Did human DRD2 haplotypes originate in India? A survey of haplotype frequencies and linkage disequilibrium in the tribes of Eastern Ghats, South India

K. Prabhakaran1,*, A. Ramesh2, M. V. Usha Rani1 and P. P. Majumder3

1Division of Genetics, Department of Environmental Sciences, Bharathiar University, Coimbatore 641 046, India
2Department of Genetics, Dr ALMPGIBMS, University of Madras, Taramani, Chennai 600 113, India
3Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata 700 035, India

In recent years, a possible role of the dopamine D2 receptor (DRD2) locus has been reported in various fields like the etiology of alcoholism, neuropsychiatric disorders, etc. Therefore, it has been the focus of considerable attention. 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) were determined in 197 unrelated individuals from four tribal populations of the Eastern Ghats, an important region of India. The three marker systems in this study are highly polymorphic in all the four tribal populations and the haplotype system showed high levels of heterozygosity than the Nilgiri Hill tribes and those in other parts of the world, except Africa. Out of the possible eight haplotypes, seven are commonly shared by all the populations. The ancestral allele B2D2A1 accounts for 0.028 to 0.166, which was present in all the groups consistently. The linkage disequilibrium was statistically significant in all the populations. The results show a chance of Indian origin or back migration of human DRD2 haplotypes. Data obtained in this study on DRD2 represent one of the small, but growing number of datasets examining disequilibrium and haplotype frequencies in human populations and also indicate that the gene flows from the Eastern Ghats to the Western Ghats. These populations might be one of the oldest among other Indian populations.

Keywords: Eastern Ghats, gene polymorphisms, haplotype frequencies, linkage disequilibrium.

The study of haplotypes and the linkage disequilibrium within them has proven fruitful in human population genetics. Haplotypes defined at the molecular level are now being used to address human evolution1–5. It is often assumed that the origin of disease mutations can be traced through the haplotypes in which they are found6. Thus, it seems important to understand how different factors, such as population history, mutation rate and recombination fraction can jointly affect haplotype evolution and linkage disequilibrium. There is a partial theoretical understanding of the effects of those factors, especially when considered alone or pairwise1. The human dopamine receptor genes and their products are of great interest in many neuropsychiatric disorders4, and are therefore extensively studied to identify genotype–phenotype relationships in neuropsychiatric disorders8–10. Several polymorphisms in the DRD2 gene have been identified in the DNA encompassing the coding sequences; most are in the introns or downstream-flanking DNA4,5,7,8, but some are in coding regions9,10 and in the promoter region11 upstream of exon 1. Many of these polymorphisms have been used to map the locus genetically to the long arm of chromosome 1112–14.

The dopamine receptors are involved in motor control, neuroendocrine regulation, cognition and emotion, and are crucial targets in the pharmacological therapy of schizophrenia, Parkinson’s disease, Tourette’s syndrome, tardive dyskinesia, and Huntington’s disease14. Hence most of the studies are performed as association studies. Almost all the studies have used only the Taq1 ‘A’ site, a single nucleotide polymorphism (SNP) in a Taq1 restriction site13,15. 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.

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. However, this study did not include the Indian population known for its rich diversity. Considering that the surviving tribes of the Eastern Ghats 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 197 tribal individuals belonging to four aboriginal populations dwelling in the important region of the Eastern Ghats on the Deccan Plateau, India (Figure 1).

We have studied a total of 394 chromosomes from 197 unrelated individuals belonging to four endogamous tribal populations of South India. The tribal groups inhabit the hilly tracts and valleys of the Eastern Ghats, a rugged hilly terrain, twin mountain ranges in South India, and this region holds the east and west regions of the Central plateau. A few peaks reach up to 3000 m (9800 ft) and it is one of the richest biodiversity regions of India (Figure 1). The names of the tribal populations (and the number of samples collected) are as follows: Kuruman (60), Sholiga (37), Malayali (58) and Uraly (41). Most of these tribes are basically hunter-gatherers, who are now practising agriculture. The population groups vary from about 15,063 (Kuruman) to 209,039 lakhs (Malayali). Among the study populations, the Malayali tribe is the largest and widespread in Tamil Nadu. Kannada is the mother tongue of all the tribes, except Malayali, who are Tamil-speaking.

The DRD2 markers typed for this study are three biallelic TaqI restriction fragment length polymorphisms (RFLPs) spanning a distance of 25.4 kb. All the markers have been described previously: TaqI ‘A’ by Grandy et al., TaqI ‘B’ by Hauge et al.; TaqI ‘D’ by Parsian et al. DNA was isolated by simple salting out procedure. All typings are PCR-based, using primers and amplification protocols, as described by Castiglione et al. and Kidd et al. After amplification, the three TaqI RFLPs were digested with TaqI restriction enzymes according to the manufacturer’s recommended conditions. Subsequently, the digested fragments were subjected to electrophoresis in 2% agarose gel 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. The assumption of Hardy–Weinberg equilibrium (HWE) was tested using goodness-of-fit test. The maximum likelihood estimates of haplotype frequencies were calculated from the multi-site marker typing data, using the program HAPLOFREQ, 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. The standardized, pairwise linkage disequilibrium value $D’$ 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 the DRD2 locus are polymorphic in all the studied tribal populations. The TaqI ‘B’ allele has a frequency greater than 60% in all the populations and with a maximum of 75%
in Sholigas. At the Taq1 ‘A’ site, the allele frequency ranges from 50 to 65%, with maximum frequency in Kurumans. The Taq1 ‘D’ site shows a maximum range of 75% in Sholigas and a minimum of 50% in Kurumans. The overall patterns of allele frequency variation at the Taq1 ‘B’ site are comparatively higher than Taq1 ‘A’ and Taq1 ‘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. Table 1 shows the estimates of heterozygosity 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 Eastern Ghats tribes are listed in Table 2 separately for each population. It is seen that all the study populations shared all the eight possible haplotypes in the three-site DRD2 locus, B2D2A2, B2D1A2, B1D2A2, B1D1A2, B2D2A1, B2D1A1 and B1D1A1, except B1D2A1 in Malayalis. Most of them (but not all the haplotypes) occurred at a frequency of ≥5% in all the studied populations. The overall dataset is graphically represented in Figure 3. Table 3 presents the three pairwise standardized $D'$ values for the three biallelic markers, Taq1 ‘B’, Taq1 ‘D’ and Taq1 ‘A’. All the comparisons were significant.

Evolutionary relationships of DRD2 haplotypes are defined by three Taq1 restriction site polymorphisms. The ancestral alleles (B2, D2 and A1) are all identical in sequence to those in chimpanzees, gorillas and orangutans. Frequency data in different human populations, especially those in Africans and also from the present study, suggest the illustrated events by which three mutational events result in four of the eight possible haplotypes. The other

### Table 1. Heterozygosity at individual sites of DRD2 loci

<table>
<thead>
<tr>
<th>Population</th>
<th>2n</th>
<th>‘B’</th>
<th>‘D’</th>
<th>‘A’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kuruman</td>
<td>114</td>
<td>0.418</td>
<td>0.500</td>
<td>0.450</td>
</tr>
<tr>
<td>Sholiga</td>
<td>74</td>
<td>0.368</td>
<td>0.368</td>
<td>0.491</td>
</tr>
<tr>
<td>Malayali</td>
<td>108</td>
<td>0.438</td>
<td>0.464</td>
<td>0.500</td>
</tr>
<tr>
<td>Uraly</td>
<td>82</td>
<td>0.470</td>
<td>0.403</td>
<td>0.497</td>
</tr>
</tbody>
</table>

2n, Number of chromosomes.

### Table 2. Haplotype frequency at DRD2 locus in tribal populations

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Kuruman</th>
<th>Sholiga</th>
<th>Malayali</th>
<th>Uraly</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2D2A2</td>
<td>0.248</td>
<td>0.461</td>
<td>0.225</td>
<td>0.392</td>
</tr>
<tr>
<td>B2D1A2</td>
<td>0.117</td>
<td>0.163</td>
<td>0.152</td>
<td>0.065</td>
</tr>
<tr>
<td>B1D2A2</td>
<td>0.183</td>
<td>0.030</td>
<td>0.100</td>
<td>0.096</td>
</tr>
<tr>
<td>B1D1A2</td>
<td>0.103</td>
<td>0.111</td>
<td>0.194</td>
<td>0.066</td>
</tr>
<tr>
<td>B2D2A1</td>
<td>0.060</td>
<td>0.037</td>
<td>0.166</td>
<td>0.028</td>
</tr>
<tr>
<td>B2D1A1</td>
<td>0.094</td>
<td>0.104</td>
<td>0.108</td>
<td>0.230</td>
</tr>
<tr>
<td>B1D2A1</td>
<td>0.160</td>
<td>0.056</td>
<td>0.000</td>
<td>0.019</td>
</tr>
<tr>
<td>B1D1A1</td>
<td>0.035</td>
<td>0.039</td>
<td>0.057</td>
<td>0.104</td>
</tr>
</tbody>
</table>

n, Number of individuals.

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

### Table 3. Pairwise linkage disequilibrium values of DRD2 locus

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Kuruman</td>
<td>0.174*</td>
<td>0.207*</td>
<td>0.050*</td>
</tr>
<tr>
<td>Sholiga</td>
<td>0.196*</td>
<td>0.357*</td>
<td>0.357*</td>
</tr>
<tr>
<td>Malayali</td>
<td>0.479*</td>
<td>0.421*</td>
<td>0.039</td>
</tr>
<tr>
<td>Uraly</td>
<td>0.058*</td>
<td>0.205*</td>
<td>0.761*</td>
</tr>
</tbody>
</table>

Standardized linkage disequilibrium coefficient ($D'$) of Lewontin 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 ($P > 0.05$).
four were then generated by at least one crossover between two of these mutually derived haplotypes. The present data are insufficient to exclude other possible evolutionary schemes.

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 dataset on DRD2 also represents one of the small, but growing number of datasets examining disequilibrium and haplotype frequencies in human populations.

The DRD2 locus has proved useful in studying genetic structure of human populations. As has been reported in the global survey and the Indian survey of DRD2 locus, 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, with the exception of the African populations. Interestingly, the average heterozygosity levels were also higher than the other global populations studied, with the exception of the African populations. Thus, the DNA markers attest that the study groups exhibit high levels of genomic diversity. Earlier studies of haplotype evolution of DRD2 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. have sequenced the regions around all the Taq1 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 the homologous sequence depending on the species compared. In 301 bp of DRD2 intron 3, Deinard and Kidd found no human or gorilla polymorphism, but reported 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 amount of time separating the species, these differences are consistent with other molecular evolution studies of these species. There is evidence that Taq1 sites are hypermutable, and the time frame for neutral polymorphic variation to persist in human populations is almost certainly less than a million years. Thus the B2D2A1 background haplotype may be concluded as the ancestral hominid haplotype, and the recurrent mutation is not a factor influencing the allele frequencies for these polymorphisms. As reported in a similar study, while a low number of alleles occurring due to recurrent mutation cannot be excluded, the phenomenon is negligible at the level of the present analysis.

In all the population samples, the disequilibrium at the DRD2 locus is highly significant (Table 3). Among the SNPs, the Taq1 ‘B’ and Taq1 ‘A’ sites showed significant pairwise disequilibrium which is clearly observed in the study of Asian-American populations, but it also exists in Africans. However, if it is more heterozygous, disequilibrium is strong and significant in the populations. Thus the distribution of haplotype frequencies and linkage disequilibrium in extant populations is 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 seven haplotypes (B2D2A2, B2D1A2, B1D2A2, B1D1A2, B2D2A1, B2D1A1 and B1D1A1) 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 0.028–0.166 frequency, which is consistently present in all the studied groups, but is found to be lesser than that in sub-Saharan African populations (0.158–0.249). According to Kidd et al., this ancestral haplotype is common in Africans, but rare or absent elsewhere. But the present study is not consistent with this because the ancestral allele B2D2A2 is significantly present in relatively high frequencies. There can be two possibilities: (1) These ancestral haplotypes could have arisen in India and then carried to other parts of the world and/or (2) these haplotypes are not specific to Africa. The three background haplotypes that differ from the ancestral haplotype by one mutation are B1D2A1, B2D2A2 and B2D1A1. The B1D2A1 and B2D1A1 haplotypes, which are uncommon in Africa but common in all the other parts of the world, including those in the present study, are seen in lower frequency and the B1D2A1 haplotype is completely absent in the Malayalis. The B2D2A2 haplotype background is common, with a frequency range of 0.225 in the Malayalis to 0.461 in the Sholiga population. The three doubly-derived background haplotypes are B1D1A1, B1D2A2 and B2D1A2 and are present at low frequencies in all the study populations. In contrast, except B1D2A1, all the haplotypes are present in all the study populations at ~5–46% (Figure 3), while it is seen with modest frequencies (8–15%) in Africans. Hence India could have contributed all DRD2 haplogroups to other parts of the world, including Africa.

Much of the variation observed today arose some time ago and was present in the ancestral African population from which modern populations descended, and 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.\textsuperscript{30} and Vishwanathan et al.\textsuperscript{31} have 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 in being a major corridor in the out-of-Africa migration\textsuperscript{22,33,39,40}. Though the present study using the same set of markers is concordant with the global survey\textsuperscript{16} and South Indian\textsuperscript{29} study of DRD2 locus, it does not affirm that India might have been in the path of this eastward migration, and strongly supports the ‘Out-of-Africa’ model of Human Evolution. However, in the examined populations the presence of three ancestral haplotype at high frequency and except B1D2A1, all the haplotypes are present in all the study populations at ~5–46% indicates the chance of Indian origin or back migration of human DRD2 haplotypes. In contrast with the regional consideration with the comparison of the same study\textsuperscript{20} of the neighbouring region, the Nilgiris of Western ghats, one of the global hotspots of India possessing the low significant results (including average heterozygosity, haplotype frequencies, LD values) indicates the gene flow from the present study area Eastern Ghats to Western Ghats. So, with this new finding of the present study, we argue that these ancestral populations of India might have settled in Eastern Ghats and later moved to Western ghats because of so many social forces like temperature, natural livelihood resources, etc. Since the gene investigated in the present study is expressed in brain and has been associated with the risk for psychiatric illness and so many others, our findings may also provide some insight into complex issues of behaviour adaptations and other significant illnesses and also may give the accurate insights in the peopling of India.


ACKNOWLEDGEMENTS. We are grateful to the Department of Biotechnology, Govt of India for the financial support. We thank Ms S. Kanthimathi, Ms M. Vijaya and the members of the laboratory of Prof. A. Ramesh, Department of Genetics, Dr ALMPGIBMS, University of Madras, Chennai, for their help at various stages. We thank Mr N. K. Biswas, Anthropology and Human Genetics Unit, Indian Statistical Institute, Kolkata for his help in statistical analysis.

Received 6 July 2007; revised accepted 2 April 2008