

GENE 07295

Characterization of the genes encoding carbonic anhydrase I of chimpanzee and gorilla: comparative analysis of 5' flanking erythroid-specific promoter sequences

(Polymerase chain reaction; primate evolution; African hominoid trichotomy; orangutan; pigtail macaque; squirrel monkey)

Bill R. Epperly, Nils C.H. Bergenhem, Patrick J. Venta* and Richard E. Tashian

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109-0618, USA

Received by J.L. Slightom: 16 March 1993; Accepted: 1 May 1993; Received at publishers: 27 May 1993

SUMMARY

The genes encoding carbonic anhydrase I (CA I) have been characterized for chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla*). In addition, 44 nucleotides (nt) at the 5' end of the noncoding first exon (exon 1a), which is unique to the erythroid CA I mRNA, together with 188 nt of the adjacent 5' flanking regions, were sequenced for the corresponding positions of the CA I of orangutan, pigtail macaque, and squirrel monkey. When these 5' flanking regions are compared, along with those published for human and mouse CA I, they were found to contain several conserved sequences that may bind factors involved in the erythroid-specific expression of CA I. Comparisons of the human, chimpanzee, and gorilla coding and noncoding CA I sequences do not significantly deviate from a pattern of trichotomy for the evolutionary origins of these three hominoid species.

INTRODUCTION

Comparative studies of the aa and nt sequences of the carbonic anhydrase (CA) isozymes have provided useful information on the evolutionary relationships of humans and other primates (Contel et al., 1981; Tashian et al., 1983; Hewett-Emmett et al., 1984; Hewett-Emmett and Tashian, 1991). Of the seven CA isozymes (CA I–CA VII) that have been described from amniotes (primarily mam-

mals), only CA I is known to be under the control of two promoters (Fraser et al., 1989; Tashian et al., 1989; Brady et al., 1991; Hewett-Emmett and Tashian, 1991). When CA I mRNA is expressed in erythrocytes, it contains an additional, short (63 and 72 bp in mouse and human, respectively), 5' noncoding exon (exon 1a) which replaces part of the untranslated region of exon 1. The CA I mRNA expressed in nonerythroid tissue (e.g., colon) lacks the noncoding exon 1a (Fraser et al., 1989; Tashian et al., 1989; Brady et al., 1991; Bergenhem et al., 1992b). As yet, only the CA I of human, pigtail macaque, and mouse have been characterized (Fraser et al., 1989; Lowe et al., 1990; Nicewander, 1990).

The question of the relative evolutionary relationships between humans and the African great apes, chimpanzee and gorilla, continues to be of special interest to students of human evolution. Evidence from DNA sequence studies have been reported supporting for the most part a closer evolutionary relationship between human and

Correspondence to: Dr. R.E. Tashian, Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109-0618, USA. Tel. (1-313) 764-1359; Fax (1-313) 763-3784.

*Present address: Small Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, MI 48874-1314, USA. Tel. (1-517) 336-2515; Fax (1-517) 336-2514.

Abbreviations: aa, amino acid(s); bp, base pairs(s); CA, carbonic anhydrase; CA, gene encoding CA; Ch, chimpanzee; Hu, human; kb, kilobase(s) or 1000 bp; nt, nucleotide(s); PCR, polymerase chain reaction; tsp, transcription start point(s).

chimpanzee than between either of these two species and gorilla by examination of the β -globin gene clusters (Miyamoto et al., 1987, 1988; Koop et al., 1989; Bailey et al., 1991, 1992) and mitochondrial genomes (Ruvolo et al., 1991; Horai et al., 1992). Nevertheless, certain inconsistencies exist, and the problem has yet to be resolved (Dijon and Green, 1990). The present report provides DNA sequence data for the *CA I* of chimpanzee (*Pan troglodytes*) and gorilla (*Gorilla gorilla*) that can be used to examine the evolutionary patterns of these hominoid species. Also, in order to compare the erythroid-specific promoter regions of the *CA I* of primates, we have determined the partial nt sequences of exons 1a and their 5' flanking regions of orangutan (*Pongo pygmaeus*), pigtail macaque (*Macaca nemestrina*), and squirrel monkey (*Saimiri sciureus*).

EXPERIMENTAL AND DISCUSSION

(a) Structural and evolutionary aspects

The structures of the *CA I* of chimpanzee and gorilla have been determined by direct sequencing of double-stranded PCR products. Both genes consist of the 8 exons found to comprise the *CA I* of human, pigtail macaque, and mouse (Fraser et al., 1989; Lowe et al., 1990; Nicewander, 1990), beginning with the untranslated exon 1a plus the 7 exons that are also characteristic of the *CA II*, *CA III*, and *CA VII* of mammals (cf. Tashian, 1992). Since directly sequencing the PCR-amplified DNA allows both alleles to be sequenced simultaneously, polymorphisms can be readily detected by the presence of two bands at any specific position on the sequencing gel (Bergenheim et al., 1992a); no such polymorphisms were observed in any of the sequences studied. In Fig. 1, the sequences for the coding regions of chimpanzee and gorilla *CA I* are compared with those of the human *CA I*, and the differences in their nt and aa are highlighted. The intron splice junctions are shown in Fig. 2 along with flanking intron sequences which are identical with those of the human *CA I*, with the exception of 1 nt in the donor sequence of intron 1. The identification of a Phe residue at position 114 from an earlier study on the protein sequence of chimpanzee *CA I* suggests a Phe/Tyr polymorphism, or a rare variant, at this position (Contel et al., 1981).

The active site of *CA I* has been well-characterized by X-ray crystallography, and the aa residues involved in binding the active-site zinc ion, enzyme catalysis, and formation of the active-site cavity have been identified (Ericksson and Liljas, 1991; Fig. 6 in Tashian, 1992). The deduced aa sequence of *CA I* from chimpanzee and gorilla are identical to the human sequence at the posi-

Ch	AAAAGAAAACGACGACAAAGAA	ATG	CGA	ACT	CGA	GAT	TGC	GGG	TAT	GAT	GAT	AAA	AAA	AA	11				
Hu	AAAAGAAAACGACGACAAAGAA	ATG	CGA	ACT	CGA	GAT	TGC	GGG	TAT	GAT	GAT	AAA	AAA	AA	11				
Go	AAAAGAAAACGACGACAAAGAA	ATG	CGA	ACT	CGA	GAT	TGC	GGG	TAT	GAT	GAT	AAA	AAA	AA	11				
Ch	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Ch	▲																		
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn	Glu	Asn	Glu	Ser	121	
Hu	GAT	GCT	GAA	GAA	TGG	ATC	AAI	CTC	TAT	CTC	ATT	GTC	AAT	GGA	AAT	AAC	AAA	TGC	119
Go	CTC	Pro	Glu	Glu	Trp	Ser	Leu	Leu	Tyr	Pro	Ile	Ala	Asn						

Go: GACAAAAATG]gtaagagttc (intron 1, 2.8 kb) -----tcctggttag [GTCCTGAACA
 Ch: tgtttg
 Go: AACCGATCAG]gtgagctgaa (intron 2, 1.0 kb) tcctttttctccag [TGCTGAAAGG
 Ch: -----
 Go: TTCTGCCGAG]gtaatgtaat (intron 3, 3.6 kb) tagtatcatttttag [CTTCACATAG
 Ch: -----
 Go: TTTGATGAAG]gtgtgttaca (intron 4, 0.9 kb) ---ttaaatctccag [GTTGGTGAGG
 Ch: a ttc
 Go: TAAACCAAG]gtaaacacac (intron 5, 3.0 kb) attcttttctccag [GGCAAACGAG
 Ch: -----
 Go: CTCAGAGCAG]gtagagttgt (intron 6, 0.8 kb) attttatecttctag [CTGGCACAAT
 Ch: -----

Fig. 2. The nt sequence at exon/intron splice junctions of the *CA I* of chimpanzee (Ch) and gorilla (Go). The brackets indicate the junctions, with the capital and lower-case letters indicating exon and intron sequences, respectively. The only difference between the gorilla and chimpanzee sequences is shown in the donor sequence of intron 4. Dashes indicate sequences that were not determined. The two g's (bold-face type) in the donor sequence of intron 1 were reported as c's by Lowe et al. (1990); sequences of several human samples determined in our laboratory show g's at these positions. The chimpanzee sequence is identical to that of human except for an A (bold-face type) at position 7 in exon 4 which is a G in the human sequence.

tions known to coordinate the zinc ion as well as almost all positions known to comprise the active site of the molecule. Gorilla *CA I*, however, differs at residues 69 and 121 in having a Thr residue at these positions in place of Asn (residue 69) and Ala (residue 121) found in chimpanzee and human. The gorilla *CA I* protein has been partially purified and shows levels of catalytic activity similar to human and chimpanzee *CA I* (Tashian, 1977; and unpublished results); therefore, any change in the active-site environment caused by these two replacements would appear to be very small.

A matrix that compares the aa and nt sequences of the *CA I* of human, chimpanzee, gorilla, and pigtail macaque is shown in Table II. As can be seen, the human sequence

TABLE I
PCR and sequencing primers

Exon ^a	Sequence
1aD	5'-G GAATTCAATCCACACCCCAACCACTTC
1aU	5'-G GAATTCACAGCTCTGAATGAGAGAAGG
1D	5'-G TTGGAATCTTGAGTGTACAAG
1U	5'-G GGATCCGCAGACAGTTCAACAATTAACC
2D	5'-C AAACAGGTAACACTACTCCT
2U	5'-A ATGGGTGTCTATGTTTCTCG
3D	5'-G GAATTCGCAAAGATAAGCTAGAGTTTG
3U	5'-G GGATCCAGGGTAATTATCTCTCACTTAC
4D	5'-G GAATTCCTGGATAAAGGTTACATA
4U	5'-C CTTCTATTTGAGGTCTAATTGG
5D	5'-G CAGTGTGTTGATGACAATAATC
5U	5'-G GAATTCACCCCGATTTTAATACTTCA
6D	5'-G GAATTCATGACTCTTAGCTAAAATCTC
6U	5'-A ATATTCCTGCTACTATATTCCC
7D	5'-T CAGTGCCTTAGTAATCCTGTAA
7U	5'-A AAGCTTGGGCTGTGTTCTTGAGGAAGG

^aD denotes a downstream-directed primer and U an upstream-directed primer.

differs less from chimpanzee than from gorilla by about 1% at the aa level and about 0.6% at the nt level. However, when these relationships are examined statistically by constructing phylogenetic trees (with the pigtail macaque *CA I* sequence as the outgroup) by neighbor-joining (Saitou and Nei, 1987) or parsimony tree-building methods (Czelusniak et al., 1990), a trichotomous branching pattern is produced which does not significantly favor a closer relationship of human to chimpanzee than to gorilla (data not shown).

(b) Comparative aspects of the 5' flanking sequences of exon 1a

In Fig. 3, the partial nt sequences for exons 1a and their upstream flanking regions are compared for the *CA I* of human, three great apes (chimpanzee, gorilla, orangutan), an Old World monkey (pigtail macaque), a New World monkey (squirrel monkey), and the house mouse. As can be seen in Table III, the human, great apes, and macaque sequences are highly conserved.

Several potential transcription factor consensus binding sequences common to primates and mouse (i.e., CACCC, GATA-1, Oct-1, TATA) are identified in Fig. 3 which may play a role in the erythroid-specific expression of *CA I* mRNA. Of these, the GATA-1 sequence at -190 has been implicated by footprinting analysis in the erythroid expression of the mouse *CA I* (Butterworth et al., 1991). Two DNase I hypersensitivity regions have been described upstream from exon 1a in the mouse and human genes (Thiefelder et al., 1991; Sowden et al., 1992). The most distal of these are 1.0 and 1.5 kb from the *tsp* of exon 1a in human and mouse, respectively, while the other two are about 200 bp from the cap site of exon 1a in both species. It is possible that the GATA-1 site at nt position -190 coincides with the distal DNase I hypersensitivity region. In addition, three sequences of 5 or more nt that have been conserved in all species are indicated in Fig. 3.

TABLE II
Comparison of aa and nt sequences of *CA I* from human (Hu), chimpanzee (Ch), gorilla (Go), orangutan (Or), and pigtail macaque (Pm)^a

	Hu	Ch	Go	Or	Pm
Hu	-	2 (0.77)	5 (1.9)	7 (2.7)	12 (4.6)
Ch	6 (0.77)	-	6 (2.3)	7 (2.7)	12 (4.6)
Go	11 (1.4)	8 (1.0)	-	7 (2.7)	12 (4.6)
Or				-	17 (6.5)
Pm	29 (3.7)	24 (3.0)	23 (2.9)		

^aAbove diagonal, aa difference followed by the percentage difference in parentheses; below diagonal, nt difference followed by the percentage difference. The nt sequence is not available for orangutan. Sources: human *CA I* (Lowe et al., 1990), orangutan *CA I* (R.E. Ferrell in Tashian, 1977), pigtail macaque *CA I* (Nicewander, 1990).

TABLE III

Comparison of nt differences in exon 1a and 5' flanking sequences for the *CAI* of human (Hu), chimpanzee (Ch), gorilla (Go), orangutan (Or), pigtail macaque (Pm), squirrel monkey (Sm), and mouse (Mo)^a

	Ch	Go	Or	Pm	Sm	Mo
Hu	4 (1.7)	6 (2.5)	10 (4.2)	10 (4.2)	25 (12.4)	87 (37.1)
Ch	—	7 (2.9)	7 (2.9)	11 (4.7)	27 (13.4)	88 (37.6)
Go		—	7 (2.9)	11 (4.7)	22 (10.9)	88 (37.6)
Or			—	13 (5.5)	26 (11.1)	88 (37.6)
Pm				—	25 (12.4)	77 (32.9)
Sm					—	78 (38.8)

^aThe nt differences between species are given followed by the percentage difference in parentheses. Gaps and insertions are counted as a single mutation. Alignments are for the nt positions between the vertical arrows in Fig. 3 except for squirrel monkey, which lacks the first 25 5' nt.

```

          CACCC          GATA-1      A          B
Hu  AAATCCACACCCCAACCACTTCCTTATCAGGTTCTCACACTCTGGGG-CCACTATGTA- -154
Ch  .....-154
Go  .....-154
Or  .....C.....-144
Pm  .....T.....C.....-154
Sm  .....C.....C.....A.....-143
Mo  ..G.....TG...TG...G...TGT...C.....G.-159
      ▲

```

```

          GATA-1      C
Hu  CCCACTCTAATCACCACAGGGCCAGACATCAGACAATTAAGGACAGCGCCCATGCCCAA -93
Ch  .....-93
Go  .....-93
Or  .....T.....-83
Pm  .....A.....A.....-93
Sm  .....C.....C.....T.....-84
Mo  T.T..C.....TTCTGAT...G...A.T...T..C.TT..G...TC.TTGG...C -98

```

```

          Oct-1          CACCC
Hu  AGCCGGCCAAATATGCAAATATTCAAAATATTCAACCTAGCTAACCCACCCCTTTT -32
Ch  .....-32
Go  .....T.....-32
Or  .....T.....-32
Pm  .....G.....-32
Sm  .....C.....T.....-32
Mo  ...TTA...G...TATT.A.CAG.C..TC.GGG.-C.A...T..C.A..G...CC -38
      CCAAT          Sp1

```

```

          TATA          Exon 1a
Hu  TGCTGTACATAAGCTGGCCATTCCCC---CTCCAGCCTGTGGTACCAGTCC---TCAG 22
Ch  .....T.....T.....G...A.....22
Go  .....T.....T.....G...A.....22
Or  .....C.....T.....G.....22
Pm  .....A.....T.....T.....22
Sm  .....G.....C.....T.....G.....22
Mo  ..T...T.....TACTG...CAA...CT...C...ATTG...20
      ▲

```

```

          Intron 1a
Hu  GTGCAACCCCTGCGTGGTCTCTCTGTCAGAGCCTTCTCTCATTGAGTAAAC 81
Ch  .....
Go  .....
Or  .....T.....
Pm  .....G...-AG.
Sm  ..A...A...A...C.....
Mo  .A.GG...TTA...GA...A...GGCCA...T.....A...G.....72
      ▲

```

Fig. 3. Comparative nt sequences of exons 1a and 5' flanking regions of the *CAI* of human (Hu), chimpanzee (Ch), gorilla (Go), orangutan (Or), pigtail macaque (Pm), squirrel monkey (Sm), and mouse (Mo). Dots indicate identical nt; dashes, deletions; blank spaces, regions not sequenced. **Materials:** Genomic DNA from chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), squirrel monkey (*Saimiri sciureus*), and orangutan (*Pongo pygmaeus*) was provided by Dr. Jerry Slightom (Upjohn Co., Kalamazoo, MI), and the DNA from pigtail macaque (*Macaca nemestrina*) was prepared in our laboratory from tissue samples provided by the University of Washington Regional Primate Center (Seattle, WA). **Methods:** PCR amplification from genomic DNA and direct sequencing of PCR products were as described in the legend to Fig. 1. Exon 1a from squirrel monkey was amplified by PCR primers with artificial linkers (Table I), and the PCR products were sequenced after purification and subcloning into pBSM13⁻.

In Table III, the nt sequences at the 5' end of exons 1a and adjacent 5' flanking regions (i.e., nt positions between the vertical arrows in Fig. 3) are compared for the *CAI* of human, three great apes, an Old World monkey, a New World monkey, and house mouse. Although the human sequence differed less from chimpanzee than from gorilla by 0.8%, when all seven sequences were analyzed statistically by the tree-building computer programs discussed in section a above, a similar pattern of trichotomy was seen for the branching of the human, chimpanzee, and gorilla lineages from a common ancestor (data not shown).

ACKNOWLEDGEMENTS

We thank Jerry Slightom as well as the personnel at the University of Washington Regional Primate Center for kindly providing, respectively, the genomic DNA and tissue samples that were used in these studies. We also thank David Hewett-Emmett, Morris Goodman, and John Czelusniak for their help in statistically testing the trichotomy branching patterns of humans and the African apes. The contribution of M. Eloise Canfield in determining the orangutan and squirrel monkey sequences is also gratefully acknowledged. These studies were supported by NIH grant GM-24681.

REFERENCES

- Bachmann, B., Luke, W. and Hunsmann, G: Improvement of PCR amplified DNA sequencing with the aid of detergents. *Nucleic Acids Res.* 18 (1990) 1309.
- Bailey, W.J., Fitch, D.H.A., Tagle, D.A., Czelusniak, J., Slightom, J.L. and Goodman, M.: Molecular evolution of the $\psi\beta$ -globin gene locus: gibbon phylogeny and the hominoid slowdown. *Mol. Biol. Evol.* 8 (1991) 155-184.
- Bailey, W.J., Hayasaka, K., Skinner, C.G., Kehoe, S., Sieu, L.C., Slightom, J.L. and Goodman, M.: Reexamination of the African

- hominoid trichotomy with additional sequences from the primate β -globin gene cluster. *Mol. Phylogenet. Evol.* 1 (1992) 97–135.
- Bergenheim, N.C.H., Venta, P.J., Hopkins, P.J. and Tashian, R.E.: Variation in coding exons of two electrophoretic alleles at the pigtail macaque carbonic anhydrase I locus as determined by direct, double-stranded sequencing of PCR products. *Biochem. Genet.* 30 (1992a) 279–287.
- Bergenheim, N.C.H., Venta, P.J., Hopkins, P.J., Kim, H.J. and Tashian, R.E.: Mutation creates an open reading frame within the 5' untranslated region of macaque erythrocyte carbonic anhydrase I (CA I) mRNA that suppresses CA I expression and supports the scanning model for translation. *Proc. Natl. Acad. Sci. USA* 89 (1992b) 8798–8802.
- Brady, H.J.M., Sowden, J.C., Edwards, M., Lowe, N. and Butterworth, P.H.W.: Multiple GF-1 binding sites flank the erythroid specific transcription unit of the human carbonic anhydrase I gene. *FEBS Lett.* 257 (1989) 451–456.
- Brady, H.J.M., Lowe, N., Sowden, J.C., Edwards, M. and Butterworth, P.H.W.: The human carbonic anhydrase I gene has two promoters with different tissue specificities. *Biochem. J.* 277 (1991) 903–905.
- Butterworth, P.H.W., Barlow, J.H., Brady, H.J.M., Edwards, M., Lowe, N. and Sowden, J.: The structure and regulation of the human carbonic anhydrase I gene. In: Dodgson, S.J., Tashian, R.E., Gros, G. and Carter, N.D. (Eds.), *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*. Plenum Press, New York, NY, 1991, pp. 197–207.
- Casanova, J.-L., Pannetier, C., Jaulin, C. and Kourlisky, P.: Optimal conditions for directly sequencing double-stranded PCR products with Sequenase. *Nucleic Acids Res.* 18 (1990) 4028.
- Contel, E.P.B., Hewett-Emmett, D., Stroup, S.K. and Tashian, R.E.: Amino acid sequence of chimpanzee carbonic anhydrase I (CA I): evolutionary implications for the origins of humans and great apes. *Isozyme Bull.* 14 (1981) 44.
- Czelusniak, J., Goodman, M., Moncrief, N.D. and Kehoe, S.M.: Maximum parsimony approach to construction of evolutionary trees from aligned homologous sequences. *Methods Enzymol.* 183 (1990) 601–615.
- Dijon, P. and Green, H.: The involucrin gene of the gibbon: the middle region shared by the hominoids. *Mol. Biol. Evol.* 7 (1990) 220–227.
- Ericksson, A.E. and Liljas, A.: X-ray crystallographic studies of carbonic anhydrase isozymes I, II and III. In: Dodgson, S.J., Tashian, R.E., Gros, G. and Carter, N.D. (Eds.), *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*. Plenum Press, New York, NY, 1991, pp. 33–48.
- Fraser, P., Cummings, P. and Curtis, P.: The mouse carbonic anhydrase I gene contains two tissue-specific promoters. *Mol. Cell Biol.* 9 (1989) 3308–3313.
- Hewett-Emmett, D. and Tashian, R.E.: Structure and evolutionary origins of the carbonic anhydrase multigene family. In: Dodgson, S.J., Tashian, R.E., Gros, G. and Carter, N.D. (Eds.), *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*. Plenum Press, New York, NY, 1991, pp. 15–32.
- Hewett-Emmett, D., Hopkins, P.J., Tashian, R.E. and Czelusniak, J.: Origins and molecular evolution of the carbonic anhydrase isozymes. *Ann. N.Y. Acad. Sci.* 429 (1984) 338–358.
- Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S. and Takahata, N.: Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J. Mol. Evol.* 35 (1992) 32–43.
- Koop, B.F., Tagle, D.A., Goodman, M. and Slightom, J.L.: A molecular view of primate phylogeny and important systematic and evolutionary questions. *Mol. Biol. Evol.* 6 (1989) 580–612.
- Lowe, N., Brady, H.J.M., Barlow, J.H., Sowden, J.C., Edwards, M. and Butterworth, P.H.W.: Structure and methylation patterns of the gene encoding human carbonic anhydrase I. *Gene* 93 (1990) 277–283.
- Miyamoto, M.M., Slightom, J.L. and Goodman, M.: Phylogenetic relations of humans and African apes from DNA sequences in the β -globin region. *Science* 238 (1987) 369–373.
- Miyamoto, M.M., Koop, B.F., Slightom, S.L. and Goodman, M.: Molecular systematics of higher primates: genealogical relations and classification. *Proc. Natl. Acad. Sci. USA* 85 (1988) 7627–7631.
- Nicewander, P.H.: Sequence and Organization of a *Macaca nemestrina* carbonic Anhydrase I Gene. Ph.D. Dissertation, University of Michigan, Ann Arbor, 1990.
- Ruvolo, M., Disotell, T.R., Allard, M.W., Brown, W.M. and Honeycutt, R.L.: Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. *Proc. Natl. Acad. Sci. USA* 88 (1991) 1570–1574.
- Saitou, N. and Nei, M.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (1987) 406–425.
- Sowden, J., Edwards, M., Morrison, K., Butterworth, P.H.W. and Edwards, Y.H.: Erythroid expression and DNAaseI-hypersensitive sites of the carbonic anhydrase gene. *Biochem. J.* 288 (1992) 545–551.
- Tashian, R.E.: Evolution and regulation of the carbonic anhydrase isozymes. In: Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (Eds.), *Isozymes: Current Topics in Biological and Medical Research*, Vol. 2. Alan R. Liss, New York, NY, 1977, pp. 21–62.
- Tashian, R.E.: Genetics of the mammalian carbonic anhydrases. *Adv. Genet.* 26 (1992) 321–356.
- Tashian, R.E., Hewett-Emmett, D. and Goodman, M.: On the evolution and genetics of carbonic anhydrases I, II, and III. In: Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. (Eds.), *Isozymes: Current Topics in Biological and Medical Research*, Vol. 7. Alan R. Liss, New York, NY, 1983, pp. 79–100.
- Tashian, R.E., Venta, P.J., Nicewander, P.H. and Hewett-Emmett, D.: Evolution, structure, and expression of the carbonic anhydrase multigene family. *Prog. Clin. Biol. Res.* 344 (1989) 159–175.
- Thierfelder, W., Cummings, P., Fraser, P. and Curtis, P.J.: Expression of carbonic anhydrases I and II in mouse erythrocytes. In: Dodgson, S.J., Tashian, R.E., Gros, G. and Carter, N.D. (Eds.), *The Carbonic Anhydrases: Cellular Physiology and Molecular Genetics*. Plenum Press, New York, NY, 1991, pp. 209–214.