J Mol Evol (2006) 63:54–68
DOI: 10.1007/s00239-005-0154-8

JOURNAL OF MOLECULAR EVOLUTION

SSADH Variation in Primates: Intra- and Interspecific Data on a Gene with a Potential Role in Human Cognitive Functions

Paola Blasi,[1]* Francesca Palmerio,[1]* Aurora Aiello,[2] Mariano Rocchi,[3] Patrizia Malaspina,[1] Andrea Novelletto[1,2]

[1] Department of Biology, University "Tor Vergata," via della Ricerca Scientifica, snc, 00133, Rome, Italy
[2] Department of Cell Biology, University of Calabria, Rende, Italy
[3] DAPEG, Section of Genetics, University of Bari, Bari, Italy

Received: 24 June 2005 / Accepted: 22 December 2005 [*Reviewing Editor:* Dr. Martin Kreitman]

Abstract. In the present study we focus on the nucleotide and the inferred amino acid variation occurring in humans and other primate species for mitochondrial NAD^+-dependent succinic semialdehyde dehydrogenase, a gene recently supposed to contribute to cognitive performance in humans. We determined 2527 bp of coding, intronic, and flanking sequences from chimpanzee, bonobo, gorilla, orangutan, gibbon, and macaque. We also resequenced the entire coding sequence on 39 independent chromosomes from Italian families. Four variable coding sites were genotyped in additional populations from Europe, Africa, and Asia. A test for constancy of the nonsynonymous vs. synonymous rates of nucleotide changes revealed that primates are characterized by largely variable d_N/d_S ratios. On a background of strong conservation, probably controlled by selective constraints, the lineage leading to humans showed a ratio increased to 0.42. Human polymorphic levels fall in the range reported for other genes, with a pattern of frequency and haplotype structure strongly suggestive of nonneutrality. The comparison with the primate sequences allowed inferring the ancestral state at all variable positions, suggesting that the c.538(C) allele and the associated functional variant is indeed a derived state that is proceeding to fixation. The unexpected pattern of human polymorphism compared to interspecific findings outlines the possibility of a recent positive selection on some variants relevant to new cognitive capabilities unique to humans.

Key words: ALDH5A1 — GABA metabolism — Evolutionary neutrality — Positive selection — Primate evolution

Introduction

An ever-increasing number of studies are being accumulated on DNA sequence differences between human and other primates. This information allows defining phylogenetic relationships and calibrating the times of divergence between primate species. Protein-coding gene comparisons also reveal that the tree topology of human, chimpanzee, and gorilla can differ from gene to gene, a contradictory result that reflects different realizations of the evolutionary path in the short time span between gorilla and the common ancestor of the human-chimpanzee clade (Kitano et al. 2004).

More recently, the complete human and chimpanzee genomic sequences are revealing new differences and similarities between the two species. Although the sequence divergence is strongly dependent on the kind of sequences considered, the average

*These authors contributed equally to the work.

Correspondence to: A. Novelletto; *email:* novelletto@bio.uniroma2.it

difference between humans and our closest relatives is low (Chimpanzee Sequencing and Analysis Consortium 2005).

The hypothesis that the low percentage of differences between human and chimpanzee genomes could reveal the determinants of human specific traits has been proposed by several authors (for a general view of the issue, see Cyranosky [2002] and Dennis [2005]). According to this view, few novel gene variants involved in behavioral and cognitive functions must have arisen after the divergence of the last common ancestor with chimpanzee, and played a major role in human lineage evolution. This argument has been put forward by Enard et al. (2002), who speculated that a variant in the otherwise highly conserved FOXP2 gene was driven to fixation by positive selection associated with the development of a proficient spoken language in the human lineage. Recently, the same argument has been developed for numerous brain-related genes by Dorus et al. (2004), who used the macaque instead, to reduce the stochastic uncertainty in the estimation of the evolutionary rates. Also, gene silencing has been claimed to be involved in anatomical changes specific for the genus *Homo* (Stedman et al. 2004). Finally, it has been suggested that some phenotypic differences, such as those displayed in brain organization and specialization, could be primarily attributed to elevated gene expression levels rather than to sequence differences (King and Wilson 1975; Caceres et al. 2003; Hill and Walsh 2005).

Analyses based on comparisons at the nucleotide level of protein-coding genes suggest different roles of selection on particular genes during primate evolution. The extent of sequence divergence at different loci can be evaluated in terms of nonsynonymous/synonymous substitution rate ratio (d_N/d_S). This kind of data can highlight the evolutionary relevant amino acids of a protein, and variable d_N/d_S values among lineages can provide evidence on the lineage-specific selection pressures (Yang and Nielsen 2002). Clark et al. (2003) applied a model allowing for the estimation of branch-specific d_N/d_S values in a genome-wide survey of human, chimpanzee, and mouse orthologues and identified subsets of genes that were subject to positive selection in the human lineage. Furthermore, estimates of nonsynonymous vs. synonymous rates of change contribute to human intraspecific analysis and allow an evaluation of the evolutionary forces that acted during modern human evolution. Genes involved in drug transport and metabolism are among the best candidates (Bamshad and Wooding 2003).

In the present study, we focus on the nucleotide and the inferred amino acid variation occurring in humans and six primate species for a key enzyme in metabolism, i.e., mitochondrial NAD^+-dependent succinic semialdehyde dehydrogenase (SSADH; ALDH5A1, EC 1.2.1.24). SSADH belongs to the aldehyde dehydrogenase superfamily, a related group of enzymes that metabolize a wide spectrum of endogenous and exogenous aldehydes (Sophos and Vasiliou 2003). SSADH catalyzes the oxidation of succinate semialdehyde (SSA) to succinate, which is the final catabolite of the γ-aminobutyric acid (GABA) shunt. In addition, it has been demonstrated that in the rat central nervous system (CNS) SSADH is also responsible for the oxidation of 4-hydroxy-2-nonenal (HNE), a cytotoxic product of lipid peroxidation (Murphy et al. 2003). We mapped SSADH as a single-copy gene in the 6p22 region in humans (Malaspina et al. 1996) and characterized its genomic and coding structures. The gene consists of 10 exons encompassing over 38 kb. The complete ORF is 1605 bp (accession no. Y11192) encoding for 535 amino acids, with the first 47 residues recognized as mitochondrial targeting peptide (Chambliss et al. 1998).

Initial evidence that SSADH normal activity is relevant to cognitive abilities derives from the analysis of patients affected by SSADH deficiency (OMIM 271980). Since early childhood, they invariably show various degrees of mental retardation and other neurological consequences affecting psychomotor, speech, and language development. Carrier parents, in whom SSADH activity is reduced from a moderate to a severe extent, have also been reported to show EEG abnormalities (Dervent et al. 2004).

Analysis of the SSADH coding region in a panel of random healthy subjects of European origin revealed the presence of several missense and samesense variants at polymorphic frequencies. In vitro expression of missense variants showed remarkable reductions in enzymatic activity in comparison with the most common form of the enzyme, leading to the suggestion that large variations among subjects exist for the enzyme activity on the substrate SSA. Whether this leads to an altered GABA and/or other metabolites balance remains to be ascertained (Blasi et al. 2002). Indeed, a possible involvement of SSADH in higher cognitive functions has been suggested by Plomin et al. (2004), who showed that the most common allele is significantly associated with higher performances.

Here we report on the variation detected in SSADH gene in human, chimpanzee, bonobo, gorilla, lar gibbon, orangutan, and rhesus macaque. We also investigated variation in Old World human populations by resequencing the entire coding region in 39 independent gene copies from Italian families and genotyping a total of 302 additional subjects. The two data sets enabled us to identify the ancestral state at each of the polymorphic residues and to reconstruct a possible phylogeny of the variant haplotypes.

This study is aimed at evaluating interspecific differences in the rate of accumulation of nonsynony-

mous vs. synonymous nucleotide substitutions both inter- and intraspecifically and thus to understand the evolutionary processes and constraints that acted on the SSADH gene. Also, it explores the pattern of human polymorphism in light of recent evidences for adaptive evolution of the same genes both inter- and intraspecifically (Evans et al. 2004a, b, 2005; Mekel-Bobrov et al. 2005).

Materials and Methods

Samples

DNA was extracted from lymphoblastoid cell lines of chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), lar gibbon (*Hylobates lar*), gorilla (*Gorilla gorilla*), and Sumatran and Bornean orangutans (*Pongo pygmaeus abelii* and *Pongo pygmaeus pygmaeus*, hereafter referred to as PPY1 and PPY6, respectively).

Blood or buccal swabs from human subjects were collected in 107, 19, and 29 individuals from Italy, the United Kingdom, and Nigeria, respectively. In addition, biological material was obtained by Italian nuclear families collected in Rome, which produced 39 independent SSADH gene copies. Informed consent was obtained in all cases. We also examined 147 subjects of Asian and African origin included in the HGPD panel and supplied by CEPH.

Nomenclature

Throughout the text the first 47 amino acids and the remaining portion (aa positions 48–535) are referred to as "mitochondrial entry peptide" and "mature peptide," respectively. Nucleotide positions preceded by "g." refer to the human genomic sequence AL031230; nucleotide positions preceded by "c." refer to the human cDNA reference sequence Y11192 (den Dunnen and Antonarakis 2000). The human c.106, c.538, and c.545 polymorphisms reported here are published at http://www.ncbi.nlm.nih.gov/projects/SNP/ as rs4646832, rs2760118, and rs3765310, respectively.

DNA Analysis

Genomic DNA was extracted from cultured cells, fresh blood, or buccal swab by standard techniques, and all the exons were specifically amplified. Primer pairs flanking each of the SSADH exons 2–10 successfully amplified primate DNA under the same conditions reported for humans (Blasi et al. 2002). For both human and primate DNA, exon 1 was amplified in a 50-μl volume, including 10% DMSO with flanking primers S77 (5′-GCGGTGCAGCGA GAAAGA-3′) and S93 (5′-GTGTCACTTTGGGTAAAGC-3′) and the following PCR conditions: 40 cycles at 95°C for 1 min, 52°C for 1 min, and 72°C for 1 min.

The rapid assay for the c.538C > T (exon 3) was as described (Blasi et al. 2002). The assays for the c.106G > C (exon 1), c.545C > T (exon 3), and c.709G > T (exon 4) were performed by the same method. The ASO probes were S105 (5′-CTGCCTCC GGGCCTG-3′) and S106 (5′-CTGCCTCCCGGCCTG-3′; hybridization and washing temperature, 54°C) for c.106G > C, S102 (5′-ACACCCCGGCAAAGGAC-3′) and S103 (5′-ACA CCCTGGCAAAGGAC-3′; hybridization and washing temperature, 60°C) for c.545C > T, and S114 (5′-CCTTCTCCGCCC TGGCC-3′) and S115 (5′-CCTTCTCCTCCCTGGCC-3′; hybridization and washing temperature, 58°C) for c.709G > T.

Total RNA was extracted from human lymphocytes only when blood samples were available, using the TRIzol Reagent (In Vitrogen) on lymphocytes separated by Ficoll gradient or by using the RNeasy Protect Midi kit (Qiagen) as recommended.

First-strand cDNA was generated using Superscript first-strand synthesis System (InVitrogen) with random hexamers. PCRs were performed in a 50-μl reaction using Taq Polymerase (Promega) for 40 cycles at 94°C for 1 min, 52°C for 1 min, and 72°C for 1 min. Exons 2 to 10 were obtained as three partially overlapping amplicons as follows: exons 2, 3, and 4 (with primers S23 [5′-CGCTGCCTACGAGGCTTTC-3′] and Tcb [5′- GTGTCTTCGG CAGGCTTC-3′]); exons 4, 5, 6, and 7 (with primers S2 [5′-TCCCCAGTGCCATGATCAC-3′] and S26 [5′-ACCGCTTTTT CATTAATTAATG-3′]); and exons 7, 8, 9 and 10 (with primers S25 [5′-GTTTGCTCAAACCAATTCTTG-3′] and L3b [5′-AATA ATGGATGGCATGTACC-3′)].

Since the high GC content at the 5′ of the SSADH mRNA prevents the synthesis of a complete cDNA, exon 1 was obtained by PCR on genomic DNA, as described above.

DNA Sequencing and Base Calling

PCR products were purified using the Marligen Bioscience Inc. kit, sequenced on both strands using BigDye dideoxy terminators, and analyzed by automatic sequencer ABI310 (PE Applied Biosystem).

Electropherograms were visually inspected. Heterozygous positions were identified as those producing two peaks, confirmed on both strands, with a height ratio not lower than 0.7:1.

Data Analysis

All genomic primate sequences experimentally obtained were subjected to multiple alignment by Clustal X v1.8 (Thompson et al. 1997) to the human orthologue as represented in AL031230. The alignments were trimmed to the length of the shortest sequence, concatenated, and reformatted with Mega2 software (Kumar et al. 2001).

The orthologous sequence from rhesus macaque (*Macaca mulatta*) was assembled from sequence traces deposited at http://www.ncbi.nlm.nih.gov/traces/cgi isolated by BLAST analysis with human coding sequence (accession no. Y11192). When variable positions were found among traces, the single trace carrying the minimum number of differences compared to the consensus generated by us was chosen. Coding and noncoding sequences were obtained from gi's 540672067, 567336349, 563938008, 448817534, 540484286, 540484248, 503180319, 497840534, 503052585, 486967564, 541131502, 517214160, and 497840420. The reliability of these sequences was also cross-checked with a partial *Macaca fascicularis* sequence experimentally obtained in our lab (not shown). The rhesus macaque sequence was then aligned to those reported above.

Mouse and rat nucleotide coding sequences were obtained by blasting Y11192 to the genomic contig sequences NT_039578.2 and NW_047492.1, respectively. A large deletion in the portion encoding the mitochondrial entry peptide of mouse and rat was positioned based on the results of multiple alignment of the corresponding amino acid sequences (NP_766120 and XP_214478, respectively) with primate (this work) and human sequences.

When necessary, ambiguous primate sequences containing heterozygous positions were resolved into two sequences, the first of which retained the consensus nucleotide at all positions. Under the assumption that at each variable nucleotide position the consensus represents the ancestral state, this method returns a putative ancestral array of variants compatible with the sampled chromosomes. Conservatively, only this sequence was used in further analyses. No heterozygous positions were observed in gorilla, bonobo, or lar gibbon. Three heterozygous positions were found in the single chimpanzee individual (g.46939, g.47060, and g.75725), all of which included the consensus nucleotide. Six and eleven heterozygous positions were found in the Bornean (PPY6)

Table 1. DNA substitutions in functional regions of SSADH DNA sequence in primate species compared with human reference sequence AL031230

	Sequence length (bp)	Chimpanzee	Bonobo	Gorilla	Orangutan (Borneo)	Orangutan (Sumatra)	Lar gibbon	Rhesus macaque
5′ UTR	132 5.2%	0	0	0	4	4	4	7
Coding[a]	1608 63.6%	7	7	8	26	23	30	58
mt entry peptide[b]	141 5.6%	0	1	0	6	4	4	6
Introns	757 29.9%	8	9	12	15	13	19	37
3′ UTR	30 1.2%	0	0	0	1	1	3	1
Total	2527	15	16	20	46	41	56	103

[a]Includes the stop codon.
[b]Included in the coding sequence.

(g.38929, g.46206, g.46219, g.58972, g.64022, and g.76293) and the Sumatran (PPY1) (g.38566, g.38629, g.38736, g.38767, g.38862, g.46206, g.46219, g.48484, g.48563, g.58972, and g.71752) orangutans, respectively.

UPGMA, neighbor joining (based on distances obtained with the two-parameter method of Kimura [1980]) and maximum parsimony trees were constructed by using the coding and noncoding sequences with Mega2, without considering positions deleted in one or more sequences (reducing the number of comparable positions to 2486). A test of phylogeny was performed in all cases by bootstrapping with 1000 replicates.

A likelihood ratio test for constancy of the nonsynonymous vs. synonymous rates of nucleotide changes in a tree including primates, rat, and mouse was performed as described by Yang and Nielsen (2000, 2002) with the program PAML (Yang 1997).

SubPSEC scores for each derived amino acid position in the *Hominoidea* subtree were calculated as described (Thomas et al. 2003) by downloading from the PANTHER database (https://panther.appliedbiosystems.com) the alignment PHTR11699.sf76 and counting the frequency of each amino acid in the relevant positions. This score measures the likelihood of the transition of one amino acid to another in annotated protein sequence alignments. When SubPSEC = 0, the substitution is interpreted as functional neutral, whereas more negative values predict more evolutionarily deleterious substitutions.

For intraspecific human resequencing data, heterozygous positions could be unambiguously resolved and phased by family studies. Haplotypic reconstruction of SNPs genotyping data were obtained with the program PHASE 2 (Stephens et al. 2001b) and further analyzed with Arlequin 2.000 (Schneider et al. 1997).

The haplotype test (Hudson et al. 1994) was performed using the program PSUBS on the set of 39 chromosomes with phase unambiguously determined by family studies. Extended haplotype homozygosity (EHH) was calculated as described (Sabeti et al. 2002). This measure is aimed at detecting the transmission of an extended haplotype without recombination. EHH null distribution was obtained from 1000 simulations generated by SIMCOAL2 (Excoffier 2000), assuming an effective population size of 5000, a population growth rate of 0.03, and recombination at 0.01/Mb.

The DNAsp package (Rozas and Rozas 1999) was used to obtain the polymorphism parameters $\theta\pi$, θs, and D (Tajima 1989), to perform the test of neutrality of McDonald and Kreitman (1991) (based on the null expectation of an equal ratio of nonsynonymous and synonymous substitutions in and between species), and to perform a sliding window analysis of sequence diversity. This latter was obtained by measuring the number of variable sites in consecutive intervals of 100 nt.

All the coding sequences generated for PPY1, PPY6, gorilla, chimpanzee, bonobo, and lar gibbon were submitted to EMBL and given accession numbers AJ621749, AJ621750, AJ621751, AJ621752, AJ891037, and AJ891038, respectively. The entire DNA sequences used in this paper are reported in the Appendix.

Results

Interspecific Comparisons

Overall, we obtained information on 2527 nucleotide positions, from g.38518 to g.77395 (Appendix). Multiple species alignment showed no gross rearrangement. A 14-bp deletion was observed in both orangutan samples from g.63986 to g.63999 at the 3′ end of the fifth intron (IVS5); a 10 bp-deletion was observed in the gibbon (g.48662–g.48671), and a 10-bp deletion in the macaque (g.77134–g.77143); other single-nucleotide deletions were observed at g.75763 in the two *Pan* species, g.75896 in all species except gorilla and human, and g.75918 in orangutan. A single C insertion is shared among orangutan, gibbon, and macaque at g.77151/2. Other insertions are observed only in the macaque at g.63999/64000 and g.75690/1.

We calculated pairwise distances among the conservative sequences using the human orthologue (AL031230). The results (Table 1) showed that the two *Pan* sequences are the most similar to the human one, followed by the gorilla and the two orangutan sequences. Compared to the human, the macaque showed a nearly doubled number of substitutions compared to other *Hominoidea*.

The construction of trees with different methods produced contrasting results. In the UPGMA tree based on distances obtained with the Kimura two-parameter model, human and the two *Pan* species were most closely related. On the other hand, in the neighbor-joining and maximum parsimony trees human and gorilla clustered together, to the exclusion of

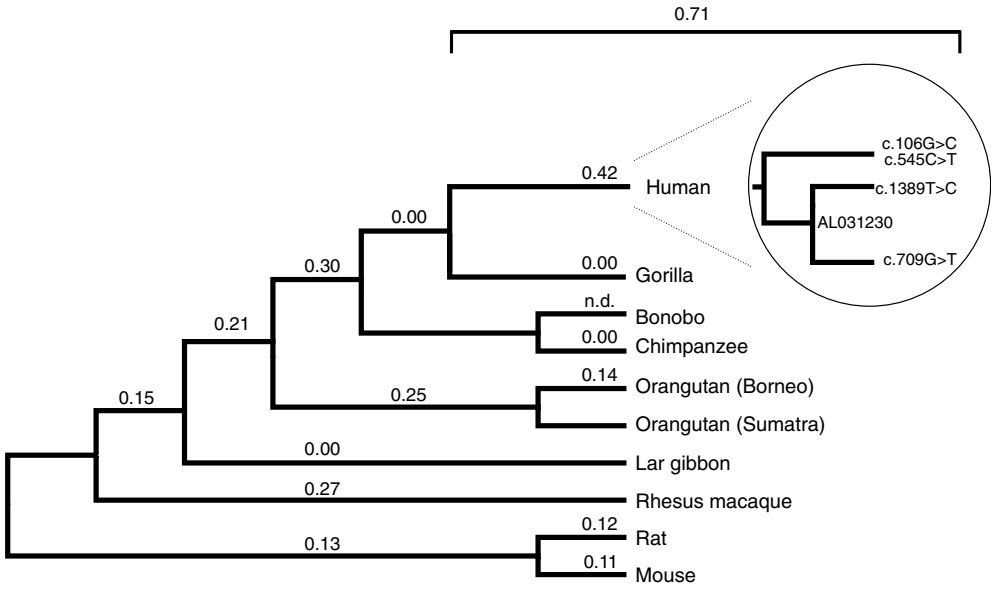

Fig. 1. Phylogenetic tree of the SSADH sequence obtained by maximum parsimony and neighbor joining on the matrix of Kimura two-parameter distances. Branch length is not drawn proportional to distances. The d_N/d_S ratio is reported for each branch when different from 0/0 (n.d. = 0.0018/0). One of the phylogenetic reconstructions discussed in the text for the human polymorphic sequences is shown in the inset. When included in the analysis, this produced the d_N/d_S ratio reported above the bracket.

all other species (Fig. 1). Indeed, no instances were found of substitutions shared by human and chimpanzee only or by chimpanzee and gorilla only. Conversely, two positions are shared between human and gorilla, i.e., g.38946(C) and g.38997(C), plus the indel at position g.75896. The g.38946(C), which is shared by human, gorilla, and other three species, appears to be a retained ancestral state, while the derived (T) is most easily interpreted as a recurrent transition in the chimpanzee and *Pongo* lineages. Despite the overall greater similarity between human and chimpanzee, these three mutations account for the clustering of human and gorilla in the maximum parsimony tree. In the three phylogenetic reconstructions the last node in the human lineage is poorly supported by the bootstrap analyses (60 to 75% vs. $>92\%$ in all remaining cases).

Six nucleotide positions uniquely identify the human reference sequence AL031230 compared to all other species (in parentheses), i.e., g.38797C(G), g.47015C(T), g.48455A(C), g.48474G(T), g.75711C(G), and g.75817T(C). In three positions (g.48451, g.48452, and g.58849), the human sequence differs from other *Hominoidea* but is similar to *M. mulatta*. We could confirm the similarity between *M. mulatta* and *M. fascicularis* at these three positions (not shown), making recurrent mutation in the human lineage the most likely explanation.

Four of the above substitutions cluster in 24 bp at the 3′ end of IVS3. In addition, one of these positions (g.48455) has also been reported as polymorphic in humans (SNP rs2744883), with two alleles (A/G) both different from the primate one (C). In a sliding window analysis of sequence diversity, this region and the adjacent exon 4 produce the highest peak of variable sites, equaled only by the amino terminus of the coding region (see below). Taken together, this observation points to an enhanced divergence of this segment on the human lineage.

DNA Substitutions in Noncoding Regions

We analyzed the occurrence of DNA substitutions separately for the 5′ UTR, intronic sequences, 3′ UTR, and coding sequence (Table 1). Human, chimpanzee, bonobo, and gorilla sequences were 100% identical in the 5′UTR. The other species showed four to seven substitutions. In intronic sequences, all species showed an increased number of substitutions with respect to the amount of intronic sequences analyzed in this study. This excess was significant in gorilla, which showed 60% of the substitutions in introns ($p = 0.02$), and borderline in bonobo. In the 30 bp downstream of the stop codon, one, three, and one substitutions were observed in the orangutan, gibbon, and macaque, respectively.

DNA Substitutions in the Coding Sequence and Amino Acid Replacements

The overall number of nucleotide substitutions within the 1605 bp of coding sequence varied between 7 and 58 (Table 1) compared to human. The number of substitutions per site was slightly reduced in the human/gorilla comparison.

The first 141 bp of the SSADH coding region encodes a 47-aa mitochondrial entry peptide. This portion turned out to be identical in human, gorilla, and chimpanzee; in orangutan, six substitutions were found, increasing to seven when considering polymorphic variants. These account for about 25% of all coding substitutions between human and orangutan, compared to less than 10% of the total surveyed sequence ($p = 0.012$). In addition, five of these six substitutions were nonsynonymous.

In the mature peptide, an overall number of 70 nucleotide substitutions was found across all species, evenly distributed among exons. Of these, 53 were synonymous and 17 nonsynonymous. Interestingly, the nonsynonymous substitutions appear to be nonrandomly distributed, as exons 2, 4, and 10 are invariant across species (4.3 substitutions expected; $p < 0.05$). Two subsets of substitutions deserve special attention. First, two positions are shared by human and gorilla. These are two synonymous changes (g.38946; c.297 and g.38997; c.348) that underlie the clustering of these two species in the maximum parsimony tree (see above). Second, of the four human specific substitutions, two (50%) are synonymous (c.756 and c.1389) and two nonsynonymous (c.148 and c.538).

The above results raised the possibility of an acceleration in the rate of nonsynonymous substitutions on the human lineage that may be consistent with either relaxed selective constraints or directional selection. To test this hypothesis, we ran the program PAML using the phylogeny shown in Fig. 1 and dictated by the neighbor-joining and maximum parsimony trees. The d_N/d_S ratio was 0.18 when considering only primate species and dropped to 0.13 when including also the rodents. When the model allowing for branch-specific rates was applied, we obtained the values reported in Fig. 1. Overall, the branch specific model fitted the data much better than the single-ratio model ($\Delta l = 9.8$, $p << 10^{-5}$). The highest ratio (0.424) was found in the human branch, followed by the branch ancestral to human/gorilla/*Pan* (0.30). Interestingly, downstream to the latter branch, only the human lineage showed a measurable d_N/d_S ratio, due to the paucity of nonsynonymous changes; only in the bonobo were two nonsynonymous (c.50C > G and c.331C > T) vs. no synonymous substitutions found. The value of 0.25 for the branch ancestral to both orangutan sequences is consistent with the excess of amino acid substitutions in the mitochondrial entry peptide (see above). Although in no instance did the d_N/d_S ratio exceed 1, these data support a heterogeneity among branches with a larger possibility for amino acid substitutions to become fixed in lineages eventually leading to human.

We also measured the d_N/d_S ratio in each exon for three pairwise comparisons, i.e., human vs. gibbon, human vs. macaque, and mouse vs. rat. In all cases the ratio was below 0.30, with the exception of the mitochondrial entry peptide in human vs. macaque and mouse vs. rat (1.17 and 0.82, respectively), and exon 3 in human vs. gibbon (0.0079/0) and human vs. macaque (0.0241/0.0229). In the latter case the ratio exceeding 1 is attributable to a nonsynonymous substitution peculiar to the human lineage (g. 47015) and two additional ones shared by all *Hominoidea* as derived states (g. 47034 and g. 47060).

Intraspecific Comparisons

Human Polymorphic Variation. We resequenced the entire coding region of 39 chromosomes of Italian origin. Of 1605 positions, 5 segregating sites were found, three of which were common variants (Table 2a). All of these sites were included among those reported as polymorphic in independent series (Blasi et al. 2002; Saito et al. 2002; Akaboshi et al. 2003). Polymorphism levels per base pair, summarized by $\theta\pi$ and θs, were $3.8 \pm 1.2 \times 10^{-4}$ and $7.4 \pm 3.8 \times 10^{-4}$, respectively. A greater degree of polymorphism was observed in nonsynonymous than in synonymous positions ($\theta\pi = 4.6$ vs. 1.3×10^{-4}; $\theta s = 7.9$ vs. 5.8×10^{-4}).

We unambiguously resolved the haplotype phase of alleles at the five positions by analyzing their segregation in families (Table 2b). Two of the common variants turned out to be strongly associated, c.106(C) and c.545(T) always being found in *cis*. Of 32 haplotypes expected, only 5 were found. In order to explore the repertoire of segregating haplotypes, we genotyped an additional 302 subjects from Europe, Africa, and Asia for the most common variants and inferred the haplotype phase with the program PHASE. The association was fully confirmed, as all genotypes were interpreted without ambiguities. In Italy, the c.709G > T turned out to be polymorphic, generating 16 possible haplotypes, of which only 4 were found. In the other populations, only three positions were polymorphic. In all population samples the haplotypes ranked in the order GCCG, GTCG, CTTG, according to their frequency. Haplotype frequency heterogeneity was assayed by AMOVA (Excoffier et al. 1992), which produced a Fst of 0.09 ($p < 10^{-4}$). Among individual loci, position c.538 contributed most of the diversity and also the highest individual Fst (0.09; $p < 10^{-4}$).

The comparison with the primate sequences allowed inferring the ancestral state at each of the five variable positions (Table 2a). It is interesting to note that at two positions (c.538 and c.1389) the most common allele is the derived one. In particular, for c.538 the C allele appears to be by far the most common in many geographically dispersed populations, except those from sub-Saharan Africa.

Table 2a. Relative frequencies (±SE) of alleles at five polymorphic SNPs in a sample of Italian families by resequencing and in five populations from three continents by genotyping (numbers in parentheses refer to gene copies)

Position	State	Allele	Families (n = 39)	Italy (n = 214)	UK (n = 38)	Nigeria (n = 58)	Sub-Saharan Africa (n = 122)	Asia (n = 172)
c.106	Anc.	G	0.949 ± 0.036	0.963 ± 0.013	0.921 ± 0.044	0.983 ± 0.017	0.975 ± 0.014	0.948 ± 0.017
g.38755	Der.	C	0.051 ± 0.036	0.037 ± 0.013	0.079 ± 0.044	0.017 ± 0.017	0.025 ± 0.014	0.052 ± 0.017
c.538	Anc.	T	0.179 ± 0.061	0.238 ± 0.029	0.395 ± 0.080	0.379 ± 0.064	0.492 ± 0.045	0.157 ± 0.028
g.47015	Der.	C	0.821 ± 0.061	0.762 ± 0.029	0.605 ± 0.080	0.621 ± 0.064	0.508 ± 0.045	0.843 ± 0.028
c.545	Anc.	C	0.949 ± 0.036	0.963 ± 0.013	0.921 ± 0.044	0.983 ± 0.017	0.975 ± 0.014	0.948 ± 0.017
g.47022	Der.	T	0.051 ± 0.036	0.037 ± 0.013	0.079 ± 0.044	0.017 ± 0.017	0.025 ± 0.014	0.052 ± 0.017
c.709	Anc.	G	0.974 ± 0.026	0.991 ± 0.007	1.00	1.00	1.00	1.00
g.48621	Der.	T	0.026 ± 0.026	0.009 ± 0.007	0.00	0.00	0.00	0.00
c.1389	Anc.	C	0.026 ± 0.026	n.t.	n.t.	n.t.	n.t.	n.t.
g.75817	Der.	T	0.974 ± 0.026					

Table 2b. Relative frequencies (±SE) of haplotypes at five polymorphic SNPs (corresponding to alleles in Table 2a) in a sample of Italian familes by resequencing and in five populations from three continents by genotyping (numbers in parentheses refer to gene copies)

Haplotype	c.106 g.38755	c.538 g.47015	c.545 g.47022	c.709 g.48621	c.1389 g.75817	Families (n = 39)	Italy (n = 214)	UK (n = 38)	Nigeria (n = 58)	Sub-Saharan Africa (n = 122)	Asia (n = 172)
AL031230	G	C	C	G	T						
1	G	C	C	G	T	0.769 ± 0.067					
2	G	C	C	T	T	0.026 ± 0.026					
3	G	C	C	G	C	0.026 ± 0.026					
4	G	T	C	G	T	0.128 ± 0.052					
5	C	T	T	G	T	0.051 ± 0.035					
	G	C	C	G	n.t.		0.752 ± 0.030	0.605 ± 0.080	0.621 ± 0.064	0.508 ± 0.045	0.843 ± 0.028
	G	C	C	T	n.t.		0.009 ± 0.007	0	0	0	0
	G	T	C	G	n.t.		0.201 ± 0.027	0.316 ± 0.076	0.362 ± 0.064	0.467 ± 0.045	0.105 ± 0.023
	C	T	T	G	n.t.		0.037 ± 0.013	0.079 ± 0.044	0.017 ± 0.017	0.025 ± 0.014	0.052 ± 0.017

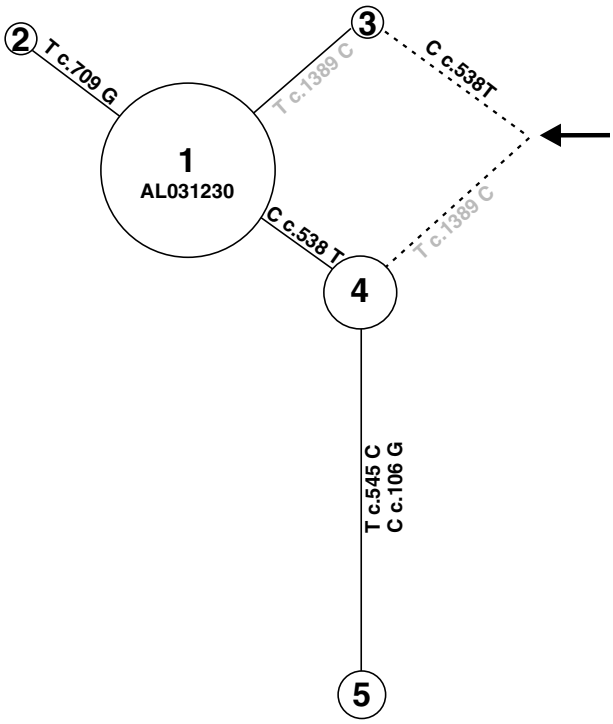

Fig. 2. Unrooted tree for haplotypes defined by five coding SNPs in humans. Numbers at each node correspond to haplotypes as in Table 2b. Circles are proportional to haplotype frequency. The most likely position for the root is marked by the arrow. The two possible mutational steps leading from the ancestral (not found) to the observed haplotypes are shown as dashed lines. For each branch the corresponding mutation is written with the allelic states oriented toward the haplotypes carrying them. Nonsynonymous and synonymous mutations are reported in black and gray, respectively.

All haplotypes could be connected on the unrooted tree shown in Fig. 2. As no haplotype carrying the ancestral state at all variable positions was found, two equally parsimonious solutions for the underlying phylogeny can be put forward. The first posits an initial c.538T > C mutation, leading from the root to haplotype 3, followed by c.1389C > T and then by either a c.538C > T reversion (branch 1–4) or a recombination with an haplotype carrying the ancestral c.538(T) allele anywhere in the genomic stretch of 28.8 kb between c.538 and c.1389. The second one (also reported in the inset in Fig. 1) posits an initial c.1389C > T mutation leading to haplotype 4, followed by c.538T > C and then by either a c.1389T > C reversion (branch 1–3) or a recombination with an haplotype carrying the ancestral c.1389(C) allele to yield haplotype 3. Under either scenario, at least 33 of 39 of the sequenced haplotypes would carry one or two derived amino acid positions, and 31 of these haplotypes are characterized by c.538(C). Among the genotyped populations, this proportion ranges between 53% (sub-Saharan Africa) and 89% (Asia). This strongly supports the view of an increase toward fixation of at least one haplotype in the extant human population.

Neutrality Tests. We applied to the sample of the resequenced gene copies four tests of the equilibrium neutral model that use information from different aspects of the data, i.e., haplotype structure, frequency spectrum, and comparison of polymorphic sites within species and fixed substitutions between species.

The haplotype test asks if a subset of haplotypes at a particular frequency contains fewer segregating sites than expected by simulating samples under neutrality, examining all subsets at the given frequency in each simulated sample, and asking how often a subset of haplotypes at the same frequency containing the same number or fewer segregating sites can be found (Hudson et al. 1994). We asked whether haplotypes carrying the c.538(C) allele showed lower than expected variation, given the frequency of the C allele, under neutrality. We obtained the borderline probability of 0.07 for observing by chance a number of segregating sites less than or equal to that observed. In order to properly account for each variant's frequency we calculated EHH on both sides of c.538. From the data in Table 2b, it is apparent that haplotypes characterized by c.538(C) have EHH = 1 at c.106 and c.545 in all populations. In the two Italian samples, EHH at c.709 is reduced to 0.938 and 0.976. Conversely, in haplotypes with c.538(T) EHH is always lower than before, ranging between 0.91 and 0.54 at both c.106 and c.545. In 1000 coalescent simulations chromosomes attaining a frequency of 0.821 or more (as c.538[C] in Italian families) yet retaining the observed EHHs were 3.6%, 4.2%, and 7.6%, for c.106, c.545, and c.709, respectively. Thus, c.538(C) is associated with a significantly reduced degree of variation on both sides.

Tajima's (1989) D compares the number of nucleotide polymorphisms with the mean pairwise difference between sequences (Bamshad and Wooding 2003). This test produced a nonsignificant though negative value of −1.25.

Furthermore, in all comparisons between human and primates except the bonobo, an excess of nonsynonymous substitutions was observed in human polymorphism, compared to the pattern observed in fixed positions (McDonald and Kreitman 1991) (Table 3). The G test showed significant results in five of seven comparisons. By summing the log(likelihood) for all tests we obtained a figure of 25.88 (χ^2_7, $p < 0.001$), with a highly constant odd ratio across tests (Breslow-Day test, $p > 0.95$), which identifies the partitioning of human polymorphic changes as the outlying value.

In order to better understand the evolutionary significance of the human polymorphisms, we determined the d_N/d_S ratios including the human polymorphic sequences (inset in Fig. 1). The result for the entire human lineage was 0.71, i.e., five times higher

Table 3. Results of McDonald and Kreitman test performed on SSADH coding region inclusive of the mitochondrial entry peptide

	Human polymorphic	Fixed in comparison with						
		Chimpanzee	Bonobo	Gorilla	Orangutan (Borneo)	Orangutan (Sumatra)	Lar gibbon	Rhesus macaque
Synonymous	1	4	2	5	16	13	21	37
Nonsynonymous	4	1	3	1	8	7	7	19
Significance of test								
P(Fisher)		0.20	1.00	0.08	0.13	0.13	0.03	0.06
P(*G*)		0.05	0.48	0.03	0.05	0.06	0.02	0.04

Table 4. subPSEC scores for amino acid substitutions fixed in *Hominoidea* and polymorphic in human

Genomic position (see Appendix)	Position in cDNA	Amino acid most represented in PHTR1699-sf76	Derived amino acid in *Hominoidea* or human	SubPSEC score
Fixed				
g.38980	c.331	A	R	−1.47
g.38980	c.331	A	C	−2.56
g.44168	c.985	L	M	0.00
g.66439	c.1033	R	Q	−3.93
g.66511	c.1105	K	R	−1.10
g.71692	c.1216	V	V	0.00
g.75776	c.1348	D	N	−2.77
Polymorphic				
g.47015	c.538	P	H	−2.89
g.47022	c.545	P	L	−2.30
g.48621	c.709	A	S	−3.91

than the rate estimated (0.13) on the rest of the tree ($\Delta l = 9$). This further reinforces the hypothesis of an accelerated rate of amino acid substitutions at least at some SSADH positions.

We also compared the pattern of amino acid replacement between and within species by means of the subPSEC score. We counted scores for human polymorphic positions and all the derived positions fixed in *Hominoidea*. In this case, we used *M. mulatta* as outgroup, since rat and mouse are too distantly related to primates to be reliable outgroups. The sequence alignment represented in the PANTHER database excludes the mitochondrial entry peptide and the first 11 amino acid residues of the mature peptide. Interestingly, the three polymorphic positions in the alignment (c.538, c.545, and c.709) turned out to be associated with largely negative scores (range, −3.91 to −2.30) (Table 4), on the low side of the subPSEC distribution for substitutions in SSADH (rank test, $p = 0.21$) and in the lowest quartile of the neutral distribution built by Thomas et al. (2004).

Discussion

SSADH is a key enzyme in the catabolism of GABA, the most important inhibitory neurotransmitter in the mammalian CNS. The efficiency of the enzyme in metabolizing SSA contributes to control of the endogenous production of γ-hydroxybutyric acid (GHB), a compound with relevant neuroactive properties. The interplay between GABA and GHB has only partially been elucidated and their joint effects on behavior might be complex (Wong et al. 2003). Clinical features of the complete enzyme deficiency have been fully described (Pearl et al. 2003) and show involvement of specific cerebral areas (Ziyeh et al. 2002). An array of molecular lesions leading to null enzyme activity has been reported (Akaboshi et al. 2003). Recently, Dervent et al. (2004) have shown that also the heterozygous state might be associated with epileptic features, probably as the result of GHB accumulation. Thus, it could be expected that different enzyme activities associated with isoenzyme genotypes might result in different clinical or subclinical phenotypes mediated by endogenous GABA and GHB levels.

In humans, we indeed observed variants with suboptimal in vitro enzyme activity and suggested that these are common polymorphic variants (Blasi et al. 2002), possibly found as multiple amino acid replacements in the same polypeptide. We tested a synthetic protein carrying both the c.538(T) and the c.545(T) mutations, resulting in an activity reduced to 36% (Akaboshi et al. 2003). Here we show that these replacements turn out to be always associated also with c.106(C), with a likely further reduction of activity. In addition, SSADH resides in a chromo-

somal region repeatedly identified in genomic scans for the genetic determinants of reading disabilities (Ahn et al. 2002; Fisher and DeFries 2002; Londin et al. 2003). Taken together, the above data make SSADH a reliable candidate, both positionally and functionally, for this condition. A more recent report (Deffenbacher et al. 2004) has indeed produced evidence for a nonrandom transmission of SSADH alleles and reading abilities. By analyzing the c.538 polymorphism, Plomin et al. (2004) have shown that the C allele is associated with higher cognitive abilities, suggesting SSADH as a contributor to the quantitative IQ phenotype.

A number of recent studies claimed the detection of natural selection based on the pattern of inter- or intraspecific sequence variation. In all cases, the most robust conclusions referred to genes with clarified biology, which cooperates in making the evolutionary picture reliable (Kreitman and Di Rienzo [2004] and references therein). In view of the putative role of SSADH in capabilities unique to humans, we sought to complement the information on its biology with an investigation of its interspecific divergence during hominoid radiation. We also related this data set with those collected from different continental human populations, to understand the pattern and origins of the extant human variation and the constraints acting on it.

As far as interspecific comparisons are concerned, the sequence explored here brings the signature of strong conservation. The overall pattern of DNA substitutions supports a gene genealogy with human closer to gorilla than chimpanzee (Fig. 1), also found in 24% of the genes explored by Kitano et al. (2004), but reveals a weak power of SSADH in resolving the *Gorilla-Pan-Homo* phylogeny, due to a very low number of substitutions.

Among primates, 25 nonsynonymous substitutions were found. None of them affects any of the five amino acidic motifs functionally important in SSADH, thus confirming strong evolutionary constraints, already detected in the alignment of amino acid sequences from largely divergent taxa (Busch and Fromm 1999). A relative accumulation of DNA and amino acid substitutions was found only in the mitochondrial entry peptide, more pronounced in the orangutan. Its possible functional significance deserves experimental investigation.

An overall strong conservation is in line with the low d_N/d_S ratio found here among primates (0.18), at the lower boundary of the distributions reported by Shi et al. (2003) and Gimelbrandt et al. (2004) in extensive human-chimpanzee comparisons. Thus, SSADH appears to be a difficult target for detecting acceleration of amino acid replacements and, more so, to distinguish between the possible effects of relaxed vs. directional selection. In the present study, the human sequence is part of a wide primate phylogeny which enabled a large number of comparisons, whose effect was to increase the power of analysis. The overall pattern is that of a largely variable d_N/d_S ratio in different primate taxa. This can be interpreted as either a noncoherent evolutionary path in different branches or, more simply, the result of stochastic fluctuations associated with the low absolute number of substitutions. Upon this general background, the human lineage displays a d_N/d_S ratio three times higher than the background. SSADH was included in the survey by Clark et al. (2003), who used sequences spanning only exons 2–4 and 6–9 in a tripartite phylogeny. These authors reported a strong increase in d_N/d_S ratio specifically in the human branch, with borderline significance. More importantly, they showed that the group of genes for amino acid catabolism, to which SSADH belongs, showed among the most significant increases in d_N/d_S ratio.

Several lines of evidence can be used to discriminate relaxed vs. directional selection in the case of increased d_N/d_S ratios (Dorus et al. 2004). The presence of several known pathological mutations and the absence of duplicated SSADH copies in the genome rule out the possibility of the generalized lack of selective pressure. More direct evidence in favor of directional selection, based on the value of d_N/d_S being the highest in the human-specific branch (0.42), is subject to the caveat mentioned above. Finally, a possible role of SSADH in brain development, a feature often associated with the highest d_N/d_S ratios, is still to be explored. Nevertheless, it is worth noting that our value for primates is higher than the average reported by Dorus et al. (2004) for nervous system-related genes, with the rodent value well within the appropriate range.

The pattern of human polymorphism is informative to attempt inferences on recent SSADH evolution. We resequenced the gene in order to exclude that assaying previously known variants (Blasi et al. 2002; Akaboshi et al. 2003) in population surveys could result in ascertainment bias (Kreitman and Di Rienzo 2004). As far as the European sample is concerned, we excluded the presence of novel variants at high frequencies. Diversity parameters equal those reported by Stephens et al. (2001a) for the coding regions. The gene is affected with a higher degree of polymorphism ($\theta\pi$) at nonsynonymous positions, with a number of nonsynonymous polymorphisms that exceeds synonymous polymorphisms and brings the overall d_N/d_S ratio to 0.71.

The derived allele at c.538 has increased its frequency and now represents the vast majority worldwide. In particular, it determines the replacement of a tyrosine conserved from rodents to apes by a histidine, to generate the allele with the highest activity recorded so far. In summary, in the range of populations tested

here, 70–80% haplotypes are associated with this maximal activity. We also showed an additional variant which nearly reached fixation, i.e., c.1389(T).

Two non-mutually exclusive processes can be envisaged to interpret our intraspecific findings. The first is a purely neutral replacement of the genealogy of genes characterized by the ancestral c.538(T) with the newly arisen c.538(C), possibly enhanced by a demographic bottleneck leading to an increase in the haplotype carrying c.538(C) and/or its overrepresentation in populations exiting out of Africa. The second process is an increase in the haplotype(s) carrying c.538(C) driven by natural selection, possibly as a result of their altered biochemical properties. Under the latter hypothesis, the selective factors possibly acting on the favored haplotype 1 (Table 2b) would either predate the splitting of European, Asian, and African populations or be convergent on the three continents. In this scenario, the possibility that the positively selected locus(i) is(are) not SSADH itself but one in strong LD with it cannot be excluded, in view of the recombination suppression in the region immediately centromeric to SSADH (Ahn et al. 2002).

We used four tests to seek which of the two scenarios best fits our data. Tajima's *D* is indicative of an excess of too common and too rare alleles, showing that the data indeed bring the signature of either population expansion or directional selection. Here, a bias introduced by testing only a sample of European ancestry cannot be dismissed, as this population may represent only a subset of the variation present in Africa. The McDonald and Kreitman test confirms that the human polymorphism is enriched in amino acid replacements, compared to most of the primates considered here. The borderline *p* value for the haplotype test is indicative of reduced diversity on the haplotypes carrying c.538(C). This is also confirmed by the significant increase in EHH, which denotes a shallow genealogy for these haplotypes despite their overall high frequency, as opposed to the less common haplotypes carrying c.538(T). In addition, the strongly negative subPSEC score for c.538(C) identifies this as an evolutionary drastic change. This further weakens the hypothesis that its frequency might be explained by neutral drift. The finding of two other variants at frequencies 2–7% in continentally dispersed populations is also at odds with drift effects acting worldwide.

Taken together, our data show a pattern of human intraspecific diversity which is compatible with selection, in continuity with interspecific divergence, although the latter is characterized by strong conservation across distant taxa (Bush and Fromm 1999). In suggesting that the c.538(C) allele and the associated functional variant are indeed a derived state that is proceeding to fixation, our data lead one to speculate that cognitive performance and/or behavior were the phenotypic traits on which selective forces acted, perhaps at different stages in the lineage eventually leading to humans. This has been hypothesized in view of the key role for GABAergic neurotransmission in many cognitive processes (Plomin et al. 2004). Recent papers (Evans et al. 2004a, b; Kouprina et al. 2004; Stedman et al. 2004; Wang and Su 2004) have produced genetic data that bear the signature of a punctuation in the evolution of the brain and the braincase features during different phases of primate radiation. This trend has been further supported by wide surveys on intraspecific human variation for the same genes (Evans et al. 2005; Mekel-Bobrov et al. 2005). The hypothesis of selective pressure on SSADH alleles/haplotypes might involve a late and final step of selection, only when cognitive capabilities became relevant traits for recent human evolution (Enard et al. 2002).

Acknowledgments. We gratefully acknowledge Dr. A. Di Rienzo and two anonymous reviewers for their helpful comments on the first draft of this work. We thank Dr. M. Basile for computational support. This work was supported by grants MIUR 60% to Prof. Carla Jodice, PRIN 2003 and 60% to A.N., and CEGBA (Centro di Eccellenza Geni in campo Biosanitario e Agroalimentare) and European Commission (INPRIMAT, QLRI-CT-2002-01325) to M.R.

Appendix

Alignment of sequences obtained in the present paper (see Materials and Methods) to human AL031230 is shown in Fig. A1. Noncoding and coding sequences are in lowercase and uppercase, respectively. Numbering refers to AL031230. Boundaries between noncontiguous genomic sequences concatenated here are shown. The arrow marks the first nucleotide of the mature peptide.

References

Ahn J, Won T-W, Kaplan DE, Londin ER, Kuzmic P, Gelernter J, Gruen JR (2002) A detailed physical map of the 6p reading disability locus, including new markers and confirmation of recombination suppression. Hum Genet 111:339–349

Akaboshi S, Hogema BM, Novelletto A, Malaspina P, Salomons GS, Maropoulos GD, Jakobs C, Grompe M, Gibson KM (2003) Mutational spectrum of the succinate semialdehyde dehydrogenase (ALDH5A1) gene and functional analysis of 27 novel disease-causing mutations in patients with SSADH deficiency. Hum Mutat 22:442–450

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

Blasi P, Pilo Boyl P, Ledda M, Novelletto A, Gibson KM, Jakobs C, Hogema B, Akaboshi S, Loreni F, Malaspina P (2002) Structure of human succinic semialdehyde dehydrogenase gene: identification of promoter region and alternatively processed isoforms. Mol Genet Metab 4:348–462

Busch KB, Fromm H (1999) Plant succinic semialdehyde dehydrogenase. Cloning, purification, localization in mitochondria, and regulation by adenine nucleotides. Plant Physiol 121:589–597

Caceres M, Lachuer J, Zapala MA, Redmond JC, Kudo L, Geschwind DH, Lockhart DJ, Preuss TM, Barlow C (2003) Elevated

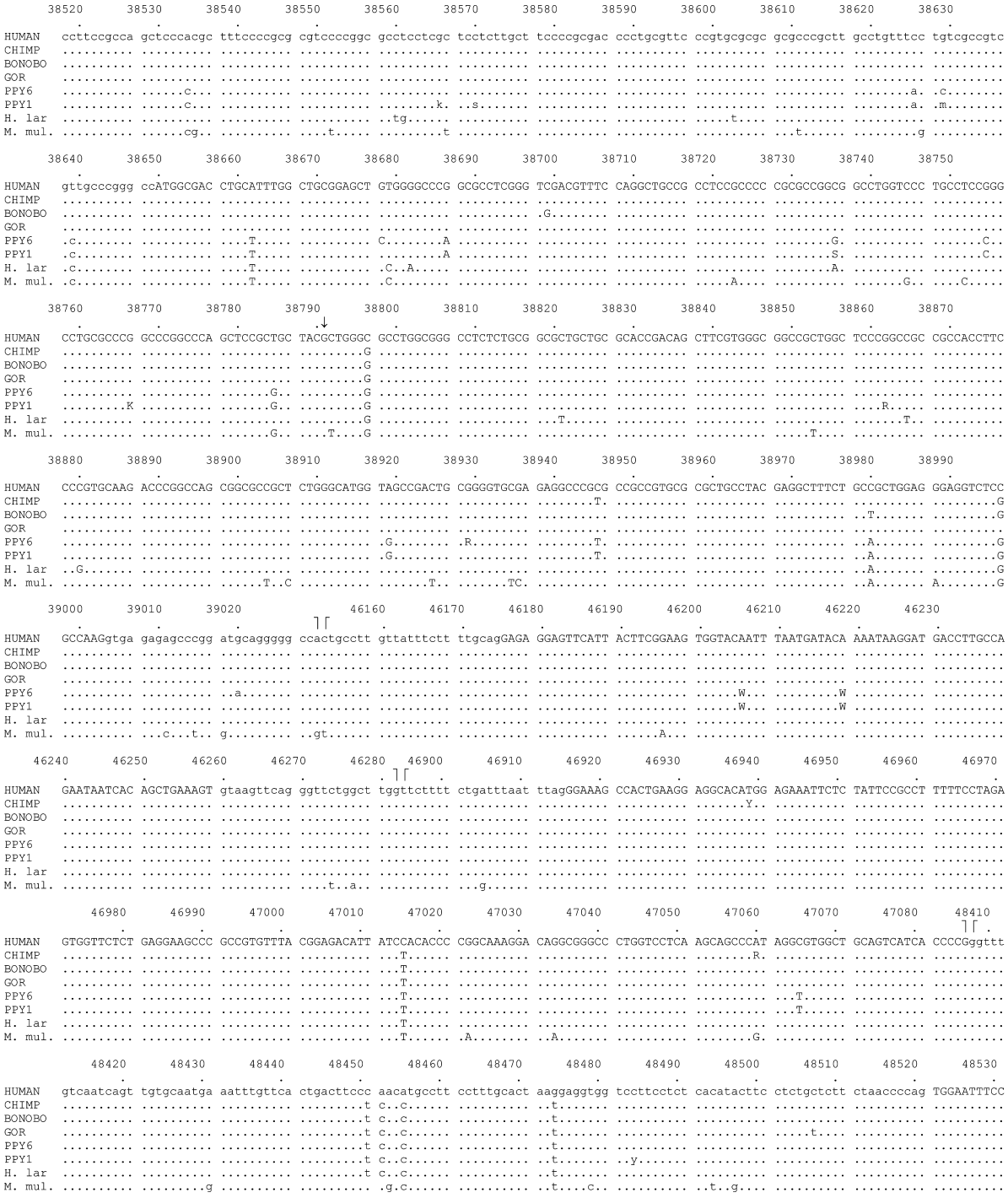

Fig. A1. Alignment of sequences obtained in the present paper (see Materials and Methods) to human AL031230. Non coding and coding sequences are in lower and upper case, respectively. Numbering refers to AL031230. Boundaries between non-contiguous genomic sequences here concatenated are shown. The arrow marks the first nucleotide of the mature peptide.

gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci USA 100:13030–13035

Chambliss KL, Hinson DD, Trettel F, Malaspina P, Novelletto A, Jakobs C, Gibson KM (1998) Two exon-skipping mutations as the molecular basis of succinic semialdehyde dehydrogenase deficiency (4-hydroxybutyric aciduria). Am J Hum Genet 63:399–408

Chimpanzee Sequencing, Analysis Consortium(2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M (2003) Inferring nonneutral evolution from human-chimpanzee-mouse orthologous gene trios. Science 302:1960–1963

Cyranoski D (2002) Almost human. Nature 418:910–912

Deffenbacher KE, Kenyon JB, Hoover DM, Olson RK, Pennington BF, DeFries JC, Smith SD (2004) Refinement of the 6p21.3 quantitative trait locus influencing dyslexia: linkage and association analyses. Hum Genet 115:128–138

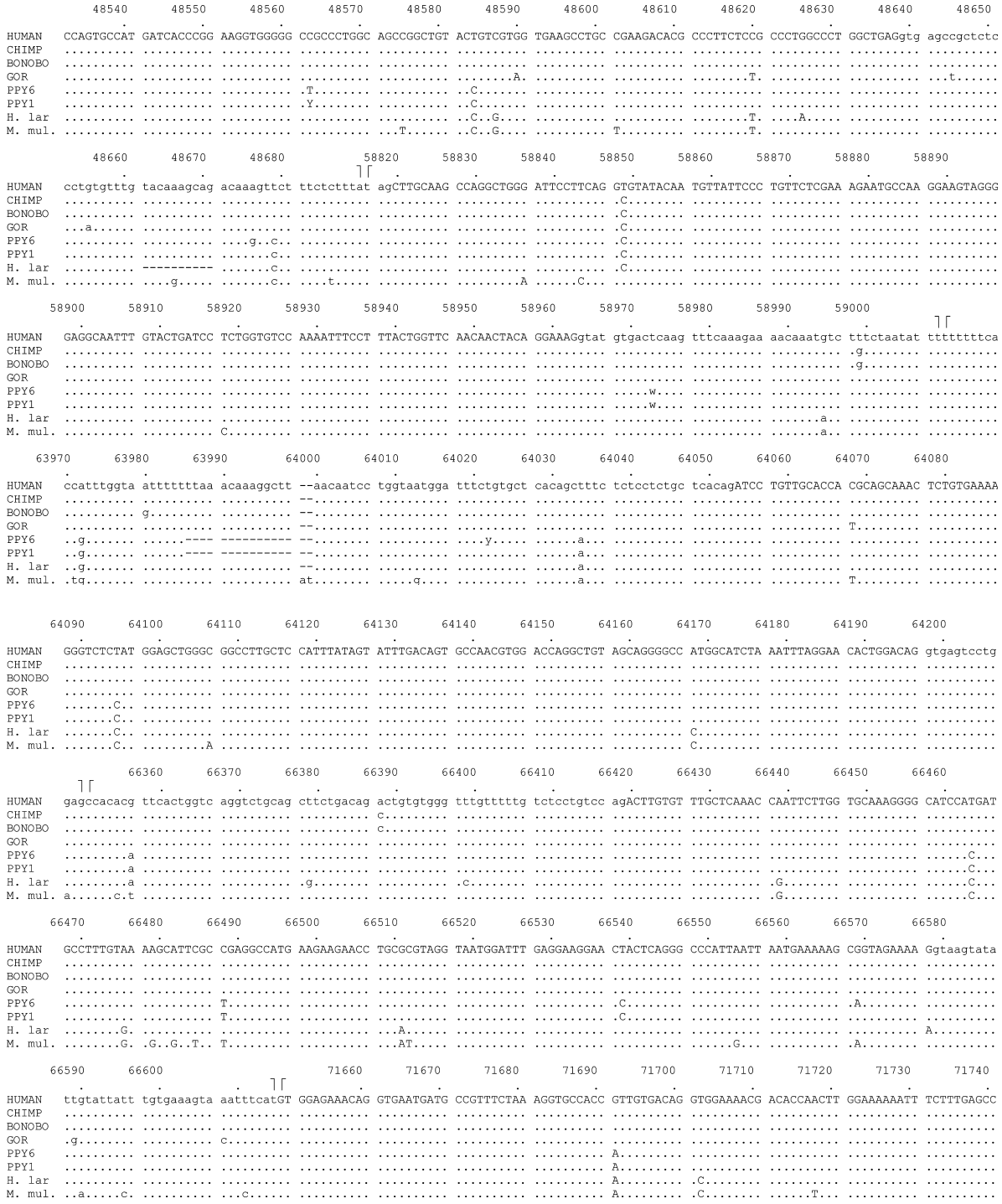

Fig. A1. Continued.

den Dunnen JT, Antonarakis SE (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12 (Erratum: Hum Mutat 20:403, 2002)

Dennis C (2005) Chimp genome: branching out. Nature 437:17–19

Dervent A, Gibson KM, Pearl PL, Salomons GS, Jakobs C, Yalcinkaya C (2004) Photosensitive absence epilepsy with myoclonias and heterozygosity for succinic semialdehyde dehydrogenase (SSADH) deficiency. Clin Neurophysiol 115:1417–1422

Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, Lahn BT (2004) Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. Cell 119:1027–1040

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869–872

Evans PD, Anderson JR, Vallender EJ, Choi SS, Lahn BT (2004a) Reconstructing the evolutionary history of microcephalin, a gene controlling human brain size. Hum Mol Genet 3:1139–1145

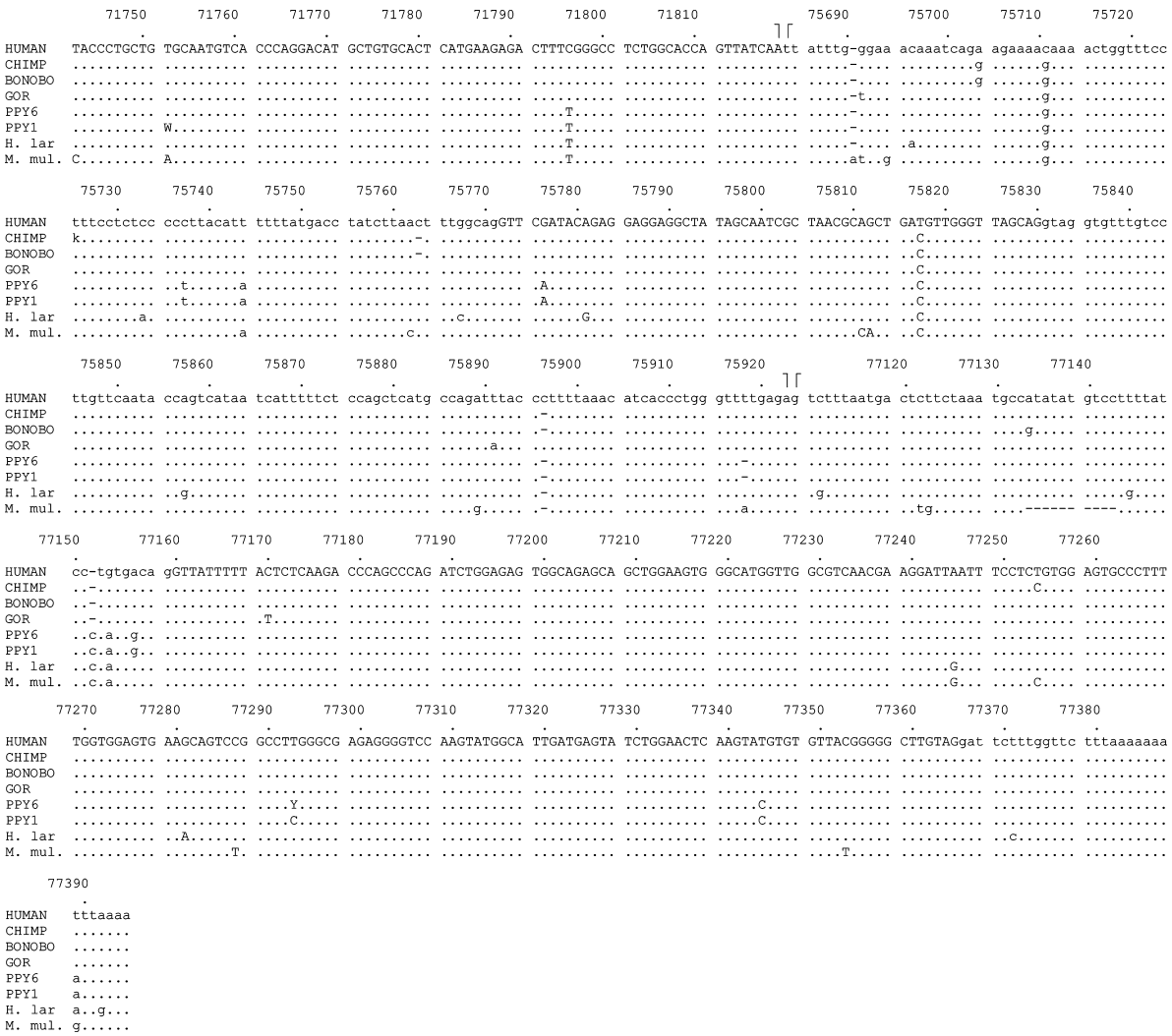

Fig. A1. Continued.

Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcom CM, Dorus S, Lahn BT (2004b) Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. Hum Mol Genet 13:489–494

Evans PD, Gilbert SL, Mekel-Bobrov N, Vallender EJ, Anderson JR, Vaez-Azizi LM, Tishkoff SA, Hudson RR, Lahn BT (2005) Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. Science 309:1717–1720

Excoffier L, Smouse PE, Quattro JN (1992) Analysis of molecular variance inferred from metric distance among DNA haplotypes: application to human mitochondrial restriction data. Genetics 131:479–491

Excoffier L, Novembre J, Schneider S (2000) SIMCOAL: a general coalescent program for the simulation of molecular data in interconnected populations with arbitrary demography. J Hered 91:506–509

Fisher SE, DeFries JC (2002) Developmental dyslexia: genetic dissection of a complex cognitive trait. Nat Rev Neurosci 3:767–780

Gimelbrant AA, Skaletsky H, Chess A (2004) Selective pressures on the olfactory receptor repertoire since the human-chimpanzee divergence. Proc Natl Acad Sci USA101:9019–9022

Hill RS, Walsh CA (2005) Molecular insights into human brain evolution. Nature 437:64–67

Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ (1994) Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. Genetics 136:1320–1340

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107–116

Kitano T, Liu YH, Ueda S, Saitou N (2004) Human-specific amino acid changes found in 103 protein-coding genes. Mol Biol Evol 21:936–944

Kouprina N, Pavlicek A, Mochida GH, Solomon G, Gersch W, Yoon YH, Collura R, Ruvolo M, Barrett JC, Woods CG, Walsh CA, Jurka J, Larionov V (2004) Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. PLoS Biol 2:E126

Kreitman M, Di Rienzo A (2004) Balancing claims for balancing selection. Trends Genet 20:300–304

Kumar S, Tamura K, Jakobsen B, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software Arizona State University, Tempe

Londin ER, Meng H, Gruen JR (2003) A transcription map of the 6p22.3 reading disability locus identifying candidate genes. BMC Genomics 4:25–33

Malaspina P, Roetto A, Trettel F, Jodice C, Blasi P, Frontali M, Carella M, Franco B, Camaschella C, Novelletto A (1996) Construction of a YAC contig covering human chromosome 6p22. Genomics 36:399–407

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654

Mekel-Bobrov N, Gilbert SL, Evans PD, Vallender EJ, Anderson JR, Hudson RR, Tishkoff SA, Lahn BT (2005) Ongoing adaptive evolution of ASPM, a brain size determinant in *Homo sapiens*. Science 309:1720–1722

Murphy TC, Amarnath V, Gibson KM, Picklo MJ Sr (2003) Oxidation of 4-hydroxy-2-nonenal by succinic semialdehyde dehydrogenase (ALDH5A). J Neurochem 86:298–305

Pearl PL, Novotny EJ, Acosta MT, Jakobs C, Gibson KM (2003) Succinic semialdehyde dehydrogenase deficiency in children and adults. Ann Neurol 54 (Suppl 6):S73–S80

Plomin R, Turic DM, Hill L, Turic DE, Stephens M, Williams J, Owen MJ, O'Donovan MC (2004) A functional polymorphism in the succinate-semialdehyde dehydrogenase (aldehyde dehydrogenase 5 family, member A1) gene is associated with cognitive ability. Mol Psychiatry 9:582–586

Rozas J, Rozas R (1999) DNAsp v.3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Sabeti PC, Reich DE, Higgins JM, Levine HZ, 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

Saito S, Iida A, Sekine A, Ogawa C, Kawauchi S, Higuchi S, Ohno M, Nakamura Y (2002) 906 variations among 27 genes encoding cytochrome P450 (CYP) anzymes and aldehyde dehydrogenases (ALDHs) in the Japanese population. J Hum Genet 47:419–444

Schneider S, Kueffer J-M, Roessli D, Excoffier L (1997) Arlequin ver.1.1: a software for population genetic data analysis Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland,

Shi J, Xi H, Wang Y, Zhang C, Jiang Z, Zhang K, Shen Y, Jin L, Zhang K, Yuan W, Wang Y, Lin J, Hua Q, Wang F, Xu S, Ren S, Xu S, Zhao G, Chen Z, Jin L, Huang W (2003) Divergence of the genes on human chromosome 21 between human and other hominoids and variation of substitution rates among transcription units. Proc Natl Acad Sci USA 100:8331–8336

Sophos NA, Vasiliou V (2003) Aldehyde dehydrogenase gene superfamily: the 2002 update. Chem Biol Interact 143–144:5–22

Stedman HH, Kozyak BW, Nelson A, Thesier DM, Su LT, Low DW, Bridges CR, Shrager JB, Minugh-Purvis N, Mitchell MA (2004) Myosin gene mutation correlates with anatomical changes in the human lineage. Nature 428:415–418

Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, Jiang R, Messer CJ, Chew A, Han JH, Duan J, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu C, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF (2001a) Haplotype variation and linkage disequilibrium in 313 human genes. Science 293:489–493(Erratum: Science 293:1048, 2001)

Stephens M, Smith NJ, Donnelly P (2001b) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

Tajima F (1989) Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Thomas PD, Kejariwal A (2004) Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc Natl Acad Sci USA 101:15398–15403

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003) PANTHER: a library of protein families and subfamilies indexed by function. Genome Res 13:2129–2141

Thompson JD, Gibson T, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Wang YQ, Su B (2004) Molecular evolution of microcephalin, a gene determining human brain size. Hum Mol Genet 13:1131–1137

Wong CG, Bottiglieri T, Snead OC 3rd (2003) GABA, gamma-hydroxybutyric acid, and neurological disease. Ann Neurol 54 (Suppl 6):S3–S12

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555–556

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol 19:908–917

Ziyeh S, Berlis A, Korinthenberg R, Spreer J, Schumacher M (2002) Selective involvement of the globus pallidus and dentate nucleus in succinic semialdehyde dehydrogenase deficiency. Pediatr Radiol 32:598–600