Hum Genet (1998) 103 : 211–227

ORIGINAL INVESTIGATION

Kenneth K. Kidd · Bharti Morar
Carmela M. Castiglione · Hongyu Zhao
Andrew J. Pakstis · William C. Speed
Batsheva Bonne-Tamir · Ru-Band Lu · David Goldman
Chaeyoung Lee · Yong Suk Nam · David K. Grandy
Trefor Jenkins · Judith R. Kidd

A global survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus

Received: 14 January 1998 / Accepted 19 March 1998

Abstract A four-site haplotype system at the dopamine D2 receptor locus (DRD2) has been studied in a global sample of 28 distinct populations. The haplotype system spans about 25 kb, encompassing the coding region of the gene. The four individual markers include three TaqI restriction site polymorphisms (RSPs) – TaqI "A", "B", and "D" sites – and one dinucleotide short tandem repeat polymorphism (STRP). All four of the marker systems are polymorphic in all regions of the world and in most individual populations. The haplotype system shows the highest average heterozygosity in Africa, a slightly lower average heterozygosity in Europe, and the lowest average heterozygosities in East Asia and the Americas. Across all populations, 20 of the 48 possible haplotypes reached a frequency of at least 5% in at least one population sample. However, no single population had more than six haplotypes reaching that frequency. In general, African populations had more haplotypes present in each population and more haplotypes occurring at a frequency of at least 5% in that population. Permutation tests for significance of overall disequilibrium (all sites considered simultaneously) were highly significant ($P<0.001$) in all 28 populations. Except for three African samples, the pairwise disequilibrium between the outermost RSP markers, TaqI "B" and "A", was highly significant with D' values greater than 0.8; in two of those exceptions the RSP marker was not polymorphic. Except for those same two African populations, the 16-repeat allele at the STRP also showed highly significant disequilibrium with the TaqI "B" site in all populations, with D' values usually greater than 0.7. Only four haplotypes account for more than 70% of all chromosomes in virtually all non-African populations, and two of those haplotypes account for more than 70% of all chromosomes in most East Asian and Amerindian populations. A new measure of the amount of overall disequilibrium shows least disequilibrium in African populations, somewhat more in European populations, and the greatest amount in East Asian and Amerindian populations. This pattern seems best explained by random genetic drift with low levels of recombination, a low mutation rate at the STRP, and essentially no recurrent mutation at the RSP sites, all in conjunction with an "Out of Africa" model for recent human evolution.

Kenneth K. Kidd (✉) · Carmela M. Castiglione · Andrew J. Pakstis
William C. Speed · Judith R. Kidd
Department of Genetics, Yale University School of Medicine,
New Haven, CT 06520-8005, USA
e-mail: kidd@biomed.med.yale.edu, Tel.: (203) 785 2654,
Fax: (203) 785 6568

Hongyu Zhao
Department of Epidemiology, Yale University School of Medicine,
New Haven, CT 06520-8005, USA

Bharti Morar · Trefor Jenkins
Department of Human Genetics, SAIMR and
University of the Witwatersrand, Johannesburg, South Africa

Batsheva Bonne-Tamir
Department of Human Genetics, Sackler School of Medicine,
Tel Aviv University, Israel

Ru-Band Lu
Department of Psychiatry, Tri-Service General Hospital,
National Defense Medical Center, Taipei, Taiwan

David Goldman
Laboratory of Neurogenetics, NIAAA/NIH, 12501 Washington Ave,
Rockville, MD 20852, USA

Chaeyoung Lee
Laboratory of Statistical Genetics, Institute of Environment and
Life Science, Hallym University, Chuncheon, 200-702, Korea

Yong Suk Nam
Department of Legal Medicine, Korea University,
Seoul, 136-701, Korea

David K. Grandy
Department of Physiology and Pharmacology, Oregon Health
Sciences University, Portland, OR 92701-3098, USA

Introduction

The human dopamine D2 receptor has a high affinity for many antipsychotic drugs (for examples see reviews Seeman and VanTol 1994; Picetti et al. 1997) and, since its cloning (Grandy et al. 1989a; DalToso et al. 1989; Araki et

al. 1992), neuroscientists, psychiatrists, and geneticists have shown great interest in DRD2, the gene for this receptor. DRD2 is one of five different dopamine receptor genes that have been identified in humans and are expressed in the central nervous system. Several polymorphisms have been identified in the DNA encompassing the coding sequences of this gene; most are in the introns or downstream flanking DNA (Grandy et al. 1989b; Bolos et al. 1990; Hauge et al. 1991; Sarkar et al. 1991; Parsian et al. 1991; Seeman et al. 1993; Suarez et al. 1994; Castiglione et al. 1995), but some are in coding regions (Gejman et al. 1994; Cravchik et al. 1996) and in the promoter region upstream of exon 1 (Arinami et al. 1997). Various of these polymorphisms have been used to map the locus genetically to the long arm of chromosome 11 (Gelernter et al. 1989; Grandy et al. 1989b; Litt et al. 1995); some of the polymorphisms have been used in linkage studies that have generally excluded this locus as causative for several neuropsychiatric disorders (e.g., Moises et al. 1991; Su et al. 1993; Gelernter et al. 1990, 1994).

The $Ser^{311} \rightarrow Cys^{311}$ amino acid variant, identified by Itakowa et al. (1993) and Gejman et al. (1994), is the only common (>1%) protein variant identified to date and has been shown to differ significantly in function from the common ("normal") receptor (Cravchik et al. 1996). Arinami et al. (1997) have recently reported promoter region variation that alters expression of a reporter gene in cultured cells transiently expressing transfected D2 receptor sequences. The Arinami et al. (1997) association study of two varying sites in this region with schizophrenia in Japanese was nominally significant for one of the sites, a single nucleotide insertion/deletion polymorphism. In summary, the role of genetic variation at this locus is under intense scrutiny for a variety of reasons (Gejman et al. 1994, for review; Arinami et al. 1997; Goldman et al. 1997).

Most of the studies of DRD2 and different disorders (primarily alcoholism and substance abuse: for reviews, see Gejman et al. 1994; Niswanger et al. 1995) have been association studies comparing the frequencies of the presumably neutral non-coding DNA polymorphisms in samples of patients and either random or unaffected controls. Almost all of those studies have used only the TaqI "A" site, a single nucleotide polymorphism (SNP) in a TaqI restriction site, located 10 kb downstream of the gene (Grandy et al. 1989b; Kidd et al. 1996; Iyengar et al. 1998). The results of these association studies have been inconsistent, at best (Gejman et al. 1994). Association studies of the Ser^{311}/Cys^{311} polymorphism have been consistently negative for schizophrenia and alcoholism (Goldman et al. 1997).

As discussed in Kidd et al. (1996) and Suarez et al. (1994), use of haplotypes of multiple genetic markers distributed through and around the gene is a powerful tool for resolving the controversial issues of such association studies based on individual polymorphisms. Haplotypes provide information on evolutionary histories, beyond what can be learned from individual markers, and we need to understand the evolutionary histories of normal allelic variants at the locus because, if a locus really has genetic variation affecting susceptibility to a complex trait, that variation also has an evolutionary history that is tied to the history of the entire region, as revealed by the adjacent normal genetic variation. By this logic, if genetic variation at DRD2 does affect some neuropsychiatric disorder(s), the measured genotype approaches of Templeton et al. (1988; see also Templeton and Sing 1993 and Templeton 1996), based on the cladistic relationships of the haplotypes, might then become relevant in identifying and quantifying that effect.

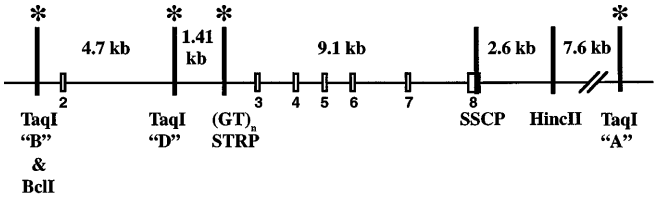

Fig. 1 Molecular locations of seven DRD2 polymorphisms. The four polymorphisms studied here (marked *) are shown relative to each other, to the exons of DRD2, and to several other known polymorphisms (Iyengar et al. 1998). The $(CA)_n$ short-tandem repeat polymorphism (STRP) in this study is shown as $(GT)_n$ since that is the DNA sequence, 5' to 3', on the sense strand. The TaqI "B" site is 913 bp upstream of the initiation codon in exon 2, and the TaqI "A" site is 10,542 bp downstream of the termination codon in exon 8

As an aid in the interpretation of haplotype data, we have precisely mapped several of the known polymorphisms in the DRD2 region (Fig. 1) and converted them to polymerase chain reaction (PCR) based typing (Castiglione et al. 1995; Kidd et al. 1996; Iyengar et al. 1998) to facilitate obtaining allele frequencies in diverse human populations. In this study, we present the haplotype frequencies and examine the linkage disequilibrium relationships of four of the DRD2 polymorphisms for a sample of 28 populations distributed around the world. These data are a superset of the data on three polymorphisms in 15 populations presented by Castiglione et al. (1995) and the data on A1-containing haplotypes presented by Kidd et al. (1996). The global pattern of variation revealed by these data reinforces the growing consensus from nuclear DNA studies that African populations have significantly more genetic variation than non-African populations (Bowcock et al. 1994; Deka et al. 1995; Armour et al. 1996; Tishkoff et al. 1996, 1998; Calafell et al. 1997). The data also indicate that, except for rare haplotypes, a relatively small number of DRD2 haplotypes occurs outside of Africa, but, especially in populations of European origin, no single polymorphic site can serve as a surrogate for the evolutionarily distinct haplotypes.

Materials and methods

Populations

The 28 populations we have typed for four markers in the DRD2 locus include seven African (Sekele San Bushmen, Central San Bushmen, Northern Sotho, Tsonga, Biaka, Mbuti, Ethiopians); five European and Southwest Asian (Yemenite Jews, Druze, Adygei, Danes, Finns), henceforth referred to collectively as European; eight East Asian (two Chinese groups – one from San Francisco and the other from Taiwan –Koreans, Japanese, Ami, Atayal, Yakut, Cambodians);

one Australomelanesian (Nasioi); four North American (Cheyenne, Jemez Pueblo, Pima, Maya); and three South American (Karitiana, Rondonia Surui, Ticuna). Most of the populations and samples have been described elsewhere (Castiglione et al. 1995; Tishkoff et al. 1996).

The samples of the Sekele San, Sotho, and Tsonga groups are described in Spurdle and Jenkins (1992). The Central San are a group of sedentary G/wi- and G//ana- speakers from the Khutse area of the central Kalahari, who have been the subjects of a longitudinal health and hematological study initiated by Dr. Susan Kent (Kent and Dunn 1993). The Korean samples are unrelated residents of Seoul. The Coriell Institute for Medical Research (NIGMS Human Genetic Mutant Cell Line Repository), Camden, New Jersey, has at least five to ten cell lines from sixteen of these samples available for distribution: Biaka, Mbuti, Adygei, Chinese (SF), Japanese, Cambodians, Ami, Atayal, Yakut, Maya, R. Surui, Karitiana, Nasioi, Druze, Jemez Pueblo, and Cheyenne. All samples were collected with informed consent from the participants and approval from the appropriate institutional review boards. The DNA in this study was purified by means of standard phenol-chloroform extraction and ethanol precipitation (Sambrook et al. 1989) from Epstein-Barr virus-transformed lymphoblastoid cell lines for most samples and from lymphocytes for other samples.

Polymorphism typing

The four DRD2 markers typed for this study are three bi-allelic TaqI restriction fragment length polymorphisms (RFLPs) and a four- to six-allele $(CA)_n$ short tandem repeat polymorphism (STRP), altogether spanning a distance of 25.4 kb (Fig. 1). All of these markers have been described previously: TaqI "A" by Grandy et al. (1989b); TaqI "B" by Hauge et al. (1991); TaqI "D" by Parsian et al. (1991); and the STRP by Hauge et al. (1991). All typings are PCR-based, using the primers and protocols described by Castiglione et al. (1995), Kidd et al. (1996), and Iyengar et al. (1998). Allele frequencies at the individual sites were calculated by gene counting. The assumption of Hardy-Weinberg ratios was tested for the separate sites in each sample by means of an auxiliary program, FENGEN (unpublished by A.J. Pakstis), which also creates the input file for the program HAPLO from raw data records.

Haplotype frequency estimates

The population samples studied consisted mainly of unrelated individuals; therefore, family data could not be used to determine haplotypes in multiply heterozygous individuals. Instead, maximum likelihood estimates of haplotype frequencies and the standard errors (jackknife method) were calculated from the multi-site marker typing data, using the program HAPLO (Hawley and Kidd 1995), which implements the EM algorithm (Dempster et al. 1977). HAPLO accommodates individuals with either missing data at some sites or partial phase information, by giving them unique phenotypes corresponding to the set of underlying genotypes compatible with the information available, as explained in Hawley and Kidd (1995). In a few cases, first-degree relatives could be used to determine the haplotypes from transmission patterns and the genotyped, unrelated members of the family were included in the data set. In most other cases, relationships were distant, or the set of relatives did not allow resolution of haplotypes and all the relatives were included as though they were unrelated individuals. This does not bias the estimates, but it does increase the sampling error somewhat. Heterozygosities for individual sites and for the haplotype have been estimated as $1-\Sigma p_i^2$, where p_i represents the allele frequencies for any given system.

Disequilibrium

The standardized, pairwise linkage disequilibrium value D' (Lewontin 1964) was calculated for each pair of markers, and the null hypothesis of linkage equilibrium (D'=0) was tested with an asymptotically chi-square statistic (see Eq 3.10 in Weir 1996), by means of the computer program LINKD (Pakstis, unpublished program) which uses the sample sizes and the haplotype frequency estimates from HAPLO as input. HAPLO also calculates a likelihood-ratio statistic that can, under some circumstances, be interpreted as an asymptotically chi-square statistic to test for significance of overall linkage disequilibrium. By "overall disequilibrium", we mean deviation of observed (i.e., estimated) haplotype frequencies from the expected frequencies, considering all sites simultaneously and not just pairwise.

Permutation tests

Because the likelihood-ratio statistic may not closely approximate a chi-square distribution for multi-site haplotype data, we also estimated significance of the overall disequilibrium for all 28 populations using a permutation test (Good 1995). The use of permutation tests for linkage disequilibrium has been proposed by Zaykin et al. (1995) and by Slatkin and Excoffier (1996). In brief, the permutation approach uses all the phenotypic and genotypic information on individuals in a sample to determine significance of linkage disequilibrium by independently permuting, 1000 times, the phenotypes at all sites in a haplotype for all individuals in the sample. For each permutation, the data are entered into HAPLO to obtain the maximum-likelihood estimates of haplotype frequencies and the likelihood-ratio statistic. Under the null hypothesis of no disequilibrium, any observed phenotype (or genotype) at one site is equally likely to occur with any observed phenotype at a second site. Therefore, any set of randomly reconstructed multi-site phenotypes, with the same numbers of individuals and the same marginal phenotype frequencies at each site, is equally likely to occur by chance alone. An exact test, not dependent on the assumption of a chi-square distribution, of the significance of the overall disequilibrium in the observed sample is approximated by the proportion of the permuted samples with more extreme likelihood ratio statistics than the observed sample.

Quantifying overall disequilibrium

Since the *significance* of the disequilibrium does not quantify the *amount* of disequilibrium, we have developed a heuristic measure of the relative amounts of disequilibrium in different populations. We proceeded via two steps. First, we calculated a standard deviation coefficient, D_{sd}, which is the number of standard deviation units that the observed likelihood ratio statistic falls from the mean of the distribution under the null hypothesis, as estimated by the 1000 permutations. Among the tests for which none of the 1000 permuted samples is more extreme than the observed, those qualitative results imply only highly significant disequilibrium ($P<0.001$). The actual permutation distribution, however, does allow further quantification of the relative significance levels of disequilibrium in different data sets, for the moment simply approximated by D_{sd}. Note, however, that the distribution of the likelihood ratio statistic for more complex multi-site haplotype systems is usually not a chi-square distribution (nor, so far as we have found, any other distribution with closed form) and, so, this D_{sd} value does not translate into a significance level.

As a second step, we further standardized the D_{sd} in a manner suggested by simulation studies (Zhao et al. 1997; Zhao et al. unpublished). In these simulations, the average D_{sd} is nearly linear with sample size over the range of sizes present in this study. This is analogous to the linear relationship of chi-square with sample size. Therefore, we have used the value $D_{sd}/2N$ for each population to allow the amount of overall disequilibrium between populations to be compared independently of sample size.

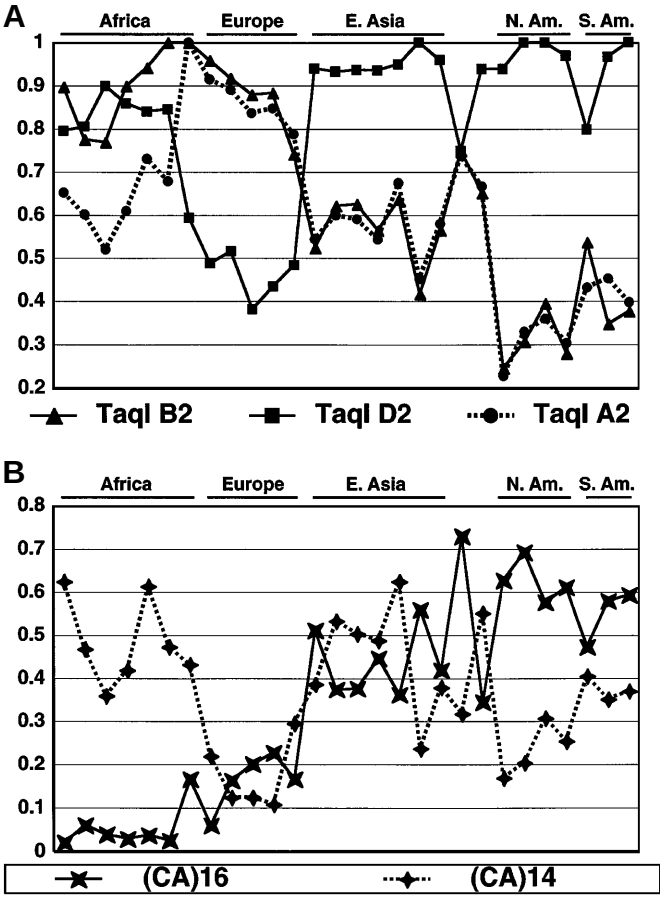

Fig. 2a, b Allele frequencies at the individual sites for all 28 populations. The populations are ordered from left to right, exactly as given from top to bottom in Table 1. The *bars* across the top indicate geographic groupings, with the single Australo-Melanesian population (Nasioi) and the single North East Asian population (Yakut) in the gap between the bars for East Asia and North America. **a** Frequencies of the "enzyme-cutting" alleles (B2, D2, and A2) at the three TaqI restriction sites. **b** Frequencies of the two most common short tandem repeat polymorphism (STRP) alleles, 14 repeats and 16 repeats

Results

Molecular mapping

The molecular distances given in Fig. 1 are more accurate than in previous studies because we have now completed sequencing of the entire region from exon 8 to the TaqI "A" site (GenBank accession #AF050737). This sequencing confirms that the TaqI "A" site is 10 kb downstream from the stop codon in exon 8: 10,542 bp, to be precise, based on the sequence of cloned material and some templates that were amplified by means of PCR from genomic DNA.

Allele frequencies at individual sites

The individual site allele frequency estimates by gene counting (Fig. 2) were virtually identical (i.e., either identical or else differences in estimates began to appear at the third decimal position) to those obtained by summing frequencies of all the haplotypes containing a given allele. This was, however, not true for the Arizona Pima population, in which the EM algorithm in HAPLO estimated from a data set that included a number of cases where typings were missing for an individual site. However, even in this case, the difference was not statistically significant ($P>0.8$) and amounted to a shift of about six chromosomes between the tallies for the two most common alleles.

Each of the four DRD2 sites was tested in each sample for agreement with the underlying assumption of Hardy-Weinberg ratios and no systematic departures were found. Of the 112 possible tests (4 sites × 28 samples), 7 were not possible because only a single allele was present. Of 105 tests, 5 were significant at the 5% level; this is approximately the expected number, given the large number of tests, and these appeared to be randomly distributed across sites and samples. There were, however, two other tests nominally significant ($P<0.005$), involving different sites and samples. One of these results from a modest excess of homozygotes at the TaqI "D" site in the Ticuna sample, while the other is due to an excess of heterozygotes at the TaqI "A" site in the Atayal sample. Neither of these seems biologically meaningful.

All four marker systems are polymorphic in nearly every population studied (Fig. 2). The "B" site is fixed (for the B2 allele) only in the samples of the Mbuti and Ethiopians; the "D" site is fixed (for the D2 allele) only in the Atayal, Jemez Pueblo, Arizona Pima, and Karitiana; the "A" site is fixed (for the A2 allele) only in the small Ethiopian sample. The B2 allele has a frequency greater than 74% in all of the African and European populations. In contrast, in all other populations, the B2 allele has a frequency of less than 65%, except for the Nasioi with a frequency of 75%. The overall pattern of frequency variation at the TaqI "B" site is of high frequencies in Africa and Europe, frequencies closer to 50% in the East Asian populations, and generally lower frequencies in the Amerindian populations.

At the TaqI "A" site, the pattern of frequency variation is virtually identical to that of the "B" site, in agreement with the strong disequilibrium between them; except that the A2 allele frequencies are somewhat lower in Africa than those of the B2 allele. Outside of Africa, the A2 allele frequencies are virtually identical to the B2 allele frequencies in most populations (Fig. 2a). In Africa, the A2 allele frequencies range from 52% to 73% (except for the small Ethiopian sample, in which it is 100%), whereas the B2 allele frequencies range from 77% to 94% and to 100% in the Mbuti and the Ethiopians. The TaqI "D" site shows a different pattern, with moderately high frequencies from 60% to 90% for the D2 allele in African populations, frequencies from 38% to 52% in European populations and frequencies in most Asian, Pacific, and New World populations of greater than 90%. Two exceptions are the Nasioi at 75% and the Ticuna at 80%.

Table 1 Frequency estimates for the 20 common DRD2 haplotypes. The four polymorphic sites are the TaqI "B" site, the TaqI "D" site, a $(CA)_n$ short-tandem repeat polymorphism (STRP), and the TaqI "A" site. The 28 populations studied are given in the first column, with populations grouped by broad geographic region: sub-Saharan and North East Africa; Europe and South West Asia ("Europe" for short); East and South East Asia ("East Asia" for short); Australo-Melanesia; North East Asia; North America; and South America. The frequency estimates for the 20 haplotypes with frequency estimates of ≥5% in at least one population are given in the columns under the haplotype labels. Estimates that do not correspond to discrete observations are an inherent aspect of using the EM algorithm to obtain maximum likelihood estimates of haplotype frequencies. Haplotypes are listed across the top and read vertically as the alleles at the four sites. The B1, D1, and A1 alleles represent the site-absent state for the RFLP sites while B2, D2, and A2 are the site-present alleles. STRP alleles are labeled according to the number of CA repeats present

A

Population sample	B1 D2 17 A1	B1 D2 16 A1	B1 D2 15 A1	B1 D2 14 A1	B1 D2 12 A1	B1 D2 16 A2	B1 D2 14 A2	B2 D1 15 A1	B2 D1 13 A1	B2 D1 16 A2	B2 D1 15 A2
Sekele San	0	0.020	0	0	0.010	0	0.028	0.011	0.085	0	0.086
C. San	0	0.061	0.081	0.007	0.030	0	0.027	0	0.031	0	0.152
N. Sotho	0	0.040	0.010	0	0.169	0	0	0	0.021	0	0.079
Tsonga	0.010	0.030	0.010	0	0.040	0	0	0	0.051	0	0.079
Biaka	0	0.042	0	0	0.008	0	0.008	0	0.008	0	0.089
Mbuti	0	0	0	0	0	0	0	0	0.110	0.013	0
Ethiopians	0	0	0	0	0	0	0	0	0	0.156	0.156
Yemenites	0	0.037	0	0	0	0	0	0	0.012	0.024	0.407
Druze	0	0.070	0	0	0	0	0	0.037	0	0.086	0.353
Adygei	0	0.085	0	0	0	0.011	0	0.024	0	0.069	0.467
Danes	0	0.105	0	0	0	0	0	0.034	0	0.030	0.477
Finns	0	0.161	0.021	0	0	0	0.020	0	0	0.018	0.305
Han (SF)	0	0.463	0	0.012	0	0	0	0	0	0.006	0.036
Han (TWN)	0	0.378	0	0	0	0	0	0	0	0	0.067
Koreans	0.004	0.332	0.016	0.004	0	0.008	0	0	0	0	0.038
Japanese	0	0.360	0.031	0.023	0	0	0.011	0	0	0.022	0.021
Ami	0	0.286	0	0.013	0	0.051	0	0	0	0	0.013
Atayal	0	0.509	0	0	0	0.050	0.012	0	0	0	0
Cambodians	0	0.420	0	0	0	0	0	0	0	0	0.017
Nasioi	0	0.227	0	0	0	0	0	0	0	0.023	0.159
Yakut	0.019	0.284	0	0	0	0.046	0	0	0	0.015	0.045
Cheyenne	0.098	0.607	0.020	0.011	0	0	0	0	0	0	0.041
J. Pueblo	0	0.670	0	0	0	0.023	0	0	0	0	0
Pima (AZ)	0	0.605	0	0	0	0	0	0	0	0	0
Maya	0.047	0.504	0.010	0.087	0.021	0.013	0.012	0	0	0.018	0
Ticuna	0	0.416	0	0	0	0.016	0	0.070	0	0	0.018
R. Surui	0	0.531	0.011	0.011	0	0	0.066	0	0	0.022	0
Karitiana	0	0.595	0	0	0	0	0.028	0	0	0	0

B

Population	B2 D1 14 A2	B2 D1 13 A2	B2 D2 14 A1	B2 D2 13 A1	B2 D2 12 A1	B2 D2 16 A2	B2 D2 15 A2	B2 D2 14 A2	B2 D2 13 A2	Residual: sum of 17 uncommon haplotypes
Sekele San	0	0	0.104	0.042	0	0	0.013	0.456	0.026	0.119
C. San	0	0	0.051	0.083	0.021	0	0.048	0.375	0	0.033
N. Sotho	0	0	0.102	0.086	0.041	0	0.041	0.258	0.142	0.011
Tsonga	0	0	0.036	0.153	0.060	0	0.011	0.384	0.126	0.010
Biaka	0.015	0	0.103	0.047	0.008	0	0	0.484	0.133	0.055
Mbuti	0	0	0.049	0.130	0	0.013	0.051	0.395	0.208	0.031
Ethiopians	0.039	0.055	0	0	0	0	0	0.411	0.183	0
Yemenites	0.038	0.031	0	0.010	0	0	0.067	0.181	0.166	0.027
Druze	0.008	0	0	0	0	0.008	0.032	0.117	0.281	0.008
Adygei	0.010	0	0.021	0	0	0	0	0.094	0.183	0.036
Danes	0	0	0	0	0	0.068	0.011	0.109	0.141	0.025
Finns	0.070	0.042	0	0	0	0	0	0.149	0.128	0.086
Han (SF)	0	0	0.014	0	0	0.026	0	0.362	0.063	0.018
Han (TWN)	0	0	0.022	0	0	0	0	0.500	0.033	0
Koreans	0.006	0.003	0.032	0	0	0.017	0.008	0.459	0.048	0.025
Japanese	0.021	0	0	0	0	0.032	0.012	0.434	0	0.033
Ami	0.025	0	0.013	0	0	0.013	0	0.574	0	0.012
Atayal	0	0	0.039	0	0	0	0	0.188	0.190	0.012
Cambodians	0.023	0	0	0	0	0	0.043	0.357	0.140	0
Nasioi	0.049	0	0	0	0	0	0	0.269	0.227	0.046
Yakut	0	0	0	0.015	0	0	0	0.543	0.017	0.016
Cheyenne	0	0	0	0	0	0	0	0.156	0.010	0.057
J. Pueblo	0	0	0	0	0	0	0	0.205	0.102	0
Pima (AZ)	0	0	0	0	0	0	0.016	0.262	0.082	0.035
Maya	0	0	0.011	0	0	0.069	0	0.142	0.040	0.026
Ticuna	0.042	0.022	0.010	0.010	0	0.006	0	0.309	0.018	0.063
R. Surui	0	0	0	0	0	0.012	0	0.268	0.035	0.044
Karitiana	0	0	0	0	0	0	0	0.377	0	0

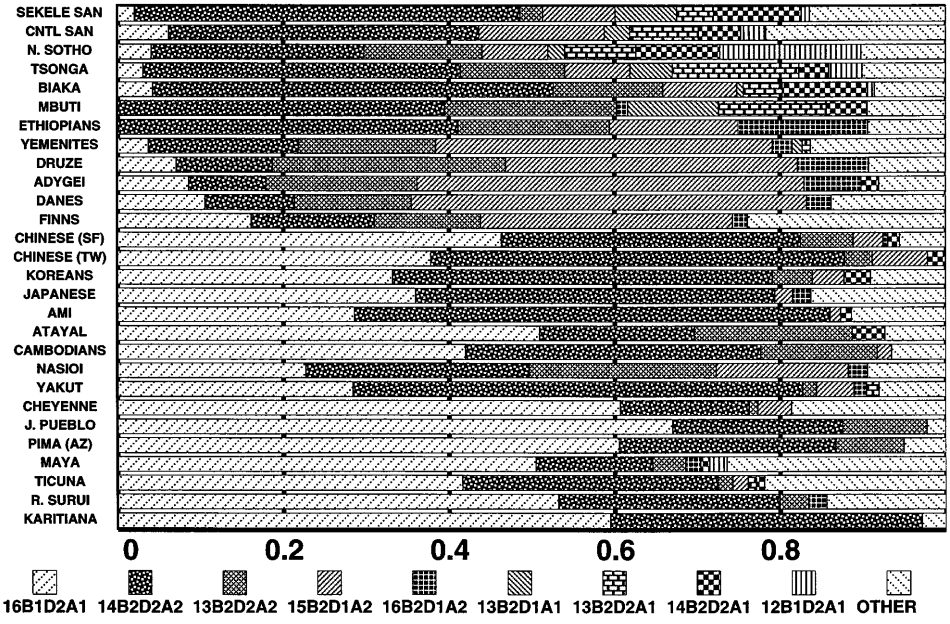

Fig. 3 Relative frequencies of the nine most common DRD2 haplotypes shown as "*stacked*" *bars* for 28 populations. The key for the shading is given across the bottom in run-on allele symbols, where B1, D1, and A1 represent the "non-cutting" alleles at the three TaqI sites; B2, D2 and A2 the three "cutting" alleles; and the initial number (12–16) gives the short-tandem repeat polymorphism (STRP) allele, out of genomic order for ease of reading

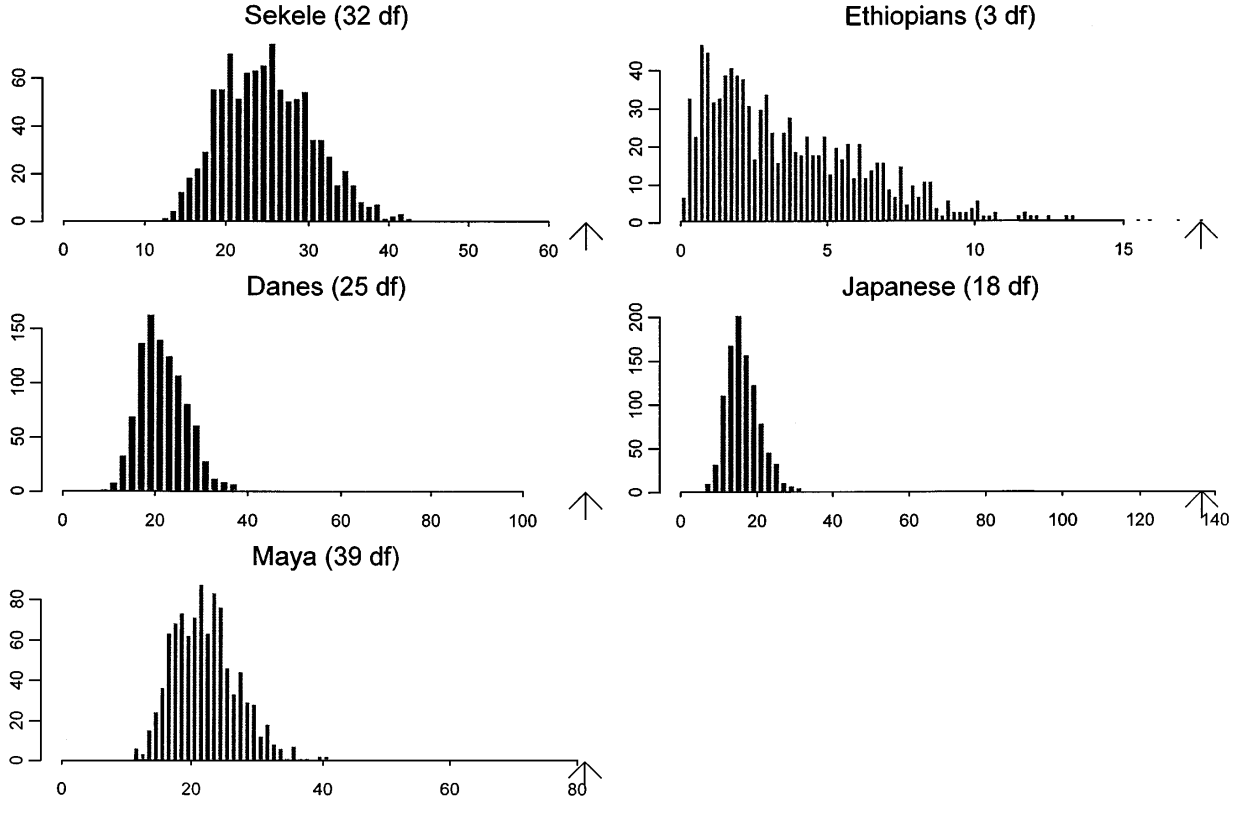

Fig. 4 Permuted-sample distributions (histograms) of the likelihood ratio statistic for the null hypothesis of no overall disequilibrium are shown for five populations. Each distribution is based on evaluation of 1000 independent permutations of the observed phenotypes at the four markers. The relative position of the likelihood ratio statistic for the observed data is shown by the *arrow* to the right of each distribution. The number of degrees of freedom are shown in parentheses (*n df*) next to the name of each population sample, so that it is convenient to observe how different the mean of the 1000 permutations can be from its expectation. Summary statistics for these distributions are given in Table 5

Haplotype frequencies and distributions

Frequency estimates of the 20 "most common" of the possible 48 haplotypes, those estimated to occur at a frequency of ≥5% in at least one of the 28 populations, are given in Table 1. To save space, frequency estimates of the remaining rare and absent haplotypes, and all of the jackknife standard errors are not given (but are available from the authors).

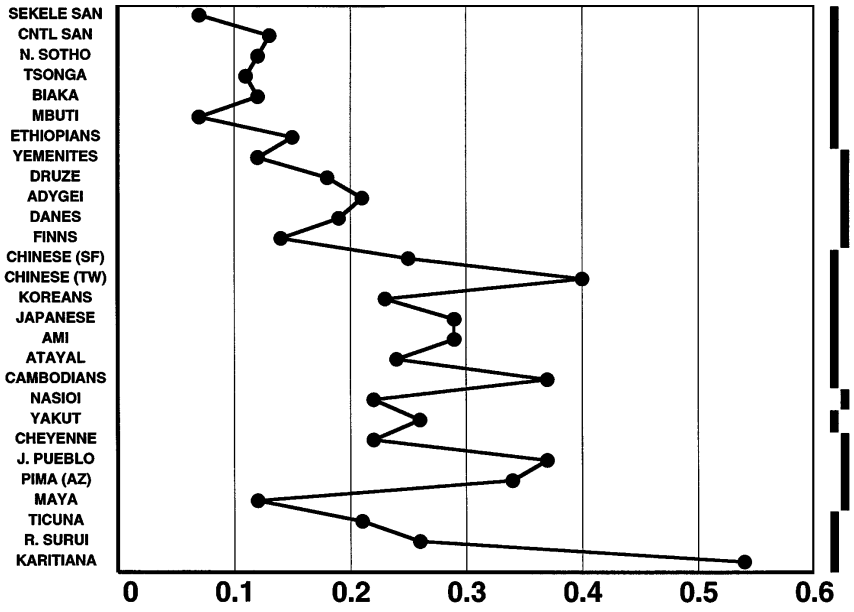

Fig. 5 The standardized amounts of overall linkage disequilibrium. The values are calculated as $D_{sd}/2N$ using the numbers in Table 5. The populations are arranged as in Tables 1–5 and Fig. 3 by geographic region from sub-Saharan at the *top* to South American at the *bottom*, an ordering generally correlated with increasing disequilibrium. The *staggered vertical bars* at the right identify the geographic groupings as in Fig. 2

Table 2 Haplotypes not in Table 1. Of the 48 theoretically possible multi-site haplotypes in this study, 11 haplotypes have frequency estimates of zero (non-occurring list) and another 17 are at "uncommon" frequencies (<5% but non-zero) in at least one of the populations sampled

Uncommon	"Non-occurring"
B1-D1-16-A1	B1-D1-17-A1
B1-D1-14-A1	B1-D1-15-A1
B1-D1-13-A1	B1-D1-12-A1
B1-D1-16-A2	B1-D1-17-A2
B1-D1-15-A2	B1-D1-13-A2
B1-D1-14-A2	B1-D2-17-A2
B1-D1-12-A2	B1-D2-12-A2
B1-D2-13-A1	B2-D1-17-A1
B1-D2-13-A2	B2-D1-17-A2
B2-D1-16-A1	B2-D1-12-A2
B2-D1-14-A1	B2-D2-17-A1
B2-D1-12-A1	
B2-D1-14-A2	
B2-D2-16-A1	
B2-D2-15-A1	
B2-D2-17-A2	
B2-D2-12-A2	

The jackknife estimates were generally somewhat larger than the corresponding binomial standard error estimates. In addition to the 20 haplotypes in Table 1, 17 other haplotypes (listed in Table 2) are estimated to occur only rarely, at a frequency of less than 5%, but may occur in more than one of the samples. Eleven of the possible haplotypes do not appear (zero frequency estimates) in this study, but some of them may exist at very low frequencies and might be detectable in larger samples. One of the 17 "rare" haplotypes (B1-D1-13-A1 in the Sekele San) has a frequency estimate corresponding to half an occurrence, indicating that one phenotype could not be resolved by the EM algorithm and one of the possible genotypes involved that haplotype; it was absent in all other samples. Such estimates that do not correspond to discrete observations are an inherent aspect of using the EM algorithm to obtain maximum likelihood estimates for multi-site haplotypes.

Figure 3 shows graphically the haplotype frequency data (from Table 1) with the rarest haplotypes clustered in "other". This figure makes it clear that a subset of nine haplotypes accounts for a minimum of 70% of all chromosomes, in all samples, and for more than 80% of all chromosomes in all but four populations (Central San, Finns, Maya, and Ticuna). In most populations from East Asia and the Americas, only three haplotypes – B1-D2-16-A1, B2-D2-14-A2, and B2-D2-13-A2 – account for the majority of all chromosomes. In Europe, a fourth haplotype becomes quite common, B2-D1-15-A2. These four most common haplotypes account for between 52% and 74% of all chromosomes in the African populations, between 74% and 83% in the European populations, between 82% and 97% in the east Asian populations, and between 67% and 95% in the Amerindian populations.

In total, more than 3100 chromosomes were examined, including some related individuals excluded from gene frequency estimates. Table 3 gives the number of chromosomes typed in each population for at least three of the four markers.

Genetic diversity

For each population, Table 3 gives the within-population genetic diversity, estimated as the expected heterozygosity for the observed haplotype frequencies and the genetic diversity (heterozygosity) for each of the four component polymorphisms. Heterozygosity for the haplotyped multi-site system ranges from 50% to 86% and, except in Africa

Table 3 Haplotype statistics. For each population, the table gives the number of chromosomes with typing results for at least three of the four sites, the expected heterozygosities based on the intact haplotypes and on each of the four sites individually, the number of haplotypes present, and the number found at common (greater than 5%) frequencies (frq)

Population sample	2*n*	Expected heterozygosity					Count of haplotypes with:	
		Haplotype	"B"	"D"	(CA)n	"A"	Frq >0%	Frq >5%
Sekele San	100	0.76	0.18	0.33	0.57	0.46	17	4
C. San	98	0.81	0.35	0.31	0.67	0.48	15	6
N. Sotho	100	0.86	0.35	0.18	0.74	0.50	12	6
Tsonga	100	0.80	0.18	0.24	0.69	0.48	13	6
Biaka	138	0.72	0.11	0.27	0.57	0.39	13	4
Mbuti	78	0.76	0	0.26	0.57	0.44	9	6
Ethiopians	32	0.74	0	0.48	0.70	0	6	5
Yemenites	82	0.76	0.08	0.50	0.65	0.16	11	4
Druze	128	0.77	0.15	0.50	0.70	0.19	10	5
Adygei	104	0.73	0.21	0.47	0.67	0.27	11	5
Danes	92	0.72	0.21	0.49	0.64	0.26	10	5
Finns	66	0.83	0.38	0.50	0.72	0.33	12	5
Han (SF)	84	0.65	0.50	0.11	0.58	0.50	11	3
Han (TWN)	90	0.60	0.47	0.13	0.57	0.48	5	3
Koreans	254	0.68	0.47	0.12	0.60	0.48	17	2
Japanese	94	0.68	0.49	0.12	0.56	0.50	11	2
Ami	80	0.58	0.46	0.10	0.48	0.44	9	3
Atayal	84	0.67	0.49	0	0.59	0.49	7	4
Cambodians	50	0.67	0.49	0.08	0.66	0.49	6	3
Nasioi	44	0.80	0.38	0.37	0.74	0.39	8	4
Yakut	66	0.62	0.45	0.12	0.57	0.44	9	2
Cheyenne	114	0.60	0.37	0.11	0.56	0.35	10	3
J. Pueblo	88	0.50	0.43	0	0.47	0.44	4	3
Pima (AZ)	86	0.56	0.48	0	0.56	0.46	5	3
Maya	100	0.71	0.40	0.06	0.56	0.42	14	4
Ticuna	134	0.72	0.50	0.32	0.62	0.49	15	3
R. Surui	92	0.64	0.45	0.06	0.54	0.50	11	3
Karitiana	106	0.50	0.47	0	0.48	0.48	3	2

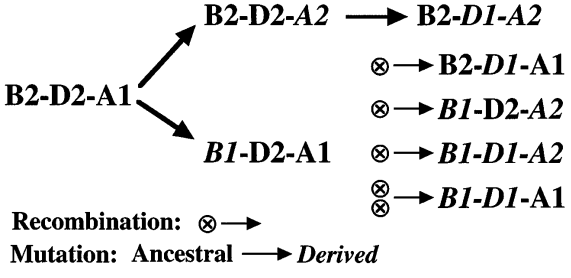

Fig. 6 Evolutionary relationships of DRD2 haplotypes defined by three Taq I restriction site polymorphisms. The ancestral alleles (B2, D2, and A1) are all identical in sequence to those in chimpanzees, gorillas, and orangutans (Iyengar et al. 1998). The frequency data in the different human populations, especially those in Africa, suggest the illustrated sequence of events by which three mutational events resulted in four of the eight possible haplotypes. The other four were then generated by at least one crossover between two of these mutationally derived haplotypes. The data are insufficient to exclude other possible evolutionary schemes

and Europe, is often not much greater than the heterozygosity of the STRP. Table 3 also gives the total number of haplotypes estimated to be present in each population and the number of those that have common (≥5%) frequencies.

Disequilibrium

Table 4 presents the three pairwise standardized D' values for the three bi-allelic markers, TaqI "B", TaqI "D", and TaqI "A", and the six pairwise D' values between each of those bi-allelic markers and the two most common STRP alleles; those with 16 and 14 repeats. With one exception, all populations in which both sites are polymorphic have highly significant disequilibrium (with D'>0.8) between the "B" and "A" sites; the most distantly separated sites, physically. The exception to statistical significance is the Sekele San with a non-significant D'=0.41. In contrast, the D' values and patterns of significance for the "B" to "D" and the "D" to "A" site comparisons are less consistent and only occasionally significant, with no clear pattern.

Pairwise disequilibrium comparisons of the bi-allelic markers with specific STRP alleles are more consistent.

Table 4 Pairwise linkage disequilibrium values for the three two-allele restriction site polymorphisms and the two common alleles at the short tandem repeat polymorphism (STRP). The standardized linkage disequilibrium coefficient (D') of Lewontin (1964) is shown. When no test was possible, because the sample was monomorphic at one or both of the sites, "-" appears instead of a value. The sign of D' is relative to the "1–1" and "2–2" haplotypes, to the "1–16" and "2–not 16" haplotypes, and to the "1–14" and "2–not 14" haplotypes. Significance levels to test whether the disequilibrium value differs from zero (one degree of freedom) are indicated as follows: * for $P<0.05$, ** for $P<0.01$, and *** for $P<0.001$. No asterisk indicates that the test was not significant ($P>0.05$)

	TaqI "B" TaqI "D" D'		TaqI "B" TaqI "A" D'		TaqI "D" TaqI "A" D'		16 TaqI "B" D'		16 TaqI "D" D'		16 TaqI "A" D'		14 TaqI "B" D'		14 TaqI "D" D'		14 TaqI "A" D'	
Sekele San	–0.76		0.41		0.34	*	1.00	***	–1.00		1.00		–0.56	*	–0.80	***	–0.40	***
C. San	–1.00	**	0.80	***	–0.46		1.00	***	–1.00		1.00	**	–0.68	***	–0.89	***	–0.64	***
N. Sotho	–1.00		1.00	***	–0.56		1.00	***	–1.00		1.00	*	–1.00	***	–1.00	*	–0.41	**
Tsonga	–0.29		0.84	***	–0.07		1.00	***	–1.00		1.00	*	–1.00	**	–1.00	***	–0.78	***
Biaka	–1.00		0.81	***	0.12		1.00	***	–1.00		1.00	***	–0.77	**	–0.85	***	–0.36	**
Mbuti	—		—		0.88	***	—		0.41		–1.00		—		–0.58	*	–0.47	**
Ethiopians	—		—		—		—		1.00	**	—		—		–0.79	***	—	
Yemenites	–1.00		1.00	***	–0.73	*	1.00	***	–0.23		0.57	***	–1.00		–0.66	**	–1.00	
Druze	–1.00	**	1.00	***	–0.33		0.88	***	0.08		0.53	***	–1.00		–0.87	***	–1.00	
Adygei	–0.73	***	0.88	***	–0.40	*	1.00	***	–0.13		0.66	***	–1.00		–0.87	***	0.00	
Danes	–0.67	**	0.90	***	–0.45	*	1.00	***	–0.58	***	0.71	***	–1.00		–1.00	***	–1.00	
Finns	–0.58	**	0.80	***	–0.72	**	0.87	***	–0.81	**	0.87	***	–0.44		–0.19		–0.50	
Han (SF)	–0.59		0.95	***	–0.80		0.95	***	–0.22		0.89	***	–0.93	***	–1.00		–0.87	***
Han (TWN)	–1.00	*	1.00	***	–1.00	*	1.00	***	–1.00	*	1.00	***	–1.00	***	–1.00	**	–0.90	***
Koreans	–0.57		0.96	***	–0.57		0.88	***	–0.57		0.87	***	–0.96	***	–0.68	**	–0.81	***
Japanese	–1.00	*	0.95	***	–1.00	*	0.72	***	–0.23		0.78	***	–0.84	***	–0.33		–0.90	***
Ami	–0.30		0.94	***	–0.22		0.94	***	–0.30		0.87	***	–0.94	***	–0.22		–0.87	***
Atayal	—		0.83	***	—		1.00	***	—		0.84	***	–0.91	***	—		–0.70	***
Cambodians	–1.00		1.00	***	–1.00		1.00	***	–1.00		1.00	***	–1.00	***	0.32		–1.00	***
Nasioi	–1.00	*	0.88	***	–1.00	*	0.87	***	–0.63		1.00	***	–1.00	**	–0.33		–1.00	**
Yakut	–1.00		0.86	***	–1.00		0.91	***	–0.31		0.84	***	–1.00	***	–1.00	*	–1.00	***
Cheyenne	–0.57	**	0.88	***	–1.00	***	0.83	***	–1.00	***	1.00	***	–0.91	***	–1.00		–0.91	***
J.Pueblo	—		1.00	***	—		1.00	***	—		1.00	***	–1.00	***	—		–1.00	***
Pima (AZ)	—		1.00	***	—		1.00	***	—		1.00	***	–1.00	***	—		–1.00	***
Maya	–0.44		0.94	***	–0.42		0.50	***	1.00		0.47	***	–0.45	***	–1.00		–0.44	***
Ticuna	–0.68	***	0.88	***	–0.01		0.97	***	–0.77	***	0.89	***	–1.00	***	–0.19		–0.86	***
R. Surui	–0.49		0.94	***	–0.42		0.77	***	0.21		0.87	***	–0.64	***	–0.06		–0.89	***
Karitiana	—		1.00	***	—		1.00	***	—		1.00	***	–1.00	***	—		–1.00	***

Table 5 Summary statistics for the permutation test distributions. For each population sample the table shows, for the permutation test of overall disequilibrium, the corresponding mean and variance of the distribution of the likelihood ratio statistic, the maximum value found in 1000 permutations, the value calculated from the observed data, the D_{sd} value, and the degrees of freedom. The "Figure" column indicates the five geographically representative population distributions displayed in Fig. 4. The "Obs" value in this table could be interpreted as a chi-square with the degrees of freedom indicated if one can validly make the assumption that the distribution under the null hypothesis is a chi-square distribution. That assumption appears to be invalid for most samples in this study

Population	Figure	Mean	Variance	Maximum	Observed	D_{sd}	DF
Sekele San	**4a**	25.0	32.8	45.0	64.7	6.9	32
C. San		31.0	36.8	58.4	111.1	13.2	32
N. Sotho		28.8	32.2	49.1	98.4	12.3	32
Tsonga		28.9	34.4	53.0	92.3	10.8	39
Biaka		25.6	35.2	54.7	125.8	16.9	39
Mbuti		9.8	14.2	28.9	31.6	5.8	10
Ethiopians	**4b**	3.8	8.3	17.0	17.7	4.8	3
Yemenites		16.1	19.4	37.0	59.3	9.8	25
Druze		21.3	27.9	42.3	143.2	23.1	25
Adygei		23.3	29.1	43.2	143.3	22.2	25
Danes	**4c**	21.8	26.8	45.4	114.0	17.8	25
Finns		27.0	27.3	48.3	76.3	9.4	25
Han (SF)		19.8	25.8	46.6	125.1	20.8	25
Han (TWN)		19.5	24.1	39.5	194.2	35.6	25
Koreans		25.6	34.6	48.1	374.7	59.3	32
Japanese	**4d**	16.3	19.7	33.3	136.5	27.1	18
Ami		12.5	15.3	28.6	101.8	22.9	18
Atayal		11.3	18.6	33.8	98.9	20.3	7
Cambodians		18.2	21.0	36.1	102.9	18.5	25
Nasioi		24.2	21.1	42.3	68.2	9.6	25
Yakut		20.6	23.0	44.9	101.7	16.9	32
Cheyenne		22.3	27.1	46.5	155.1	25.5	32
J. Pueblo		8.6	15.5	24.6	136.8	32.6	7
Pima(AZ)		10.7	13.4	28.1	118.1	29.4	10
Maya	**4e**	22.2	25.8	40.6	81.1	11.6	39
Ticuna		27.2	39.3	48.6	206.4	28.6	25
R. Surui		13.3	17.1	36.3	114.1	24.4	25
Karitiana		4.5	9.6	20.7	182.5	57.5	4

The 14-repeat allele shows significant to highly significant disequilibrium with both the "B" and "A" sites in all except the European and Southwest Asian populations. Even for those five populations, the D' values are generally negative, often -1.0, indicating a low frequency or absence of one of the possible haplotypes in the same general pattern of association seen elsewhere, though it is not statistically significant. In contrast, the 14-repeat allele and the TaqI "D" site show highly significant disequilibrium in the European (except for the Finns) and Southwest Asian populations. That disequilibrium is also strongly negative and significant to highly significant in the African populations; it is rarely significant in the remaining populations. The 16-repeat allele gives highly significant positive disequilibrium values with the "B" site in all populations and with the "A" site in all non-African populations. Disequilibrium with the "D" site is only occasionally statistically significant with no pattern across the populations.

Statistical tests of overall disequilibrium for each sample are given in Table 5 – both the observed likelihood ratio statistic, with its degrees of freedom if interpreted as chi-square, and the values from the permutation test. For every population, both tests of overall linkage disequilibrium gave statistical significance with a probability value of less than 0.001. For the chi-square test, the significance level was often much less than 0.001. For the permutation test, all observed sets of haplotype frequencies show disequilibrium that is nominally significant at $P<0.001$ because, in all cases, none of the 1000 random permutations was more extreme than the observed. The D_{sd} values are often quite large, suggesting that the significance level is much less than 0.001 in those cases, but much larger numbers of permutations would be required to properly estimate lower values.

The empiric distributions of the likelihood ratio statistic, generated by the 1000 permutations of the four-site phenotype data for five representative populations (Ethiopians, Sekele San Bushmen, Danes, Japanese, Maya) are shown in Fig. 4. These distributions show that there can be major differences in how deviant the observed value is, as indicated by the D_{sd} values in Table 5, even though all are assigned the same significance level, $P<0.001$, because none of the 1000 random permutations was more extreme than the observed.

Figure 5 graphs the $D_{sd}/2N$ values for the 28 populations, arranged in the same order as in Fig. 3. This measure of the amount of overall disequilibrium shows a pattern of low disequilibrium in Africa, somewhat higher disequilibrium in Europe, and still higher disequilibrium in East Asia and the Americas. This pattern is the "inverse" of the pat-

tern of number of common (5% or more) haplotypes in each population (Table 3).

Discussion

By combining into the same study (1) multi-site haplotypes based on non-coding regions encompassing the coding sequences of the DRD2 locus and (2) many samples representing indigenous populations from around the world, we can begin to learn both about some of the genetic characteristics of the locus (mutation rate, recombination history, etc) and about the population histories (migration, etc.) of the ethnic groups. The tools for this include examining the pattern of haplotype frequencies, linkage disequilibrium of the various haplotypes in populations, and knowledge of ancestral haplotype. This data set on DRD2 also represents one of the small, but growing, number of data sets examining disequilibrium and haplotype frequencies in human populations on a global scale. There are similarities to some of the others, as well as differences.

Global haplotype frequency distributions

The 28 populations studied here are sufficient in number with diverse enough geographic origins to allow many general conclusions about the DRD2 locus, in most indigenous populations, in most regions of the world. As discussed in several of the papers that describe the sampling of the individual populations, efforts were made to sample only individuals whose ancestors, so far as could be determined, were all from the same area and ethnic group. Thus, while recognizing that there is no such thing as a "pure" human population, these samples should be representative of the different geographic regions. Additional support for the general representativeness of these populations comes from the observation that geographically close populations tend to have similar sets of alleles and haplotypes at similar frequencies. This pattern can be seen in Figs. 2 and 3 and is also seen in analyses of many of these same samples for other loci (Tishkoff et al. 1996, 1998; Kidd and Kidd 1996; Chang et al. 1996; Calafell et al. 1997).

In all the populations studied here, the DRD2 locus shows highly significant overall linkage disequilibrium for this set of markers spanning the coding region. This is seen using both the likelihood ratio chi-square and the permutation test. However, significance tests of overall disequilibrium define neither where in the haplotype the disequilibrium exists, nor the patterns of disequilibrium. Therefore, to develop an evolutionary perspective for DRD2 variation, both the pattern and significance of the overall disequilibrium must be considered.

One approach to examine the pattern is, simply, to inspect the distributions of the various haplotypes. Of the 48 possible haplotypes, 36 have been estimated to be present in at least some population(s) and one other is possibly present in one population (B1-D1-13-A1 in Sekele San). The maximum likelihood estimates of the frequencies (Table 1) were obtained using the EM algorithm; such estimates and gene-counting estimates based on molecular haplotyping are generally quite similar (Tishkoff et al. unpublished analyses), but it is possible that some rare haplotypes were misidentified by the EM algorithm.

All regions of the world share the same set of common haplotypes, with only four haplotypes (B1-D2-16-A1, B2-D1-15-A2, B2-D2-14-A2, and B2-D2-13-A2) accounting for at least half of the chromosomes in every population (Fig. 3). However, only one haplotype is found in every population studied – B2-D2-14-A2 – with a frequency range of 9.4% in the Adygei and up to 57.4% in the Ami. This haplotype occurs at frequencies greater than 25% in all sub-Saharan populations. In previous studies, we showed that the B2, D2, and A1 alleles are the ancestral hominid alleles that are shared with other hominoids (Iyengar et al. 1998). Therefore, given that the small STRP alleles occur in both *Pan* species and that the 13-repeat allele is found essentially only in Africa, it seems likely that the B2-D2-13-A1 and B2-D2-14-A1 haplotypes represent the most ancient human haplotypes. Both of these are rare outside of Africa, but are found in all sub-Saharan populations at frequencies of up to 15%.

The globally common haplotype, B2-D2-14-A2, differs from one of these ancestral haplotypes by the mutation creating the TaqI "A" restriction site, the A2 allele. Another of the globally common haplotypes, B2-D2-13-A2, differs from the globally ubiquitous haplotype just at the STRP site, but shows the least clear geographic pattern. In most African and European populations, it occurs at frequencies between 12% and 30%, though it is uncommon to absent in the San populations. Elsewhere, it occurs with no obvious frequency pattern. The differences in frequencies between otherwise similar populations may reflect mutation interacting with drift.

The other two common haplotypes show stronger region-specific patterns. In Europe (and Southwest Asia), there is one common haplotype (B2-D1-15-A2 at 30–48%) that is not as common in any other population in any other geographic region. This haplotype is uncommon or absent in the New World populations. This argues for some ancestry common to the European populations, but distinct from that of any other populations in this study, during which this haplotype drifted to a higher frequency than found elsewhere in the world. The B1-D2-16-A1 haplotype was seen in almost all populations, but is least common (even rare) in African and European populations. In the rest of the world, this haplotype is usually seen at frequencies greater than 30% and has become the most common haplotype in the New World. This pattern argues for the East Asian and Amerindian populations sharing some common ancestry, distinct from that of Europeans and Africans, during which drift preserved and/or increased the frequency of this haplotype at the expense of others.

As noted above, the population samples available for this study are a diverse sampling from the different regions of the world, though they are not uniformly representative of any geographic region. There are obvious similarities

Table 6 Haplotypes that are common in only one region of the world. "Common" is defined as occurring in at least one population at ≥5%

Region	Haplotype	Maximum frequency (%)	Number of Populations ≥5%
Africa:	B1-D2-12-A1	16.9	1
	B2-D1-13-A1	11.0	3
	B2-D1-13-A2	5.5	1
	B2-D2-14-A1	10.6	4
	B2-D2-13-A1	15.3	4
	B2-D2-12-A1	6.0	1
E. Asia:	B1-D2-16-A2	5.1	2
N. America:	B1-D2-17-A1	9.8	1
S. America:	B1-D2-14-A2	6.6	1
	B2-D1-15-A1	7.0	1

among many of the populations within each region. In sub-Saharan Africa, in addition to the single haplotype (B2-D2-14-A2) which occurs at a frequency greater than 25% in all of the populations, several others occur at frequencies of 5–20% in most populations. Most of the haplotypes seen at frequencies of greater than 5% in any population in the region are found in most of the populations. These data argue that all of these populations have had large effective population sizes for a long time and that drift has not resulted in distinctly different sets of haplotypes in the different populations. Since the study includes both Pygmies and San Bushmen, populations thought to be quite remotely related based on several lines of evidence, the occurrence of similar sets of haplotypes argues that both sets have had large effective population sizes since they separated because other data argue against significant gene flow between them.

The mitochondrial (mt) data from Vigilant et al. (1989) showed considerable difference between the locations of the San Bushmen and Pygmies in the mtDNA gene tree. Using classical genetic polymorphisms, Cavalli-Sforza et al. (1994) failed to show any close affinity between the Pygmies and San Bushmen in tree analyses, although principal component analysis showed "clustering" with the second principal component. mtDNA studies have shown that a 9-bp deletion polymorphism attains significant frequencies in the Pygmies and the Central and Southeast Africans, although it is not present in the Khoisan populations tested (Soodyall et al. 1996). Other nuclear haplotype data show the converse pattern: a high frequency of a haplotype in the San Bushmen that was not seen in the Pygmies (Tishkoff et al. 1998). All these data suggest that Pygmies and San Bushmen have had minimal contact in recent times.

As can be seen from Table 1, we observe very few haplotypes that reach a frequency of more than 5% in only one region of the world; these are summarized in Table 6. Most such cases, six of the ten, occur in Africa. Outside of Africa, chance fluctuations – random genetic drift or sampling error – is sufficient to explain these observations. Thus, most haplotypes are either common in multiple populations in several parts of the world or are rare to absent everywhere. This is clearly shown by the pattern in Fig. 3. Most of the rarest and "missing" haplotypes are those containing the smallest or largest (12 or 17 repeats) allele at the $(CA)_n$.

Disequilibrium

In all the populations sampled, the overall disequilibrium at this locus is highly significant. Among the SNPs, the TaqI "B" and the TaqI "A" sites always show significant pairwise disequilibrium, with a proportionate excess of B1-A1 and B2-A2 haplotypes. That excess is most clearly seen in the Asian and American populations, but it exists also in Africa. A previous study of African Americans also found the same pattern of disequilibrium for these two sites (O'Hara et al. 1993).

The "D" site does not show such a universal pattern of disequilibrium, partly because it has low heterozygosity in many populations and there is not sufficient power in those cases to detect weaker disequilibrium. However, where it is more heterozygous, in Europe especially, disequilibrium is strong and significant; the D1 allele is primarily found on B2-A2 haplotypes. By means of pairwise tests, the $(CA)_n$ STRP is shown to be in linkage disequilibrium with at least one of the other polymorphisms in all populations. The 14-repeat allele shows significant to highly significant disequilibrium, with both the "B" and "A" sites in all except the European populations. In most European populations, the D' value is at -1.0, indicating absence of the B1-14 haplotype, but the result is not significant. In contrast, the 14-repeat allele and the TaqI "D" site show highly significant disequilibrium in the European populations, except for the Finns. That disequilibrium is also strongly negative and significant to highly significant in the African populations; it is rarely significant in the remaining populations. The 16-repeat allele shows highly significant disequilibrium with the "B" site uniformly and, outside of Africa, with the "A" site. Clearly, this pattern of disequilibrium indicates both a low recombination rate across the 25 kb of this locus and a low mutation rate at the STRP.

The analyses of disequilibrium show a somewhat different pattern than we have found for other loci in that there is significant disequilibrium across the 25 kb interval in all populations that have variation at both sites. At other loci, studies of many of the same populations have found no disequilibrium between markers more than 10 kb apart in African populations, but highly significant disequilibrium between those markers in all non-African populations. At CD4 disequilibrium was greatly reduced to absent across 10 kb in sub-Saharan African populations, but was essentially complete in all non-African populations that had variation for the bi-allelic deletion polymorphism (Tishkoff et al. 1996). The PAH locus also shows that pattern of disequilibrium, extending across a shorter distance in African populations than in non-African populations (unpublished data). At the DM locus, the pattern of bi-allelic markers across 25 kb involves, essentially, complete disequilibrium with only two haplotypes in all non-African populations and similar, but less, disequilibrium in the sub-Saharan African

populations (Tishkoff et al. 1998). When we quantify overall disequilibrium using $D_{sd}/2N$ (Fig. 5), we see a pattern that is similar to the PAH and DM patterns. There are three populations, Taiwanese Chinese, Maya, and Karitiana, that appear to deviate from the general pattern. While the statistical significance of these "deviations" is questionable, two of them are expected from what is known of these population samples: Maya and Karitiana.

Haplotype evolution

We have shown previously (Castiglione et al. 1995; Kidd et al. 1996; Iyengar et al. 1998) that the B2, D2, and A1 alleles are the ancestral alleles, because they are the DNA sequences present in chimpanzees, gorillas, and orangutans. We have sequenced the regions around all of the TaqI sites in other great apes (Iyengar et al. 1998). Ignoring the sites that are polymorphic in humans, the differences among *Homo*, *Pan*, and *Gorilla* vary from 1.2% to 1.8% in 550–750 bp of homologous sequence, depending on the species compared. In 301 bp of DRD2 intron 3, Deinard and Kidd (1997) found no human or gorilla polymorphism, but extensive chimpanzee polymorphism. The nucleotide differences between species were 1–2% for *Homo–Pan*, 1–3% for *Homo–Gorilla*, and 2–3% for *Pan–Gorilla*. Given the relative amounts of time separating the species, these differences are consistent with other molecular evolution studies of these species (Koop et al. 1989). There is no evidence that the TaqI sites are hypermutable, and the time frame for neutral polymorphic variation to persist in human populations is almost certainly less than a million years (Takahata 1993). Considering the SNPs as defining background haplotypes (upon which STRP alleles are found) and considering estimates for mutation rates for single nucleotides on the order of 10^{-8} (Crow 1995; Li et al. 1996), we conclude that the B2-D2-A1 background haplotype (ignoring the STRP) is the ancestral hominid haplotype, and that recurrent mutation is not a factor influencing the allele frequencies for these polymorphisms. While we cannot exclude a very low number of alleles occurring due to recurrent mutation, the phenomenon is negligible at the level of our analysis.

If we consider each derived allele to have arisen once and then drifted to high frequency, how did the different haplotypes arise? We have published one likely possibility to explain the origins of the background haplotypes (Kidd et al. 1996), but our additional data on African populations now suggest a different scenario (Fig. 6). The ancestral background human DRD2 haplotype, B2-D2-A1, is common in sub-Saharan Africa (0.158 to 0.249), but rare to absent elsewhere. In Africa, three STRP allele sizes are frequently found on this background, the 12-, 13-, and 14-repeat alleles. The EM algorithm estimates two occurrences of the 15-repeat allele on this background in the Sekele San and the possibility of one occurrence of that STRP allele with this background in the Central San. No occurrences of a 16-repeat or 17-repeat allele occurred on this background haplotype in Africa. This would suggest that one of the three smaller alleles is the ancestral human allele for the STRP. This is supported by the finding that, in chimpanzees, the STRP allele sizes ranged from six repeats to twelve repeats based on the size of the PCR product (Castiglione et al. 1995). However, since more sequencing could reveal differences among the repeat alleles across primates, as we demonstrated elsewhere (Deinard and Kidd 1995), it cannot be considered as certain that the ancestral hominoid only had STRP alleles with low repeat numbers. Among the uncommon occurrences of this ancestral background haplotype outside of Africa, the 12-repeat STRP allele has not been seen, the 13-repeat STRP allele has been seen only three times (once each in the Yemenites, Yakut, and Ticuna), and the 14- and 16-repeat alleles have been seen most often.

The three background haplotypes that differ from the ancestral haplotype by one mutation are B1-D2-A1, B2-D2-A2, and B2-D1-A1. The B1-D2-A1 background is uncommon in Africa (and absent in the Mbuti and Ethiopian samples), but common in all other parts of the world. The 12-repeat and 13-repeat STRP alleles are both seen on this background, but essentially only in Africa; the 14-, 15-, and 17-repeat alleles are seen occasionally on this background, but the 16-repeat allele is seen on this background in every population, often exclusively. Outside of Africa, the frequency of this background haplotype ranges from 0.037 (Yemenites) to greater than 0.7 (Cheyenne). This haplotype has the greatest diversity of STRP alleles in sub-Saharan Africa with the highest average frequency of 12-repeat alleles and even a rare 17-repeat allele. Thus, the B2 to B1 mutation appears to be quite old.

The B2-D2-A2 background is common in all parts of the world, with frequencies ranging from a low of 0.166 (in Cheyenne) to a high of 0.666 (Mbuti). Both the 13-repeat and 14-repeat STRP alleles occur commonly on this background; the 12-repeat allele has been seen on this background only in the Sekele San; and 15-repeat and 16-repeat alleles are seen sporadically in all parts of the world. This haplotype and the A1 to A2 mutation that created it also appear to be quite old.

The B2-D1-A1 haplotype, in contrast with the other two, is uncommon in sub-Saharan Africa (0.02 to 0.14), rare in Europe and southwest Asia (0.012 to 0.042), and virtually absent everywhere else. In sub-Saharan Africa, the predominant STRP allele on this background is 13 repeats, whereas in Europe it is 15 repeats, suggesting a different ancestry for this haplotype in Europe than in Africa. The rarity of this background haplotype and the lower diversity of STRP alleles suggest that its origin is more recent than the other two singly derived background haplotypes.

The three doubly derived background haplotypes are B1-D1-A1, B1-D2-A2, and B2-D1-A2. The first two of these and the triply derived haplotype, B1-D1-A2, are absent or only sporadically present at low frequencies in all regions of the world. In contrast, B2-D1-A2, is present at modest frequencies (8% to 15%) in all African samples, except the Mbuti, and becomes quite common in Europe. The 15-repeat STRP allele seems to be the common one on this background in African and non-African populations. This

pattern suggests that the D2 to D1 mutation occurred on the B2-D2-A2 background, rather than on the ancestral background, on a chromosome that also had the 15-repeat allele. According to this scenario, the B1-D1-A1 haplotype would have arisen from at least two successive crossovers, starting with the B2-D1-A2 haplotype and others with the B1 and A1 alleles. These observations also suggest that larger sized STRP alleles have arisen more recently and tend to occur as part of haplotypes with one or more derived alleles at the three TaqI sites.

STRP mutation

It is clear from Fig. 3 that haplotype frequencies tend to be similar within broad geographic regions (continents to hemispheres). Therefore, as one way to view the STRP distributions, we have pooled all samples in each region and plotted the four resulting histograms and the global distribution (not shown, but largely retrievable from Table 1). Given the diversity among continents, the global distribution is dependent on the numbers of samples in each area. Within an area, however, the individual population distributions were sufficiently similar that the pooled distributions are reasonably accurate continental estimates. It is clear that Africa has the highest proportion of smaller alleles, Europe has significantly more larger alleles than Africa, and East Asia and the Americas have bimodal distributions. These bimodal distributions reflect the preponderance of only two haplotypes in these populations, one with a 16-repeat STRP allele and the other with a 14-repeat STRP allele (Fig. 3). Thus, the distributions of STRP alleles reflect random genetic drift of whole haplotypes.

The nearly complete association of the 15-repeat STRP allele with the B2-D1-A2 background haplotype, which is very common only in Europe and Southwest Asia, and the strong to nearly complete association of the 16-repeat STRP allele with the B1-D2-A1 haplotype in most non-African populations both argue for a relatively low mutation rate of this STRP over the span of time since humans have been out of Africa. The strong association of the 14-repeat allele with the B2-D2-A2 haplotype in all regions of the world, but especially in the non-African populations, also supports this conclusion.

Implications for testing the significance of disequilibrium

We conclude that the overall disequilibrium is statistically significant in every population, whether evaluated by a likelihood ratio chi-square test or a permutation test. Our objective with the permutation test is not to estimate the significance level accurately when it is less than 0.001, as it clearly is for all populations in this study. Greater numeric accuracy would require that severalfold more random permutations of the data be evaluated and would not alter the conclusion that disequilibrium does exist in all populations. Even applying a correction for multiple independent tests, and these are not independent because of shared ancestry, the least significant value is still less than 0.028.

Table 5 gives the details of all the permutation test distributions and Fig. 4 shows the actual distributions for a representative sample of five populations with different heterozygosities and different apparent patterns of disequilibrium. As noted before, none of the permutations showed more disequilibrium than the observed. Thus, at this first level of comparison, the permutation test gives significance levels comparable to those using a likelihood ratio chi-square test, i.e., $P<0.001$. However, the validity of the likelihood ratio chi-square test rests on the assumption that the underlying distribution of the likelihood ratio statistic follows a chi-square distribution. The mean of that distribution should approximate the degrees of freedom and the variance should be twice the mean. Only three of the 28 permutation distributions even came close: the Ethiopians, Jemez Pueblo, and Karitiana. Interestingly, these are the populations with the three lowest means of the permutation test, and all have at least one site that has gone to fixation. These are also the populations with the lowest estimated degrees of freedom (see following discussion). In general (Table 5), the variance of the distribution of likelihood ratio statistics from 1000 permutations was only somewhat larger than the mean and smaller than the variance expected for a chi-square distribution. Thus, if this observation is generalizable, the implication is that the likelihood ratio chi-square test, which assumes an asymptotic chi-square distribution, is in general not valid for multi-site haplotype systems but, if used, gives a conservative result.

A likelihood ratio chi-square test requires determination of the degrees of freedom, but it is not clear how to estimate the degrees of freedom for a complex haplotype system correctly. The set of multi-site phenotypes can have many potential observations that do not occur, and many haplotypes that could potentially occur are estimated to be absent. HAPLO estimates the degrees of freedom as the number of possible haplotypes in the specific population minus the number of independent parameters estimated minus one. The number of possible haplotypes is the product across all polymorphic sites of the numbers of alleles with non-zero frequency estimates at each site. HAPLO prints a clear warning that this estimate may not be valid for any particular data set. For example, if the product of an allele frequency and sample size is such that the allele would occur fewer times than there are haplotypes defined by the other sites, not all possible haplotypes could appear in a sample. Since HAPLO allows for missing data at a site, the number of observed occurrences of an allele is not necessarily the criterion when there are missing data.

Our experience with simpler systems (Castiglione et al. 1995; Tishkoff et al. 1996) has suggested that the degrees of freedom were accurately estimated and that the likelihood ratio chi-square test was a valid statistic for significance of overall disequilibrium. These results support that experience in those cases where fixation at some sites resulted in much simpler haplotype systems. However, for the generally much more complicated system in this study, it is clear that (1) the degrees of freedom are not, in general, estimat-

ed accurately and (2) the underlying distribution does not follow a chi-square distribution. We note that a simple comparison of the means from the permutation analyses often differ considerably from the degrees of freedom estimated as described above (Table 5).

Population evolution

One obvious observation from these data is that the non-African populations tend to have fewer haplotypes than African populations. In general, only three haplotypes account for most of the non-African chromosomes. From Africa to Europe and then across to East Asia, we see a pattern of the average heterozygosity in the regions decreasing and the numbers of haplotypes and common (>5%) haplotypes also decreasing (Table 3). The values for the Americas are roughly comparable with those for East Asia, but there is more variation in these values (except the number of common haplotypes). The overall disequilibrium (Fig. 5) increases as the number of common haplotypes decreases (Table 3). These observations are in general agreement with the observations we have made for CD4 and DM (Tishkoff et al. 1996, 1998) of decreasing genetic variation, in a pattern starting with highest variation in Africa, then decreasing from west to east in Eurasia, then decreasing separately into the Pacific and into the Americas, with still less variation in South America. Though the decrease in variation coming out of Africa is less dramatic for these markers at DRD2 than for the markers studied at CD4 and DM, the general pattern is the same and supports the Out of Africa model of recent human origins (see Takahata (1993) and Tishkoff et al. (1996) for discussions of the different models).

We found no evidence for distinct old haplotypes in East Asia, which would have been expected if there was substantial genetic continuity from *Homo erectus* and/or archaic *Homo sapiens* and modern populations. Additional independent loci will need to be studied, of course, to be more certain of which patterns are general and which differ. The β-globin locus, for instance, has certainly been much studied for extended haplotypes, initially because of mutant alleles involved in thalassemias and the opportunity to assess selective forces in various human populations. A recent study by Harding et al. (1997) of a 3-kb region of β-globin in nine population samples differs in methodology, populations sampled, and sample sizes compared with our studies of DRD2, but the similarities of the results are as interesting as the differences. Their sub-Saharan African samples display the largest number of different haplotypes compared with the other regions sampled, but a simple decreasing pattern of variation as one moves from West to East is not found. In their single samples from Europe and the Americas, there is indeed decreased variation, while in their East-Asian and Pacific-Island samples, there is relatively high variation, although not as great as in Africa.

Other evidence exists for additional DRD2 variation in African populations. O'Hara et al. (1993) studied the TaqI "A" RFLP system and described two additional uncommon alleles found in African Americans. These apparently arose from variation at the next TaqI site on either side of the "A" site and were detected by Southern blotting. These variants are presumably of African origin, but could not be detected by the PCR typing system used for this study. DNA sequence across the upstream TaqI site is now available (GenBank #AF050737) and could be used to confirm the nature of those variant RFLP alleles and to evaluate their origins, but that was not done in this study. The great rarity of these alleles outside of Africa is documented by their absence in an earlier large survey of the TaqI "A" site in primarily non-Africans using Southern blotting (Barr and Kidd 1993).

The large number of haplotypes in each of the sub-Saharan African populations and the similar sets of common haplotypes in those populations suggest both that much of the variation observed today arose some time ago and was present in the ancestral African population from which these modern populations descended and that all of these populations have had large effective population sizes, allowing them to maintain all the different haplotypes. The differences among these populations are most likely attributable to recent effects of random genetic drift, but the smallness of the differences also argues that effective population sizes have been large for the entire period since divergence from the common ancestor. In contrast, in the non-African populations, there are fewer common haplotypes, the frequencies of those haplotypes vary across the different regions, and, though their frequencies differ, the same few haplotypes account for the majority of chromosomes in all of the non-African populations. This small subset of the haplotypes in sub-Saharan African populations is consistent with the hypothesis of a significant founder effect associated with a single migration of modern *Homo sapiens* out of Africa and additional loss of variation as that initial non-African founder population grew and expanded to the East and later into the Americas. Both the recombination rate among the sites and the mutation rate at the STRP appear to be sufficiently low that significant numbers of new haplotypes have not been generated since that founder event.

Acknowledgements We want to acknowledge and thank all of the individuals who volunteered to provide blood samples for this and similar studies of normal genetic variation. We also want to thank the several colleagues who helped collect the samples. These studies were primarily supported by USPHS grants AA09379 and MH39239 to KKK, and by NIMH grant MH30929 in support of the Mental Health Clinical Research Center at Yale and by NSF grant SBR9632509 to JRK. Additional support was provided by the South Africa Medical Research Council, the University of the Witwatersrand, and the South African Institute for Medical Research to TJ; by the National Science Council of Taiwan (NSC86–2314-B016–116Y) and the Department of Health of Taiwan (DOH86-IIR-612) to RBL; and by a grant from the Hallym Academy of Sciences, Hallym University to CL. Collection of some of the samples was also supported by a grant from the Alfred P. Sloan Foundation to KKK.

References

Araki K, Kwano R, Morii K, Hayashi S, Minoshima S, Shimizu N, Katagiri T, Usui H, Kumanishi T, Takahashi Y (1992) Structure and expression of human and rat D2 dopamine receptor genes. Neurochem Int 21:91–98

Arinami T, Gao M, Hamaguchi H, Toru M (1997) A functional polymorphism in the promoter region of the dopamine D2 receptor gene is associated with schizophrenia. Hum Mol Gen 6:577–582

Armour JA, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, Bertranpetit J, Paabo S, Jeffreys AJ (1996) Minisatellite diversity supports a recent African origin for modern humans. Nat Genet 13:154–160

Barr C, Kidd KK (1993) Population frequencies of the A1 allele at the dopamine D2 receptor locus. Biol Psychiatry 34:204–209

Bolos AM, Dean M, Lucas-Derse S, Ramsburg M, Grown GL, Goldman D (1990) Population and pedigree studies reveal a lack of association between the dopamine D_2 receptor gene and alcoholism. JAMA 264:3156–3160

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Calafell F, Shuster A, Speed WC, Kidd JR, Kidd KK (1997) Short tandem repeat polymorphism evolution in humans. Eur J Hum Genet 6:38–49

Castiglione CM, Deinard AS, Speed WC, Sirugo G, Rosenbaum HC, Zhang Y, Grandy DK, Girgorenko EL, Bonne-Tamir B, Pakstis AJ, Kidd JR, Kidd KK (1995) Evolution of haplotypes at the DRD2 locus. Am J Hum Genet 57:1445–1456

Cavalli-Sforza L, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, New Jersey

Chang F-M, Kidd JR, Livak KJ, Pakstis AJ, Kidd KK (1996) The world-wide distribution of allele frequencies at the human dopamine D4 receptor locus. Hum Genet 98:91–101

Cravchik A, Sibley DR, Gejman PV (1996) Functional analysis of the human D_2 dopamine receptor missense variants. J Biol Chem 271:26013–26017

Crow JF (1995) Spontaneous mutation as a risk factor. Exp Clin Immunogenet 12:121–128

DalToso R, Sommer B, Ewert M, Herb A, Pritchert DB, Bach A, Shivers BD, Seeburgh PH (1989) The dopamine D2 receptor: two molecular forms generated by alternative splicing. EMBO J 8:4025–4034

Deinard AS, Kidd KK (1995) Levels of DNA polymorphism in extant and extinct hominoids. In: S. Brenner, K. Hanihara (eds) The origin and past of modern humans as viewed from DNA. World Scientific, New Jersey, pp 149–170

Deinard AS, Kidd KK (1998) Evolution of a D2 dopamine receptor intron within the great apes and humans. DNA Sequence (in press)

Deka R, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995) Population genetics of dinucleotide (dC-dA)n (dG-dT)n polymorphisms in world populations. Am J Hum Genet 56:461–474

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc B 39:1–38

Gejman P, Ram A, Gelernter J, Friedman E, Cao Q, Pickar D, Blum K, Noble EP, Kranzler HR, O'Malley S, Hamer DH, Whitsitt F, Rao P, DeLisi LE, Virkkunen M, Linnoila M, Goldman D, Gershon ES (1994) No structural mutation in the dopamine D_2 receptor gene in alcoholism or schizophrenia. JAMA 271:204–208

Gelernter J, Grandy DK, Bunzow J, Pakstis AJ, Civelli O, Retief AE, Litt M, Kidd KK (1989) D2 dopamine receptor locus (probe hD2G1) maps close to D11S29 (probe L7) and is also linked to PBGD (probe PBGD0.0) and D11S84 (probe p-2–7-ID6) on 11q (A2669) Cytogenet Cell Genet 51:1002

Gelernter J, Pakstis AJ, Pauls DL, Kurlan R, Gancher ST, Civelli O, Grandy D, Kidd KK (1990) Tourette syndrome is not linked to D_2 receptor. Arch Gen Psychiatry 47:1073–1077

Gelernter J, Pauls DL, Leckman J, Kidd KK, Kurlan R (1994) D2 dopamine receptor alleles do not influence severity of Tourette's syndrome. Arch Neurol 51:397–400

Goldman D, Urbanek M, Guenther D, Robin R, Long JC (1997) Linkage and association of a functional DRD2 variant [Ser311Cys] and DRD2 markers to alcoholism, substance abuse and schizophrenia in Southwestern American Indians. Am J Med Genet Neuropsychiatr Genet 74:386–394

Good P (1995) Permutation tests. Springer-Verlag, Berlin Heidelberg New York

Grandy DK, Marchionni MA, Makam H, Stofko RE, Alfano M, Frothingham L, Fischer JB, et al. (1989a) Cloning of the cDNA and gene for a human D_2 dopamine receptor. Proc Natl Acad Sci U S A 86:9762–9766

Grandy DK, Litt M, Allen L, Bunzow JR, Marchionni M, Makam H, Reed L, Magenis RE, Civelli O (1989b) The human dopamine D_2 receptor gene is located on chromosome 11 at q 22-q23 and identifies a TaqI RFLP. Am J Hum Genet 45:778–785

Harding RM, Fullerton SM, Griffths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. Am J Hum Genet 60:772–789

Hauge X, Grandy D, Eubanks J, Evans G, Civelli O, Litt M (1991) Detection and characterization of additional DNA polymorphisms in the dopamine D_2 receptor gene. Genomics 10:527–530

Hawley ME, Kidd KK (1995) HAPLO: a program using the EM algorithm to estimate frequencies of multi-site haplotypes. J Hered 86:409–411

Itakowa M, Arinami T, Futamura N, Hamaguchi H, Toru M (1993) A structural polymorphism of human dopamine D2 receptor, D2(Ser^{311}→Cys) Biochem Biophys Res Commun 196: 1369–1375

Iyengar S, Seaman M, Deinard AS, Rosenbaum HC, Sirugo G, Castiglione CM, Kidd JR, Kidd KK (1998) Analyses of cross-species polymerase chain reaction products to infer the ancestral state of human polymorphisms. DNA Sequence (in press)

Kent S, Dunn D (1993) Etiology of hypoferremia in a recently sedentary Kalahari village. Am J Trop Med Hyg 48:554–567

Kidd KK, Kidd JR (1996) A nuclear perspective on human evolution. In: AJ Boyce, CGN Mascie-Taylor (eds.) Molecular biology and human diversity. Cambridge University Press, Cambridge, pp 242–264

Kidd KK, Pakstis AJ, Castiglione CM, Kidd JR, Speed WC, Goldman D, Knowler WC, Lu R-B, Bonne-Tamir B (1996) DRD2 haplotypes containing the TaqI A1 allele: implications for alcoholism research. Alcohol Clin Exp Res 20:697–705

Koop BF, Tagle DA, Goodman M, Slightom JL (1989) A molecular view of primate phylogeny and important systematic and evolutionary questions. Mol Biol Evol 6:580–612

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations: heterotic models. Genetics 49:49–67

Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D (1996) Rates of nucleotide substitution in primates and rodents and the generation–time effect hypothesis. Mol Phylogenet Evol 5:182–187

Litt M, Kramer P, Kort E, Fain P, Cox S, Root D, White R, Weissenbach J, Donis-Keller H, Gatti R, Wever J, Nakamura Y, Julier C, Hayashi K, Spurr N, Dean M, Mandel J, Kidd K, Kruse T, Retief A, Bale A, Meo T, Vergnaud G, Warren S, Willard HF (1995) The CEPH consortium linkage map of human chromosome 11. Genomics 27:101–112

Moises HW, Gelernter J, Giuffra LA, Zarconi V, Civelli O, Wetterberg L, Kidd KK, Cavalli-Sforza LL (1991) No linkage between D2-dopamine receptor gene region and schizophrenia. Arch Gen Psychiatry 48:643–647

Niswanger K, Hill SY, Kaplan BB (1995) Association and linkage studies of the TaqI A1 allele at the dopamine D2 receptor gene in samples of female and male alcoholics. Am J Med Genet Neuropsychiatr Genet 60:267–271

O'Hara BF, Smith SS, Bird G, Persico AM, Suarez BK, Cutting GR, Uhl GR (1993) Dopamine D2 receptor RFLPs, haplotypes and their association with substance use in black and caucasian research volunteers. Hum Hered 43:209–218

Parsian A, Fisher L, O'Malley K, Todd R (1991) A new TaqI RFLP within intron 2 of human dopamine D2 receptor gene. Nucleic Acids Res 19:6977

Picetti R, Saiardi A, Samad TA, Bozzi Y, Baik J-H, Borrelli E (1997) Dopamine D2 receptors in signal transduction and behavior. Clin Rev Neurobiol 11:121–142

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor, NY

Sarkar G, Kapelner S, Grandy DK, Marchionni M, Civelli O, Sobell J, Heston L, Sommer SS (1991) Direct sequencing of the dopamine D2 receptor (DRD2) in schizophrenics reveals three polymorphisms, but no structural change in the receptor. Genomics 11:8–14

Seeman P, O'Hara K, Ulpian C, Seeman MV, Jellinger K, Van Tol HH, Niznik HB (1993) Schizophrenia: normal sequence in the dopamine D2 receptor region that couples to G-proteins. Neuropsychopharmacology 8:137–142

Seeman P, Van Tol HHM (1994) Dopamine receptor pharmacology. Trends Pharmacol Sci 15:264–270

Slatkin M, Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity 76:377–383

Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-Specific" 9-bp deletion in sub-Saharan Africans. Am J Hum Genet 58:595–608

Spurdle A, Jenkins T (1992) Y chromosome probe p49a detects complex PvuII haplotypes and many new TaqI haplotypes in southern African populations. Am J Hum Genet 50:107–125

Su Y, Burke J, O'Neill FA, Murphy B, Nie L, Kipps B, Bray J, Shinkwin R, Nuallain MN, MacLean CJ, Walsh D, Diehl SR, Kendler KS (1993) Exclusion of linkage between schizophrenia and the D2 dopamine receptor gene region of chromosome 11q in 112 Irish multiplex families. Arch Gen Psychiatry 50:205–211

Suarez BK, Parsian A, Hampe CL,Todd RD, Reich T, Cloninger CR (1994) Linkage disequilibria at the D_2 receptor locus (DRD2) in alcoholics and controls. Genomics 19:12–20

Takahata N (1993) Allelic genealogy and human evolution. Mol Biol Evol 10:2–22

Templeton AR, Sing C, Kessling A, Humphries S (1988) A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. Genetics 120:1145–1154

Templeton AR, Sing C (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. Genetics 134:659–669

Templeton AR (1996) Cladistic approaches to identifying determinants of variability in multifactorial phenotypes and the evolutionary significance of variation in the human genome. 1996 Variation in the human genome. Ciba Foundation Symposium 197, Wiley, Chichester, pp 259–283

Tishkoff SA, Dietzsch E, Speed WC, Pakstis AJ, Cheung K, Kidd JR, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Watson E, Krings M, Pääbo S, Risch N, Jenkins T, Kidd KK (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380–1387

Tishkoff SA, Goldman A, Calafell F, Speed WC, Deinard AS, Bonne-Tamir B, Kidd JR, Pakstis AJ, Jenkins T, and Kidd KK (1998) A global haplotype analysis of the myotonic dystrophy locus: implications for the evolution of modern humans and for the origin of myotonic dystrophy mutations. Am J Hum Genet 62:1389–1402

Vigilant L, Pennington R, Harpending H, Kocher TD, Wilson AC (1989) Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci U S A 86:9350–9354

Weir BS (1996) Genetic Data Analysis II. Sinauer Associates, Inc., Sunderland, Mass

Zaykin D, Zhivotovsky L, Weir BS (1995) Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169–178

Zhao H, Pakstis AJ, Kidd KK, Kidd JR (1997) Overall and segmental significance levels of linkage disequilibrium. Am J Hum Genet 61[Suppl]:A17