Available online at www.sciencedirect.com
ScienceDirect
Gene 392 (2007) 64–76
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Conservative evolution in duplicated genes of the primate Class I *ADH* cluster

Hiroki Oota [a,*], Casey W. Dunn [b,1], William C. Speed [a], Andrew J. Pakstis [a], Meg A. Palmatier [a], Judith R. Kidd [a], Kenneth K. Kidd [a,b,*]

[a] *Department of Genetics, Yale University School of Medicine, 333 Cedar St., New Haven, CT, USA*
[b] *Department of Ecology and Evolutionary Biology, Yale University, 165 Prospect St., New Haven, CT, USA*

Received 19 July 2006; received in revised form 11 November 2006; accepted 15 November 2006
Available online 23 November 2006
Received by J.G. Zhang

Abstract

Humans have seven alcohol dehydrogenase genes (*ADH*) falling into five classes. Three out of the seven genes (*ADH1A*, *ADH1B* and *ADH1C*) belonging to Class I are expressed primarily in liver and code the main enzymes catalyzing ethanol oxidization. The three genes are tandemly arrayed within the *ADH* cluster on chromosome 4 and have very high nucleotide similarity to each other (exons: >90%; introns: >70%), suggesting the genes have been generated by duplication event(s). One explanation for maintaining similarity of such clustered genes is homogenization via gene conversion(s). Alternatively, recency of the duplications or some other functional constraints might explain the high similarities among the genes. To test for gene conversion, we sequenced introns 2, 3, and 8 of all three Class I genes (total >15.0 kb) for five non-human primates – four great apes and one Old World Monkey (OWM) – and compared them with those of humans. The phylogenetic analysis shows each intron sequence clusters strongly within each gene, giving no evidence for gene conversion(s). Several lines of evidence indicate that the first split was between *ADH1C* and the gene that gave rise to *ADH1A* and *ADH1B*. We also analyzed cDNA sequences of the three genes that have been previously reported in mouse and Catarrhines (OWMs, chimpanzee, and humans) and found that the synonymous and non-synonymous substitution (d_N/d_S) ratios in all pairs are less than 1 representing purifying selection. This suggests that purifying selection is more important than gene conversion(s) in maintaining the overall sequence similarity among the Class I genes. We speculate that the highly conserved sequences on the three duplicated genes in primates have been achieved essentially by maintaining stability of the hetero-dimer formation that might have been related to dietary adaptation in primate evolution.

Keywords: Gene duplication; Gene conversion; ADH; Primates; Negative selection; Coenzyme binding domain

1. Introduction

The alcohol dehydrogenase (ADH) family exists widely in the genomes of bacteria, insects, plants, and vertebrates (Guagliardi et al., 1996; Fischer and Maniatis, 1985; Martinez et al., 1996; Barth and Kunkel, 1979; Canestro et al., 2000; Reimers et al., 2004). All ADH classes form dimers and catalyze oxidization of various kinds and concentrations of alcohols using NAD^+/NADH as coenzyme (Eklund et al., 1976a,b; Höög et al., 2001). The ADH family is classified into five classes (I–V) based on biochemical properties, and nucleotide/amino acid sequence similarity. Humans have three Class I *ADH* genes and one each of Classes II–V (Matsuo and Yokoyama, 1989; Duester et al., 1986; Yokoyama et al., 1992; von Bahr-Lindstrom et al., 1991; Hur and Edenberg, 1992;

Abbreviations: ADH, alcohol dehydrogenase; OWM, Old World Monkey; kb, kilo base pair(s); NWM, New World Monkey; bp, base pairs; cDNA, DNA complementary to RNA; d_N/d_S, synonymous and non-synonymous substitution ratio; K^c, nucleotide sequence difference; BAC, bacterial artificial chromosome; PCR, polymerase chain reaction; np, nucleotide position; MP, maximum parsimony; ML, maximum likelihood; NJ, neighbour joining; ILD, incongruence length difference; Has, *Homo sapiens*; Ptr, *Pan troglodytes*; Ppa, *Pan paniscus*; Ggo, *Gorilla gorilla*; Ppy, *Pongo pygmaeus*; Pap, *Papio anubis*; My BP, million years before present; F_{st}, fixation index; qter, the long arm telomere; cen, the centromere.

* Corresponding authors. Oota, is to be contacted at Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Seimeitou 502, Kashiwa, Chiba 277-8562, Japan. Tel.: +81 4 7136 5421; fax: +81 4 7136 6713. Kidd, Department of Genetics, Yale University School of Medicine, 333 Cedar St., New Haven, CT 06520-8005, USA. Tel.: +1 203 785 2653; fax: +1 203 785 6568.
E-mail addresses: hiroki_oota@k.u-tokyo.ac.jp (H. Oota), kenneth.kidd@yale.edu (K.K. Kidd).
[1] Current address: PBRC University of Hawaii, 41 Ahui St., Honolulu, USA.

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.11.008

Yasunami et al., 1991; Satre et al., 1994); all *ADH* genes cluster on chromosome 4 (4q21–23) in tandem extending >380 kb (International Human Genome Sequencing Consortium, 2001; Kent et al., 2002) (Fig. 1). The high similarity among seven ADH cDNA sequences (60–90%) suggests that those genes have been generated by multiple duplications.

The Class I *ADH* genes have been the best studied in the ADH family, because all three Class I enzymes (ADH1A, ADH1B, ADH1C) of humans are expressed primarily in liver, catalyzing the oxidation of ethanol to acetaldehyde, and the variants of the enzymes have been shown to be associated with protection against alcoholism (Osier et al., 1999, 2002; Edenberg, 2000). The human Class I *ADH* gene cluster spans 80 kb in the physical order of qter-*ADH1C–ADH1B–ADH1A*-cen (Yasunami et al., 1990a,b), and the three genes are similar to each other not only in the exon–intron structure (Fig. 1) but also in the nucleotide sequences of both the exons (>90%) and the introns (>70%) (Matsuo and Yokoyama, 1989; Duester et al., 1986; Yokoyama et al., 1992). This leaves an important question: what mechanism has led to the high degree of conservation in the three Class I genes in the primate lineage? The currently favored explanation is homogenization via gene conversion(s) among the three genes. Cheung et al. (1999) compared the exons and the 5′ and 3′ flanking regions of the Class I *ADH* genes between humans and OWMs, and argued that multiple gene conversion(s) have occurred in the three genes. However, such similar exon sequences (>90%), sharing a very small number of nucleotide differences, provide little information about evolution among such close species.

To better estimate the evolutionary history of the Class I *ADH* genes, we sequenced introns from the 5′- and the 3′-sides (introns 2+3, and 8, respectively) of all three Class I genes (total >15.0 kb) for five non-human primates – four great apes (chimpanzee, bonobo, gorilla, orangutan) and one OWM (baboon) – and compared the sequences to those of humans, in order to see if the phylogenetic tree topology shows evidence of gene conversion(s). In addition, we used the entire Class I region sequences of mouse, chimpanzee, and human, derived from the whole genome sequences, to calculate nucleotide sequence difference (K^c) of the exon/intron sequences, and all the Class I cDNA sequences of mouse, baboon, rhesus macaque, chimpanzee, and humans deposited in the international database (GeneBank/EMBL/DDBJ), to calculate synonymous (d_S) and non-synonymous (d_N) substitution ratios. Here we present the data and explore the mechanisms that have maintained the high similarity of the Class I *ADH* paralogous genes in the primate lineage. We pay particular attention to gene conversion because of the conclusions of Cheung et al. (1999); but we find no evidence of gene conversion.

2. Materials and methods

2.1. Nomenclature

The nomenclature of the gene names follows the official HUGO nomenclature (http://www.gene.ucl.ac.uk/nomenclature/genefamily/ADH.shtml). The old names of the Class I *ADH* genes, *ADH1*, *ADH2* and *ADH3*, have been renamed to *ADH1A*, *ADH1B* and *ADH1C*, respectively, reflecting their coding for Class I enzymes. The traditional names of the gene products, α, β, γ, are called the "ADH1A, ADH1B, ADH1C enzymes or subunits" in this paper to simplify the descriptions.

2.2. Sample DNAs

We prepared the DNA samples from four hominoids (great apes: chimpanzee, bonobo, gorilla, orangutan) and one Old

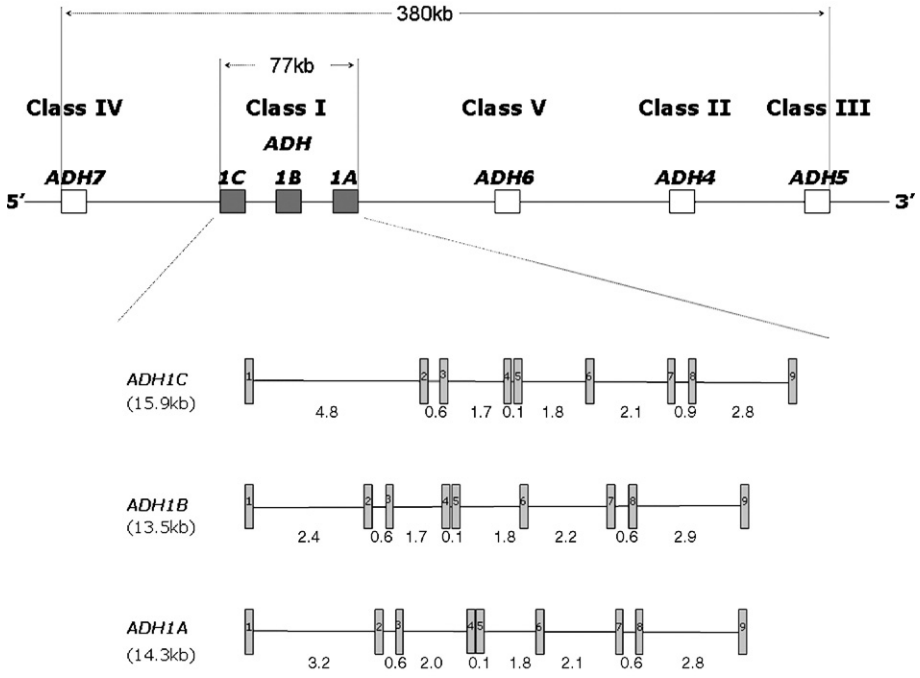

Fig. 1. Map of the human *ADH* gene family. The five classes cluster on chromosome 4 (4q21–23) in tandem.

Table 1
Nucleotide sequence similarity (%) among genes

Genes compared	In2	In3	In8	Ave.	S.E.
Human					
1A–1B	87.0	82.8	90.5		
1A–1C	82.7	82.7	84.6		
1B–1C	83.9	85.3	84.3	84.9	2.54
Chimpanzee					
1A–1B	85.5	82.6	90.7		
1A–1C	83.2	82.8	85.0		
1B–1C	84.8	85.5	84.2	84.9	2.44
Bonobo					
1A–1B	85.5	82.8	90.8		
1A–1C	83.1	83.3	85.2		
1B–1C	85.0	85.3	84.4	85.0	2.39
Gorilla					
1A–1B	86.4	83.2	90.8		
1A–1C	82.8	82.5	85.0		
1B–1C	84.6	85.6	84.5	85.0	2.52
Orangutan					
1A–1B	86.7	82.6	89.7		
1A–1C	81.4	81.8	85.4		
1B–1C	85.4	86.2	85.5	85.0	2.64
Baboon					
1A–1B	84.5	82.4	87.8		
1A–1C	84.4	82.4	84.5		
1B–1C	84.9	85.5	83.6	84.4	1.65

Note: “In,” “Ave,” and “S.E.” represent “introns,.” “average,” and “standard error, respectively.”

World Monkey (OWM: baboon). Genomic DNA of chimpanzee (*Pan troglodytes*; $n=2$), bonobo (pygmy chimpanzee) (*Pan paniscus*; $n=2$), gorilla (*Gorilla gorilla*; $n=2$), and orangutan (*Pongo pygmaeus*; $n=2$) was extracted from lymphoblastoid cell lines established from blood. The BAC (bacteria artificial chromosome) libraries purchased from Children’s Hospital Oakland Research Institute (http://bacpac.chori.org/), chimpanzee (RPCI-43) and olive baboon (*Papio anubis*; $n=1$; RPCI-41), were screened using human PCR products from *ADH1A* and the BAC clone DNAs including the *ADH* Class I region were extracted. These DNA samples from the cell lines and the BAC clones were used as templates for PCR.

2.3. Primer design for PCR and direct sequencing

To assess the argument of Cheung et al. (1999), we chose introns 2 and 3 (5′ side) and intron 8 (3′ side) because the sizes of introns 2 and 3 are almost the same in each of the three genes, and the size of intron 2 and 3 combined is almost the same as that of intron 8 in humans (see Fig. 1). We initially designed the PCR primers based on the human Class I exon/intron sequences deposited in the GenBank/EMBL/DDBJ international DNA sequence database. Many of these primers worked for chimpanzee, bonobo, and gorilla. For orangutan and baboon (especially for baboon), many of the primers based on the human sequences did not work for PCR or sequencing. Therefore, we designed additional primers based on the species-specific sequences already obtained from direct sequencing using the PCR products that had been obtained using human sequence primers. By iterating this process we were able to sequence all long PCR products and fill all gaps. The long PCRs were performed using the “Expand long Template PCR System (Roche)” and the PCR products were purified using “QIAquick PCR Purification Kit (QIAgene).” Sequencing was performed using the purified PCR products on an ABI 3730x1 capillary sequencer of the W.M. Keck facilities in Yale University. The PCR and sequencing primers are available on the web (doi:10.1016/j.gene.2006.11.008) and the PCR conditions are always available by requesting them from us directly. All

Table 2
Nucleotide sequence difference K (×100) among species

	ADH1C					*ADH1B*					*ADH1A*				
	Hsa	Ptr	Ppa	Ggo	Ppy	Hsa	Ptr	Ppa	Ggo	Ppy	Hsa	Ptr	Ppa	Ggo	Ppy
Intron2															
Hsa															
Ptr	0.5					0.3					1.2				
Ppa	0.7	0.3				0.5	0.2				1.4	0.5			
Ggo	0.7	0.8	1.2			0.8	0.7	0.7			1.1	1.2	1.4		
Ppy	2.7	2.8	3.2	3.0		3.5	3.2	3.0	3.7		3.2	3.3	3.4	3.1	
Pap	5.7	6.3	6.3	6.2	6.7	5.7	5.4	5.6	6.0	6.2	5.6	5.9	5.9	5.7	5.9
Intron3															
Hsa															
Ptr	0.4					1.3					1.4				
Ppa	0.4	0.1				1.4	0.3				1.5	0.4			
Ggo	1.1	1.2	1.2			1.3	1.0	1.1			1.3	1.9	2.0		
Ppy	2.4	2.6	2.6	3.1		3.8	3.5	3.6	3.3		2.8	2.9	3.1	3.0	
Pap	7.9	8.1	8.1	8.4	8.1	6.9	6.6	6.7	6.4	6.6	6.5	6.9	7.0	7.5	6.6
Intron8															
Hsa															
Ptr	1.2					0.7					1.0				
Ppa	1.1	0.3				0.6	0.2				0.9	0.2			
Ggo	1.2	1.3	1.2			1.3	1.3	1.2			1.0	0.9	0.8		
Ppy	4.3	4.3	3.9	4.5		2.2	2.2	2.1	2.1		2.5	2.4	2.3	2.5	
Pap	6.5	6.5	6.5	6.4	6.9	7.2	7.1	6.9	7.0	6.9	4.7	4.6	4.5	4.7	4.9

Note: Abbreviations employed are the same as those found in the legend of Fig. 2.

Table 3
The 2×3 tables comparing human and chimpanzee introns

	Observed values				Expected values				Chi-square test	
	Intron2	Intron3	Intron8	Totals	Intron2	Intron3	Intron8	Totals		
ADH1C									0.016	*p*
Number of diff sites	3	7	32	42	4.9171	14.317	22.77	42	8.300	ChiSq
Number of same sites	597	1740	2746	5083	595.08	1732.7	2755	5083	2	*df*
Total	600	1747	2778	5125	600	1747	2778	5125		
ADH1B									0.034	*p*
Number of diff sites	2	23	21	46	5.2561	15.295	25.45	46	6.735	ChiSq
Number of same sites	598	1723	2884	5205	594.74	1730.7	2880	5205	2	*df*
Total	600	1746	2905	5251	600	1746	2905	5251		
ADH1A									0.367	*p*
Number of diff sites	7	29	27	63	6.8118	23.87	32.32	63	2.006	ChiSq
Number same sites	580	2028	2758	5366	580.19	2033.1	2753	5366	2	*df*
Total	587	2057	2785	5429	587	2057	2785	5429		

nucleotide sequence data determined in this study are available in the international database GeneBank/EMBL/DDBJ [Accession numbers:AB243573–AB243602].

2.4. Published data for analyses

In addition to the new sequence data, we also used the chimpanzee, bonobo, and gorilla *ADH1A* intron 2 sequences previously reported by Jensen-Seaman et al. (2001), human, chimpanzee, mouse, and rat genomic sequences, and the human, baboon, rhesus macaque, and mouse cDNA sequences (Matsuo and Yokoyama, 1989; Duester et al., 1986; Yokoyama et al., 1992; Trezise et al., 1989; Cheung et al., 1999; Light et al., 1992; Edenberg et al., 1985) in the database. Since the chimpanzee cDNA has not been isolated, we superimposed the human exon sequences on the chimpanzee genome sequences

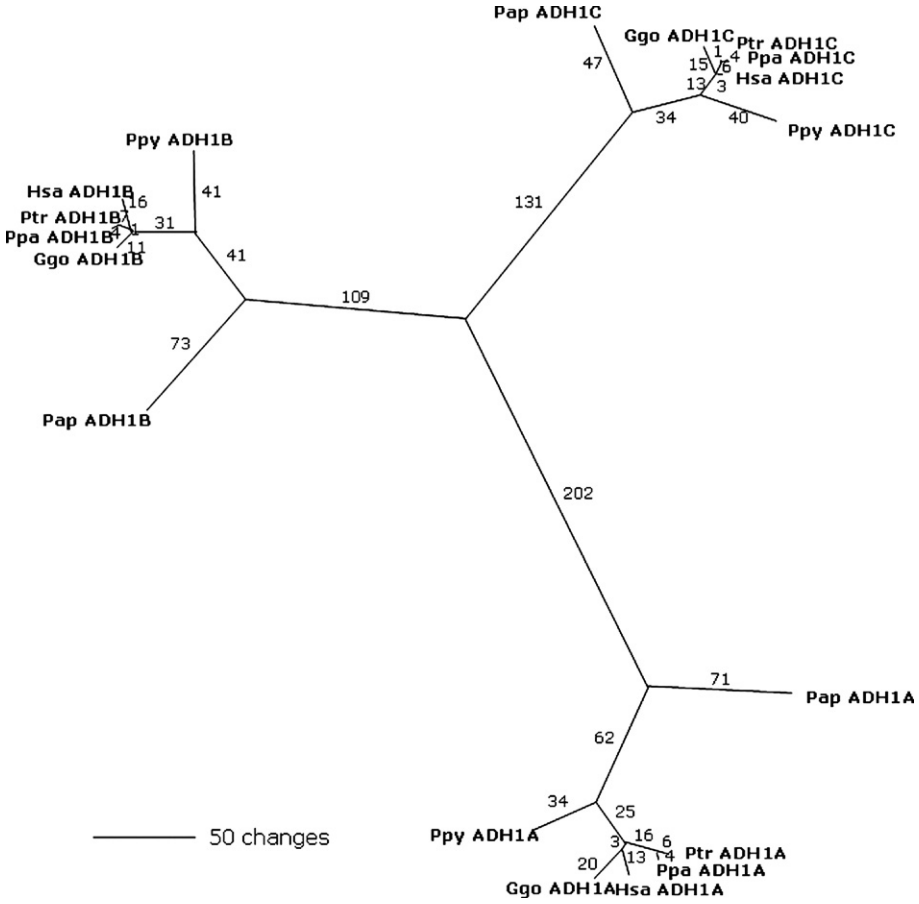

Fig. 2. The maximum parsimony (MP) tree based on nucleotide sequences (total approx. 5.1 kb) of three introns. The numbers on the branches represent the nucleotide differences between the nodes. Abbreviations are, Hsa: *Homo sapiens*, Ptr: *Pan troglodytes*, Ppa: *Pan paniscus*, Ggo: *Gorilla gorilla*, Ppy: *Pongo pygmaeus*, Pap: *Papio anubis* (to avoid confusion with genus *Pan*).

from the whole genome sequencing and obtained the chimpanzee exon sequences. Those nucleotide sequence data were used to compare with the entire human *ADH* gene cluster (Classes I to V) sequences.

2.5. Alignments and phylogenetic analyses

We constructed phylogenetic trees based on nucleotide sequences by maximum parsimony (Fitch, 1977) and maximum likelihood (Cavalli-Sforza and Edwards, 1967; Felsenstein, 1973, 1981) methods. The nucleotide and amino acid sequences were aligned by the program CLUSTAL W (Thompson et al., 1994) through the DDBJ browser on the webpage (http://www.ddbj.nig.ac.jp/). The alignment was then modified by eye using MacClade v4.06 (Maddison and Maddison, 2003). PAUP version 4.03B10 (Swofford, 1998) was used to search for the best tree as evaluated by maximum parsimony (MP) and maximum likelihood (ML). The chosen MP tree is a consensus tree with 1000 bootstrap replications. MODELTEST v3.06 was used to evaluate sequence evolution models under the likelihood ratio test (Posada and Crandall, 1998).

We tested the significance of the difference in the topologies of the trees estimated from the 5′- and 3′-regions using the incongruence length difference (ILD) test (Farris et al., 1995). The test statistic of the ILD test is the difference between the parsimony score of the best tree found from the combined dataset and the sum of the parsimony scores for the best trees for each of the independent partitions. If there is no incongruence (difference between trees) the test statistic will be zero. The significance of the test statistic is estimated by comparing the test statistic for the real partitions to a null distribution generated by randomly sampling new partitions from the combined dataset.

The SOWH parametric bootstrap test (Goldman et al., 2000; Hillis et al., 1996; Huelsenbeck et al., 1996) was used to determine the significance of the difference between the best score overall and the score of the best tree in searches constrained to be consistent with particular topological hypotheses. The test was implemented by constructing pseudoreplicate datasets with seqgen (Rambaut and Grassly, 1997), analyzing these with PAUP*, and parsing with scores from each replicate with extract scores (Dunn et al., unpublished data).

2.6. K^c, d_N and d_S estimation

The proportion of sites differing (K^c) after adjusting for gaps and ambiguities was calculated using the program PAUP

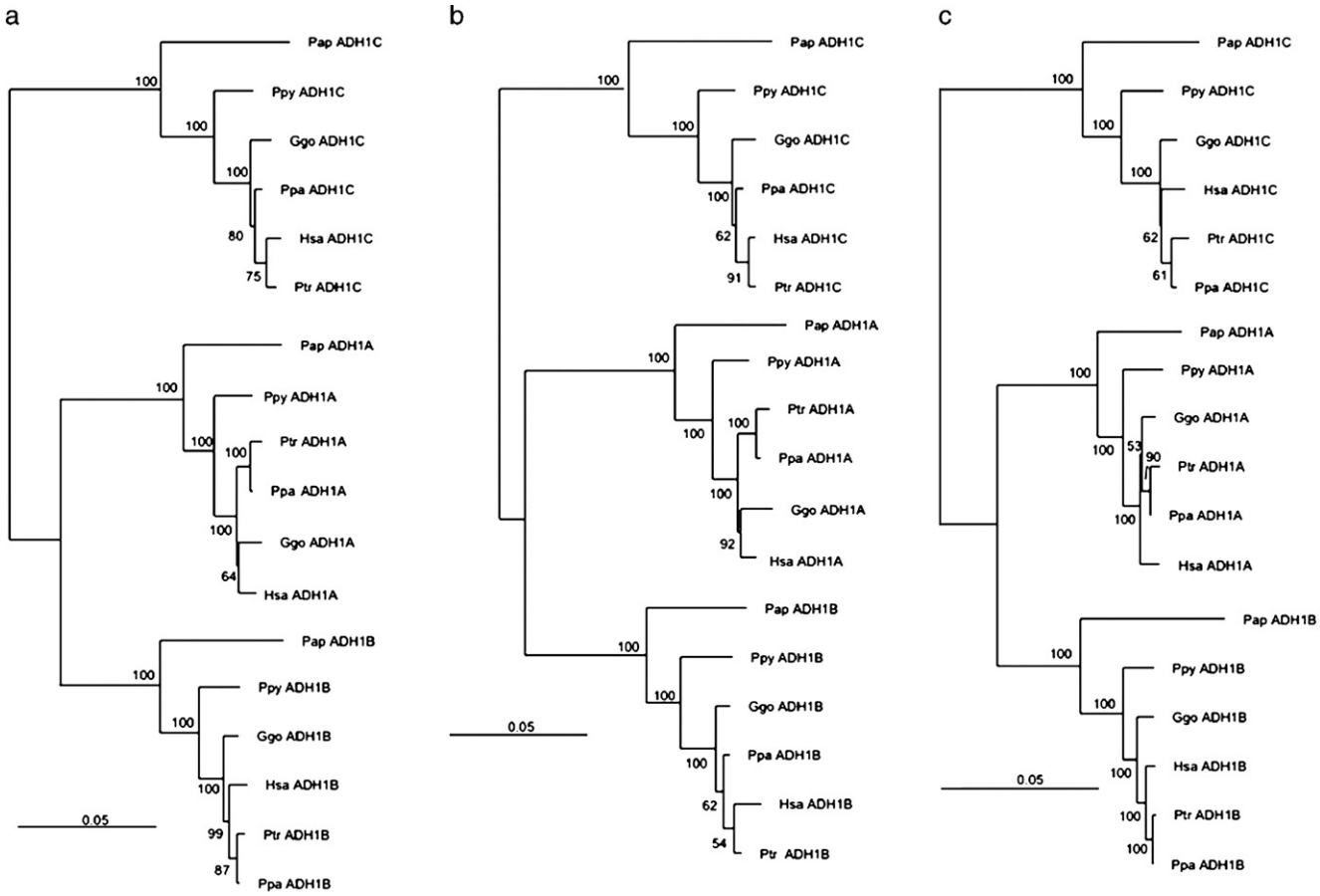

Fig. 3. The maximum likelihood (ML) trees under the GTR+G sequence evolution model (Lanave et al., 1984; Rodriguez et al., 1990). The illustrated trees are the best out of 100 heuristic searches. Bootstrap support values shown at the nodes (100 bootstrap replicates were run). a. The ML tree based on introns 2, 3, and 8 and exons 3 and 9; b. the ML tree based on the 5′-side (intron 2+3, exon 2) nucleotide sequences; c. the ML tree based on the 3′-side (intron 8, exon 9) nucleotide sequences. Abbreviations are the same as those noted in Fig. 2 legends. These searches were not rooted, but other analyses indicated that the root is along the branch that separates *ADH1C* from the other genes.

version 4.03B10 (Swofford, 1998). The non-synonymous (d_N) and synonymous (d_S) substitutions were estimated by the Nei and Gojobori (1986) method using the program MEGA3.1 (Kumar et al., 2004) and were examined if d_N is significantly different from d_S by a *Z*-test (Nei and Kumar, 2000).

3. Results

3.1. Nucleotide sequence diversity

The sequences (introns 2, 3 and 8: approximately 0.6, 2.0, and 3.0 kb in humans, respectively) obtained from all three genes clearly have similarity within each of the six species (Table 1), yet show relatively substantial differences among species and among genes (Table 2). In Table 1, all three introns have very high similarity (83–91%). The highest similarity among the sequence pairs is observed in intron 8 between *ADH1A* and *ADH1B* (90–91% in apes, and 88% in baboon); these are not so different from all other comparisons that average 85±2.4% in apes and 84±1.6% in baboons. Table 2 shows pairwise nucleotide difference (K^c) values among species. Except for orangutan, the four hominoids (human, chimpanzee, bonobo, gorilla) commonly have fewer differences from one another, while orangutan and baboon differ more from the four African "apes" and from each other.

The K^c values between human and chimpanzee are extremely small in *ADH1C* and *ADH1B* intron 2 (0.5 and 0.3, respectively), and *ADH1C* intron 3 (0.4). In a simple chi-square heterogeneity test comparing human and chimpanzee sequences, the 2×3 (the number of the different/same sites and three introns) tables for three genes (Table 3) show statistical significance in *ADH1C* and *ADH1B* ($p=0.016$ and $p=0.034$, respectively). Thus, the K^c values among species vary among Class I genes and among introns, at least for these two closely related species, though overall the Class I *ADH* introns are highly similar to one another within each species.

3.2. Phylogenetic topology and rooting

Fig. 2 shows the MP tree (1000 bootstrap replications) based on the nucleotide sequences combining the three introns 2, 3, and 8 (total approx. 5.3 kb) in each of the three Class I *ADH* genes. This tree shows that intron sequences for each species clearly do cluster within each gene, a topology with no indication of gene conversion(s). While there is no evidence of gene conversion, this does not exclude gene conversion affecting only a small segment of this 5.3 kb of sequence as discussed later. The ML analysis of the same dataset also resulted in a tree with the sequences of all the species clustered within each gene (Fig. 3a).

Several different approaches were used to determine where the root of the primate Class I *ADH* gene tree is. We first used the intron sequences for the single mouse Class I *ADH* gene (Ceci et al., 1987; Mouse Genome Sequencing Consortium, 2002) as the outgroup. The introns 2, 3, and 8 from mouse Class I *ADH* gene are approximately 0.5, 1.9, and 0.6 kb, respectively. Because the mouse *ADH* intron 8 (0.6 kb) is much shorter than those of the primates (3.0 kb) examined in this study, intron 8 was removed from the analysis, and introns 2 and 3 were used for evaluating the tree. In both introns 2 and 3, each sequence does cluster within each gene as in the tree of Fig. 2, and the mouse sequence roots the primate Class I *ADH* tree along the branch that gives rise to the *ADH1C* genes (tree not shown).

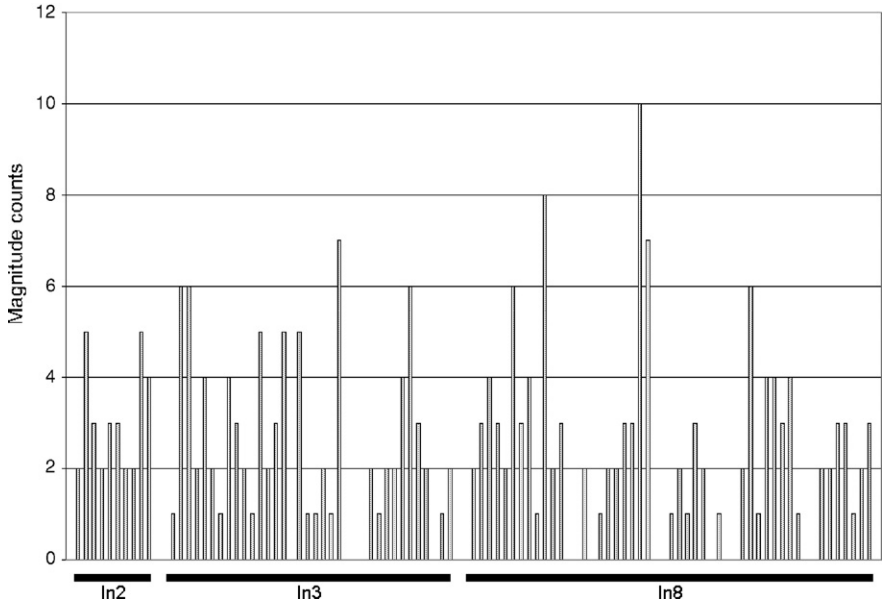

Fig. 4. The magnitude counts of "shared" sites in the three introns among three Class I ADH genes. In Appendix A, pink, light orange, and light blue colored nucleotides represents *ADH1A*, *ADH1B*, and *ADH1C* specific sites, respectively. Dark blue represents "singleton" that is a different nucleotide from others happened only in one gene in one species. Yellow represents "specific polymorphic site" that is polymorphism observed in one gene in one species. Red and brown represent "shared sites across genes," whereas orange represents "shared nucleotides within a gene. Here we counted red and brown sites as "shared" sites. The *X* axis represents the physical regions of the introns, and the *Y* axis represents the magnitude counts of the shared sites. The counting interval covers 60 bp, so one bar represents the number of the shared sites in 60 bp. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The bootstrap replicates, however, show no support for this particular topology (less than 0.1%), probably because the mouse Class I *ADH* gene is too divergent from the primate genes to be a suitable outgroup in analyses of intron data.

Next, we constructed an exon-only dataset, which was aligned without problems, to see which rooting it supports. This dataset contained all available complete coding sequences for primate Class I *ADH* genes (human and chimpanzee *ADH1A*, *ADH1B*, and *ADH1C*; rhesus *ADH1A*, baboon *ADH1B*, and baboon *ADH1C*), as well as the mouse Class I *ADH* gene as an outgroup (Refs. are in Materials and methods). MODELTEST selected the K80+G sequence evolution model for this dataset. MP (1000 branch and bound bootstrap replicates) and ML (100 branch and bound bootstrap replicates) analyses of this dataset supported a bipartition between "primate *ADH1C*+mouse Class I *ADH*" and "primate *ADH1A*+primate *ADH1B*" (MP bootstrap value: 80%; ML bootstrap value: 86%). This suggests that the root of the primate Class I *ADH* tree occurs along the branch that separates *ADH1C* from *ADH1A* and *ADH1B*. Finally, midpoint rooting of the most parsimonious primate Class I *ADH* tree in PAUP supported the same root location as the analyses that included mouse as an outgroup, i.e. along the branch that gives rise to the primate *ADH1C* genes. These separate lines of inquiry are all consistent with the hypothesis that the first split in primate Class I *ADH* genes was between *ADH1C* and a common ancestor of *ADH1A* and *ADH1B*. The second split was between *ADH1A* and *ADH1B*. These results agree with our preliminary analyses using the neighbour joining (NJ) tree (Saitou and Nei, 1987) including all ADH classes based on cDNA and amino acid sequences from human,

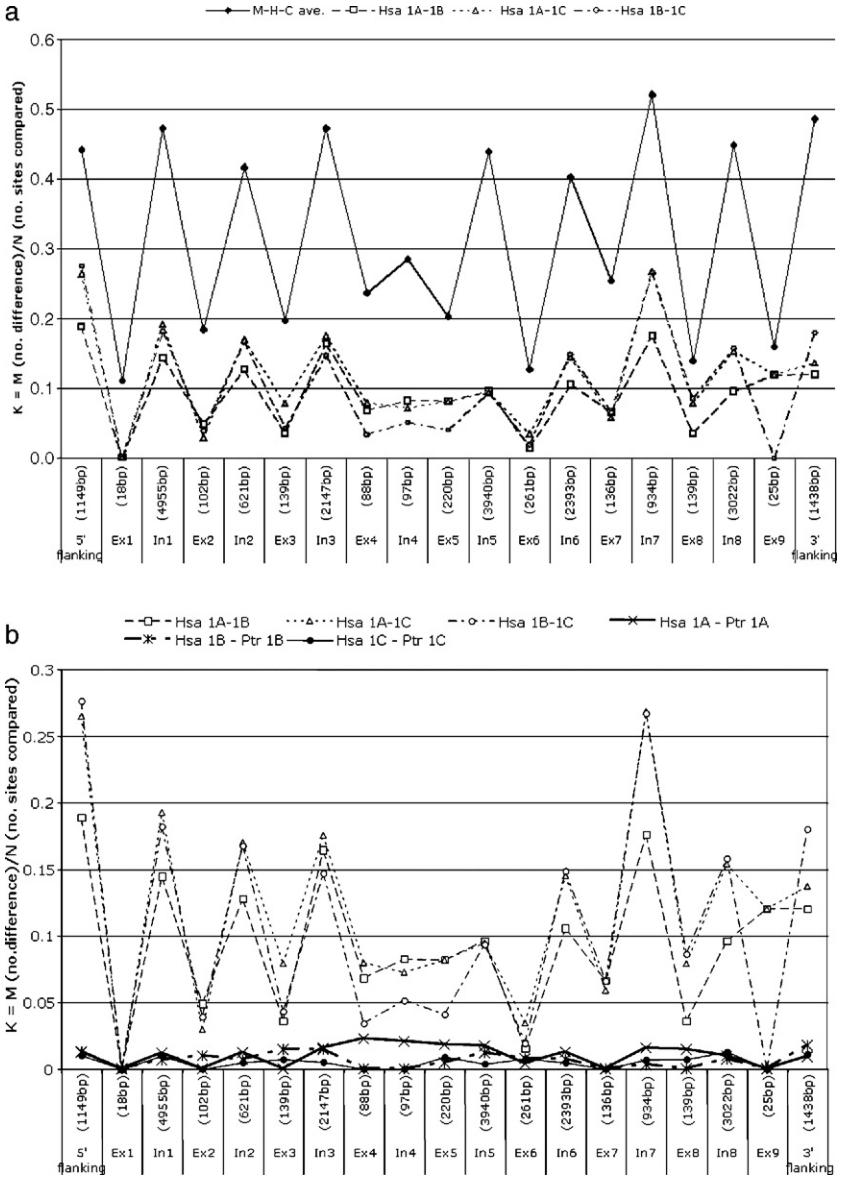

Fig. 5. The nucleotide sequence difference (K^c) values for the exons and the introns: a. the K^c values between mouse and human/chimpanzee (the average between mouse and human, and between mouse and chimpanzee) homologous, and between human Class I *ADH* paralogous genes; b. the K^c values between human Class I paralogous genes, and between human and chimpanzee orthologous genes. The K^c values are in the *Y* axis, and the physical order of exons/introns in the *X* axis. Abbreviations are, Hsa: *Homo sapiens*, Ptr: *Pan troglodytes*, M–H–C: Mouse–Human–Chimpanzee, 1A: *ADH1A*, 1B: *ADH1B*, 1C: *ADH1C*.

baboon, horse, mouse, rat, fish, amphioxus, and *E. coli* plasmid (tree not shown) derived from databases, and also with previously reported trees (Yokoyama and Yokoyama, 1987; Höög et al., 2001).

3.3. Shared-site distribution

Gene conversion(s) might have occurred in very short regions of these duplicated genes that cannot be detected from the overall topology. To detect such tiny and short gene conversion(s), we count the magnitude of "shared" sites where two nucleotides at variable sites are shared by two or more genes across the species: looking at sites potentially informative for the "phylogeny" of the three genes. These would potentially indicate gene conversion. Fig. 4 shows that the "shared" sites are distributed broadly across the three introns, but not in particular regions (all the alignments of introns 2+3, and 8 are shown in Appendix A that is available on the doi:10.1016/j.gene.2006.11.008). A non-uniform pattern of shared sites would be a signal of the short/partial gene conversion(s). The shared sites occur, however, rather uniformly in the three introns (Fig. 4), indicating no distinction between the shared sites caused by gene conversion(s) and those caused by chance in these genes. We also counted the magnitude of the "gene specific" sites where the same nucleotide is shared with all the six primates in one gene, but it is different in the other two genes. The "gene specific" sites are also distributed broadly in three introns and do display uniform patterns (see Appendix A). Thus, the nucleotide sequence alignments for those introns show no evidence for tiny/short gene conversion(s).

The 67 polymorphic sites that had two alleles in at least one species are distributed as follows: 40 were in intron 2+3 region, and 27 are in intron 8. The 39 out of 67 sites that are singletons occur only in the particular species or genes but are not shared by other species or in other genes. A total of 28 out of 67 sites are shared with other genes and/or with the other species, but none of these polymorphisms can be attributed to gene conversion(s) (see Appendix A).

3.4. Nucleotide sequence difference comparisons in exons

Regarding concerted evolution including gene conversion, an "intervening sequence mediated domain transfer" has been proposed that argues genetic information transfer is more likely to occur in introns flanking exon(s) corresponding to enzymatic functional domains (Miyata et al., 1980). Fig. 5a shows nucleotide sequence difference (K^c) values between mouse and human (or chimpanzee), and between human Class I *ADH* genes. As expected, the K^c values in the exons are always lower than those in introns, clearly showing that exons are much more conserved than introns, because functional constraint on the protein affects primarily the exons. No pairwise comparison between human Class I genes (Hsa *1A–1B*, Hsa *1A–1C*, and Hsa *1B–1C*) shows higher K^c values than those between mouse and human in either exons or the introns, indicating the split time of the Class I *ADH* genes is more recent than the divergence time between the two species.

Fig. 5b shows the K^c values between human Class I paralogous genes, and those between human and chimpanzee orthologous genes. The K^c values between human and chimpanzee orthologous genes are consistently much lower than those between human Class I paralogous genes. If the K^c values between human Class I genes were lower than those between human and chimpanzee in the same genes, it would be a strong signal of gene conversion(s) in the human Class I genes. However, the K^c values do not show a signal for gene conversion(s) in any exons or introns. When we look at K^c values between the chimpanzee Class I genes, the results are the same as what we find for the human Class I genes (data not shown). Thus, we do not find any evidence for gene conversion(s) in the K^c-value comparisons between human and chimpanzee exon/intron sequences from the Class I *ADH* genes.

4. Discussion

Our analyses thus do not provide any evidence supporting gene conversion(s) for both exon and intron sequence data (Figs. 2, 3a, 4, and 5b). This does not mean that gene conversion(s) has never occurred in the primate Class I *ADH* cluster, but it appears that gene conversion(s) is not the main factor accounting for the high degree of similarity between the three Class I genes (Fig. 1 and Table 1).

A previous study using *ADH* Class I cDNA, based on matrix phylogenetic compatibility, argued that at least three gene conversion events affecting the exons and 3′-noncoding regions had occurred in both humans and OWMs (Cheung et al., 1999). MP trees from this earlier study were constructed using nucleotide sequences from three regions, (1) 5′-noncoding, (2) exons 2–5, and (3) exons 7–9 and 3′-noncoding (see Fig. 4 in Cheung et al., 1999). The topology of their MP tree of the 5′-noncoding region shows no indication of gene conversion(s). But, in exons 2–5, and exons 7–9 and 3′-noncoding regions, the two trees show different branching and clustering patterns than the topology of the 5′-noncoding region tree, and some genes do cluster within the species rather than paralogs, indicating the conversion-like topologies in the two trees. These reticulations indicated by tree topologies were the core argument for gene conversion by Cheung et al. (1999). The problem with this interpretation, though, is that the exons are so similar that they have little phylogenetic information.

To explore this hypothesis of intragenic reticulation with our new data, we partitioned our data into a 5′-side region (intron 2+3) and a 3′-side region (intron 8). We then constructed two ML trees from these two regions (the trees from the 5′ and 3′ regions are presented in Fig. 3b and c, respectively). The topology of the tree based on the 5′ region (Fig. 3b) differs little from the tree based on both regions (intron 2+3+8, Fig. 3a), whereas that of the 3′ region (Fig. 3c) differs within both the *ADH1C* and the *ADH1A* clades. The clustering patterns in human, chimpanzee, and bonobo are slightly different between the 5′- and 3′-side introns in the *ADH1C*, the *ADH1A*, and the *ADH1B* clades in the ML trees; the branches are weakly supported by the maximum likelihood bootstrap values (Fig. 3b and c), exhibiting the discrepancy between the trees for the 5′- and

3′-side intron regions similar to those observed by Cheung et al. (1999). The ILD test (following 100 partition-homogeneity test replicates) found that the observed differences in topology between the 5′- and 3′-partitions was not significant ($p=0.14$). The null hypothesis that the topologies of the maximum parsimony trees for each partition are congruent could not be rejected. A lack of incongruence is not consistent with the hypothesis that the 5′- and 3′-regions of the gene have had different histories, suggesting that differential intragenic reticulation has not occurred.

Gene conversion homogenizes two duplicated genes, whereas positive selection and neutral fixation of mutations lead the two genes to diverge (Ohta, 1988, 1993, 2000a). In order that the two duplicated genes may diverge, the two genes need to "escape" from the homogenization caused by gene conversion (Innan, 2003). Eventually, duplicated genes can diverge when difference accumulation caused by positive selection or neutral fixation exceeds homogenization by gene conversion (Teshima and Innan, 2004). The estimation of the rates of synonymous and non-synonymous substitution is efficient for detecting natural selection (Ohta, 2000b). To assess the neutrality of the Class I *ADH* genes, we estimated the number of synonymous substitutions per synonymous site (d_S) and the number of non-synonymous substitutions per non-synonymous site (d_N) (Nei and Gojobori, 1986) in human, chimpanzee, OWMs (rhesus macaque and baboon), and mouse, for which all nine exon sequences from at least one gene are available. According to the generally accepted criteria, a d_N/d_S ratio not significantly different from 1 indicates neutral evolution; $d_N>d_S$ indicates positive selection, and $d_N<d_S$ indicates negative (purifying) selection. Table 4 shows the d_N and d_S estimates between the homologous genes in the five species. The probabilities for the significance of the difference are presented with a *Z*-test that computes the variance of (d_N-d_S) by bootstrapping (Nei and Kumar, 2000). In all of the combinations compared here, the d_N/d_S ratios of the whole cDNA sequences are always significantly different from 1, except for the ratio between human and chimpanzee in *ADH1A* ($d_N/d_S=0.382$; $p=0.119$) probably because the d_N and d_S values are so small. This suggests that overall the Class I *ADH* genes have evolved through purifying selection. At the very least, there is no evidence that positive selection has operated on the Class I *ADH* gene cluster in human, chimpanzee, OWMs, and mouse, to drive them apart and to "escape" from the possibility of homogenization by gene conversion.

The intron sequences do show several signs indicating their evolution is unlikely to be purely neutral: (1) the extremely low K^c values observed between human and chimpanzee in a couple of introns (for instance, the average between human and chimpanzee in the three genes is 0.9 but the K^c value is 0.3 in *ADH1B* intron 2, see Table 2); (2) the maximum likelihood test chooses the non-clock model for the intron sequence data; (3) the ML tree (Fig. 3a) shows discrepancies in each branch from the topology of the "species" tree; (4) the topological discrepancies between the trees of Fig. 3a and b, and between the trees of Fig. 3a and c are statistically significant in the SOWH test (Goldman et al., 2000; Hillis et al., 1996; Huelsenbeck et al., 1996). Regarding the signs (2)–(4), the most plausible explanation may be that the divergence time (4–7 My) is too short to accumulate sufficient independent neutral mutations among the four apes (human, bonobo, chimpanzee, and gorilla) who show the topological discrepancies from the "species." The two other species (baboon and orangutan) which have sufficiently longer divergence times from the other four apes (15 and 25 My, respectively) have no discrepancy in the tree topologies. The lower K^c values among the four closely related species (Table 2) overall can also be explained by the divergence time of the species being too recent. However, the remarkably low K^c values between human and chimpanzee in the *ADH1C* and the *ADH1B* intron(s) 2 and/or 3 is statistically significant (Table 3). Because the three very similar Class I *ADH* genes exist only in an ~80 kb region, the mutation rates depending on the GC contents and the effect of drift should be equal within a species. In addition, high linkage disequilibrium (LD) values are observed in the Class I *ADH* cluster for human populations from around the world (Osier et al., 1999, 2002). It is unlikely that only the recent divergence time explains the remarkably low K^c values observed in the particular parts, the *ADH1C* and the *ADH1B* intron(s) 2 and/or 3. We are unable to ignore the possibility that there might be functional constraints, such as intronic regulatory elements for gene expression (i.e., Kleinjan et al., 2004), on the *ADH1C* and the *ADH1B* introns 2 and 3 through human and chimpanzee lineages.

The exon sequences also show some indications of potential functional constraints. The ADH enzymes consist of a catalytic and a coenzyme binding domains (Eklund et al., 1976a,b; Duester et al., 1986; von Bahr-Lindstrom et al., 1986; Hurley et al., 1991; Davis et al., 1996). Fig. 6 shows the domain structure of the ADH enzymes. The coenzyme binding domain (np529–np957) covers almost all of exons 6 and 7, and a part of exon 5. Because the coenzymes and the substrates bind in the cleft between the catalytic and the coenzyme binding domains in the ADH dimers (Eklund et al., 1976b; Svensson et al., 2000; Höög et al., 2001), it is predicted that the coenzyme domain is conserved, whereas the catalytic domains from the ADH enzymes varied corresponding to the various substrates (Svensson et al., 2000). We calculated the d_N and d_S values for the N- and the C-terminal catalytic, and the coenzyme domains (Table 4). Between the human and the mouse domains, d_N values are always smaller than d_S (significantly different at the 5% level) as well as observed in the whole cDNA. In the comparisons among the primates, again d_N is always smaller than d_S but not significantly so between all pairs in the C-terminal catalytic domains, or between four pairs (*ADH1A* human–chimpanzee, *ADH1B* human–chimpanzee; human–baboon, and *ADH1C* human–chimpanzee) in the N-terminal catalytic domains, suggesting effective neutrality in those pairs. Meanwhile, the d_N/d_S ratios in the coenzyme domain are smaller than those of the whole cDNA and the catalytic domains, and the differences are statistically significant, except for human–chimpanzee in *ADH1A* and *ADH1B*, suggesting purifying selection has operated mainly in the coenzyme domain that contributes to purifying selection in the total enzyme. The higher rates of synonymous substitutions observed in the coenzyme binding

Table 4
d_S and d_N estimates among mouse, OWM, chimpanzee, and human

				Whole cDNA (np1–1125)				N-terminal catalytic domain (np1–528)				Coenzyme domain (np529–957)				C-terminal catalytic domain (np958–1125)			
				d_S	d_N	d_N/d_S	p	d_S	d_N	d_N/d_S	p	d_S	d_N	d_N/d_S	p	d_S	d_N	d_N/d_S	p
ADH1A	Human	vs	Mouse Class I *ADH*	0.786	0.100	0.127	<1E−7	0.750	0.141	0.188	<0.001	0.815	0.060	0.074	<0.001	0.834	0.076	0.091	0.033
	Chimpanzee	vs		0.750	0.102	0.136	<1E−7	0.718	0.142	0.198	<0.001	0.788	0.060	0.077	<0.001	0.756	0.085	0.112	0.035
	Rhseus	vs		0.729	0.109	0.149	<1E−7	0.700	0.156	0.223	<0.001	0.729	0.067	0.092	<0.001	0.837	0.072	0.086	0.029
ADH1B	Human	vs		0.723	0.098	0.136	<1E−7	0.754	0.127	0.168	<0.001	0.701	0.074	0.106	<0.001	0.686	0.073	0.106	0.031
	Chimpanzee	vs		0.751	0.096	0.127	<1E−7	0.773	0.121	0.156	<0.001	0.749	0.074	0.099	<0.001	0.686	0.073	0.106	0.031
	Baboon	vs		0.759	0.101	0.133	<1E−7	0.748	0.130	0.174	<0.001	0.781	0.078	0.099	<0.001	0.735	0.073	0.099	0.024
ADH1C	Human	vs		0.762	0.088	0.116	<1E−7	0.727	0.125	0.172	<0.001	0.864	0.061	0.070	<0.001	0.622	0.048	0.076	0.022
	Chimpanzee	vs		0.772	0.088	0.114	<1E−7	0.770	0.128	0.166	<0.001	0.836	0.061	0.072	<0.001	0.618	0.040	0.064	0.020
	Baboon	vs		0.710	0.096	0.136	<1E−7	0.652	0.124	0.190	<0.001	0.773	0.074	0.096	<0.001	0.736	0.068	0.092	0.029
ADH1A	Human	vs	Chimpanzee	0.015	0.006	0.382	0.119	0.016	0.010	0.613	0.293	0.009	0.000	0.000	0.142	0.028	0.008	0.268	0.250
	Human	vs	Rhesus	0.117	0.030	0.255	<0.001	0.146	0.034	0.235	0.001	0.068	0.013	0.183	0.022	0.167	0.059	0.356	0.085
	Chimpanzee	vs	Rhesus	0.099	0.031	0.312	0.001	0.126	0.034	0.272	0.004	0.058	0.013	0.215	0.038	0.134	0.068	0.506	0.148
ADH1B	Human	vs	Chimpanzee	0.019	0.002	0.130	0.022	0.024	0.005	0.209	0.089	0.019	0.000	0.000	0.090	0.000	0.000	nc	nc
	Human	vs	Baboon	0.061	0.015	0.253	0.003	0.057	0.020	0.356	0.058	0.058	0.009	0.163	0.029	0.082	0.016	0.190	0.097
	Chimpanzee	vs	Baboon	0.073	0.013	0.179	<0.001	0.083	0.015	0.183	0.008	0.058	0.009	0.163	0.031	0.082	0.016	0.190	0.097
ADH1C	Human	vs	Chimpanzee	0.026	0.002	0.088	0.007	0.024	0.003	0.104	0.060	0.028	0.000	0.000	0.034	0.027	0.008	0.288	0.242
	Human	vs	Baboon	0.139	0.026	0.184	<0.001	0.182	0.034	0.189	<0.001	0.109	0.013	0.115	0.003	0.085	0.031	0.370	0.146
	Chimpanzee	vs	Baboon	0.134	0.026	0.191	<0.001	0.193	0.037	0.192	<0.001	0.099	0.013	0.128	0.004	0.055	0.023	0.425	0.237

Note: d_S and d_N values are calculated using the Nei–Gojobori substitution model with the Jukes–Cantor model.
We show p-values with Z-test for the difference between d_S and d_N (Nei and Kumar, 2000).
The d_S/d_N values underlined are not significantly different from 1 at the 5% level in the Z-test.
The estimation of d_S and d_N and the Z-test calculations were carried out using the program MEGA3.1 (Kumar et al., 2004).
nc indicates not computable.

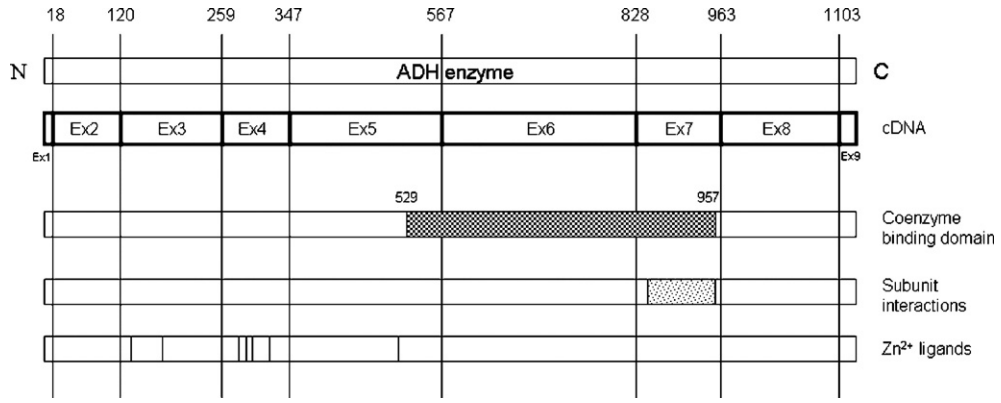

Fig. 6. The domain structure of the ADH enzymes. The style for displaying the structure imitates Fig. 2 of Duester et al. (1986). The vertical lines represent the boundary between exons and introns, and the numbers on the lines represent the nucleotide position number (np) at the end of the former exon.

domain than within the catalytic domains are also reported for duplicate *ADH* genes in plants (Yokoyama et al., 1990).

The highly conserved sequences in the binding domains described above suggest that the selection for hetero-dimer stability and the acquisition of regulatory elements determining the organ-dependent expression levels could have been keys to evolution of the Class I genes. All the human Class I ADH enzymes are expressed in liver, kidney, stomach, small intestine, ileum, colon, uterus, lung and skin (Engeland and Maret, 1993). The Class I ADH subunits combine randomly and form homo- and hetero-dimers; the different combinations have different ethanol catalytic efficiencies (von Bahr-Lindstrom et al., 1986). To metabolize various concentrations of ethanol in different organs, various combinations of the homo- and the hetero-dimers of the Class I subunits would be necessary, and the stabilities of the hetero-dimers would be very important in any subunit combinations. The highly conserved binding domain is probably maintained by such functional constraints. The duplication of the Class I ADH genes may be associated with dietary adaptations of primates. Some of the current leaf-eaters (i.e., Colobus belonging to OWMs) ferment leaves in their foregut (Kay and Davies, 1994), though smaller primates including prosimians or fossil primates are mostly insectivorous (Fleagle, 1988, 1999). In the process of fermentation, the ADH enzymes must play an important role in digesting any alcohols generated by the fermentation. When ancestral primates ate fermented fruits and/or leaves, having various dimer combinations could have been advantageous to digest various types and concentration of alcohols.

Here we have another question: when did the duplication events of the Class I occur? Mouse is too divergent to be a useful outgroup for analyses of primate Class I *ADH* intron evolution as mentioned above. This is problematic for duplication timing estimates based on primate intron data, as these analyses require rooted trees. Because multiple lines of evidence suggest that the first split in primate Class I *ADH* was between *ADH1C* and a common ancestor of *ADH1A* and *ADH1B*, and the second split was between *ADH1A* and *ADH1B*, *ADH1C* genes can be used as an outgroup to root and date the split between *ADH1A* and *ADH1B* genes. MODELTEST selected the HKY+G sequence evolution model (Hasegawa et al., 1985) for the *ADH1A* and *ADH1B* intron dataset. Human *ADH1C* was used to root further analyses of this dataset. The molecular clock was rejected ($p<0.001$). Bootstrapped likelihood analyses gave a mean age for the split between the *ADH1A* and *ADH1B* paralogs of 56.0 My BP (95% confidence interval: 55.6 My–56.3 My) when the divergence of baboon from hominoids was fixed at 25 My (Fleagle, 1988, 1999).

We also estimated the split times assuming the "linear evolutionary rate" and baboon divergence at 25 My BP; the splits of *ADH1A*/*ADH1B* and *ADH1C*/*ADH1A* and *1B* are estimated as 44–70 My BP and 54–84 My BP. This estimate of the time of the split between the *ADH1A* and *ADH1B* encompasses the previous estimate. Recent discoveries of fossil primates and molecular studies push the OWMs split time back at least 10 My (Martin, 1993; Takahata and Satta, 1997). So if we assume a divergence time of 35 My, the estimates obtained must be pushed back to much older values (data not shown). For the initial duplication, the estimated time is sufficiently imprecise to virtually encompass the entirety of the initial mammalian radiation into the classes seen today. Fossil data show that the first primates emerged 66–98 My BP, the first prosimians emerged before 66 My BP, and the divergence between New World Monkeys (NWMs) and OWMs is 36–55 My BP (Martin, 1993). Therefore, we hypothesize that the second duplication probably occurred around or just before prosimians emerged, and the first duplication occurred probably during the mammalian radiation. Thus, our current data argue that at least the second duplication event of the Class I *ADH* genes occurred within the primate lineage. It seems highly unlikely that even the earliest duplication event occurred before divergence between rodents and primates with subsequent loss of all but one copy in the rodents. However, given the still uncertain relationships of mammalian Classes, it is possible that the two copies from that initial duplication are present in *Perissodactyls* (horses). Considerable additional data will be required to resolve that question.

Gene conversion might have occurred within populations in each species, but it is not detectable in this study because we examined at most only two individuals for each species. The pattern of polymorphic sites observed in one species in one gene does not show the signal of gene conversion. A population study

would be required next to search for gene conversion(s) that occurred more recently. Human population genetic data have given strong evidence that the genes related to alcohol metabolism have undergone changes that reflect the work of adaptive processes. In studies of human populations from around the world, the Class I *ADH* gene cluster and the aldehyde dehydrogenase 2 (*ALDH2*) gene locus each show very high linkage disequilibrium (LD) values, indicating those regions have low recombination rate (Osier et al., 2002; Oota et al., 2004), while there is no indication for gene conversion in the Class I *ADH* cluster. The ALDH2 enzyme catalyzes acetaldehyde oxidization that is the second step of ethanol metabolism (Oota et al., 2004). Interestingly, both of the loci have East Asian specific haplotypes at markedly high frequencies and several single nucleotide polymorphisms (SNPs) in the loci have very high F_{st} values, suggesting that positive selection has operated in those SNPs or in the closely related loci (Osier et al., 2002; Oota et al., 2004). Thus, the *ADH* and *ALDH2* data from human global populations support the idea that the polymorphisms of the genes related to ethanol metabolism have undergone adaptive changes despite the relatively short time frame represented by the evolution of modern humans. This also implies that the evolution of the Class I *ADH* genes is associated with dietary adaptations in the primate lineage.

The sequence data suggest that purifying selection played a more important role than gene conversion for retaining high similarity between the primate Class I *ADH* genes; accordingly, we speculate that this is the result of selection for homo- and hetero-dimer stability. It might be a general phenomenon that duplicate genes survive by purifying selection in genes coding subunits to form dimers. Since there are many duplicate genes coding proteins that have binding activities in the genomes of many species, we would predict further findings of such conservative duplicated genes.

Acknowledgments

We thank Drs. N. Takahata, Y. Satta (Souken University, Japan), N. Saitou (National Institute of Genetics, Japan), J. Yamagiwa, M. Nakatsukasa, N. Nakagawa, K. Katayama (Kyoto University, Japan), and G. Wagner (Yale University, U.S.A.), for their useful comments and discussions on preliminary versions of this paper. This study was supported by NIH grant AA09379.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2006.11.008.

References

Barth, G., Kunkel, W., 1979. Alcohol dehydrogenase (ADH) in yeasts. II. NAD+- and NADP+-dependent alcohol dehydrogenases in *Saccharomycopsis lipolytica*. Z. Allg. Mikrobiol. 19, 381–390.

Canestro, C., et al., 2000. Amphioxus alcohol dehydrogenase is a class 3 form of single type and of structural conservation but with unique developmental expression. Eur. J. Biochem. 267, 6511–6518.

Cavalli-Sforza, L.L., Edwards, A.W., 1967. Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19, 233–257.

Ceci, J.D., Zheng, Y.W., Felder, M.R., 1987. Molecular analysis of mouse alcohol dehydrogenase: nucleotide sequence of the Adh-1 gene and genetic mapping of a related nucleotide sequence to chromosome 3. Gene 59, 171–182.

Cheung, B., Holmes, R.S., Easteal, S., Beacham, I.R., 1999. Evolution of class I alcohol dehydrogenase genes in catarrhine primates: gene conversion, substitution rates, and gene regulation. Mol. Biol. Evol. 16, 23–36.

Davis, G.J., Bosron, W.F., Stone, C.L., Owusu-Dekyi, K., Hurley, T.D., 1996. X-ray structure of human beta3beta3 alcohol dehydrogenase. The contribution of ionic interactions to coenzyme binding. J. Biol. Chem. 271, 17057–17061.

Duester, G., Smith, M., Bilanchone, V., Hatfield, G.W., 1986. Molecular analysis of the human class I alcohol dehydrogenase gene family and nucleotide sequence of the gene encoding the beta subunit. J. Biol. Chem. 261, 2027–2033.

Edenberg, H.J., Zhang, K., Fong, K., Bosron, W.F., Li, T.K., 1985. Cloning and sequencing of cDNA encoding the complete mouse liver alcohol dehydrogenase. Proc. Natl. Acad. Sci. U. S. A. 82, 2262–2266.

Edenberg, H.J., 2000. Regulation of the mammalian alcohol dehydrogenase genes. Prog. Nucleic. Acid. Res. Mol. Biol. 64, 295–341.

Eklund, H., Brändén, C.I., Jornvall, H., 1976a. Structural comparisons of mammalian, yeast and bacillar alcohol dehydrogenases. J. Mol. Biol. 102, 61–73.

Eklund, H., et al., 1976b. Three-dimensional structure of horse liver alcohol dehydrogenase at 2-4 A resolution. J. Mol. Biol. 102, 27–59.

Engeland, K., Maret, W., 1993. Extrahepatic, differential expression of four classes of human alcohol dehydrogenase. Biochem. Biophys. Res. Commun. 193, 47–53.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.

Felsenstein, J., 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25, 471–492.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Fischer, J.A., Maniatis, T., 1985. Structure and transcription of the *Drosophila mulleri* alcohol dehydrogenase genes. Nucleic Acids Res. 13, 6899–68917.

Fitch, W.M., 1977. On the problem of discovering the most parsimonious tree. Am. Nat. 111, 223–257.

Fleagle, J.G., 1988, 1999. Primate Adaptation and Evolution, 1st and 2nd editions. Academic Press, London.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Guagliardi, A., Martino, M., Iaccarino, I., De Rosa, M., Rossi, M., Bartolucci, S., 1996. Purification and characterization of the alcohol dehydrogenase from a novel strain of *Bacillus stearothermophilus* growing at 70 degrees C. Int. J. Biochem. Cell Biol. 28, 239–246.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating the human–ape split by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160–174.

Hillis, D.M., Mable, B.K., Moritz, C., 1996. Applications of molecular systematics: the state of the field and a look to the future, In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular systematics, Second edition. Sinauer Associates, Sunderland.

Höög, J.O., Hedberg, J.J., Stromberg, P., Svensson, S., 2001. Mammalian alcohol dehydrogenase — functional and structural implications. J. Biomed. Sci. 8, 71–76.

Huelsenbeck, J.P., Hillis, D.M., Jones, R., 1996. Parametric bootstrapping in molecular phylogenetics: applications and performance. In: Ferraris, J.D., Palumbi, S.R. (Eds.), Molecular zoology: Advances, strategies, and protocols. Wiley-Liss, Chichester.

Hur, M.W., Edenberg, H.J., 1992. Cloning and characterization of the ADH5 gene encoding human alcohol dehydrogenase 5, formaldehyde dehydrogenase. Gene 121, 305–311.

Hurley, T.D., Bosron, W.F., Hamilton, J.A., Amzel, L.M., 1991. Structure of human beta 1 beta 1 alcohol dehydrogenase: catalytic effects of non-active-site substitutions. Proc. Natl. Acad. Sci. U. S. A. 88, 8149–8153.

Innan, H., 2003. A two-locus gene conversion model with selection and its application to the human RHCE and RHD genes. Proc. Natl. Acad. Sci. U. S. A. 100, 8793–8798.

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

Jensen-Seaman, M.I., Deinard, A.S., Kidd, K.K., 2001. Modern African ape populations as genetic and demographic models of the last common ancestor of humans, chimpanzees, and gorillas. J. Heredity 92, 475–480.

Kay, R.N.B., Davies, A.G., 1994. Colobinem Monkeys: Their Ecology, Behaviour and Evolution. In: Davis, A.G., Oate, J.F. (Eds.), Cambridge University Press, Cambridge, pp. 229–250.

Kent, W.J., et al., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006.

Kleinjan, D.A., Seawright, A., Childs, A.J., van Heyningen, V., 2004. Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription. Dev. Biol. 265, 462–477.

Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief. Bioinform. 5, 150–163.

Lanave, C., Preparata, G., Saccone, C., Serio, G., 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.

Light, D.R., et al., 1992. Alpha-isoenzyme of alcohol dehydrogenase from monkey liver. Cloning, expression, mechanism, coenzyme, and substrate specificity. J. Biol. Chem. 267, 12592–12599.

Maddison, D., Maddison, W., 2003. MacClade, version 4.06. Sinauer, Sunderland, Massachusetts.

Martin, R.D., 1993. Primate origins: plugging the gaps. Nature 363, 223–234.

Martinez, M.C., et al., 1996. Arabidopsis formaldehyde dehydrogenase. Molecular properties of plant class III alcohol dehydrogenase provide further insights into the origins, structure and function of plant class p and liver class I alcohol dehydrogenases. Eur. J. Biochem. 241, 849–857.

Matsuo, Y., Yokoyama, S., 1989. Molecular structure of the human alcohol dehydrogenase 1 gene. FEBS Lett. 243, 57–60.

Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M., Honjo, T., 1980. Nucleotide sequence divergence of mouse immunoglobulin gamma 1 and gamma 2b chain genes and the hypothesis of intervening sequence-mediated domain transfer. Proc. Natl. Acad. Sci. U. S. A. 77, 2143–2147.

Mouse Genome Sequencing Consortium, 2002. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. OXFORD University press, New York.

Ohta, T., 1988. Evolution by gene duplication and compensatory advantageous mutations. Genetics 120, 841–847.

Ohta, T., 1993. Pattern of nucleotide substitutions in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. Genetics 134, 1271–1276.

Ohta, T., 2000a. Evolution of gene families. Gene 259, 45–52.

Ohta, T., 2000b. Mechanisms of molecular evolution. Philos. Trans. R. Soc. Lond., B. Biol. Sci. 355, 1623–1626.

Oota, H., et al., 2004. The evolution and population genetics of the *ALDH2* locus: random genetic drift, selection, and low levels of recombination. Ann. Hum. Genet. 68, 93–109.

Osier, M., et al., 1999. Linkage disequilibrium at the ADH2 and ADH3 loci and risk of alcoholism. Am. J. Hum. Genet. 64, 1147–1157.

Osier, M.V., et al., 2002. A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. Am. J. Hum. Genet. 71, 84–99.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rambaut, A., Grassly, N.C., 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13, 235–238.

Reimers, M.J., Hahn, M.E., Tanguay, R.L., 2004. Two zebrafish alcohol dehydrogenases share common ancestry with mammalian class I, II, IV, and V alcohol dehydrogenase genes but have distinct functional characteristics. J. Biol. Chem. 279, 38303–38312.

Rodriguez, R., Oliver, J.L., Marin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Satre, M.A., Zgombic-Knight, M., Duester, G., 1994. The complete structure of human class IV alcohol dehydrogenase (retinol dehydrogenase) determined from the ADH7 gene. J. Biol. Chem. 269, 15606–15612.

Svensson, S., Höög, J.O., Schneider, G., Sandalova, T., 2000. Crystal structures of mouse class II alcohol dehydrogenase reveal determinants of substrate specificity and catalytic efficiency. J. Mol. Biol. 302, 441–453.

Swofford, D.L., 1998. Phylogenetic analysis using parsimony. Version 4.0 beta.

Takahata, N., Satta, Y., 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. Proc. Natl. Acad. Sci. U. S. A. 94, 4811–4815.

Teshima, K.M., Innan, H., 2004. The effect of gene conversion on the divergence between duplicated genes. Genetics 166, 1553–1560.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Trezise, A.E., Godfrey, E.A., Holmes, R.S., Beacham, I.R., 1989. Cloning and sequencing of cDNA encoding baboon liver alcohol dehydrogenase: evidence for a common ancestral lineage with the human alcohol dehydrogenase beta subunit and for class I ADH gene duplications predating primate radiation. Proc. Natl. Acad. Sci. U. S. A. 86, 5454–5458.

von Bahr-Lindstrom, H., et al., 1986. cDNA and protein structure for the alpha subunit of human liver alcohol dehydrogenase. Biochemistry 25, 2465–2470.

von Bahr-Lindstrom, H., Jornvall, H., Höög, J.O., 1991. Cloning and characterization of the human ADH4 gene. Gene 103, 269–274.

Yasunami, M., Kikuchi, I., Sarapata, D., Yoshida, A., 1990a. The human class I alcohol dehydrogenase gene cluster: three genes are tandemly organized in an 80-kb-long segment of the genome. Genomics 7, 152–158.

Yasunami, M., Chen, C.S., Yoshida, A., 1990b. Multiplication of the class I alcohol dehydrogenase locus in mammalian evolution. Biochem. Genet. 28, 591–599.

Yasunami, M., Chen, C.S., Yoshida, A., 1991. A human alcohol dehydrogenase gene (ADH6) encoding an additional class of isozyme. Proc. Natl. Acad. Sci. U. S. A. 88, 7610–7614.

Yokoyama, S., Matsuo, Y., Rajasekharan, S., Yokoyama, R., 1992. Molecular structure of the human alcohol dehydrogenase 3 gene. Jpn. J. Genet. 67, 167–171.

Yokoyama, S., Yokoyama, R., Kinlaw, C.S., Harry, D.E., 1990. Molecular evolution of the zinc-containing long-chain alcohol dehydrogenase genes. Mol. Biol. Evol. 7, 143–154.

Yokoyama, S., Yokoyama, R., 1987. Molecular evolution of mammalian class I alcohol dehydrogenase. Mol. Biol. Evol. 4, 504–513.