A survey of haplotype frequencies and linkage disequilibrium at the DRD2 locus in the Nilgiri hill tribes, South India

H. Vishwanathan[†], Deepa Edwin[†], M. V. Usha Rani[†,*] and P. P. Majumder[#]

[†]Division of Genetics, Department of Environmental Sciences, Bharathiar University, Coimbatore 641 046, India
[#]Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata 700 035, India

DNA analysis has made it easier to study haplotypes, arrays of alleles at closely linked loci along the chromosome. These regions are short enough to show little or no recombination, and behave as blocks that might have ancient origins. Scoring these markers as haplotypes, allows analysis both in terms of haplotype frequencies and identity in terms of linkage disequilibrium. The human dopaminergic system is an important focus of study in the fields of neuropsychiatry and pharmacology; it is also a promising nuclear DNA marker in studies of human genome diversity. Haplotype frequencies and linkage disequilibrium for the dopamine D2 receptor gene (*DRD2*) was determined in 250 unrelated individuals from five tribal populations. The three marker systems in this study are highly polymorphic in all the five tribal populations and the haplotype system showed high level of heterozygosities. Out of the possible eight haplotypes, four are commonly shared by all the populations. The ancestral allele B2D2A1 accounts for 0.021 to 0.080, which was present in all the groups consistently. The linkage disequilibrium was statistically significant in all the populations. Data obtained in this study on *DRD2* represent one of the small, but growing number of data sets examining disequilibrium and haplotype frequencies in human populations.

THE human dopaminergic system is an important focus of study in the fields of neuropsychiatry and pharmacology, and therefore genes involved in dopaminergic transmission and metabolism have been extensively studied to identify genotype–phenotype relationships in neuropsychiatric disorders[1–3]. Among the five known human dopamine receptors, the *DRD2* receptor gene has been studied most extensively because it is the site of action of neuroleptic drugs and it is also believed to be involved in the pathophysiology of various neuropsychiatric diseases. Several polymorphisms in the *DRD2* gene have been identified in the DNA encompassing the coding sequences; most are in the introns or downstream-flanking DNA[4–7], but some are in coding regions[8,9] and in the promoter region upstream of exon 1 (ref. 10). Many of

*For correspondence. (e-mail: malla_drur@yahoo.com)

these polymorphisms have been used to map the locus genetically to the long arm of chromosome 11 (refs 4, 11).

Most of the studies of DRD2 and different disorders, such as alcoholism and addictive behaviours[12,13] 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. Indeed, positive associations have been reported between the presence of several DRD2 alleles and diseases[1–3,13,14]. Almost all the studies have used only the *Taq*1 'A' site, a single nucleotide polymorphism (SNP) in a *Taq*1 restriction site[4,15]. The use of haplotypes of multiple genetic markers distributed through and around the gene is thus 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. Scoring these markers as haplotypes allows analysis, both in terms of haplotype frequencies and identity and also in terms of linkage disequilibrium.

By and large, diversity of the DRD2 system has recently been investigated on global samples[16]. The global pattern of DRD2 haplotype variation reinforces the growing consensus from nuclear DNA studies that African populations have significantly more genetic variation than non-African populations[17,19]. Unfortunately, the Indian population known for its rich diversity was not included in the global survey. Considering that surviving Nilgiri hill tribes exhibit remarkable cultural, linguistic and biological diversity, the aim of this investigation is to disclose the full diversity of the DRD2 locus and to survey the haplotype frequencies and linkage disequilibrium for about 250 tribal individuals belonging to five aboriginal populations dwelling in the undulating sholas of the Nilgiri hills of southern India.

The Nilgiri tribal samples consisted of 250 individuals from badaga (51), irula (50), kota (45), kurumba (54) and toda (50) tribes. Additional linguistic, historical, demographic and genetic information about these populations has been reported previously[20–24].

DNA from blood samples was extracted by the salting-out procedure[25]. The DRD2 markers typed for this study are three bi-alleic *Taq*I restriction fragment length polymorphisms (RFLPs) spanning a distance of 25.4 kb (Figure 1). All the markers have been described previously: *Taq*I 'A' by Grandy *et al.*[26], *Taq*I 'B' by Hauge *et al.*[5]; *Taq*I 'D' by Parsian *et al.*[6]. All typings are PCR-based, using primers and amplification protocols, as described by Castiglione *et al.*[7] and Kidd *et al.*[15]. After amplification, the three *Taq*I RFLPs were digested with *Taq*I restriction enzymes as per the manufacturer's recommended conditions. Subsequently, the digested fragments were subjected to electrophoresis in 2% agarose gels and stained with ethidium bromide for visualization.

Allele frequencies at the individual sites were estimated by gene counting. Heterozygosity was obtained according to the method of Nei[27]. The assumption of Hardy–Weinberg equilibrium (HWE) was tested using χ^2 goodness-of-fit test. The maximum likelihood estimates of haplotype frequencies were calculated from the multi-site marker typing data, using the program HAPLOFREQ[28], which implements the EM algorithm. HAPLOFREQ accommodates individuals with either missing data at some sites or partial phase information, by giving the unique phenotype corresponding to the set of underlying genotypes compatible with the information available, as explained by Hawley and Kidd[29]. The standardized, pairwise linkage disequilibrium value D'[30] was calculated for each pair of markers by means of the computer program LINKD, which uses the sample sizes and the haplotype frequency estimates from HAPLOFREQ as input.

The individual-site allele frequency estimated at the DRD2 locus by gene counting is represented graphically in Figure 2. All the three marker systems of DRD2 locus are polymorphic in all the studied tribal populations. The *Taq*1 'B' allele has a frequency greater than 60% in all the populations and with a maximum of 91% in todas. At the *Taq*1 'A' site, the allele frequency ranges from 49 to 80%, with maximum frequency in todas. The *Taq*1 'D' site shows a similar pattern to that of *Taq*1 'B', except for the kurumbas and todas where it is the minimum (43

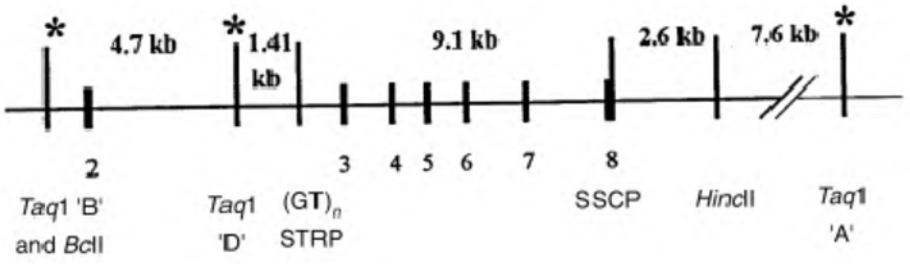

Figure 1. Molecular locations of the DRD2 polymorphisms. The three polymorphisms studied * are shown relative to each other, to the exons of DRD2, and to several other known polymorphims[31]. *Taq*1 'B' site is 913 bp upstream of the initiation codon in exon 2, and *Taq*1 'A' site is 10,542 bp downstream of the termination codon in exon 8.

to 50%). The overall patterns of allele frequency variation at the *Taq*1 'B' site are comparatively higher than *Taq*1 'A' and *Taq*1 'D'. Each of the three DRD2 sites was tested in each sample for agreement with the underlying assumption of Hardy–Weinberg ratios and no systematic departures were found, except for *Taq*I 'D' in the kota sample. This is approximately the expected number, because these appeared to be randomly distributed across sites and samples. Table 1 shows the estimates of heterozygosities for each of the three component polymorphisms. Heterozygosity for the haplotyped multi-site system showed high levels of diversity.

Haplotype frequencies at the DRD2 locus identified in the Nilgiri hill tribes are listed in Table 2 separately for each population. It is seen that all the study populations shared four haplotypes B2D2A2, B2D1A2, B2D2A1 and B1D2A1. The haplotype B2D1A1 was absent in badagas and irulas, while the haplotypes B1D2A2 and B1D1A2 were present only in irulas and badagas, kurumbas respectively. The kurumbas were the only population who shared seven out of the eight possible haplotypes in the three-site DRD2 locus with most of them, but not all occurring at a frequency of ≥ 5%. The overall data set is graphically represented in Figure 3. Table 3 presents the three pairwise standardized D' values for the three biallelic markers, *Taq*1 'B', *Taq*1 'D' and *Taq*1 'A'. All the comparisons were significant with the exception to statistical significance in the todas with a non-significant $D' = 0.008$ for 'A' to 'D' and the badagas with a non-significant $D' = -0.025$ for 'D' to 'B' comparison.

Evolutionary relationships of DRD2 haplotypes are defined by three *Taq*1 restriction site polymorphisms. The ancestral alleles (B2, D2, and A1) are all identical in sequence to those in chimpanzees, gorillas, and orangutans[31]. Frequency data in different human populations, especially those in Africans and also from the present

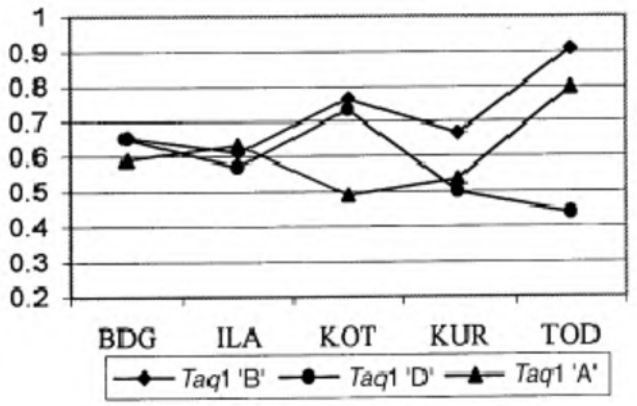

Figure 2. Allele frequency at individual sites of DRD2 for the five populations.

Table 1. Heterozygosity at individual sites of DRD2 loci

Population	2*n*	Heterozygosity		
		'B'	'D'	'A'
Badaga	102	0.449	0.455	0.484
Irula	100	0.475	0.490	0.466
Kota	90	0.354	0.394	0.500
Kurumba	108	0.442	0.500	0.497
Toda	100	0.165	0.493	0.319

2*n*, Number of chromosomes.

Table 2. Haplotype frequency at DRD2 locus in tribal populations

Haplotype	Badaga	Irula	Kota	Kurumba	Toda
B2D2A2	0.313	0.162	0.441	0.145	0.300
B2D1A2	0.273	0.429	0.048	0.381	0.495
B1D2A2	0.000	0.042	0.000	0.000	0.000
B1D1A2	0.012	0.000	0.000	0.011	0.000
B2D2A1	0.078	0.021	0.059	0.080	0.033
B2D1A1	0.000	0.000	0.219	0.060	0.082
B1D2A1	0.262	0.346	0.233	0.275	0.090
B1D1A1	0.063	0.000	0.000	0.047	0.000
n	46	49	45	54	39

n, Number of individuals.
Haplotypes are listed as B1, D1 and A1 alleles for the site-absent state for the RFLP sites, while B2, D2 and A2 for the site-present alleles.

Table 3. Pairwise linkage disequilibrium values of DRD2 locus

Population	*Taq*1 'A' and *Taq*1 'B'	*Taq*1 'A' and *Taq*1 'D'	*Taq*1 'D' and *Taq*1 'B'
Badaga	0.189*	–0.089*	–0.025
Irula	0.204*	–0.157*	–0.166*
Kota	0.109*	0.078*	–0.062*
Kurumba	0.168*	–0.119*	–0.103*
Toda	0.072*	0.008	–0.052*

Standardized linkage disequlibrium coefficient (D') of Lewontin[30] is shown. Significant levels to test whether the disequilibrium value differs from zero (one degree of freedom) are indicated by asterisks. No asterisk indicates that the test was not significant ($p > 0.05$).

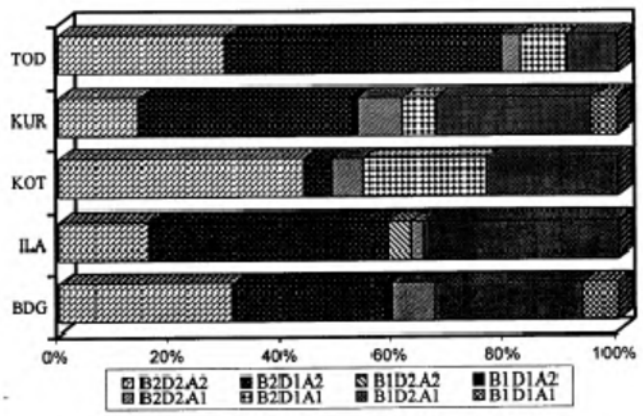

Figure 3. Relative frequency of eight possible DRD2 haplotypes shown as 'stacked' bars for five populations.

study, suggest the illustrated events by which three mutational events result 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 present data are insufficient to exclude other possible evolutionary schemes (Figure 4).

By combining into the same study (1) the multi-site haplotypes based on non-coding regions encompassing the coding sequences of the DRD2 locus, and (2) the samples representing indigenous populations, we can learn about some of the genetic characteristics of the locus (mutation rate, recombination history, etc.) and also about the population histories (migration, etc.) of the ethnic groups. The tools for this include examining the pattern of haplotype frequencies, linkage disequilibrium and the knowledge of ancestral haplotypes. The present data set on DRD2 also represents one of the small, but growing number of data sets examining disequilibrium and haplotype frequencies in human populations.

The DRD2 locus has proved useful in studying genetic structure of human populations[16]. As has been reported in the global survey of DRD2 locus[16], the tribal populations from the present study also show high levels of polymorphism for all the individual site markers. In the present study, the estimated levels of average heterozygosities are consistently high in all the populations. The heterozygosity levels are substantially higher than the average heterozygosity levels of other Indian populations as reported by Majumder *et al.*[32] and Mukherjee *et al.*[33] using nuclear DNA markers. Interestingly, the average heterozygosity levels were also higher than the other global populations studied, with exception to African populations[34,35]. Thus, the DNA markers attest that the study groups exhibit high levels of genomic diversity.

Earlier studies of haplotype evolution of DRD2[7,15,16] have shown that the B2, D2 and A1 alleles are the ancestral alleles, because they are the DNA sequences present in chimpanzees, gorillas and orangutans. Iyengar *et al.*[31] have sequenced the regions around all of the *Taq*1 sites in other great apes. 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

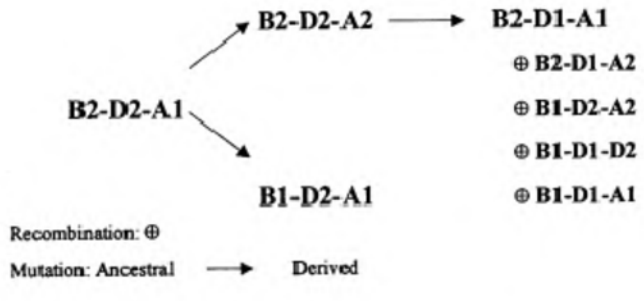

Figure 4. Haplotype evolution (Kidd *et al.*[16]).

301 bp of DRD2 intron 3, Deinard and Kidd[36] found no human or gorilla polymorphism, but reported extensive chimpanzee polymorphisms. 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[37]. There is evidence that *Taq*1 sites are hypermutable, and the time frame for neutral polymorphic variation to persist in human populations is almost certainly less than a million years[38]. Thus the B2D2A1 background haplotype may be concluded as the ancestral hominid haplotype, and the recurrent mutation is not a factor influencing the alelle frequencies for these polymorphisms. While a very low number of alleles occurring due to recurrent mutation cannot be excluded, the phenomenon is negligible at the level of present analysis.

In all the population samples, the disequilibrium at the DRD2 locus is highly significant (Table 3). Among the SNPs, the *Taq*1 'B' and *Taq*1 'A' sites showed significant pairwise disequilibrium which is most clearly seen in the Asian American populations, but it also exists in Africans[16]. However, if it is more heterozygous, disequilibrium is strong and significant in the populations. Thus the distribution of haplotype frequency and linkage disequilibrium in extant populations are the result of several processes, such as mutation at the SNPs, recombination between the two sites, random genetic drift and gene flow among populations.

All the populations of the present study share the same set of common haplotypes, with only four haplotypes (B2D2A2, B2D1A2, B2D2A1 and B1D2A1) accounting for at least half of the chromosomes in all the populations (Figure 3). The ancestral B2D2A1 haplotype background in the human DRD2 loci accounts for 0.021 to 0.080 frequency, which is consistently present in all the studied groups but it is found to be lesser than the sub-Saharan Africa (0.158 to 0.249)[15]. Kidd *et al.*[15] state that this ancestral haplotype is common in the Africans, but rare or absent elsewhere. The three background haplotypes that differ from the ancestral haplotype by one mutation are B1D2A1, B2D2A2 and B2D1A1 (Figure 4). The B1D2A1 and B2D1A1 haplotypes, which are uncommon in Africa but common in all the other parts of the world, including the present study, are seen in lower frequency and are completely absent in the badagas and irulas. The B2D2A2 haplotype background is common, with a frequency range of 0.145 in kurumba to 0.441 in kota population. The three doubly-derived background haplotypes are B1D1A1, B1D2A2 and B2D1A2. The first two of these and the triply-derived haplotype B1D1A2 are absent or only sporadically present at low frequencies in the study population. In contrast, the B2D1A2 haplotype is present in all the study populations at ~5–49% (Figure 3), while it is seen with modest frequencies (8–15%) in Africans[16].

Kidd *et al.*[15] also state that much of the variation observed today arose some time ago and was present in the ancestral African population from which modern populations descended, and that all of these populations have had large effective population size, allowing them to maintain all the different haplotypes. This is consistent with the single migration of modern *Homo sapiens* out of Africa, and additional loss of variation as that initial non-African founder populations grew and expanded to the east and later into the Americas. Using nuclear DNA markers, Majumder *et al.*[32] also found that a major population expansion has taken place in India. It is also clear from the recent reports on Indian populations that India has played a vital role of being a major corridor in the out-of-Africa migration[24,39]. By and large, the present study using the same set of markers is concordant with the global survey of DRD2 locus[16], affirming that India might have been in the path of this eastward migration. Since the gene investigated in the present study is expressed in the brain and has been associated with the risk for psychiatric illness, our findings may also provide some insight into complex issues of behaviour adaptations.

1. Comings, D. E. *et al.*, *JAMA*, 1991, **266**, 1793–1800.
2. Seeman, P., *Sci. Am. Sci. Med.*, 1995, **2**, 28–37.
3. Inada, T. *et al.*, *Am. J. Med. Genet.* (*Neuropsychiatr. Genet.*), 1996, **67**, 406–408.
4. Grandy, D. K. *et al.*, *Am. J. Hum. Genet.*, 1989, **45**, 778–785.
5. Hauge, X., Grandy, D., Eubanks, J., Evans, G., Givelli, O. and Litt, M., *Genomics*, 1991, **10**, 527–530.
6. Parsian, A., Fisher, L., O'Malley, K. and Todd, R., *Nucleic Acids Res.*, 1991, **19**, 6977.
7. Castiglione, C. M. *et al.*, *Am. J. Hum. Genet.*, 1995, **57**, 1445–1456.
8. Gejman, P. *et al.*, *JAMA*, 1994, **271**, 204–208.
9. Cravchik, A., Sibley, D. R. and Gejman, P. V., *J. Biol. Chem.*, 1996, **271**, 26013–26017.
10. Arinami, T., Gao, M., Hamaguchi, H. and Toru, M., *Hum. Mol. Genet.*, 1997, **6**, 577–582.
11. Litt, M. *et al.*, *Genomics*, 1995, **27**, 101–112.
12. Niswanger, K., Hill, S. Y. and Kaplan, B. B., *Am. J. Med. Genet.* (*Neuropsychiatr. Genet.*), 1995, **60**, 267–271.
13. Noble, E. P., *Alcohol*, 1998, **16**, 33–45.
14. Seeman, P. and Van Tol, H. H. M., *Trends Pharmacol. Sci.*, 1994, **15**, 264–270.
15. Kidd, K. K. *et al.*, *Alcohol Clin. Exp. Res.*, 1996b, **20**, 697–705.
16. Kidd, K. K. *et al.*, *Hum. Genet.*, 1998, **103**, 211–227.
17. Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. and Cavalli-Sforza, L. L., *Nature*, 1994, **368**, 455–457.
18. Deka, R. *et al.*, *Am. J. Hum. Genet.*, 1995, **56**, 461–474.
19. Tishkoff, S. A. *et al.*, *Science*, 1996, **271**, 1380–1387.
20. Thurston, E., *Castes and Tribes of Southern India*, Government Press, Madras, 1909, vol. IV.
21. Saha, N. *et al.*, *Hum. Hered.*, 1976, **26**, 175–197.
22. Breeks, J. W., *Primitive Tribes and Monuments of the Nilgiris*, Cultural Publishing House, Delhi, 1983.
23. Singh, K. S., *The People of India*, Oxford University Press, 1994.
24. Roychoudhury, S. *et al.*, *Curr. Sci.*, 2000, **79**, 1182–1192.
25. Miller, S. A., Dykes, D. D. and Polesky, H. F., *Nucleic Acids Res.*, 1988, **16**, 1215.
26. Grandy, D. K., Zhang, Y. and Civelli, O., *Hum. Mol. Genet.*, 1993, **12**, 2197.
27. Nei, M., *Molecular Evolutionary Genetics*, Columbia University Press, New York, 1987.
28. Majumder, P. and Majumder, P. P., Tech. Rep. No. ISI/AHGU/1-2000, Indian Statistical Institute, Kolkata.
29. Hawley, M. E. and Kidd, K. K., *J. Hered.*, 1995, **86**, 409–411.
30. Lewontin, R. C., *Genetics*, 1964, **49**, 49–67.
31. Iyengar, S. *et al.*, *DNA Sequence*, 1998, **8**, 317–327.
32. Majumder, P. P. *et al.*, *Eur. J. Hum. Genet.*, 1999, **7**, 435–446.
33. Mukherjee, N., Mitra, M., Chakraborty, M. and Majumder, P. P., *J. Genet.*, 2000, **79**, 41–46.
34. Stoneking, M. *et al.*, *Genome Res.*, 1997, **7**, 1061–1071.
35. Novick, G. E. *et al.*, *Hum. Biol.*, 1998, **70**, 23–39.
36. Deinard, A. S. and Kidd, K. K., *DNA Sequence*, 1998, **8**, 289–301.
37. Koop, B. F., Tagle, D. A., Goodman, M. and Slightom, J. L., *Mol. Biol. Evol.*, 1989, **6**, 580–612.
38. Takahata, N., *Mol. Biol. Evol.*, 1993, **10**, 2–22.
39. Roychoudhury, S. *et al.*, *Hum. Genet.*, 2001, **109**, 339–350.

ACKNOWLEDGEMENTS. Project grant from the Department of Biotechnology, Govt. of India is gratefully acknowledged. D.E. thanks the Council of Scientific and Industrial Research for award of Senior Research Fellowship. We also thank the members of the laboratory of the Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata, for help at various stages of this work.

Received 8 July 2002; accepted 28 November 2002