Am. J. Hum. Genet. 74:931–944, 2004

The Genetic Architecture of Selection at the Human Dopamine Receptor D4 (*DRD4*) Gene Locus

E. Wang,[1,*] Y.-C. Ding,[1,*] P. Flodman,[3] J. R. Kidd,[4] K. K. Kidd,[4] D. L. Grady,[1] O. A. Ryder,[5] M. A. Spence,[3] J. M. Swanson,[2,3] and R. K. Moyzis[1,2]

[1]Department of Biological Chemistry, and [2]Child Development Center, University of California, Irvine; [3]Department of Pediatrics, University of California at Irvine, Medical Center, Orange, CA; [4]Department of Genetics, School of Medicine, Yale University, New Haven; and [5]Center for Reproduction of Endangered Species, The Zoological Society of San Diego, San Diego

Associations of the seven-repeat (7R) allele of the human dopamine receptor D4 (*DRD4*) gene with both the personality trait of novelty seeking and attention deficit/hyperactivity disorder have been reported. Recently, on the basis of the unusual DNA sequence organization of the *DRD4* 7R 48-bp tandem repeat (VNTR), we proposed that the 7R allele originated as a rare mutational event that increased to high frequency by positive selection. We now have resequenced the entire *DRD4* locus from 103 individuals homozygous for 2R, 4R, or 7R variants of the VNTR, a method developed to directly estimate haplotype diversity. DNA from individuals of African, European, Asian, North and South American, and Pacific Island ancestry were used. 4R/4R homozygotes exhibit little linkage disequilibrium (LD) over the region examined, with more polymorphisms observed in DNA samples from African individuals. In contrast, the evidence for strong LD surrounding the 7R allele is dramatic, with all 7R/7R individuals (including those from Africa) exhibiting the same alleles at most polymorphic sites. By intra-allelic comparison at 18 high-heterozygosity sites spanning the locus, we estimate that the 7R allele arose prior to the upper Paleolithic era (~40,000–50,000 years ago). Further, the pattern of recombination at these polymorphic sites is the pattern expected for selection acting at the 7R VNTR itself, rather than at an adjacent site. We propose a model for selection at the *DRD4* locus consistent with these observed LD patterns and with the known biochemical and physiological differences between receptor variants.

Introduction

The human dopamine receptor D4 (*DRD4*) gene, located near the telomere of chromosome 11p, exhibits an unusual amount of expressed polymorphism (Lichter et al. 1993; Ding et al. 2002; Grady et al. 2003). Much of this variation is the result of length and of SNP changes in a 48-bp tandem repeat (VNTR) in exon 3, encoding the third intracellular loop of this D2-like receptor. Alleles containing 2-11 repeats (2R-11R) have been found, with over 67 different haplotype variants uncovered to date (Ding et al. 2002; Grady et al. 2003). The three most common variants—2R, 4R, and 7R—represent >90% of the observed population allelic diversity. In most geographical locations, the 4R allele is the most common, whereas 2R and 7R allele frequencies vary widely (Chang et al. 1996; Ding et al. 2002).

Received December 23, 2003; accepted for publication February 25, 2004; electronically published April 9, 2004.

Address for correspondence and reprints: Dr. Robert K. Moyzis, Department of Biological Chemistry, College of Medicine, Sprague Hall, University of California at Irvine, Irvine, CA 92697. E-mail: rmoyzis@uci.edu

* The first two authors contributed equally to this article.

0002-9297/2004/7405-0014$15.00

The functional significance of these length/sequence changes in the DRD4 protein, in a region that couples to G proteins and mediates intercellular cAMP levels, has been studied extensively (Asghari et al. 1995; Jovanovic et al. 1999; Oak et al. 2000). In particular, the 7R variant exhibits a blunted ability to reduce cAMP levels, in comparison with that of the common 4R variant (Asghari et al. 1995). The DRD4 protein is expressed in a number of brain regions, with a high level of expression in the prefrontal cortex, a region thought to be involved in cognition, attention, and other higher brain functions (Oak et al. 2000). Significantly, *DRD4* knockout mice display better performance on complex motor tasks; are supersensitive to cocaine, ethanol, and methamphetamine; and exhibit reduced exploration of novel stimuli (Rubinstein et al. 1997; Dulawa et al. 1999). Taken together, these results are consistent with the proposal that *DRD4* receptors act as inhibitors of neuronal firing, especially in the prefrontal cortex (Oak et al. 2000; Rubinstein et al. 2001).

On the basis of these biochemical and physiological observations, a number of investigations have looked for associations between behavioral phenotypes and particular alleles of this highly variable gene (Faraone et al. 2001; Swanson et al. 2001; Klugar et al. 2002; Grady et al. 2003). Although some studies have sug-

gested that the *DRD4* 7R allele might be associated with the personality trait of novelty seeking (MIM 601696) (Klugar et al. 2002), the most reproduced association is between the 7R allele and attention deficit/hyperactivity disorder (ADHD) (MIM 126452) (Faraone et al. 2001; Swanson et al. 2001; Grady et al. 2003). ADHD is the most prevalent disorder of childhood (~5% incidence) and is defined by symptoms of developmentally inappropriate inattention, impulsivity, and hyperactivity. The ~2-fold greater prevalence of the *DRD4* 7R allele in ADHD probands ($\lambda = 1.9$), calculated from a recent meta-analysis (Faraone et al. 2001), indicates that this allele is associated with a significant fraction (25%–50%) of the attributable genetic risk for the disorder (Grady et al. 2003).

Elsewhere, we showed by DNA resequencing/haplotyping of 600 *DRD4* VNTRs, which represented a worldwide population sample, that the origin of most haplotype variants could be explained by simple one-step recombination/mutation events (Ding et al. 2002). In contrast, the 7R allele is not related simply to the other common alleles, differing by >6 recombinations/mutations. This unusual sequence architecture of the 7R VNTR, suggesting it arose as a rare mutational event, led to exploratory measures of linkage disequilibrium (LD) between the 4R and 7R alleles. Large discrepancies between allele ages, estimated from low intra-allelic variability and high population frequency, are taken as evidence that selection has increased the frequency of an allele beyond that expected by chance (Slatkin and Rannala 2000; Tishkoff et al. 2001; Ding et al. 2002; Sabeti et al. 2002; Bamshad and Wooding 2003; Swanson 2003). Strong LD was found between the 7R allele and four surrounding *DRD4* polymorphisms, suggesting this allele is significantly "younger" than the common 4R allele. Our preliminary estimates placed the most recent common ancestor of the *DRD4* 7R allele around the time of major human population expansion out of Africa and the appearance of radical new technology (the upper Paleolithic era) (Harpending and Rogers 2000; Ingman et al. 2000; Underhill et al. 2000; Ding et al. 2002). We speculated that these events may be related to the appearance of and selection for the *DRD4* 7R allele (Ding et al. 2002).

If the *DRD4* 7R allele arose recently and underwent strong positive selection, why is it currently disproportionately represented in individuals diagnosed with ADHD? One possibility is that an adjacent polymorphism in strong LD with the 7R VNTR is actually associated with ADHD or may be the target of selection. We have argued that selection for an adjacent site was unlikely, given the distinct and unusual DNA sequence organization of the *DRD4* 7R allele itself (Ding et al. 2002), but, owing to the high density of SNPs in the human genome, selection for an adjacent site remained a possibility. Obviously, even if the *DRD4* VNTR is the site of selection, strong LD in the region could have carried an adjacent ADHD-predisposing polymorphism along with it. Such "hitch-hiking" events should be common (Fay and Wu 2000), again, given the high density of SNPs in human DNA (International SNP Map Working Group 2001).

Alternatively, the biochemical and physiological properties of the DRD4 protein (discussed above) suggest a more direct relationship. We have proposed that the 7R–ADHD association is an example consistent with the Common Variant–Common Disorder hypothesis (Risch and Merikangas 1996; Grady et al. 2003), in which common predisposing alleles result in deleterious effects only when combined with other environmental/genetic factors. We have speculated that the same traits that may be selected for in individuals with a *DRD4* 7R allele also may predispose behaviors that are deemed inappropriate in the typical classroom setting and hence diagnosed as ADHD (Ding et al. 2002; Grady et al. 2003). In this environmental-mismatch hypothesis (Hartman 1993; Jensen et al. 1997), the *DRD4* 7R subset of individuals diagnosed with ADHD is assumed to have a different, evolutionarily successful behavioral strategy, rather than a disorder. It is also possible, however, that *DRD4* 7R, although selected for in human populations, could have deleterious effects when combined with genetic variants in other genes (Ding et al. 2002; Grady et al. 2003).

To clarify some of these issues, we report an extensive analysis of polymorphisms surrounding the *DRD4* VNTR, by genomic resequencing. Remarkably, we show that the 7R allele exhibits strong LD worldwide, in geographic locations as diverse as sub-Sahara Africa and South American rainforests. By intra-allelic comparison at 18 high-heterozygosity sites spanning the locus, we refine the age estimate of the most recent 7R-allele common ancestor to ~40,000–50,000 years. Further, the pattern of inferred recombination at these sites is the pattern expected for selection acting at the VNTR itself, rather than at an adjacent polymorphism.

Material and Methods

Establishing Cell Lines and DNA Purification

Lymphoblastoid cell lines were established for all individuals. Methods for transformation, cell culture, and DNA purification have been described elsewhere (Chang et al. 1996; Ding et al. 2002; Grady et al. 2003). All individuals gave their informed consent before their inclusion in this study, which was carried out under protocols approved by the human subjects committees at the participating institutions. The geographical and eth-

nic origins of the 103 individuals who were part of this study, grouped by genotype, are:

1. *DRD4* 4R/4R—24 European (11 unspecified European, 5 Irish, 3 English, 3 German, 1 Greek, and 1 Italian), 20 African (11 Biaka, 3 Chagga, 3 Mbuti, 2 Hausa, and 1 African American), and 7 Asian (5 Han Chinese and 2 Japanese);
2. *DRD4* 7R/7R—19 American (6 Karitiana, 5 Ticuna, 4 Maya, and 4 Surui), 16 European (6 unspecified European, 3 European/Hispanic, 2 Irish, 1 Dane, 1 Druze, 1 English, 1 German, and 1 Italian), 6 African (2 Biaka, 2 Hausa, 1 African American, and 1 Chagga), and 2 Pacific (both Nasioi); and
3. *DRD4* 2R/2R—6 Asian (5 Han Chinese and 1 Yakut) and 3 European (1 Russian and 2 unspecified European).

Of these individuals, 20 (19%) (15 European, 4 Asian, and 1 African) were ADHD probands (Grady et al. 2003), including 1 2R/2R, 14 4R/4R, and 5 7R/7R genotypes.

Primate DNA was obtained from 5 chimpanzees (*Pan troglodytes*), 5 bonobos (*P. paniscus*), and 5 western lowland gorillas (*Gorilla gorilla gorilla*), all unrelated individuals.

PCR Amplification and DNA Sequencing

The entire *DRD4* allelic region was PCR amplified as three overlapping fragments (totaling 6.3 kb), which cover positions 140173 to 146480 in GenBank accession number AC 021663. The current Human Genome Project (HGP) assembly contains a 9-kb unordered fragment containing the *DRD4* locus (from BAC RP11-496I9), but the terminal *DRD4* upstream region of this contig contains 1.9 kb of *Alu* DNA (International Human Genome Sequencing Consortium 2001). Forward and reverse primers for these amplifications were 140173^{F1} (5′-GTGGTCGCAGACATCTTGG-3′), 142075^{R1} (5′-TAGACGAAGAGCGGCAGCA-3′), 142057^{F2} (5′-TGCTGCCGCTCTTCGTCTA-3′), 145072^{R2} (5′-ATGCTGCTGCTCTACTGGG-3′), 144901^{F3} (5′-CCTGCTGTGCTGGACGCCCT-3′), and 146480^{R3} (5′-TAGTCGGAGAAGGTGTCCTG-3′). PCR amplification and the removal of excess primer and dNTP was as described elsewhere (Ding et al. 2002; Grady et al. 2003). Additional primer sequences for forward and reverse sequencing of the *DRD4* amplification products are available on our Web site. DNA cycle sequencing on automated sequencers ABI 3100 and 3700 was conducted as described elsewhere (Riethman et al. 2001; Ding et al. 2002; Grady et al. 2003).

Analysis of Sequence Data

Analysis of sequence data was aided by Phred, Phrap, Polyphred, and Consed (Nickerson et al. 1997). The collection and editing of SNPs into a relational database was done via an in-house software package we designated "SNPMAN" (Grady et al. 2003). Visual displays of SNP data were performed using visual genotyping (VG) (Nickerson et al. 1998).

LD and Population Analyses

Analysis and display of LD was conducted using the GOLD program (Abecasis and Cookson 2000). Tajima's *D* test (Tajima 1989), Fu and Li's *F** test, and Fu and Li's *D** test (Fu and Li 1993) were run using DnaSP v4.0 (Rozas et al. 2003). Our original panel of 103 samples contained a disproportionate number of individuals homozygous for the 2R and 7R alleles. An additional panel of samples, representing a realistic percentage of allelic frequencies found in natural populations (Chang et al. 1996; Ding et al. 2002), was constructed from a random subset of the original 103 samples. The new panel consists of 73 homozygotes (146 chromosomes) with 51, 15, and 7 individuals who were homozygous for 4R, 7R, and 2R, respectively.

Allele Age Calculations

Allele age calculations were conducted by standard methods (Slatkin and Rannala 2000; Ding et al. 2002). In brief: $t = [1/\ln(1-c)]\ln[(x(t)-y)/(1-y)]$, where t = allele age (in generations), c = recombination rate, $x(t)$ = frequency in generation t, and y = frequency on ancestral chromosomes. We assumed the origin of the 7R allele was on a specific 4R haplotype, and calculations utilized the extreme values of c, determined from the telomeric recombination (including 11p) frequencies obtained by Kong et al. (2002) (2 cM/Mb to 4 cM/Mb). For example, the T/C polymorphism at position 140,692 in the *DRD4* consensus sequence is 3,889 bp upstream of the VNTR, and, hence, c values ranging from 0.0000778 to 0.0001556 were used (calculated from the average recombination rate per Mb times the VNTR-SNP distance). In all cases, the frequency on ancestral chromosomes (y) was assumed to be that observed on chromosomes obtained from 4R/4R African individuals. Similar results were obtained with the use of y values from the entire 4R/4R population sample. The frequency of the derived allele, $x(t)$, was the frequency observed in the total population of 7R/7R individuals. For example, the T/C polymorphism at position 140,692 has $y = 5.3\%$ (the percent of the C variant in African 4R/4R individuals) and $x(t) = 84.9\%$ (the percent of the C variant in all 7R/7R individuals). For conversion of time in generations, t, into

time in years, a generation time of 20 or 25 years was assumed (Harpending and Rogers 2000; Ding et al. 2002).

Results

The unusual nature of the sequence architecture of the *DRD4* 7R VNTR, suggesting it arose as a recent rare mutational event (Ding et al. 2002), led us to determine whether differences in LD exist between 4R and 7R chromosomes. We resequenced 6,307 bp of contiguous DNA surrounding the *DRD4* VNTR from 103 individuals (1.5 Mb total) chosen from previous screenings (Chang et al. 1996; Ding et al. 2002; Grady et al. 2003) as homozygous for the VNTR (4R/4R, 7R/7R, and 2R/2R) (fig. 1). *DRD4* loci from 51 4R/4R individuals, 43 7R/7R individuals, and 9 2R/2R individuals were resequenced. 7R/7R and 2R/2R individuals were highly oversampled in comparison to their frequency in the population (Chang et al. 1996; Ding et al. 2002). This approach was developed as a direct and efficient method to estimate the haplotype diversity surrounding the putative ancestral 4R allele, in comparison to the recent 7R allele.

Figure 1 displays the polymorphism distribution of individuals grouped by genotype (4R/4R, 7R/7R, and 2R/2R) and by geographic origin (African, European, etc.). Individuals were intentionally chosen from diverse populations. For example, the African sample consisted of 13 Biaka, 4 Chagga, 4 Hausa, 3 Mbuti, and 2 African American individuals (see "Methods"). Our sample reflects the wide variation in 7R allele distribution (Chang et al. 1996; Ding et al. 2002). For example, 7R/7R individuals of North and South American ancestry are common, but 7R/7R individuals from Asia are rare. In our 4R/4R sample, we intentionally included a large fraction (39%) of individuals of African ancestry, to estimate the "ancestral" frequency of polymorphisms (see below).

Not including VNTR variants (Ding et al. 2002; Grady et al. 2003), a total of 70 SNPs/polymorphisms were detected (on average, one per 90 bp), many at low heterozygosity (fig. 1). As expected, most of these low heterozygosity SNPs were not in current databases (International SNP Map Working Group 2001; International HapMap Consortium 2003). As can be seen in figure 1, the polymorphism spectral distribution of the 4R/4R homozygotes exhibits little LD over the region examined. Traditional measures of LD and other population parameters (below) were not developed for use on such nonrandom samples. Nevertheless, calculations of D' (Abecasis and Cookson 2000) for the 4R/4R homozygotes indicate that few regions have D' values >0.6, which is the average value expected at these close (<6 kb) distances (data not shown; Kruglyak 1999). In addition, 28% (20/70) of the observed SNPs were polymorphic only in African samples. These results are consistent with many studies on other genomic regions (Harpending and Rogers 2000; Ingman et al. 2000; Underhill et al. 2000; Tishkoff and Verrelli 2003), and they likely reflect the "fingerprint" of an out-of-Africa expansion of modern humans and a genetic bottleneck ancestral to European and Asian populations.

In contrast, the evidence for strong LD surrounding the 7R allele is dramatic, with all 7R/7R individuals exhibiting the same alleles at most polymorphic sites (fig. 1). It is especially noteworthy that all 7R/7R African individuals have the same alleles as the non-African individuals. Such striking worldwide allelic homogeneity surrounding a single high-frequency variant (*DRD4* 7R), in comparison to its more common counterpart (*DRD4* 4R), is unprecedented (Stephens et al. 2001; Carlson et al. 2003) (fig. 1). By resequencing this same genomic region in 15 primate genomes, the likely ancestral variant could be determined unambiguously for most SNP pairs. Of the most common variants, 76% (13/17) were inferred to be ancestral in origin (table 1), with one SNP (144,842) having both variants in primate DNAs. Four of the most common variants in the population were "human specific" (table 1). Of the observed polymorphisms in tight LD with the 7R VNTR, 41% (7/17) were the rarer human-specific allele (fig. 1; table 1).

As compared to the 4R–allele-specific low-heterozygosity variants, ~3-fold fewer variants were 7R allele specific (7 vs. 26) (fig. 1). Only one new high-frequency variant was found near the 7R alleles examined, located in the downstream region of the gene (146,033) (marked by an asterisk [*] in fig. 1). The majority of individuals containing this variant allele were of North or South American ancestry (Karitiana, Ticuna, Maya, and Surui), suggesting a possible New World origin. This allele was not found in our African population samples and was at low frequency in our European populations, which included some individuals with partial Hispanic (and likely North/South American) ancestry (fig. 1).

One exception to the strong LD found at the *DRD4* 7R locus is in a small 288-bp region at the promoter (−809 G/A to −521 T/C; fig. 1), where eight high-heterozygosity SNPs are found in comparable, nearly random allelic combinations in both 4R and 7R alleles. Five of these SNPs are clustered in a region of only 95 bp. The −809 to −521 nomenclature (nucleotides 140,989–141,277) is designated with respect to the start of translation (fig. 1), for consistency with prior work (Okuyama et al. 2000). It is difficult to understand this specific breakdown of LD in a single region representing <5% of the locus, unless numerous gene-conversion events have occurred (Ardlie et al. 2001). Similar high-

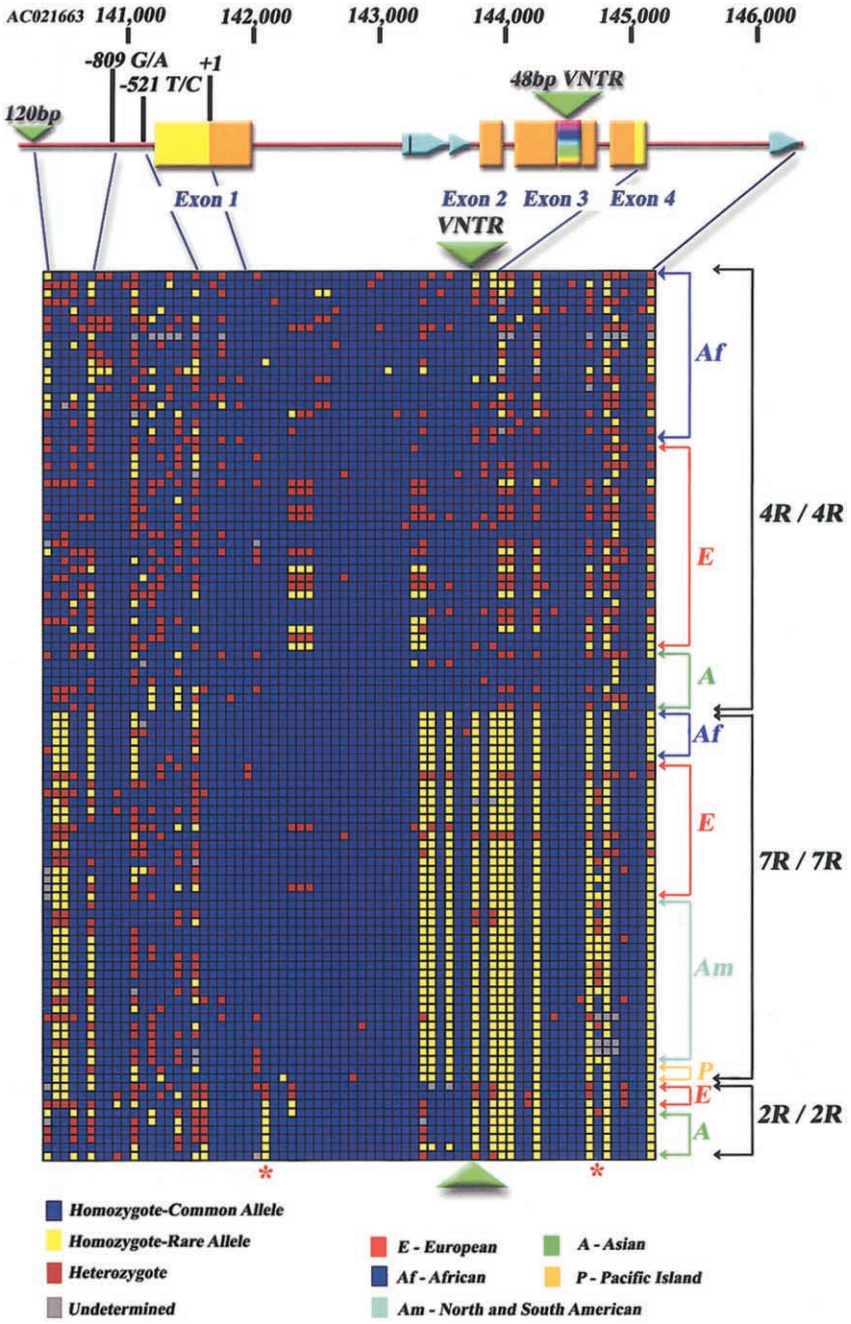

Figure 1 Polymorphism distribution at the *DRD4* locus. Using VG (Nickerson et al. 1998), 70 *DRD4* polymorphisms are displayed, with variants aligned along the horizontal axis. Approximate locations of the variants along the *DRD4* locus (GenBank AC021663) are indicated by blue lines reaching to the diagrammatic representation of the gene. In this representation, exon positions are represented by blocks (yellow = noncoding; orange = coding; +1 = translation start), and the positions of *Alu* repetitive sequences are represented by pointed blue blocks. The positions of a 120-bp upstream duplication and the 48-bp VNTR in exon 3 are indicated by green triangles. A 288-bp site (designated −809 G/A to −521 C/T) at the promoter region that contains an anomalously high number of SNPs is indicated. These SNPs exhibit little difference in 4R vs. 7R frequency. Individuals (on the vertical axis) are grouped by VNTR length (4R/4R, 7R/7R, and 2R/2R) and by geographic origin (African, European, etc.) as indicated. Homozygotes for the allele with the highest relative frequency (common allele) are indicated by blue squares, homozygotes for alternative (rare) alleles by yellow squares, and heterozygotes by red squares. The 7R/7R and 2R/2R individuals were greatly oversampled in comparison to their population frequency; hence, common and rare alleles were defined by the frequencies in a randomly sampled population. SNP and genotype information is available at dbSNP and the authors' Web site.

Table 1

Calculated Allele Age for *DRD4* 7R

Polymorphism	Alleles	Distance To Exon 3 VNTR	SNP Frequency (%) 4R/4R	SNP Frequency (%) 7R/7R	Allele Age (years) Age 1	Allele Age (years) Age 2
140,438	L/S (120 bp)*	−4143	61.9	90.8**	33,361	66,715
140,582	G/del	−3999	95.0	13.9**	19,175	39,557
140,692	T/C	−3889	94.7	15.1**	22,269	44,653
140,892	T/C	−3689	70.0	95.4	22,558	45,119
141,507	C/T	−3074	73.7	97.7	14,884	29,771
142,426	G/A	−2155	90.0	97.7	28,961	57,922
143,578	A/del	−1003	78.4	98.80	28,493	56,989
143,766	C/A*	−815	77.5	4.6	37,539	75,078
143,862	G/del	−719	95.0	2.3**	17,037	34,075
144,842	G/A	261	88.7	1.6	34,865	69,731
144,862	G/C	281	87.9	1.6**	32,686	65,373
145,239	del/G*	658	43.8	2.3	31,637	63,221
145,295	T/C	714	79.0	2.3**	20,690	41,381
145,353	del/G	772	41.7	2.3**	27,615	55,181
145,684	A/C*	1103	69.4	3.5	23,457	46,915
146,041	T/C	1460	42.1	3.8	26,699	53,397
146,056	T/C	1475	57.9	96.3	26,989	53,966
146,293	C/A	1712	44.7	4.6	28,187	56,367
Average					26,505 (SD = 6,112)	53,078 (SD = 12,139)

Note.— High-heterozygosity polymorphisms in the *DRD4* sequence are arranged in upstream to downstream order. The most frequent allele, in most cases the "ancestral" variant determined from primate *DRD4* resequencing, is listed first. Four alleles noted with an asterisk (*) are the less common yet ancestral variant. The frequency of the common variant in African 4R/4R individuals and all 7R/7R individuals is given. All values were obtained from the data displayed in fig. 1, except for polymorphisms 140,438; 144,842; and 144,862, which were obtained from a much larger sample set (>2,000 individuals) (Chang et al. 1996; Ding et al. 2002; Grady et al. 2003). Two asterisks (**) indicate "human-specific" alleles tightly linked to the 7R allele. The low–high range for allele age was calculated from the extreme values of telomeric recombination reported in Kong et al. (2002) (Age 1 = 4 cM/Mb and Age 2 = 2 cM/Mb), by use of standard methods (Slatkin and Rannala 2000; Ding et al. 2002). The average value obtained from all polymorphisms is ~40,000 years (the average of Age 1 and Age 2), with maximum likely limits of 20,000–65,000 years. The assumption of a generation time of 25 years, rather than 20 years, increases the average age to ~50,000 years, with maximum likely limits of 25,000–81,000 years.

frequency variations are found in this region in the limited primate samples examined, including more extensive deletions in chimp and gorilla (data not shown). These highly variable SNPs are not included in table 1, their ancestral origin cannot be determined, and they cannot be used in allele-age determinations. Regardless of the mechanism of homogenization at this small *DRD4* promoter region, the strong LD observed at the 7R allele continues upstream of this region, with four high-heterozygosity SNPs/polymorphisms in strong association with the VNTR (fig. 1 and table 1). Genotyping of other VNTRs associated with three *DRD4*-adjacent loci (*PTDSS2, HRAS,* and *SCT*) (Riethman et al. 2001) indicates that the region of strong LD surrounding the *DRD4* 7R allele extends for at least 50–100 kb (data not shown).

The resequencing of a sample of 2R/2R homozygotes (fig. 1) confirmed our proposal (Ding et al. 2002) that the 2R allele is a recombination product between a 4R and a 7R allele. The 2R VNTR downstream region contains a polymorphism pattern identical to that found in 7R alleles, whereas the VNTR upstream region is similar to 4R alleles. The upstream variability suggests that there were multiple origins for this proposed recombination (fig. 1). Most of the examined 2R alleles, however, contain a variant in the first intron (142,115; marked by an asterisk [*] in fig. 1) found only in one 4R individual of African ancestry, suggesting a common origin and expansion for these 2R alleles.

Under the assumption of random assortment of variants, the probability that these radically different LD patterns (fig. 1) could have occurred by chance is infinitesimally small, requiring multiple low-probability events (10^{-3}–10^{-43}). Indeed, it would be difficult to construct alleles that differ more. Calculations of allele age made on the basis of the relatively high worldwide population frequency of the *DRD4* 2R, 4R, and 7R alleles suggest that these alleles are ancient (>300,000–500,000 years old) (Ding et al. 2002). On the other hand, calculations of allele age on the basis of the ob-

served intra-allelic variability suggest that the 7R allele is 10-fold "younger" (table 1). Although it is useful (and common) to refer to the "allele age" calculated from intra-allelic variability, it should be noted that such calculations more accurately estimate the most recent common ancestor (Slatkin and Rannala 2000). Such large discrepancies between the allele ages calculated by these two methods are usually taken as evidence that strong selection has increased the frequency of the allele to higher levels than expected by random genetic drift (Slatkin and Rannala 2000; Ding et al. 2002). The absolute values of these estimates are greatly affected by the assumptions on which they base their computations—for example, the assumed recombination-frequency. For the calculations in table 1, we have used the extreme values of estimates of the recombination frequency observed for the telomeric regions of human chromosomes (including 11p) (Kong et al. 2002). All 18 high-heterozygosity *DRD4* SNPs used to estimate allele age yield comparable results (table 1). This suggests that the average of these values (40,000–50,000 years) is a reasonable current estimate of the most recent common ancestor for *DRD4* 7R. The use of the extremes of assumed recombination frequency (Age 1 and Age 2) and of generation time, ± SD, yields an estimate of 20,000–81,000 years (table 1) for the limits of allele age. The proposed origin of the 2R allele as a 7R allele derivative (fig. 1) indicates that it also must be a young allele. The discrepancy between the observed high frequency of the 2R allele (especially in Asian populations) and its recent 7R origin (fig. 1) suggests that it, too, has increased in frequency by positive selection.

Although these observed LD differences are the best indicator of selection at this locus, other statistical tests of selection also confirm this hypothesis. These ad hoc tests condense information into a summary statistic and therefore lose information present in LD displays (fig. 1); nevertheless, they have been widely applied in prior work and are presented for comparison (Kreitman 2000). One of the most commonly applied tests, Tajima's *D*, compares the difference between two estimators of neutrality: (1) the number of segregating sites in a sample and (2) the average number of nucleotide differences between two sequences (Tajima 1989). Tajima's *D* applied separately to 4R/4R and 2R/2R homozygote data did not reach statistical significance, whereas 7R/7R homozygote data yielded a statistically significant negative value (-2.117, $P < .05$), usually interpreted as an indicator of selection (Kreitman 2000). Tajima's *D* applied to all 103 individuals yielded a statistically significant value indicating selection (-1.827, $P < .05$). The application of Fu and Li's D^* and of Fu and Li's F^* (Fu and Li 1993), other widely used tests for selection, to the total population sample also yielded statistically significant indications of selection (-2.958, $P < .05$ and -2.975, $P < .02$, respectively). As discussed above, the application of these tests to nonrandom data sets, such as ours, may be inappropriate. However, calculating the same statistics on a modified data set (see "Material and Methods"), in which the frequency of 2R, 4R, and 7R alleles was comparable to that expected from random population sampling (Ding et al. 2002), still yielded statistically significant signs of selection ($P < .05$).

We suggested elsewhere (Ding et al. 2002) that the pattern of amino-acid changes at the *DRD4* locus (K_a/K_s ratio >1) and the unusual distribution of VNTR variants was consistent with a balanced selection, rather than a purely directional selection of 7R variants. Though it is common to distinguish between these two modes of selection (directional vs. balanced), in actuality, all balanced selection systems, unless they are ancient, must also have the "fingerprint" of a directional component (i.e., one of the alleles must be a "younger" variant). Plots of site-frequency spectra can help distinguish between directional and balanced selection (Harpending and Rogers 2000), by comparison of the actual distribution-frequency of variant sites to the expected values, under a model of neutral evolution. Indeed, the above-mentioned statistical tests are measures of departures from equilibrium-neutral expectations in the direction either of too many low-frequency polymorphisms or of too many intermediate-frequency polymorphisms (Kreitman 2000). In a simple directional selection, an advantageous allele, once it arises, rapidly increases in frequency. If we observe this process at a random time, we are likely to find one allele at a low frequency and the other at a high frequency. This process increases the fraction of singletons in a sample of sites (Harpending and Rogers 2000). Balanced selection, on the other hand, maintains two or more alleles in the population at polymorphic frequencies, and the observed result is an excess of alleles with frequencies ~0.5 and a deficit of singletons. Plots of site frequency spectra for the data in figure 1 (or a modified random dataset, as described above) show ~50% (0.554) reduction of singletons and a 3.85-fold increase in variants >0.25 in frequency as compared to expected frequencies (data not shown), consistent with the expectations of a balanced selection system.

The data in figure 1 and table 1 can also be used to test whether the *DRD4* VNTR itself, rather than an adjacent SNP, is the target of selection. Ideally, one should observe an increase in historical recombination (and lower LD), as distance from the selected polymorphism is increased. This result is predicted because random recombination events, which separate adjacent variants from the selected allele, are counterbalanced by the selection itself. There are few ways to produce a recombinant chromosome near the selected allele and

still retain the allele (Fay and Wu 2000). Figure 2 plots distance from the *DRD4* 7R VNTR versus percent inferred recombination (from table 1). The observed association with adjacent variants is greatest near the VNTR and decreases with distance in both directions. Grouping this same population sample by homozygosity at any adjacent SNP resulted in largely random association patterns, in comparison to the 4R/7R division (data not shown). For example, dividing the sample on the basis of homozygosity for the 140,892 T/C SNP at position −3689, rather than the 4R/7R VNTR, yielded minimum association with adjacent variants in both T/T and C/C homozygotes. Further, none of these other genotype groupings yielded recombinant-chromosome frequency plots indicating statistically significant selection at the SNP, unlike the 4R/7R division (fig. 2). Although the observed fraction of recombinant chromosomes is low and there is significant scatter in the data (fig. 2), these results support the hypothesis that the *DRD4* 7R VNTR is the target of selection.

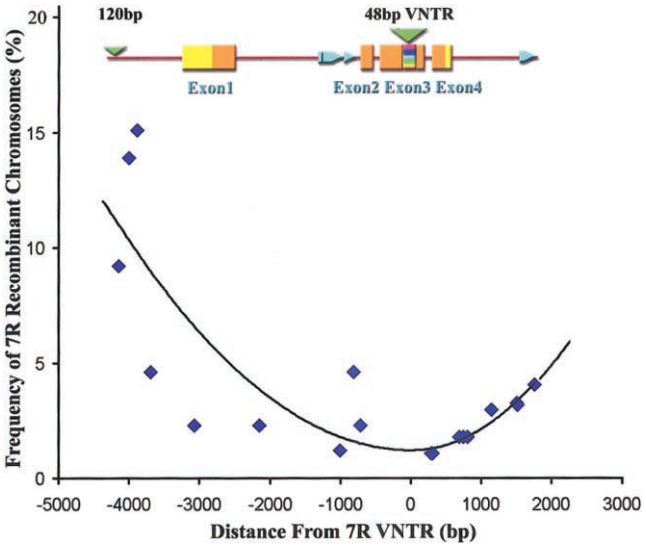

Figure 2 Frequencies of *DRD4* 7R recombinant chromosomes. The graph shows the observed percentage of recombinant chromosomes at the 18 SNPs (table 1) versus the distance from the 7R VNTR. The curve is an empirically determined least-squares fit to the data. The diagrammatic representation of the *DRD4* locus is as described in fig. 1.

Discussion

In this study, we have expanded our LD analysis of the *DRD4* gene by resequencing the entire locus in 2R, 4R, and 7R homozygous individuals. This method was chosen as an accurate and efficient approach to determine the comparative LD of two alleles, requiring little statistical manipulation to infer haplotype differences. By use of this approach (fig. 1), the pattern of LD surrounding the *DRD4* 4R allele is the pattern expected for an ancient gene locus (300,000–500,000 years old), in which haplotype diversity is greatest in African populations and more restricted outside Africa (Slatkin and Rannala 2000; Tishkoff and Verrelli 2003).

In contrast, the evidence for strong LD surrounding the 7R allele is dramatic (fig. 1). Such worldwide LD for a single human allele is remarkable. For example, in one of the best-characterized examples of selection in humans, the frequencies of low activity alleles of glucose-6-phosphate dehydrogenase are highly correlated with the prevalence of malaria, yet many regional variants have been selected for (Tishkoff et al. 2001; Sabeti et al. 2002). There is no worldwide "malaria resistant" variant, presumably because the introduction of agriculture 10,000 years ago (and the *Plasmodium* parasite) selected for independent regional mutations (Tishkoff et al. 2001; Sabeti et al. 2002). By intra-allelic comparison at 18 high-heterozygosity sites, we can estimate that the *DRD4* 7R allele arose prior to the upper Paleolithic era (~40,000–50,000 years ago, with likely limits of 20,000–81,000 years ago) (table 1).

We emphasize that this estimate is only as reliable as the assumptions on which its calculation is based, and further studies will likely refine this number. However, allele-age estimates calculated by alternative methods (for example, by assuming a fixed mutation rate on Y-chromosome DNA; see Ingman et al. [2000], Slatkin and Rannala [2000], and Underhill et al. [2000]) are also subject to considerable uncertainty. Hence, all such estimates should be considered provisional. In particular, it should be noted that no theory has been developed to predict the average age of a selected allele in a growing population and that strong selection may alter the calculated age in such a population. Current estimates suggest, however, that such an effect is not dramatic (Slatkin and Rannala 2000). Regardless of the absolute origin-date for the *DRD4* 7R allele, the unusual worldwide LD observed for this allele (fig. 1) suggests either that it predates the last major out-of-Africa human expansion or that there has been significant backflow of this allele from a non-African origin into Africa. Evidence for such a backflow has been presented for other loci (Cruciani et al. 2002). Further population sampling may resolve these possibilities; however, a pre-out-of-Africa 7R origin seems more likely.

Elsewhere, we have argued that population bottlenecks and local admixture cannot explain the observed bias toward nonsynonymous amino-acid changes, the unusual VNTR structure, and the high LD surrounding the *DRD4* locus (Ding et al. 2002). Further, proposing unusual genetic mechanisms that reduce recombination on 7R chromosomes relative to 4R chromosomes, to explain the strong 7R LD, seems equally unlikely. For

example, chromosome inversions suppress or eliminate local recombination. However, the 7R allele recombines with other alleles at comparable (or greater) frequency, as inferred from haplotype analysis (Ding et al. 2002; Grady et al. 2003). Further, inversions should not alter the frequency of new mutations in the inverted region, but 7R alleles have ~3-fold fewer rare variants adjacent to the VNTR in comparison with 4R alleles (7 vs. 26) (fig. 1). The simplest explanation, then, for this worldwide LD is the most straightforward: the 7R allele was strongly selected for at about the time of the last major out-of-Africa exodus, currently estimated at 44,000–47,000 years ago (with 95% probability intervals of 35,000–89,000 years) (Harpending and Rogers 2000; Ingman et al. 2000; Underhill et al. 2000).

How did the 7R variant arise, and what DNA sequence was selected for? The alleles on 7R chromosomes at 41% of adjacent SNPs are human specific (table 1). This argues for the derivation of this variant by mutation from the common human 4R allele, rather than importation from a related hominid lineage, an alternative we proposed previously (Ding et al. 2002). If importation (and hence a long coalescence time) is unlikely, what can account for the multiple changes necessary to convert a 4R allele into a 7R allele? It is possible that the 7R allele arose as a multistep mutational process over long evolutionary time and was selected for only during the last 40,000–50,000 years (our measured coalescence time). This model proposes that proto-7R alleles were, at best, selectively neutral and that the random LD that was present surrounding these ancient alleles was eliminated 40,000–50,000 years ago by coalescence at a single allele. Alternatively, the documented instability of tandem repetitive DNA could have generated the 7R variant in a single step, immediately prior to selection. We have shown that unequal recombination at this site is common and that single recombination events can generate multiple point mutations, most likely by imperfect DNA repair (Ding et al. 2002; Grady et al. 2003). The origin of this allele may never be known with certainty. More importantly, regardless of its origin, we suggested that it is unlikely that selection for an adjacent variant can account for the proposed selection, given the distinct and unusual DNA sequence of the *DRD4* 7R VNTR itself (Ding et al. 2002). We have now shown that the observed association with adjacent SNPs is likely centered on the VNTR in 7R alleles (fig. 2), consistent with the hypothesis that it is the target of selection. Strong LD with the *DRD4* 7R allele can be detected $\geqslant$50–100 kb from the VNTR (near the *PTDSS2, HRAS,* and *SCT* loci; data not shown). However, since the current HGP assembly in this subtelomeric region contains many gaps and ambiguous contig orders (Riethman et al. 2001), it is impossible at present to refine these LD studies. Further work to define the limits of LD for this locus will help clarify both the estimates of allele age and the evidence for VNTR selection.

The breakdown of this strong LD, at a small region representing <5% of the 7R locus, is surprising (fig. 1). One can only speculate as to what mechanisms might be involved. It is especially intriguing that this homogenization occurs at the promoter. It has been reported that the −521 polymorphism in this region affects transcriptional efficiency (Okuyama et al. 2000). Further, some studies have reported an association of this same polymorphism with novelty-seeking behavior (Okuyama et al. 2000; Ronai et al. 2001). Given that this region (including the −521 polymorphism) is not in strong LD with the 7R VNTR (fig. 1), it is unlikely these association results are due to 7R linkage. The CpG frequency at this site is not significantly higher than the remainder of this GC-rich gene; hence, a higher mutation rate seems unlikely. Further, high frequency variations are found at this site in primate DNA, suggesting that this region is a hotspot for such changes. We suggest, therefore, that high frequency gene conversion might explain this homogenization. Small hotspots for gene conversion have been proposed to exist at many loci in the human genome (Ardlie et al. 2001). The overall strong LD associated with the 7R allele continues upstream of this anomalous region (fig. 1). It is possible that the association of a polymorphism, in this upstream region, with ADHD (McCracken et al. 2000) reflects this strong LD with the 7R VNTR. The small sample of ADHD probands (20 individuals), included in the current analysis (fig. 1), did not exhibit any variations in adjacent polymorphisms different from the remaining 2R/2R, 4R/4R, and 7R/7R individuals.

A Model

Extensive biochemical analyses of DRD4 protein variants have been conducted (Asghari et al. 1995; Jovanovic et al. 1999; Oak et al. 2000). The 7R protein has a blunted response for cAMP reduction, requiring a three-fold increase in dopamine concentration for reductions comparable to the 4R protein (Asghari et al. 1995). This "suboptimal" response of the 7R allele to dopamine was hypothesized to underlie its association with the personality trait of novelty seeking (Klugar et al. 2002) and with ADHD (Swanson et al. 2001). It was suggested that the inhibitory neurons utilizing the DRD4 7R receptor would require increased dopamine for "normal" function (Swanson et al. 2000). Such increased dopamine levels were hypothesized to be the result of risk-taking behavior (in the case of novelty seeking) or methylphenidate (in the case of ADHD). Methylphenidate is thought to act by binding to the dopamine trans-

porter and raising the levels of dopamine at the synapse (Swanson et al. 2001; Volkow and Swanson 2003).

We propose a simple model integrating the known biochemical, physiological, and genetic data regarding the common *DRD4* alleles (fig. 3). The 4R allele appears to be the most common allele throughout most of human prehistory. This ancestral allele has the fewest amino–acid-changing variants, implying strong purifying selection (Ding et al. 2002). The 7R allele arose as a rare mutation that significantly blunted the receptor's response to dopamine (fig. 3). We hypothesize, however, that this blunted response led to behaviors that were selected for in certain environments, and the 7R allele increased in frequency (starting ~40,000–50,000 years ago). Rather than spreading in a purely directional pattern, the 7R allele, we have proposed, exists as a balanced polymorphism (Ding et al. 2002). Evidence for this model includes: (1) the less likely probability that a purely directional selection will be observed in the process of spreading (Harpending and Rogers 2000; Ding et al. 2002); (2) the high worldwide allele frequencies of 2R, 4R, and 7R alleles, suggesting selection for all alleles (Chang et al. 1996; Ding et al. 2002); (3) the strong nonhomologous amino-acid replacement (K_a/K_s ratio >1) observed for all alleles (2R, 4R, and 7R) (Ding et al. 2002); (4) the unusual geographic distribution of 2R and 7R alleles, not consistent with strong directional selection but consistent with "deme" variation (Chang et al. 1996; Kreitman 2000; Ding et al. 2002); and (5) a site frequency distribution of adjacent variants not consistent with directional selection but consistent with balanced selection (i.e., a deficit of singletons and an excess of intermediate frequency variants; see fig. 1). Such a proposed multiallelic adaptive strategy is likely common (Sinervo and Lively 1996; Harpending and Rogers 2000), and is predicted by evolutionary game theory (Smith 1982).

The genetic data suggest that most 2R alleles are 7R derivatives and likely had limited (yet multiple) origins (fig. 3). Interestingly, the 2R variant also has a blunted cAMP response, but one that is midway between those of the 4R allele and 7R allele (fig. 3). We hypothesize that individuals with 2R alleles exhibit behaviors "intermediate" between those manifested by individuals with 4R and 7R alleles. This "nonlinear" response (i.e., cAMP reduction capability is not linearly related to *DRD4* VNTR-repeat length) (Asghari et al. 1995; Jovanovic et al. 1999) is consistent with the genetic evidence and suggests a typical biochemical "optimum" strategy (fig. 3). We have suggested that this biochemical/genetic data implies that characterizing individuals by *DRD4* 4R/4R versus non-4R genotype may be more appropriate for testing various gene/behavior associations (Grady et al. 2003).

What could be the behavioral differences that are selected for? By observing current genetically influenced differences in human personality (summarized in Bouchard and Loehlin [2001]), it has been suggested that resource-depleted, time-critical, or rapidly changing environments might select for individuals with "response ready" adaptations, whereas resource-rich, time-optimal, or little-changing environments might select against such adaptations (Jensen et al. 1997). We have speculated that such a "response ready" adaptation might have played a role in the out-of-Africa exodus and that allele frequencies of genes associated with such behavior certainly would be influenced, subsequently, by the local cultural milieu (Ding et al. 2002; Harpending and Cochran 2002). For example, the wide variation in allele frequency in different geographic locations (Chang et al. 1996) is expected for balanced polymorphisms, where individual demes have different equilibrium states based on local conditions (Kreitman 2000). Consistent with this "response ready" behavior hypothesis is the significantly better performance of *DRD4* knockout mice on tests of complex coordination (Rubinstein et al. 1997) and the observed faster reaction times exhibited by individuals with ADHD who have a 7R allele, in comparison to non-7R individuals (Swanson et al. 2000; Langley et al. 2004). Counterbalancing

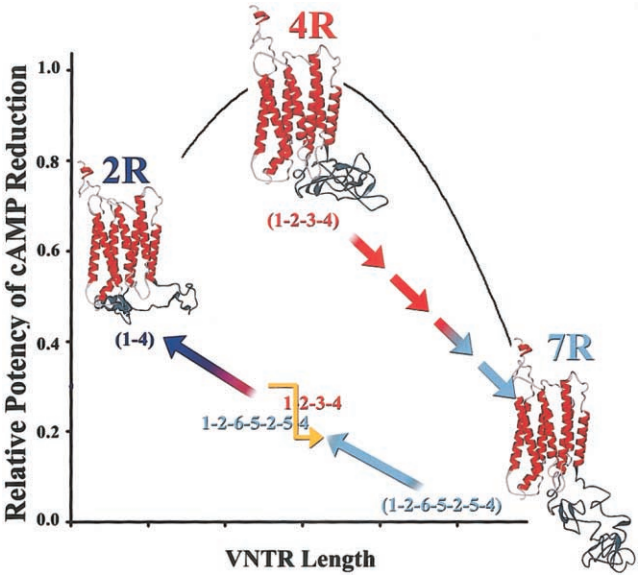

Figure 3 A diagrammatic model for *DRD4* variant selection. DRD4 2R, 4R, and 7R protein variants are shown diagrammatically, aligned on a scale of relative efficiency for cAMP reduction. These values were calculated from the data of Asghari et al. (1995), normalized to $4R = 1.0$. Haplotype nomenclature (i.e., 1-2-3-4) appears as proposed elsewhere (Ding et al. 2002). The unusual derivation of the 7R allele from the ancestral 4R allele (~40,000–50,000 years ago) and its increase in prevalence are indicated by red to turquoise arrows. The subsequent derivation of the 2R allele from a 7R/4R recombination is indicated by turquoise to blue arrows.

the presumed benefits of faster reaction times may be the increased risk of cortical hyperexcitability (including convulsions) exhibited by *DRD4* knockout mice (Rubinstein et al. 2001). It is possible that the trade-offs between faster reaction time and hyperexcitability are influencing the hypothesized balanced selection at this locus.

In this model, the 4R variant has been honed for hundreds of thousands of years to function optimally, whereas the new 7R and 2R variants are suboptimal yet confer a behavioral advantage in some environments. Though the "response ready" hypothesis was proposed as an environmental adaptation, sexual selection has long been proposed as another source of human variation (Darwin 1871). For example, there is ample evidence that polygyny has been a historical norm in most human societies and that some degree of risk-taking behavior is an expected path to reproductive success (for reviews, see Betzig [1993], Buss [1999], and Fehr and Fischbacher [2003]). We suggest that sexual selection, influenced by the local cultural norms and by the response to behaviors affected by these VNTR variants, has contributed to the unusual *DRD4* allelic distribution.

In particular, we suggest that this model helps explain the unusual geographic distribution of the 7R allele, which is at low frequency in Asian populations but at high frequency in the Americas (Chang et al. 1996; Ding et al. 2002). The 2R and 7R alleles are genetically and functionally related, each exhibiting suboptimal dopamine-signaling, in comparison with that of 4R alleles (fig. 3). One possibility is that the 7R allele never existed at high frequency in Asian populations, but its "role" in the hypothesized balanced-selection system was filled by the "comparable" 2R allele. This hypothesis implies that there are no major selective differences acting on this locus in Asian populations, that it is merely a functionally irrelevant deme variation. Consistent with this hypothesis is our recent study indicating that the 2R allele frequency in Chinese ADHD probands is increased by an amount comparable to the 7R allele increase observed in European-ancestry ADHD probands (Leung et al. 2004). Alternatively, the frequency of the 7R allele in Asian populations might at one time have been higher. In this hypothesis, after the initial migration into Asia, the 7R frequency was reduced while the derivative 2R allele frequency increased, implying the two alleles may not be "functionally" equivalent. One can only speculate as to what environmental/cultural factors could have influenced such a replacement. Most Asian cultures, however, have a much longer recorded history of effective societal reproductive intervention/selection than those of European and African cultures (Betzig 1993), and hence are more likely to have influenced gene frequency by such a mechanism.

In either hypothesis, the severe bottlenecks North and South American ancestry populations underwent during their migration from Asia can explain adequately their high incidence of 7R alleles. Either the proposed strong selection for the 7R allele in certain "response-ready" environments (above) or random chance could influence allele frequency in such small migrating populations.

Although this speculative model (fig. 3) is consistent with available genetic, biochemical, and physiological data, only further work can test, refine, and modify these ideas. Clearly, further genetic, biochemical, and behavioral studies are needed. The evidence for selection acting at the *DRD4* locus is strong, however (figs. 1 and 2), and challenges us to determine the specific mechanism driving it. Regardless of the ultimate details, is it reasonable to think that a single gene variation can modify human behavior yet be shaped by cultural diversity? We argue that just such single-gene changes regulating complex social behavior have been identified in other social organisms (Krieger and Ross 2002). We see no reason to think humans should be exempt from similar Darwinian selection (Darwin 1871). The evolutionary payoff of an individual's behavior in complex human societies, obviously, will depend strongly on the reaction of others to that behavior (Betzig 1993; Buss 1999; Ding et al. 2002; Harpending and Cochran 2002; Fehr and Fischbacher 2003). We suggest the exciting possibility that the *DRD4* locus is a prime candidate for investigating such gene and culture interactions.

Acknowledgments

This work was supported by grant MH60660 (to R.K.M.) from the National Institute of Mental Health.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Authors' Web site, http://www.genome.uci.edu(for genotype information)

dbSNP, http://www.ncbi.nlm.nih.gov/SNP/ (for DRD4-140582 [accession number 20399264]; DRD4-140692 [accession number 20399265]; DRD4-140892 [accession number 20399266]; DRD4-140926 [accession number 20399267]; DRD4-140989 [accession number 20399268]; DRD4-141034 [accession number 20399269]; DRD4-141044 [accession number 20399270]; DRD4-141085 [accession number 20399271]; DRD4-141102 [accession number 20399272]; DRD4-141182 [accession number 20399273]; DRD4-141183 [accession number 20399274]; DRD4-141195 [accession number 20399275]; DRD4-141204 [accession number 20399276]; DRD4-141198 [accession number 20399277]; DRD4-141203 [accession number 20399278]; DRD4-141270 [accession number 20399279]; DRD4-141277 [accession number 20399280]; DRD4-

141422 [accession number 20399281]; DRD4-141434 [accession number 20399282]; DRD4-141507 [accession number 20399283]; DRD4-141745 [accession number 20399284]; DRD4-141784 [accession number 20399285]; DRD4-141828 [accession number 20399286]; DRD4-142115 [accession number 20399287]; DRD4-142203 accession number 20399288]; DRD4-142347 [accession number 20399289]; DRD4-142426 [accession number 20399290]; DRD4-142494 [accession number 20399291]; DRD4-142495 [accession number 20399292]; DRD4-142496 [accession number 20399293]; DRD4-142497 [accession number 20399294]; DRD4-142732 [accession number 20399295]; DRD4-142775 [accession number 20399296]; DRD4-142940 [accession number 203992697]; DRD4-142976 [accession number 20399298]; DRD4-143118 [accession number 20399299]; DRD4-143318 [accession number 20399300]; DRD4-143348 [accession number 20399301]; DRD4-143578 [accession number 20399302]; DRD4-143766 [accession number 20399303]; DRD4-143862 [accession number 20399304]; DRD4-143870 [accession number 20399305]; DRD4-143867 [accession number 20399306]; DRD4-143058 [accession number 20399307]; DRD4-143060 [accession number 20399308]; DRD4-144054 [accession number 20399309]; DRD4-144746 [accession number 20399310]; DRD4-144842 [accession number 20399311]; DRD4-144843 [accession number 20399312]; DRD4-144862 [accession number 20399313]; DRD4-145239 [accession number 20399314]; DRD4-145295 [accession number 20399315]; DRD4-145298 [accession number 20399316]; DRD4-145334 [accession number 20399317]; DRD4-145353 [accession number 20399318]; DRD4-145365 [accession number 20399319]; DRD4-145417 [accession number 20399320]; DRD4-145430 [accession number 20399321]; DRD4-145568 [accession number 20399322]; DRD4-145651 [accession number 20399323]; DRD4-145684 [accession number 20399324]; DRD4-146033 [accession number 20399325]; DRD4-146041 [accession number 20399326]; DRD4-146056 [accession number 20399327]; DRD4-146140 [accession number 20399328]; DRD4-146158 [accession number 20399329]; DRD4-146289 [accession number 20399330]; and DRD4-146293 [accession number 20399331]; dbSNP handle: rmoyzis)

References

Abecasis GR, Cookson WO (2000) GOLD—graphical overview of linkage disequilibrium. Bioinformatics 16:182–183

Ardlie K, Liu-Cordero SN, Eberle MA, Daly M, Barrett J, Winchester E, Lander ES, Kruglyak L (2001) Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. Am J Hum Genet 69:582–589

Asghari V, Sanyal S, Buchwaldt S, Paterson A, Jovanovic V, Van Tol HH (1995) Modulation of intercellular cyclic AMP levels by different human dopamine D4 receptor variants. J Neurochem 65:1157–1165

Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111

Betzig L (1993) Sex, succession and stratification in the first six civilizations. In: Ellis L (ed) Social stratification and socioeconomic inequality. Praeger, Westport, CT, pp 37–74

Bouchard TJ, Loehlin JC (2001) Genes, evolution, and personality. Behav Genet 31:243–273

Buss DM (1999) Evolutionary psychology: the new science of mind. Allyn and Bacon, Needham Heights, MA

Carlson CS, Eberle MA, Reider MJ, Smith JD, Kruglyak L, Nickerson DA (2003) Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. Nat Genet 33:518–521

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

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-Sforza LL, Scozzari R, Underhill PA (2002) A back migration from Asia to sub-Sahara Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am J Hum Genet 70: 1197–1214

Darwin C (1871) The descent of man and selection in relation to sex. London, J. Murry

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Moyzis RK (2002) Evidence of positive selection acting at the human dopamine receptor D4 gene locus. Proc Natl Acad Sci USA 99:309–314

Dulawa SC, Grandy DK, Low MJ, Paulas MP, and Geyer MA (1999) Dopamine D4 receptor-knock-out mice exhibit reduced exploration of novel stimuli. J Neurosci 19:9550–9556

Faraone SV, Doyle AE, Mick E, Biederman J (2001) Meta-analysis of the association between the 7-repeat allele of the dopamine D4 receptor gene and attention deficit hyperactivity disorder. Am J Psychiatry 158:1052–1057

Fay JC, Wu CI (2000) Hitchhiking under positive selection. Genetics 155:1405–1413

Fehr E, Fischbacher U (2003) The nature of human altruism. Nature 424:785–791

Fu YX, Li WH (1993). Statistical tests of neutrality of mutations. Genetics 133:693–709

Grady DL, Chi HC, Ding YC, Smith M, Wang E, Schuck S, Flodman P, Spence MA, Swanson JM, Moyzis RK (2003) High prevalence of rare dopamine receptor D4 alleles in children diagnosed with attention-deficit hyperactivity disorder. Mol Psychiatry 8:536–545

Harpending H, Rogers A (2000) Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet 1:361–385

Harpending H, Cochran G (2002) In our genes. Proc Natl Acad Sci USA 99:10–12

Hartman T (1993) Attention deficit disorder: a different perception. Underwood Books

Ingman M, Kaessmann H, Paabo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. Nature 408:708–713

International HapMap Consortium (2003) The international HapMap project. Nature 426:789–796

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933

Jensen PS, Mrazek D, Knapp PK, Steinberg L, Pfeffer C, Schowalter J, Shapiro T (1997) Evolution and revolution in child psychiatry: ADHD as a disorder of adaptation. J Am Acad Child Adolesc Psychiatry 36:1672–1679

Jovanovic V, Guan HC, Van Tol HH (1999) Comparative pharmacological and functional analysis of the human dopamine D4.2 and D4.10 receptor variants. Pharmacogenetics 9:561–568

Klugar AN, Siegfried Z, Ebstein RP (2002) A meta-analysis of the association between DRD4 polymorphism and novelty seeking. Mol Psychiatry 7:712–717

Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K (2002) A high-resolution recombination map of the human genome. Nat Genet 31:241–247

Krieger MJ, Ross KG (2002) Identification of a major gene regulating complex social behavior. Science 295:328–332

Kreitman M (2000) Methods to detect selection in populations with applications to the human. Annu Rev Genomics Hum Genet 1:539-559

Kruglyak L (1999) Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nat Genet 22:139–144

Langley K, Marshall L, van den Bree M, Thomas H, Dphil O, O'Donovan M, Thapar A (2004) ADHD children with and without the dopamine D4 receptor 7-repeat allele: evidence of differences in performance on neuropsychological tests. Am J Psychiatry 161:133-138

Lichter JB, Barr CL, Kennedy JL, Van Tol HH, Kidd KK, Livak KJ (1993) A hypervariable segment in the human dopamine receptor D4 (DRD4) gene. Hum Mol Genet 2:767–773

Leung PWL, Lee CC, Hung SF, Ho TP, Tang CP, Kwong SL, Leung SY, Yuen ST, Lieh-Mak F, Oosterlaan J, Grady D, Harxhi A, Ding YC, Chi HC, Flodman P, Schuck S, Spence MA, Moyzis R, Swanson J. The dopamine receptor D4 (*DRD4*) gene in Han Chinese children with attention-deficit/hyperactivity disorder (ADHD): increased prevalence of the 2-repeat allele. Am J Med Gen (in press)

McCracken JT, Smalley SL, McGough JJ, Crawford L, Del'Homme M, Cantor RM, Liu A, Nelson SF (2000) Evidence for linkage of a tandem duplication polymorphism upstream of the dopamine D4 receptor gene (DRD4) with attention deficit hyperactivity disorder (ADHD). Mol Psychiatry 5:531–536

Nickerson DA, Tobe VO, Tayler SL (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 14:2745–2751

Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Steingard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240

Oak JN, Oldenhof J, Van Tol HHM (2000) The dopamine D4 receptor: one decade of research. Eur J Pharmacol 404:303–327

Okuyama Y, Ishiguro H, Nankai M, Shibuya H, Watanabe A, Arinami T (2000) Identification of a polymorphism in the promoter region of DRD4 associated with the human novelty seeking personality trait. Mol Psychiatry 5:64–69

Riethman HC, Xiang Z, Paul S, Morse E, Hu XL, Flint J, Chi HC, Grady DL, Moyzis RK (2001) Integration of telomeric sequences with the draft human genome sequence. Nature 409:948–951

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Ronai Z, Szekely A, Nemoda Z, Lakatos K, Gervai J, Staub M, Sasvarri-Szekely (2001) Association between novelty seeking and the −521 C/T polymorphism in the promoter region of the DRD4 gene. Mol Psychiatry 6:35–38

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Rubinstein M, Phillips TJ, Bunzow JR, Falzone TL, Dziewczapolski G, Zhang G, Fang Y, Larsen JL, McDougall JA, Chester JA, Saez C, Pugsley TA, Gershanik O, Low MJ, Grandy DK (1997) Mice lacking dopamine D4 receptors are supersensitive to ethanol, cocaine, and methamphetamine. Cell 90:991–1001

Rubinstein M, Cepeda C, Hurst RS, Flores-Hernandez J, Ariano MA, Falzone TL, Kozell LB, Meshul CK, Bunzow JR, Low MJ, Levine MS, Grandy DK (2001) Dopamine D4 receptor-deficient mice display cortical hyperexcitability. J Neurosci 21:3756–3763

Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837

Sinervo B, Lively CM (1996) The rock-scissors-paper game and the evolution of alternative male strategies. Nature 380: 240–243

Slatkin M, Rannala B (2000) Estimating allele age. Annu Rev Genomics Hum Genet 1:225–249

Smith JM (1982) Evolution and the theory of games. Cambridge University Press, Cambridge, United Kingdom

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R et al. (2001) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293: 489–493

Swanson JM, Oosterlaan J, Murias M, Schuck S, Flodman P, Spence MA, Wasdell M, Ding YC, Chi HC, Smith M, Mann M, Carlson C, Kennedy JL, Sergeant JA, Leung P, Zhang YP, Sadeh A, Chen C, Whalen CK, Babb KA, Moyzis R, Posner MI (2000) Attention deficit/hyperactivity disorder children with a 7-repeat allele of the dopamine receptor D4 gene have extreme behavior but normal performance on critical neuropsychological tests of attention. Proc Natl Acad Sci USA 97:4754–4759

Swanson J, Deutsch C, Cantwell D, Posner M, Kennedy J, Barr C, Moyzis R, Schuck S, Flodman P, and Spence MA (2001) Genes and attention-deficit hyperactivity disorder. Clinical Neuroscience Research 1:207–216

Swanson WJ (2003) Adaptive evolution of genes and gene families. Curr Opin Genet Dev 13:617–622

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595

Tishkoff SA, Varkonyl R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Touma EH, Williams SM, Clark AG (2001) Haplotype diversity and linkage disequilibrium at human *G6PD*: recent origin of alleles that confer malarial resistance. Science 293:455–462

Tishkoff SA, Verrelli BC (2003) Patterns of human genetic diversity: Implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Volkow ND, Swanson JM (2003) Variables that affect the clinical use and abuse of methylphenidate in the treatment of ADHD. Am J Psychiatry 160:1909–1918