Am. J. Hum. Genet. 46:85–91, 1990

Cloning and Sequencing of a Processed Pseudogene Derived from a Human Class III Alcohol Dehydrogenase Gene

Yoshinori Matsuo and Shozo Yokoyama

Department of Ecology, Ethology, and Evolution, University of Illinois at Urbana-Champaign

Summary

Current information on the molecular structure of human alcohol dehydrogenase (ADH) genes is fragmentary. To characterize all ADH genes, we have isolated 63 ADH clones from human genomic libraries made from one individual. Fifty-nine clones have been classified into five previously known loci: ADH1 (18 clones), ADH2 (20 clones), and ADH3 class I (16 clones), ADH4 class II (4 clones), and ADH5 class III (1 clone). Sequencing of one of the remaining four unclassified clones, SYλADHE38, about 1.1 kb in length, shows no introns and three frameshift mutations in the coding region, with a total of 10 internal termination codons. When its deduced amino acid sequence was compared with those of the class I, class II, and class III ADHs, the proportions of identical amino acids were 56.7%, 55.5%, and 88.7%, respectively, suggesting that the processed pseudogene was derived from an ADH5 gene. The duplication event seems to have occurred about 3.5 million years ago, and the pseudogene has undergone a rapid change since then.

Introduction

Human alcohol dehydrogenase (ADH) is a dimeric metalloenzyme and is classified into three classes by different electrophoretic properties and different substrate specificities (e.g., see Smith 1986). ADH subunits within each class make heterodimer enzymes, but those from different classes do not. Human class I ADH consists of the subunits α, β, and γ, which are encoded by three loci designated ADH1, ADH2, and ADH3, respectively (Smith 1986; Smith et al. 1971, 1972). Class I ADHs migrate cathodically in starch gels at pH 7–8 (Bosron et al. 1983) and have relatively low K_m for ethanol at a near physiological pH of 7.5 (Bosron et al. 1983; Yin et al. 1984). Class II (subunit π) and class III (subunit χ) are encoded by ADH4 and ADH5 loci, respectively. Both of these subunits migrate toward the anode and do not oxidize ethanol efficiently (Bosron et al. 1979), and only long-chain alcohols such as 1-pentanol or 16-OH-hexadecanoic acid or aromatic alcohols such as cinnamyl alcohol are efficiently oxidized by χ-ADH (Wagner et al. 1984).

With the exceptions of ADH1 (Matsuo et al. 1989) and ADH2 (Duester et al. 1986; Matsuo and Yokoyama 1989) genes, information on the molecular structure of the human ADH genes is still lacking. In order to characterize all of these human ADH genes at the molecular level, we have isolated 63 ADH-positive clones and classified them into different genetic loci. Here we report the nucleotide sequence of a processed pseudogene in the human genome which was most probably derived from an ADH5 gene.

Received June 12, 1989; revision received July 27, 1989.

Address for correspondence and reprints: Dr. Shozo Yokoyama, Department of Ecology, Ethology, and Evolution, University of Illinois at Urbana-Champaign, Shelford Vivarium, 606 East Healey Street, Champaign, IL 61820.

0002-9297/90/4601-0010$02.00

Material and Methods

Human Genomic Library

Two sets of genomic libraries were constructed by using genomic DNA partially digested with either *Mbo*I or *Eco*RI and by ligating it with λEMBL3 or λEMBL4 DNA, respectively. About 120 μg genomic DNA from one of us (S. Y.), obtained from peripheral blood leukocytes (Kan and Dozy 1978), was partially digested and fractionated on an agarose gel. The DNA in the

size range of 9–23 kb was electroeluted from the gel and ligated with λEMBL vector DNA which has been double-digested with *Bam*HI and *Eco*RI. This double digestion was performed to minimize self-ligation. The ligated DNA was packaged in vitro into phage particles by using Gigapack packaging extract (purchased from Stratagene) and plated on the nonpermissive *Escherichia coli* host NM539.

Selection of Positive Clones

Plaque hybridization was carried out using the method of Benton and Davis (1977) for 24–36 h at 68°C in 4 × SETDS (4 × SETDS = 0.6 M NaCl pH 7.5, 8 mM EDTA, 10 × Denhardt, 0.1% SDS) with nick-translated (Rigby et al. 1977) cDNA probe of β-ADH (Yokoyama et al. 1987) and 50 μg heat-denatured herring sperm DNA/ml. After screening of about 1 million recombinant plaques from the two human genomic libraries, 63 positive clones were obtained.

Analysis of ADH Clones

For the 63 positive clones, restriction mapping, using *Eco*RI, *Bam*HI, and *Hin*dIII, and olignucleotide hybridization were conducted. Oligonucleotides which are specific for the three class I subunits α (3′-GCA TGG ACC TTC CCT CGA TAA GAA CCA-5′), β (3′-GCG TGG ACC TTC CCC CGA CAA ATA CCA-5′), and γ (3′-GCG TGC ACC TTT CCT CGA TAA AAA CCT-5′) between amino acids 312 and 320 were synthesized and used for locus-specific hybridization. Dot blots containing DNA from the λ clones were hybridized to the ^{32}P-end-labeled oligonucleotides and to heat-denatured herring sperm DNA at 37°C for 12 h in 4 × SETDS (Matsuo et al. 1989).

From these analyses, 17, 13, and 13 clones were assigned to ADH1, ADH2, and ADH3 loci, respectively. Eleven of the remaining 20 positive clones were subcloned by the shotgun method with *Sau*3A. Those subclones hybridizing to a cDNA probe of β-ADH (Yokoyama et al. 1987) were sequenced using the dideoxy chain-termination method (Sanger et al. 1977; Hattori et al. 1985). DNA sequences and the deduced amino acids were compared with those of human class I (Bühler et al. 1984; Hempel et al. 1984; Ikuta et al. 1986), class II (Höög et al. 1987), and class III (Kaiser et al. 1988) ADHs. With this information, one, two, two, one, and one clones were assigned to ADH1, ADH2, ADH3, ADH4, and ADH5 loci, respectively, but the remaining four clones could not be assigned to any of these loci. Furthermore, with information of restriction maps, an additional five, one, and three clones were assigned to ADH2, ADH3, and ADH4, respectively. Thus, a total of 18, 20, 16, 4, and 1 clones were assigned to ADH1, ADH2, ADH3, ADH4, and ADH5 loci, respectively.

Results

Isolation of SYλADHE38, an ADH Pseudogene

Two of the four unclassified clones, SYλADHE38 and SYλADHE49, showed identical restriction maps and represented the same gene. Subcloning and SYλADHE38 was performed using restriction enzymes *Bam*HI, *Bgl*II, *Hin*dIII, *Hin*PI, *Kpn*I, *Pst*I, *Pvu*II, *Sau*3AI, and *Taq*I (for its restriction map, see fig. 1). Digested DNA fragments were ligated into the plasmid Bluescript vector (Stratagene). The DNA sequence of SYλADHE38 contains 10 internal termination codons and, therefore, is not able to code for a functional polypeptide.

Origin of SYλADHE38 Clone

Since the complete DNA sequence of the ADH5 gene is currently not available, the amino acid sequence deduced from the DNA sequence of SYλADHE38 and those of the three classes of human ADHs were compared using the method of Needleman and Wunsch (1970) to evaluate the magnitudes of sequence similarity between them. After the alignment, three frameshift mutations were found: two of them were single base deletions at amino acid positions 109 and 244 (or 243), and another was a single base insertion at the amino acid position between 87 and 88 (fig. 2). The proportions of identical amino acids between SYλADHE38 and class I (Bühler et al. 1984; Hempel et al. 1984; Ikuta et al. 1986; Höög et al. 1987; Kaiser et al. 1988), class II (Höög et al. 1987), and class III (Kaiser et al. 1988) ADHs were 56.7% (211/372), 55.5% (207/373),

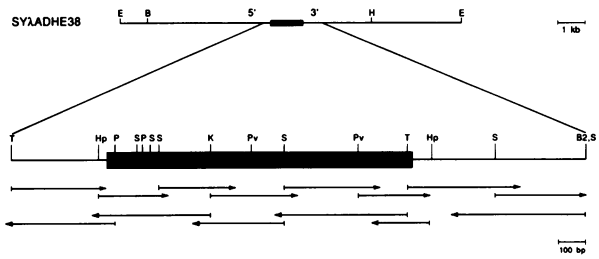

Figure 1 Restriction map and sequencing strategy of SYλADHE38. B = *Bam*HI; B2 = *Bgl*II, E = *Eco*RI; H = *Hin*dIII; Hp = *Hin*PI; K = *Kpn*I; P = *Pst*I, Pv = *Pvu*II; S = *Sau*3A; and T = *Taq*I.

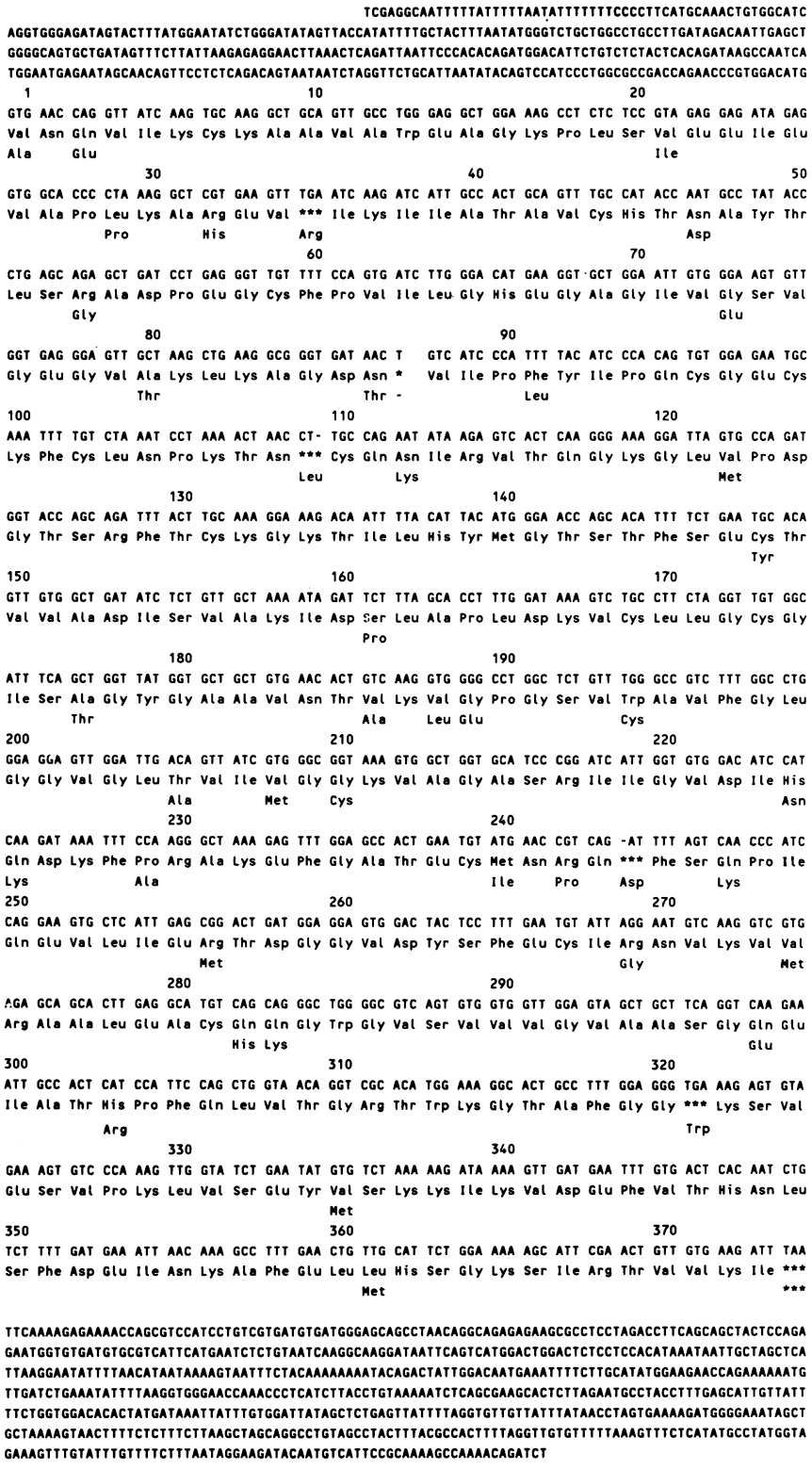

Figure 2 DNA sequence of SYλADHE38 (*upper row*) shown together with deduced amino acids (*middle row*). The amino acid sequence of human χ-ADH (Kaiser et al. 1988) is also shown (*third row*). Numbers above the nucleotide sequence show the residue position of the χ-ADH (Kaiser et al. 1988). Dash (–) indicates the deletion of a nucleotide.

and 88.7% (331/373), respectively. Thus, SYλADHE38 is most closely related to the ADH5 gene.

Processed ADH Pseudogene in Human Genome

Human ADH1 (Matsuo and Yokoyama 1989) and ADH2 (Duester et al. 1986; Matsuo et al. 1989) genes contain eight introns. Within the class I ADH genes, the positions and sizes of these introns seems to be well conserved (Matsuo and Yokoyama 1989). From the screening of two genomic libraries, we have obtained one ADH5 gene, SYλADHE33. This clone has been partially sequenced. Within the sequenced region, one intron was detected, and its position was identical to the position of intron 6 of ADH1 and ADH2 genes (fig. 3). The proportion of identical nucleotides between SYλADHE33 and SYλADHE38 in this region was 161/170 = .91, again showing the close evolutionary relationship of SYλADHE38 to an ADH5 gene.

The simplest explanation for the lack of introns in SYλADHE38 is that this gene was probably derived from the transcript of an ADH5 gene by reverse transcription. A large number of such processed genes from many organisms have been described (Vanin 1985; Weiner et al. 1986). However, this is the first case of a processed ADH pseudogene in vertebrates.

Discussion

Discovery of different processed pseudogenes—such as mouse α-globin (Nishioka et al. 1980; Vanin et al. 1980), human α- and β-globin genes (Modiano et al. 1981), α-tubulin (Lemischka and Sharp 1982) and cytochrome c (Scarpulla et al. 1982; Scarpulla 1984) genes in rat, and β-tubulin (Wilde et al. 1982*a*, 1982*b*), β-actin (Moos and Gallwitz 1982), and keratin (Savtchenko et al. 1988) genes in human—has provided significant information for the study of gene organization and molecular evolution (for reviews, see Vanin 1985; Weiner et al. 1986).

To date, χ-ADH is the only ADH isozyme found in human brain and testes (Beisswenger et al. 1985; Dafeldecker and Vallee 1986). It is not known whether the enzyme activity is present in the germ line. For many of the processed genes, however, their functional counterparts are also expressed in the germ-line cell. In such a process, it seems unlikely that transcriptional promoters are correctly positioned 5′ to the resultant processed gene and that the latter presumably will have lost its function at a very early stage of divergence (Vanin 1985; Weiner et al. 1986).

Recently, Keung (1988) found in hamster a genuine testes-specific ADH isozyme which is different from the human class III ADH. The corresponding human ADH has not yet been reported, but similar ADH isozymes may also exist in human testes. If the functional genes, encoding such isozymes, are expressed in the germ-line cell, then other types of processed ADH pseudogenes may also exist.

Since a pseudogene was identified, it is of interest to evaluate this pseudogene's time of divergence from a functional ADH5 gene, as well as its evolutionary rate. Figure 4 shows a probable evolutionary scheme for human χ-ADH ($ADH\chi_H$), a hypothetical protein deduced from SYλADHE38 ($ADH\psi\chi_H$), and rat χ-ADH ($ADH\chi_R$). In figure 4, $K_{H\psi}$, $K_{R\psi}$, and K_{HR} are the numbers of amino acid substitutions per residue between $ADH\chi_H$ and $ADH\psi\chi_H$, between $ADH\chi_R$ and $ADH\psi\chi_H$, and between $ADH\chi_H$ and $ADH\chi_R$, respec-

```
          218     220                                 230                                 240
          Ile Ile Gly Val Asp Ile Asn Lys Asp Lys Phe Ala Arg Ala Lys Glu Phe Gly Ala Thr Glu Cys Ile Asn
SYλADHE33 ATC ATT GGT GTG GAC ATC AAT AAA GAT AAA TTT GCA AGG GCC AAA GAG TTT GGA GCC ACT GAA TGT ATT AAC
SYλADHE38                         C   C               C         T                                   G
                                  His Gln             Pro                                         Met

                             250                                 260
          Phe Ser Lys Pro Ile Gln Glu Val Leu Ile Glu Met Thr Asp Gly Gly Val Asp Tyr Ser Phe Glu Cys Ile
SYλADHE33 TTT AGT AAA CCC ATC CAG GAA GTG CTC ATT GAG ATG ACC GAT GGA GGA GTG GAC TAT TCC TTT GAA TGT ATT
SYλADHE38         C                                 CG    T                       C
                  Gln                               Arg

                  274
          Lys Val Met      Intron 6
SYλADHE33 AAG GTC ATG GTGAGTATGGGCTTCATTCCTTTT
SYλADHE38         G   ------------------------
                  Val
```

Figure 3 Comparison of the DNA sequence of SYλADHE38 and the ADH5 gene, SYλADHE33. Dash (—) indicates the deletion of a nucleotide in SYλADHE38. The position of a intron of SYλADHE33 is identical with those of the intron 6 of ADH1 and ADH2 genes.

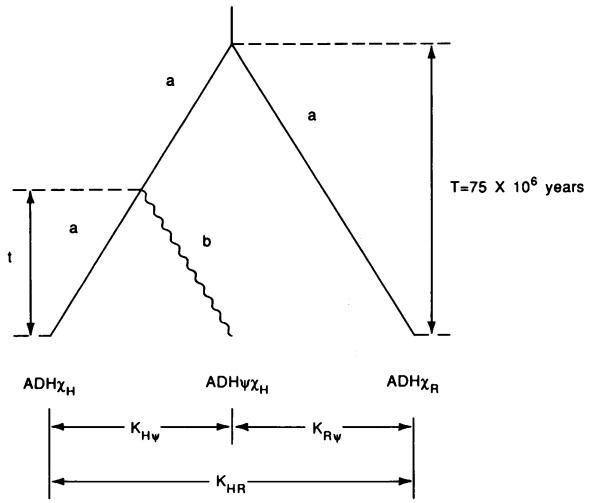

Figure 4 A phylogenetic tree for human χ-ADH (ADHχH), the amino acid sequence deduced from SYλADHE38 (ADHψχH), and rat class III ADH (ADHχR). T = divergence time between human and rat; t = time since duplication between ADHχH and ADHψχH; a = rate of amino acid substitution per site per year in the χ-ADH in human and rat; b = effective rate of amino acid substitution in SYλADHE38.

tively. Let *a* be the rate of amino acid substitutions per site per year in both ADHχH and ADHχR. Once the ADHψχH has been derived from ADHχH, the evolutionary rate of nucleotide substitution can be expected to change. Since a pseudogene cannot be translated into amino acid sequences, the biological significance of the rate of amino acid substitution in SYλADHE38 becomes somewhat obscure. However, since this rate can be compared with that of a functional gene, we will use the *effective rate of amino acid substitution* for this pseudogene and denote it by *b*. In figure 4, the time (*T*) since divergence between ADHχH and ADHχR is known to be about 75×10^6 years, and *t* denotes the divergence time between ADHψχH and ADHχH. From figure 4,

$$\begin{aligned} K_{H\psi} &= (a+b)\,t\,; \\ K_{HR} &= 2aT\,; \\ K_{R\psi} &= 2aT + (b-a)\,t. \end{aligned} \qquad (1)$$

From these relations,

$$\begin{aligned} a &= K_{HR}/(2T)\,; \\ b &= a(K_{H\psi} - K_{HR} + K_{R\psi})/(K_{H\psi} - K_{HR} - K_{R\psi})\,; \\ t &= T(K_{H\psi} + K_{HR} - K_{R\psi})/K_{HR}\,. \end{aligned} \qquad (2)$$

To evaluate *a*, *b*, and *t*, we will use the amino acid sequence data from the human χ-ADH (Hempel et al. 1984), the processed pseudogene SYλADHE33, and the rat class III ADH (Julia et al. 1988). Proportions of different amino acids (*p*) between ADHχH and ADHψχH, between ADHχH and ADHχR, and between ADHψχH and ADHχR were .113 (42/373), .056 (21/373), and .158 (59/373), respectively. The numbers of amino acid substitutions per site (*K*) were estimated by $K = -\ln(1-p-p^2/5)$ (Kimura 1983). Thus, $K_{H\psi}$, K_{HR}, and $K_{R\psi}$ were .123, .059, and .178, respectively.

Substituting these values into equation (2), we obtain $a = .39 \times 10^{-9}$, $b = 35 \times 10^{-9}$, and $t = 3.5 \times 10^6$. The *a* value for class III ADH is about 1/5–3/5 that of the class I ADHs, whose rate of amino acid substitution varies from $.69 \times 10^{-9}$ to 2.11×10^{-9} (Yokoyama and Yokoyama 1987; Yokoyama et al., in press). The estimated *b* value is about 90 times higher than *a*, reflecting the effect of the nonfunctionalization of the pseudogene. The *t* value obtained shows a rather recent origin of the pseudogene.

In this computation, we assumed that the gene became nonfunctional at a very early stage of divergence. Once the DNA sequences of both functional class III ADH genes from human and rat are obtained, it will be possible to evaluate the exact structural changes which occurred in an ADH pseudogene, SYλADHE38.

Acknowledgments

This work was partially supported by grants from the National Institutes of Health and the National Science Foundation. Comments by anonymous reviewers were greatly appreciated.

References

Beisswenger TB, Holmquist B, Vallee BL (1985) χ-ADH is the sole alcohol dehydrogenase isozyme of mammalian brains: implications and inferences. Proc Natl Acad Sci USA 82:8369–8373

Benton WD, Davis RW (1977) Screening λgt recombinant clones by hybridization to single plaques in situ. Science 196:180–182

Bosron WF, Li T-K, Dafeldecker WP, Vallee BL (1979) Human liver pi alcohol dehydrogenase kinetic and molecular properties. Biochemistry 18:1101–1105

Bosron WF, Magnes LJ, Li T-K (1983) Kinetic and electrophoretic properties of native and recombined isoenzymes of human liver alcohol dehydrogenase. Biochemistry 22: 1852–1857

Bühler R, Hempel J, Kaiser R, De Zalenski C, Von Wartburg

J-C, Jörnvall H (1984) The primary structure of the γ_1 protein chain of human liver alcohol dehydrogenase. Eur J Biochem 145:447–453

Dafeldecker WP, Vallee BL (1986) Organ-specific human alcohol dehydrogenase: isolation and characterization of isozymes from testis. Biochem Biophys Res Commun 134: 1056–1063

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

Hattori M, Hidaka S, Sakaki Y (1985) Sequence analysis of a Kpn I family member near the 3′ end of human β-globin gene. Nucleic Acids Res 13:7813–7827

Hempel J, Bühler R, Kaiser R, Holmquist B, DeZalenski C, Von Wartburg IP, Vallee B, et al (1984) Human liver alcohol dehydrogenase 1: the primary structure of the $\beta_1\beta_1$ isozyme. Eur J Biochem 145:437–445

Höög J-O, von Bahr-Lindström H, Heden L-O, Holmquist B, Larsson K, Hempel J, Vallee BL, et al (1987) Structure of the class II enzyme of human liver alcohol dehydrogenase: combined cDNA and protein sequence determination of the π subunit. Biochemistry 26:1926–1932

Ikuta E, Szeto S, Yoshida A (1986) Three human alcohol dehydrogenase subunits: cDNA structure and molecular and evolutionary divergence. Proc Natl Acad Sci USA 83: 634–638

Julia P, Pares X, Jörnvall H (1988) Rat liver alcohol dehydrogenase of class III: primary structure, functional consequences and relationships to other alcohol dehydrogenases. Eur J Biochem 172:73–83

Kaiser R, Holmquist B, Hempel J, Vallee BL, Jörnvall H (1988) Class III human liver alcohol dehydrogenase: a novel structural type equidistantly related to the class I and class II enzymes. Biochemistry 27:1132–1140

Kan YW, Dozy AM (1978) Polymorphism of DNA sequence adjacent to human β-globin structural gene: relationship to sickle mutation. Proc Natl Acad Sci USA 75:5631–5635

Keung, W-M (1988) A genuine organ specific alcohol dehydrogenase from hamster testes: isolation, characterization and developmental changes. Biochem Biophys Res Commun 156:38–45

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge

Lemischka I, Sharp PA (1982) The sequences of an expressed rat α-tubulin gene and pseudogene with an increased repetitive element. Nature 300:330–335

Matsuo Y, Yokoyama R, Yokoyama S (1989) Human alcohol dehydrogenases β1 and β2 can be specified by a single nucleotide substitution. Eur J Biochem 183:317–320

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

Modiano G, Battistuzzi G, Motulsky AG (1981) Nonrandom patterns of codon usage and of nucleotide substitutions in human α- and β-globin genes: an evolutionary strategy reducing the rate of mutations with drastic effects. Proc Natl Acad Sci USA 78:1110–1114

Moos M, Gallwitz D (1982) Structure of a human β-actin related pseudogene which lacks intervening sequences. Nucleic Acids Res 10:7843–7849

Needleman SB, Wunsch CD (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. J Mol Biol 48:443–453

Nishioka Y, Leder A, Leder P (1980) An unusual alpha globin-like gene that has clearly lost both globin intervening sequences. Proc Natl Acad Sci USA 77:2807–2809

Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J Mol Biol 113:237–251

Sanger F, Nicklen S, Coulsen AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Savtchenko ES, Schiff TA, Jiang C-K, Freedberg IM, Blumenberg M (1988) Embryonic expression of the human 40-kD keratin: evidence from a processed pseudogene sequence. Am J Hum Genet 43:630–637

Scarpulla RC (1984) Processed pseudogenes for rat cytochrome c are preferentially derived from one of three alternate mRNAs. Mol Cell Biol 4:2279–2288

Scarpulla RC, Agne KM, Wu R (1982) Cytochrome c gene-related sequences in mammalian genomes. Proc Natl Acad Sci USA 83:4167–4171

Smith M (1986) Genetics of human alcohol and aldehyde dehydrogenases. Adv Hum Genet 14:249–290

Smith M, Hopkinson DA, Harris H (1972) Alcohol dehydrogenase isozymes in stomach and liver: evidence for activity of the ADH3 locus. Ann Hum Genet 35:243–253

——— (1971) Developmental changes and polymorphism in human alcohol dehydrogenase. Ann Hum Genet 34: 251–271

Vanin EF (1985) Precessed pseudogenes: characteristics and evolution. Annu Rev Genet 19:253–272

Vanin EF, Goldberg GI, Tucker PW, Smithies O (1980) A mouse α-globin-related pseudogene lacking intervening sequences. Nature 286:222–226

Wagner FW, Pares X, Holmquist B, Vallee BL (1984) Physical and enzymatic properties of a class III isozyme of human liver alcohol dehydrogenase: χ-ADH. Biochemistry 23: 2193–2199

Weiner AM, Deninger PL, Efstratiadis A (1986) Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu Rev Biochem 55:631–661

Wilde CD, Crowther CE, Cowan NJ (1982*a*) Diverse mechanisms in the generation of human β-tubulin pseudogenes. Science 217:549–552

Wilde CD, Crowther CE, Cripe TP, Gwo-Shu Lee M, Cowan NJ (1982*b*) Evidence that a human β-tubulin gene is derived from its corresponding mRNA. Nature 297:83–84

Yin S-J, Bosron WF, Magnes LJ, Li T-K (1984) Human liver alcohol dehydrogenase: purification and kinetic characterization of the $\beta_2\beta_2$, $\beta_2\beta_1$, $\alpha\beta_2$ and $\beta_2\gamma$ "Oriental" isoenzymes. Biochemistry 23:5847–5853

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

Yokoyama S, Yokoyama R, Kinlaw CS, Harry DE. Molecular evolution of the zinc-containing long-chain alcohol dehydrogenase genes. Mol Biol Evol (in press)

Yokoyama S, Yokoyama R, Rotwein P (1987) Molecular characterization of cDNA clones encoding the human alcohol dehydrogenase β1 and the evolutionary relationship to the other class I subunits α and γ. Jpn J Genet 62:241–256