted as evidence for population growth (Ingman et al. 2000; Thomson et al. 2000; Stephens et al. 2001), yet populations of chimpanzee are rapidly dwindling. However, the decline of chimpanzee populations is likely to be a recent phenomenon (Teleki 1989), not yet visible in patterns of genetic variation. Furthermore, a number of other models also predict a high proportion of rare alleles, including a sudden reduction in population size followed by a recovery (Fay and Wu 1999) and some models of fine-scale population structure (Wakeley and Aliacar 2001; Ptak and Przeworski 2002), which may be more realistic for chimpanzees (e.g., Whiten et al. 1999).

In contrast to central chimpanzee samples, western samples do not harbor a high proportion of rare alleles. In fact, the mean *D* statistic value in western chimpanzees is very close to the standard neutral model expectation of  $\sim 0$ . As can be seen in table 2, this observation is consistent with patterns at other autosomal loci but not with data from Xq13.3. Because Xq13.3 is only one locus, the discrepancy may be the result of chance; alternatively, it may reflect a difference between autosomal and sex-linked loci.

The difference in allele frequencies observed in samples of central and western chimpanzee points to

divergent evolutionary histories of the two subspecies. Under a simple model of population history (see *Materials and Methods*), we estimate a split time of 430,000 to 650,000 years (depending on which data are used [see *Results*]) for western and central populations, substantially less than the 1.3 Myr estimated previously from the mtDNA (Morin et al. 1994). Based on the data collected by Yu et al. (2003), the split time for the *Pan* species does not appear much older (800,000 years).

Our estimate of the bonobo/chimpanzee split is more recent than previously published estimates: 930,000 years for Xq13.3 (Kaessmann, Wiebe, and Pääbo 1999), 1.5 Myr for data from the Y chromosome (Stone et al. 2002), 1.6 Myr for mtDNA data (Morin et al. 1994), and 1.8 Myr for the same data (Yu et al. 2003). However, with the exception of Stone et al. (2002), these estimates are based on pairwise differences between species, which reflects not only differences accumulated since the split but also ancestral polymorphism. In other words, these are actually estimates of the coalescence time of a chimpanzee and a bonobo sequence and not of the split time of the two species. If the split time was recent or the ancestral population size was large, as appears to be the case here, these two times will be substantially different (Rosenberg and Feldman 2002). As an illustration, when we reanalyzed the data of Yu et al. (2003), we obtained estimates of 43,000 for the ancestral effective population size and 800,000 years for the split time, less than half their estimate of the coalescence time. Thus, the discrepancy between estimates reflects, in part, a difference in what is being estimated. Nonetheless, there remains substantial variability in the split time estimates based on different regions (e.g., between ours and those obtained by Stone et al. [2002] for the NRY). More data (and methods that use more of the data) are clearly needed for an accurate estimate of these parameters.

On the basis of existing data, it appears that divergence levels between bonobo and common chimpanzee at random loci are not much higher than what is observed between subspecies of chimpanzees. In contrast, phenotypic differences are much greater between the two species, consistent with their taxonomic designations. For example, while the species can be distinguished in captivity (by humans), the subspecies cannot. Moreover, morphometric studies of craniofacial variation find much larger differences between species than between subspecies (Guy et al. 2003; Taylor and Groves 2003). In this respect, it is interesting to note that in humans, where there is much less genetic differentiation between populations than in chimpanzees, ancestry from different continents can often be reasonably well predicted from craniofacial features (e.g., Lynch, Wood, and Luboga 1996; Relethford 2002). This uncoupling of phenotypic differentiation and genetic differentiation at random markers may reflect greater genetic drift in humans and bonobos relative to chimpanzee subspecies or natural selection associated with the more diverse habitats exploited by humans (Akey et al. 2002; Kayser, Brauer, and Stoneking 2003) as well as possibly by bonobos (Myers-Thompson 2003).

## Outlook

With the imminent publication of the chimpanzee genome, there is great interest in contrasting patterns of variation in humans and chimpanzees, in particular to estimate recombination rates and identify targets of natural selection. Both of these aims rely on a demographic model for the history of chimpanzees. These data suggest that the standard model for population genetic analyses is a poor description of the demographic history of chimpanzees. Indeed, given the complex demographic history apparent in these data, it appears that an understanding of the population history of chimpanzees will require extensive data collected with careful attention to geography.

## Acknowledgments

Thanks to Kathrin Koehler for providing the cell cultures, to Yoav Gilad for help in the lab, to Ines Hellmann for estimation of recombination rates, to Susan Ptak for help with simulations and to Susan Ptak, Jonas Eriksson, and two anonymous reviewers for comments on an earlier version of this manuscript. Support for this work was provided by the Max Planck Society and the Bundesministerium für Bildung und Forschung.

## Literature Cited

- Akey, J. M., G. Zhang, K. Zhang, L. Jin, and M. D. Shriver. 2002. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**:1805–1814.
- Andolfatto, P., and M. Nordborg. 1998. The effect of gene conversion on intralocus associations. *Genetics* **148**:1397–1399.
- Beadle, L. C. 1981. *The inland waters of tropical Africa*. Longman, London.
- Chen, F. C., and W. H. Li. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* **68**:444–456.
- Chen, F. C., E. J. Vallender, H. Wang, C. S. Tzeng, and W. H. Li. 2001. Genomic divergence between human and chimpanzee estimated from large-scale alignments of genomic sequences. *J. Hered.* **92**:481–489.
- Deinard, A., and K. Kidd. 1999. Evolution of a HOXB6 intergenic region within the great apes and humans. *J. Hum. Evol.* **36**:687–703.
- Dufour, C., D. Casane, D. Denton, J. Wickings, P. Corvol, and X. Jeunemaitre. 2000. Human-chimpanzee DNA sequence variation in the four major genes of the renin-angiotensin system. *Genomics* **69**:14–26.
- Ebersberger, I., D. Metzler, C. Schwarz, and S. Pääbo. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* **70**:1490–1497.
- Enard, W., M. Przeworski, S. E. Fisher, C. S. Lai, V. Wiebe, T. Kitano, A. P. Monaco, and S. Paabo. 2002. Molecular evolution of FOXP2, a gene involved in speech and language. *Nature* **418**:869–872.
- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479–491.
- Fay, J. C., and C. I. Wu. 1999. A human population bottleneck can account for the discordance between patterns of



- mitochondrial versus nuclear DNA variation. *Mol. Biol. Evol.* **16**:1003–1005.
- Ferris, S. D., W. M. Brown, W. S. Davidson, and A. C. Wilson. 1981. Extensive polymorphism in the mitochondrial DNA of apes. *Proc. Natl. Acad. Sci. USA* **78**:6319–6323.
- Frisse, L., R. R. Hudson, A. Bartoszewicz, J. D. Wall, J. Donfack, and A. Di Rienzo. 2001. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet.* **69**:831–843.
- Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**:915–925.
- Fu, Y. X., and W. H. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* **133**:693–709.
- Gagneux, P. 2002. The genus *Pan*: population genetics of an endangered outgroup. *Trends Genet.* **18**:327–330.
- Gagneux, P., C. Willis, 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.
- Gilad, Y., C. D. Bustamante, D. Lancet, and S. Paabo. 2003. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* **73**:489–501.
- Gonder, M. K., J. F. Oates, T. R. Disotell, M. R. Forstner, J. C. Morales, and D. J. Melnick. 1997. A new West African chimpanzee subspecies? *Nature* **388**:337.
- Guy, F., M. Brunet, M. Schmittbuhl, and L. Viriot. 2003. New approaches in hominoid taxonomy: morphometrics. *Am. J. Phys. Anthropol.* **121**:198–218.
- Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics. Sinauer Associates, Sunderland, MA.
- Hey, J., and E. Harris. 1999. Population bottlenecks and patterns of human polymorphism. *Mol. Biol. Evol.* **16**:1423–1426.
- Hill, W. G., and A. Robertson. 1968. The effects of inbreeding at loci with heterozygote advantage. *Genetics* **60**:615–628.
- Hudson, R. R. 2001a. Linkage disequilibrium and recombination. Pp. 309–318 in D. J. Balding, M. Bishop, and C. Cannings, eds. *Handbook of statistical genetics*. John Wiley & Sons, Ltd, Chichester, UK.
- . 2001b. Two-locus sampling distributions and their application. *Genetics* **159**:1805–1817.
- Hudson, R. R., M. Kreitman, and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**:153–159.
- Huttley, G. A., S. Eastaile, M. C. Southey, A. Tesoriero, G. G. Giles, M. R. McCredie, J. L. Hopper, and D. J. Venter. 2000. Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Australian Breast Cancer Family Study. *Nat. Genet.* **25**:410–413.
- Ingman, M., H. Kaessmann, S. Paabo, and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**:708–713.
- Jensen-Seaman, M. I., A. S. Deinard, and K. K. Kidd. 2001. Modern African ape populations as genetic and demographic models of the last common ancestor of humans, chimpanzees, and gorillas. *J. Hered.* **92**:475–480.
- 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.
- Kaessmann, H., V. Wiebe, and S. Pääbo. 1999. Extensive nuclear DNA sequence diversity among chimpanzees. *Science* **286**:1159–1162.
- Kayser, M., S. Brauer, and M. Stoneking. 2003. A genome scan to detect candidate regions influenced by local natural selection in human populations. *Mol. Biol. Evol.* **20**:893–900.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
- Kitano, T., C. Schwarz, B. Nickel, and S. Paabo. 2003. Gene diversity patterns at 10 X-chromosomal Loci in humans and chimpanzees. *Mol. Biol. Evol.* **20**:1281–1289.
- Kong, A., D. F. Gudbjartsson, J. Sainz et al. (13 co-authors). 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* **31**:241–247.
- Lander, E. S., L. M. Linton, B. Birren et al. (255 co-authors). 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
- Lynch, J. M., C. G. Wood, and S. A. Luboga. 1996. Geometric morphometrics in primatology: craniofacial variation in *Homo sapiens* and *Pan troglodytes*. *Folia Primatol. (Basel)* **67**:15–39.
- McVean, G. A. 2002. A genealogical interpretation of linkage disequilibrium. *Genetics* **162**:987–991.
- Morin, P. A., J. J. Moore, R. Chakraborty, L. Jin, J. Goodall, and D. S. Woodruff. 1994. Kin selection, social structure, gene flow, and the evolution of chimpanzees. *Science* **265**:1193–1201.
- Myers-Thompson, J. A. 2003. A model of the biogeographical journey from Proto-*Pan* to *Pan paniscus*. *Primates* **44**(2):191–7.
- Nachman, M. W., and S. L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**:297–304.
- Napier, J. R., and P. H. Napier. 1967. A handbook of living primates. Academic Press, New York.
- Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**:5269–5273.
- Olson, M. V., and A. Varki. 2003. Sequencing the chimpanzee genome: insights into human evolution and disease. *Nat. Rev. Genet.* **4**:20–28.
- Pittman, D. L., and J. C. Schimenti. 1998. Recombination in the mammalian germ line. *Curr. Topics Dev. Biol.* **37**:1–35.
- Pluzhnikov, A., A. Di Rienzo, and R. R. Hudson. 2002. Inferences about human demography based on multilocus analyses of noncoding sequences. *Genetics* **161**:1209–1218.
- Pritchard, J. K., and M. Przeworski. 2001. Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.* **69**:1–14.
- Przeworski, M., R. R. Hudson, and A. Di Rienzo. 2000. Adjusting the focus on human variation. *Trends Genet.* **16**:296–302.
- Przeworski, M., and J. D. Wall. 2001. Why is there so little intragenic linkage disequilibrium in humans? *Genet. Res.* **77**:143–151.
- Ptak, S. E., and M. Przeworski. 2002. Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet.* **18**:559–563.
- Ptak, S. E., K. Voelpel, and M. Przeworski. 2004. Insights into recombination from patterns of linkage disequilibrium in humans. *Genetics* (in press).
- Relethford, J. H. 2002. Apportionment of global human genetic diversity based on craniometrics and skin color. *Am. J. Phys. Anthropol.* **118**:393–398.
- Rosenberg, N. A., and M. W. Feldman. 2002. The relationship between coalescence times and population divergence times. Oxford University Press, Oxford, UK.
- Rozas, J., and R. Rozas. 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**:174–175.

- Ruvolo, M. 1997. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* **14**:248–265.
- Satta, Y. 2001. Comparison of DNA and protein polymorphisms between humans and chimpanzees. *Genes Genet. Syst.* **76**:159–168.
- Simonsen, K. L., G. A. Churchill, and C. F. Aquadro. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. *Genetics* **141**:413–429.
- Slatkin, M., and R. R. Hudson. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**:555–562.
- Stephens, J. C., J. A. Schneider, D. A. Tanguay et al. (25 co-authors). 2001. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**:489–493.
- Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**:978–989.
- Stone, A. C., R. C. Griffiths, S. L. Zegura, and M. F. Hammer. 2002. High levels of Y-chromosome nucleotide diversity in the genus *Pan*. *Proc. Natl. Acad. Sci. USA* **99**:43–48.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**:585–595.
- Taylor, A. B., and C. P. Groves. 2003. Patterns of mandibular variation in *Pan* and *Gorilla* and implications for African ape taxonomy. *J. Hum. Evol.* **44**:529–561.
- Teleki, G. 1989. Population status of wild chimpanzees (*Pan troglodytes*) and threats to survival. Harvard University Press, Cambridge, Mass.
- Thomson, R., J. K. Pritchard, P. Shen, P. J. Oefner, and M. W. Feldman. 2000. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97**:7360–7365.
- Wakeley, J., and N. Aliacar. 2001. Gene genealogies in a metapopulation. *Genetics* **159**:893–905.
- Wakeley, J., and J. Hey. 1997. Estimating ancestral population parameters. *Genetics* **145**:847–855.
- Wall, J. D. 2003. Estimating ancestral population sizes and divergence times. *Genetics* **163**:395–404.
- Wall, J. D., L. A. Frisse, R. R. Hudson, and A. Di Rienzo. 2003. Comparative linkage disequilibrium analysis of the  $\beta$ -globin hotspot in primates. *Am. J. Hum. Genet.* **73**:1330–1340.
- Watterson, G. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Pop. Biol.* **7**:256–276.
- Whiten, A., J. Goodall, W. C. McGrew, T. Nishida, V. Reynolds, Y. Sugiyama, C. E. Tutin, R. W. Wrangham, and C. Boesch. 1999. Cultures in chimpanzees. *Nature* **399**:682–685.
- Wise, C. A., M. Sraml, D. C. Rubinsztein, and S. Easteal. 1997. Comparative nuclear and mitochondrial genome diversity in humans and chimpanzees. *Mol. Biol. Evol.* **14**:707–716.
- Wiuf, C., and J. Hein. 2000. The coalescent with gene conversion. *Genetics* **155**:451–462.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
- Yi, S., D. L. Ellsworth, and W. H. Li. 2002. Slow molecular clocks in Old World monkeys, apes, and humans. *Mol. Biol. Evol.* **19**:2191–2198.
- Yu, N., M. I. Jensen-Seaman, L. Chemnick, J. R. Kidd, A. S. Deinard, O. Ryder, K. K. Kidd, and W. H. Li. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* **164**:1511–1518.
- Zhao, Z., L. Jin, Y. X. Fu et al. (10 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.

David B. Goldstein, Associate Editor

Accepted November 26, 2003