Evolution of Exon 1 of the Dopamine D4 Receptor (DRD4) Gene in Primates

MICHAEL I. SEAMAN,1 FONG-MING CHANG,2,3 AMOS S. DEINARD,1 ANA T. QUIÑONES,2 AND KENNETH K. KIDD2,4*

1Department of Anthropology, Yale University, New Haven, Connecticut 06520-8277
2Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520-8005
3Department of Obstetrics and Gynecology, National Cheng Kung University Medical College, Taiwan
4Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut 06250

ABSTRACT

The dopamine D4 receptor (DRD4) gene exhibits a large amount of expressed polymorphism in humans. To understand the evolutionary history of the first exon of DRD4—which in humans contains a polymorphic 12bp tandem duplication, a polymorphic 13bp deletion, and other rare variants—we examined the homologous exon in thirteen other primate species. The great apes possess a variable number of tandem repeats in the same region as humans, both within and among species. In this sense, the 12bp tandem repeat of exon 1 is similar to the 48bp VNTR of exon 3 of DRD4, previously shown to be polymorphic in all primate species examined. The Old World monkeys show no variation in length, and a much higher conservation of amino acid sequence than great apes and humans. The New World monkeys show interspecific differences in length in the region of the 12bp polymorphism, but otherwise show the higher conservation seen in Old World monkeys. The different patterns of variation in monkeys compared to apes suggest strong purifying selective pressure on the exon in these monkeys, and somewhat different selection, possibly relaxed selection, in the apes. J. Exp. Zool. (Mol. Dev. Evol.) 288:32–38, 2000.

The dopamine D4 receptor is a 7-transmembrane G-protein-coupled dopamine receptor found in the limbic system, frontal cortex, and other areas of the brain (for reviews, see Strange, ’94; Seeman, ’95; Van Tol, ’96). It is one of three “D2-like” dopamine receptors. The D4 receptor is believed to play a role in higher brain functions, such as affection and personality, and is a candidate gene for several behavioral disorders. The DRD4 gene is located on chromosome 11p15.5 (Gelernter et al., ’92), and has four exons (Fig. 1). It harbors a large amount of expressed polymorphism in humans. Some of these allelic variants have been claimed to be associated with psychiatric disorders such as Tourette syndrome, delusional disor-der, attention-deficit hyperactivity disorder, and obsessive-compulsive disorder with tics (Catalano et al., ’93; Grice et al., ’96; LaHoste et al., ’96; Cruz et al., ’97); substance abuse, including alcoholism and heroin addiction (George et al., ’93; Muramatsu et al., ’96; Kotler et al., ’97); and variation in normal human behavior, especially novelty-seeking behavior (Benjamin et al., ’96; Ebstein et al., ’96, ’97; Ono et al., ’97).

Among the numerous expressed polymorphisms at DRD4, exon 3 contains a 48bp expressed VNTR, corresponding to a 16-amino-acid repeat in the third cytoplasmic loop (Lichter et al., ’93). Also in exon 3, a Val→Gly polymorphism at codon 194 is found in some African populations (Seeman et al., ’94); the Gly194 allele is believed to result in an essentially nonfunctional protein (Liu et al., ’96). The coding region of exon 1 contains a polymorphic 12 bp (4 amino acid) tandem repeat with three alleles in humans, a 13bp deletion polymor-
EVOLUTION OF DRD4 EXON 1 IN PRIMATES

aphic in Europeans, a rare 21 bp deletion, and a rare nonsynonymous transversion in codon 11 (Catalano et al., '93; Nöthen et al., '94; Cichon et al., '95; Chang and Kidd, '97). The 13 bp and the 21 bp deletions presumably result in nonfunctional proteins (Nöthen et al., '94). Of the exon 1 polymorphisms, only the 12 bp repeat is polymorphic worldwide (Chang et al., unpublished data). This polymorphism consists of a 12 bp unit present once, twice, or rarely thrice. The resulting four amino acid duplication (GASA) occurs near the junction of the extracellular domain of the protein and the first transmembrane domain (Seeman et al., '94). Exon one of DRD4 codes for the extracellular amino terminus domain and the first two transmembrane domains (Seeman, '95). The extracellular domain has been shown to be functionally important for agonist binding and signal transduction (Schoots et al., '96).

The evolution of DRD4 has been studied by comparative sequencing of the VNTR in exon three in several non-human primate species. This VNTR was found to be polymorphic in many primate species, both for the number of repeat units as well as for sequence variation among units. No alleles appeared to be shared between species; rather, the length polymorphism was apparently newly regenerated in each independent lineage examined (Livak et al., '95; Matsumoto et al., '95; Inoue-Murayama et al., '98).

Here we report on the comparative sequencing of the coding region of exon one in several primate species. As part of ongoing studies of the 12 bp tandem repeat polymorphism in exon one of DRD4 (Chang et al., '97; Chang and Kidd, '97), we examined the homologous regions in several non-human primate species in order to accurately determine the ancestral human allele (Iyengar et al., '98). Length variation has been regenerated several times in the great apes in exon one, similar to exon three. Also, the patterns of nucleotide substitution observed suggest different selective constraints on this region in humans and great apes compared to other primates. In this paper, we use the term "hominid" to refer to the Family Hominidae, including humans (Homo), chimpanzees (Pan), gorillas (Gorilla), and orangutans (Pongo) as per Groves ('86). While we realize that this usage is not universally accepted, it appears to be increasing in popularity, and furthermore is preferable to the more traditional—but paraphyletic—usage of Pongidae to include Pan and Gorilla.

MATERIALS AND METHODS

DNA samples

Genomic DNA of chimpanzees (Pan troglodytes; n = 11), bonobos or pygmy chimpanzees (P. paniscus; n = 6), gorillas (Gorilla gorilla; n = 6), orangutans (Pongo pygmaeus; n = 8), and spider monkey (Ateles sp.; n = 1) was extracted by standard techniques from whole blood or lymphoblastoid cell lines established from blood generously provided by Atlanta Zoo/Yerkes Regional Primate Center, New Iberia Regional Primate Center, Arizona Primate Foundation, Henry Doorly Zoo, Sacramento Zoo, Milwaukee County Zoo, and Miami Metro Zoo. Genomic DNA of the following gibbon species was extracted by standard techniques from lymphoblastoid cell lines generously provided by Dr. Johannes Weinberg: lar gibbon (Hylobates lar; n = 1), klossi's

Fig. 1. Schematic representation of the DRD4 gene, based on Van Tol et al. ('92) and Kamakura et al. ('97). Locations of the four expressed polymorphic sites and the two rare variants are shown above the gene; the locations of the PCR and sequencing primers used in this study are shown below the gene. For exact locations of the primers, compare the sequences given in Table 1 to the genomic DNA sequence of humans (GenBank accession number L12397).
gibbon (H. klossi; n = 1), and siamang (H. syndactylus; n = 1). DNA from proboscis monkey (Nasalis larvatus; n = 1) and silver langur (Presbytis cristata; n = 1) was extracted from liver tissue generously provided by Dr. George Amato of the Wildlife Conservation Society. Genomic DNA of baboons (Papio hamadryas; n = 2) and squirrel monkeys (Saimiri boliviensis; n = 2) was generously provided by Dr. Jeffrey Rogers of the Southwest Foundation for Biomedical Research. Genomic DNA of rhesus macaques (M. mulatta; n = 7) was extracted from whole blood generously provided by Dr. Goldman-Rakic of the Yale University School of Medicine.

**PCR-based typing of the exon 1 length variant**

The exon 1 12 bp tandem repeat polymorphism was typed using the PCR primers D4EX1F and D4EX1R (Table 1) as described in Chang and Kidd (‘97) using genomic DNA of all of the great apes, Old World monkeys, and New World monkeys.

**DNA sequencing of exon 1 alleles**

The sequences and locations of all primers used are given in Table 1 and Figure 1, respectively. DNAs from humans, chimps, bonobos, gorillas, and orangutans determined to be homozygous for the 12 bp repeat polymorphism using the above typing method were amplified and sequenced with the primers D4EX1E and D4EX1DW, or with D4UP2 and D4EX1DW, which flank exon 1. Additional sequencing with D4EX1C and D4EX1BK1 was performed to resolve ambiguities. Since we detected no gorillas or orangutans who were homozygous for the short allele, those alleles were sequenced by excising the band from a 4% low-melting point agarose gel, melting in 100 µl dH2O for 10 min at 65°C, and re-amplifying 1 µl of the melted solution in a 100-µl PCR using the same primers. The sequence of the remaining primates was determined by amplifying genomic DNA with the degenerate primer D4UPmod and D4DW2. An aliquot of this product was re-amplified and sequenced with D4UPmod and D4EX1R, or with D4EX1F and D4EX1R. At times, the PCR product was excised from the gel and purified prior to sequencing because several bands were present. Additional sequence was determined by amplifying genomic DNA with D4EX1D and D4EX2R, and sequencing with D4EX1D. For most species only a single individual was sequenced, although all of the different length alleles in the great apes were sequenced. Sequences were manually edited and nucleotide positions were determined to be homozygous or heterozygous after visual comparison of sequence traces in both directions with other individuals (Kwok et al., ’94). All sequencing was done with the ABI Prism cycle sequencing kit and run on an ABI373 automated sequencer. Sequences were deposited in GenBank (accession numbers AF010294-AF010301, AF125662-AF125670).

**Sequence analysis**

The MEGA software program (Kumar et al., ‘93) was used to count the numbers of nonsynonymous and synonymous substitutions between haplotypes, and to calculate the Jukes-Cantor distances and standard errors of nonsynonymous and synonymous substitutions.

**RESULTS**

All hominid (Family Hominidae, meaning the family comprised of great apes and humans) genera possess alleles of exon one of more than one size (Fig. 2). Alleles of three lengths were observed in orangutans, identical in sizes to the three hu-

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ to 3’)</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>D4EX1F</td>
<td>CGCCATGGGAAACCGCAG</td>
<td>Typing exon 1 12bp polymorphism, sequencing</td>
</tr>
<tr>
<td>D4EX1R</td>
<td>CGGCTCCTCCTCGGAGTAGA</td>
<td>Typing exon 1 12 bp polymorphism, sequencing</td>
</tr>
<tr>
<td>D4-3</td>
<td>CGGACTACGTGGTCTGTCG</td>
<td>Typing exon 3 48bp VNTR</td>
</tr>
<tr>
<td>D4-42</td>
<td>AGGACCTCATGCGCTTGG</td>
<td>Typing exon 3 48bp VNTR</td>
</tr>
<tr>
<td>D4UP2</td>
<td>CGCTACGTCCGCCGCAGTTT</td>
<td>PCR</td>
</tr>
<tr>
<td>D4EX1E</td>
<td>AGGGACTCCCCGGCTTGC</td>
<td>PCR, sequencing</td>
</tr>
<tr>
<td>D4DW</td>
<td>GAGCAGGGAGTAGTCTCG</td>
<td>PCR, sequencing</td>
</tr>
<tr>
<td>D4EX1C</td>
<td>GGGGGCGTGTGCTCATC</td>
<td>Sequencing</td>
</tr>
<tr>
<td>D4BK1</td>
<td>GAGCCCCCGCGGATGAG</td>
<td>Sequencing</td>
</tr>
<tr>
<td>D4UPmod</td>
<td>CGTGTGCAGCGGATGAG</td>
<td>PCR, sequencing</td>
</tr>
<tr>
<td>D4DW2</td>
<td>TCCGCGCGCCGCGGCTCAC</td>
<td>PCR, sequencing</td>
</tr>
<tr>
<td>D4EX1D</td>
<td>GCGGTGCAGCGGATGAG</td>
<td>PCR, sequencing</td>
</tr>
<tr>
<td>D4EX2R</td>
<td>AGGCCACCGCGACCCCTGG</td>
<td>PCR</td>
</tr>
</tbody>
</table>
man alleles (hereafter referred to as short, long, and extra long). Gorillas were observed to possess alleles identical in length to the human short and long alleles, while common chimpanzees and bonobos are each apparently monomorphic at exon one, but for different alleles, possessing only alleles identical in length to the human long and short alleles, respectively. Sequencing of each different allele confirmed that all of the length variation in all of the great apes occurred in the same region as the 12 bp tandem repeat polymorphism in humans. Several shared base pair changes have occurred within and on either side of the repeat, indicating that the length variation has been repeatedly regenerated within each taxon, rather than the individual alleles having persisted across species. No length variation was observed within or between any of the cercopithecoid species (Old World monkeys), although—except for rhesus macaques (2n = 14 chromosomes)—only one or two individuals per species were examined. Among the platyrrhines (New World primates), the spider monkey's coding region of exon 1 is 12 bp shorter than the short allele in hominids, and the squirrel monkey has 39 bp, or 13 amino acids, less than the hominid short allele.

Figure 3 gives the inferred amino acid sequence of exon 1 for the 20 primate alleles identified, as well as that inferred from the previously published rat and mouse sequences (O'Malley et al., '92; Fishburn et al., '95; GenBank accession numbers M84009, U19880). All of the insertions/deletions and most of the amino acid substitutions occur in the first half of the exon, corresponding to the amino-terminus extracellular tail. Many of the substitutions are nonconservative (e.g., G→W at codon 11 in orangutans), and at least three sites (codons 9, 11, and 19) have been repeatedly changed. The P→Q substitution in the great apes and humans at amino acid 37 in the alignment occurs at the point where the peptide chain presumably enters the cellular membrane (Seeman et al., '94).

Outside of the tandem repeat region there are more nonsynonymous than synonymous substitu-
tions between haplotypes in the hominids, while among all other primate clades examined there are more synonymous than nonsynonymous substitutions (Table 2). Only one amino acid position, at codon 11, varies among the cercopithecoids (Fig. 3). Taking into account the sequence composition of the exon, the average ratio of the rate of nonsynonymous substitutions per nonsynonymous site ($d_N$) to the rate of synonymous substitution per synonymous site ($d_S$) is 0.44 for all pairwise comparisons between observed alleles in the hominids, and only 0.13 in the cercopithecoids. The null hypothesis of neutral evolution of exon 1 cannot be rejected for the hominids, but is rejected for the cercopithecoids (one-tailed $t$-test, $P < 0.05$; Kumar et al., ’93), suggesting that purifying selection has maintained the amino acid sequence in these monkeys. The repeat region is excluded in these calculations.

**DISCUSSION**

Exon one of the DRD4 gene exhibits polymorphism in humans, gorillas, and orangutans. Recurrent, or parallel, mutations have occurred repeatedly in several primate lineages generating homoplasy in both repeat length and in amino acid substitutions. The recurrent mutation in the 12 bp repeat region makes reconstruction of the evolutionary pathway between existing haplotypes problematic, if not impossible. Orangutans appear to possess alleles that are identical by state in size, but not identical by descent. It is certainly possible that additional intraspecific variation in any of the species might be found if more individuals were examined.

Different patterns of nucleotide substitutions in the different primate lineages suggest that natural selection on exon one of DRD4 is operating differently in different lineages. Hominids are characterized by relaxed selection, while Old World and New World monkeys are characterized by strong purifying selection on this exon. Most of the amino acid substitutions in the hominids are found in the amino-terminus of the protein, corresponding to the extracellular domain—a region shown to be important for the binding of dopamine and signal transduction (Schoots et al., ’96). This apparent onset of relaxation of selection on the amino-terminus of this protein in hominids seems to have occurred concomitantly with the P→Q amino acid substitution at the junction of the extracellular and transmembrane domains. As proline residues are often involved in conferring rigid structural turns in peptide chains, the loss of a proline here in hominids may have repositioned the extracellular domain in a way that decreased its functional relevance to the overall function of the protein.

It is tempting to speculate that DRD4 may be responsible for some of the remarkable changes involved in the evolution of the hominid brain, as it has been suggested that this gene in humans may be involved in higher thought processes, including personality (Benjamin et al., ’96; Ebstein et al., ’96, ’97). Other studies have not supported that hypothesis (Malhotra et al., ’96; Pogue-Geile et al., ’98). In any case, some domains of this receptor may be functioning differently in Old World and New World monkeys than in humans and other apes, as inferred from the different patterns of natural selection observed.

Finally, although only a portion of this gene was examined, the effective use of many primate taxa here reinforces the advantages of the “bushy tree” approach to comparative sequencing, as exemplified by previous studies of primate molecular evo-

**TABLE 2. Numbers of nonsynonymous and synonymous substitutions observed in exon one**

<table>
<thead>
<tr>
<th>Non-synonymous substitutions</th>
<th>Synonymous substitutions</th>
<th>Comparison with hominids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hominids</td>
<td>90</td>
<td>84</td>
</tr>
<tr>
<td>Gibbons</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Cercopithecoids</td>
<td>7</td>
<td>23</td>
</tr>
<tr>
<td>Platyrrhines</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Cercopithecoids and platyrrhines</td>
<td>45</td>
<td>96</td>
</tr>
<tr>
<td>Non-hominids</td>
<td>109.5</td>
<td>234</td>
</tr>
</tbody>
</table>

1The 12 bp repeat region is excluded in these calculations. The observed ratios did not differ significantly between gibbons and cercopithecoids, cercopithecoids and platyrrhines, or gibbons and platyrrhines. The significance values given in the table are for each individual two-by-two comparison. However, the comparisons between hominids and gibbons and between hominids and platyrrhines are no longer significant after applying Bonferroni techniques to adjust the critical values for each of the eight individual comparisons in order to create an experimentwise error rate of $P < 0.05$ (Sokal and Rohlf, ’95). This is not surprising given this conservative approach and the small numbers involved in comparisons with gibbons and platyrrhines.

2Total numbers of substitutions between all haplotypes observed within each group.
lution (Messier and Stewart, '97; Wu et al., '97). The use of many taxa allows better definition of “conserved” amino acids. For example, a Gly→Arg variant at codon 11 of DRD4 was identified in humans (Cichon et al., '95). Since mice also have a glycine at codon 11, a human-mouse comparison would suggest that this site has been highly conserved for over 70 million years, and therefore that this amino acid was essential for proper dopamine receptor function. However, this glycine has been substituted at least three times independently in different primate lineages (Fig. 3), including a Gly→Arg change in two Old World monkey species indicating that the rare Arg11 allele in humans is more likely to be functionally normal.

The function of DRD4 has previously been examined through the use of in vitro studies, histology, laboratory animals, population genetics, and disease association studies. Here we employ a comparative sequencing approach toward understanding the function of exon 1 of DRD4, and the evolutionary history that has shaped it.

ACKNOWLEDGMENTS

We thank all of the people and institutions who generously provided us with DNA, blood, cell lines, or tissues, listed in the Materials and Methods section. We thank Sudhir Kumar, Koichiro Tamura, and Masatoshi Nei of the Institute of Molecular Evolutionary Genetics at The Pennsylvania State University for the availability of the MEGA software. We thank William C. Speed for assistance with the evolutionary history that has shaped it.

LITERATURE CITED

Kumar S, Tamura K, Nei M. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park, PA 16802.
Lichter JB, Barr CL, Kennedy JL, Van Tol HHM, Kidd KK,


