Copyright © 1999 by the Genetics Society of America

Origin of Gene Overlap: The Case of *TCP1* and *ACAT2*

Seikou Shintani, Colm O'hUigin, Satoru Toyosawa, Věra Michalová and Jan Klein

Max-Planck-Institut für Biologie, D-72076 Tübingen, Germany

Manuscript received December 10, 1998
Accepted for publication February 11, 1999

ABSTRACT

The human acetyl-CoA acetyltransferase 2 gene, *ACAT2*, codes for a thiolase, an enzyme involved in lipid metabolism. The human T-complex protein 1 gene, *TCP1*, encodes a molecular chaperone of the chaperonin family. The two genes overlap by their 3′-untranslated regions, their coding sequences being located on opposite DNA strands in a tail-to-tail orientation. To find out how the overlap might have arisen in evolution, the homologous genes of the zebrafish, the African toad, caiman, platypus, opossum, and wallaby were identified. In each species, standard or long polymerase chain reactions were used to determine whether the *ACAT2* and *TCP1* homologs are closely linked and, if so, whether they overlap. The results reveal that the overlap apparently arose during the transition from therapsid reptiles to mammals and has been retained for >200 million years. Part of the overlapping untranslated region shows remarkable sequence conservation. The overlap presumably arose during the chromosomal rearrangement that brought the two unrelated and previously separated genes together. One or both of the transposed genes found by chance signals that are necessary for the processing of their transcripts to be present on the noncoding strand of the partner gene.

THE classical notion of genes being arranged on chromosomes in a beads-on-string-like fashion had to be amended when the first genomic sequences, those of viruses, became known. One of the first such sequences, that of the bacteriophage ϕX174 (Sanger *et al.* 1977), revealed that in addition to the tandemly arranged genes, there were two pairs of genes in which different proteins were translated in two reading frames from a common DNA sequence (Barrell *et al.* 1976). This observation was the first documented instance of overlapping genes. Many other instances of overlapping genes and other deviations from strictly tandem gene arrangements in the genomes have been described since then (Normark *et al.* 1983). In some cases, a gene of a given pair is nested within another gene, as in the two ϕX174 pairs; in others, the two genes overlap partially.

Overlapping genes occur frequently in viral genomes as well as in genomes of cellular prokaryotes and prokaryote-derived organelles such as mitochondria (Normark *et al.* 1983). They occur less frequently in nuclear genomes of eukaryotes (Williams and Fried 1986; Adelman *et al.* 1987; Emi *et al.* 1988; Morel *et al.* 1989; Cawthon *et al.* 1991; Laudet *et al.* 1991; Grima *et al.* 1992; Shayiq and Avadhani 1992; Ashworth 1993; Nicoloso *et al.* 1994; Aaronson *et al.* 1996; Hadano *et al.* 1996; Swalla and Jeffery 1996; Van Bokhoven *et al.* 1996; Cooper *et al.* 1998; Joseph 1998). There are two principal types of overlapping genes: in one type, the transcribed (and translated) reading frames of the genes are on the same DNA strand; in the other, they are on complementary strands. The former type is less common than the latter, at least in nuclear genomes.

Little is known about the manner in which the overlap arises during evolution. Though overlaps have been identified, less effort has been expended on determining their origins. We consider two ways by which, theoretically, overlapping genes can come into existence. Within a gene-constituting DNA stretch, often more than one reading frame, either on the same or on complementary strands, can potentially code for a peptide. If an initiation codon and a transcription initiation site arise by chance within the stretch and in register with the extra open reading frame, two or more mRNA types may be transcribed from the same locus. Alternatively, two independently derived genes on the same or on different chromosomes can be brought together, for example, by translocation, and arranged in such a way that each derives part of its transcript from the same or complementary DNA sequence as the other.

To investigate the mode of origin of overlapping eukaryotic genes, we chose the *ACAT2-TCP1* pair. The *ACAT2* or acetyl-CoA acetyltransferase 2 gene codes for an enzyme of lipid metabolism, a member of the thiolase family (Clinkenbeard *et al.* 1973; Middleton 1973, 1974; Song *et al.* 1994). Thiolases, a widely distributed group of enzymes found in both prokaryotes and eukaryotes, are of two basic types: 3-ketoacyl-CoA thiolases (type I, EC 2.3.1.16) and acetoacetyl-CoA thiolases (type II, EC 2.3.1.9, see Middleton 1975; Igual *et al.* 1992). The former are involved in the degradation of fatty acids by β-oxidation; the latter participate in the biogenesis of

Corresponding author: Jan Klein, Max-Planck-Institut für Biologie, Abteilung Immungenetik, Corrensstr. 42, D-72076 Tübingen, Germany. E-mail: jan.klein@tuebingen.mpg.de

steroids and the formation of ketone bodies. Type I thiolases have a broad chain-length specificity; type II enzymes are specifically involved in the thiolysis of acetyl-CoA. Mammalian genomes contain at least five thiolase-encoding loci that specify three mitochondrial type I thiolases (Fukao *et al.* 1990; Abe *et al.* 1993; Kamijo *et al.* 1994), one peroxisomal type I thiolase (Fairbairn and Tanner 1989), and one cytosolic type II thiolase (Song *et al.* 1994). In humans, the locus encoding the cytosolic thiolase is designated *ACAT2*; the encoded enzyme catalyzes the condensation of two acetyl-CoA molecules into acetoacetyl-CoA, which is then converted via several steps into steroids (Middleton 1974). The rat *ACAT2* product is a homotetramer that is abundantly expressed in the liver, brain, and adrenals and poorly expressed in most other tissues (Middleton 1974).

The *TCP1* or T-complex protein 1 gene codes for a molecular chaperone that assists in the folding of proteins during their synthesis or their recovery from a denatured state (Willison *et al.* 1986; Ellis and van der Vies 1991; Horwich and Willison 1993). It is also a member of a large family of proteins, specifically a class of molecular chaperones known as chaperonins (Yaffe *et al.* 1992; Horwich and Willison 1993; Kubota *et al.* 1994). This class includes GroEL of *Escherichia coli*, the mitochondrial heat shock protein Hsp60, the plastid Rubisco subunit-binding protein, and the archaeal protein TF55. Chaperonins are involved in the folding, transport, and assembly of newly synthesized proteins that, in the case of *TCP1*-containing chaperonins, include actin and tubulin. In mammals, the *TCP1* gene codes for one subunit (α) of a particle that contains at least six other subunit types (β, γ, δ, ε, ζ, and η) that are all encoded in distinct but related genes (Kubota *et al.* 1994). Because the *TCP1* gene has apparently nothing to do with the *t*-complex phenotype, it has been renamed *CCTA*, for chaperonin-containing TCP1α (Kubota *et al.* 1994). In the mouse, the *TCP1* gene is expressed in several tissues, but most abundantly in the testes (Willison *et al.* 1986).

In mice and humans, the *ACAT2* and *TCP1* genes are located in the same chromosomal regions on chromosomes 17 and 6q25.3-q26, respectively (Willison *et al.* 1987; Ashworth 1993; Masuno *et al.* 1996). In both species, the two genes overlap in a manner shown in Figure 1. The coding sequences of the two genes are located on opposite DNA strands and are in a tail-to-tail orientation to each other. They share a DNA segment encompassing portions of their 3′-untranslated regions (UTRs) and, in one direction, also part of the translated region. In the mouse, the pairs have apparently undergone a tandem duplication so that the genes are arranged in the order *TCP1 . . . ACAT2 . . . TCP1 . . . ACAT2.* (One of the mouse *ACAT2* genes was originally erroneously designated *Tcp1x*, see Dudley *et al.* 1991.) The aim of the present study was to determine at which stage in vertebrate evolution and in what manner the overlap of the *ACAT2* and *TCP1* genes arose.

MATERIALS AND METHODS

Source and isolation of DNA: The spleen of an adult red-necked wallaby (*Macropus rufogriseus*) was obtained from an animal that died in the Hamburg-Hagenbeck Zoological Garden. Tissues from the gray short-tailed opossum (*Monodelphis domestica*) were obtained from the colony maintained by Professor W. H. Stone (Department of Biology, Trinity University, San Antonio, Texas). DNA from the duck-billed platypus (*Ornithorhynchus anatinus*) was provided by Dr. Robert W. Slade (Queensland Institute for Medical Research, Royal Brisbane Hospital, Australia). Fertilized eggs of a smooth-fronted caiman (*Paleosuchus palpebrosus*) were obtained from Dr. Hans-Peter Herrmann (Köln Zoo, Germany). African clawed toads (*Xenopus laevis*) were provided by Dr. C. Dreyer (the Max Planck Institute for Developmental Biology, Tübingen, Germany). Zebrafishes (*Danio rerio*) bred in our aquarium were used. All tissue samples were kept frozen at $-70°$ until their use. Genomic DNA was isolated from the tissues by phenol-chloroform extraction.

cDNA library construction and screening: Animals were killed under anesthesia, and their tissues were removed and frozen in liquid nitrogen. The frozen tissues were homogenized to a fine powder, and total RNA was extracted. Poly(A)$^+$ RNA isolation and cDNA synthesis were performed with the help of the mRNA purification kit (Pharmacia Biotech, Freiburg, Germany) and the TimeSaver cDNA synthesis kit (Pharmacia Biotech), respectively. The cDNA was inserted into the *Eco*RI-digested λgt10 vector (Stratagene, Heidelberg, Germany), and the cDNA library was *in vitro* packaged with the help of the Gigapack cloning kit (Stratagene) and used to transform competent *E. coli* MN514 bacteria. The opossum, caiman, toad, and zebrafish libraries were amplified once to titers of 4.0×10^{10}, 1.5×10^{11}, 1.8×10^{11}, and 1.0×10^{11} pfu, respectively.

PAC clone screening: PAC zebrafish library filters (library BUSMP706) were obtained from the Resource Center of the German Human Genome Project at the Max-Planck-Institut for Molecular Genetics (Berlin). Probes for *TCP1* were prepared by polymerase chain reaction (PCR) amplification of the zebrafish library using the zebrafish-specific primers Tcp F5 and Tcp F7, as well as Acat F12 and Acat F14, which yielded products of ~600 and 450 bp, respectively, covering the bulk of the translated regions of the corresponding gene transcripts. Filters were hybridized at 65° in 7% sodium dodecylsulfate (SDS), 0.5 m sodium phosphate, pH 7.2, and 1 mm ethylenediaminetetraacetic acid (EDTA), and were washed twice in 40 mm sodium phosphate containing 0.1% SDS. Positive hybridization signals from two *ACAT2*-containing PAC clones (numbers G1276Q2 and G23214), as well as two *TCP1*-containing clones (numbers H0274Q2 and O23263), were confirmed by PCR amplification of the probe regions from the PAC clones.

PCR amplification: *ACAT2* sequences of the opossum, caiman, toad, and zebrafish were amplified from cDNA libraries using primers based on a comparison of human and mouse sequences (Table 1) in combination with a vector primer in an anchored PCR. Genomic DNA or lysate of the cDNA libraries (1 μl) was amplified by PCR in 50 μl PCR buffer (1.5 mm $MgCl_2$, 200 μm dNTP, 10 mm Tris buffer, pH 8.5) in the presence of the two primers and 2.5 units of *Taq* polymerase (Pharmacia Biotech). Amplification was performed in the PTC-100 Thermal Cycler (MJ Research Inc., Watertown, MA). After the first cycle at 95° for 3 min, 35 cycles followed, each consisting of 1 min denaturation at 95°, 1 min annealing at

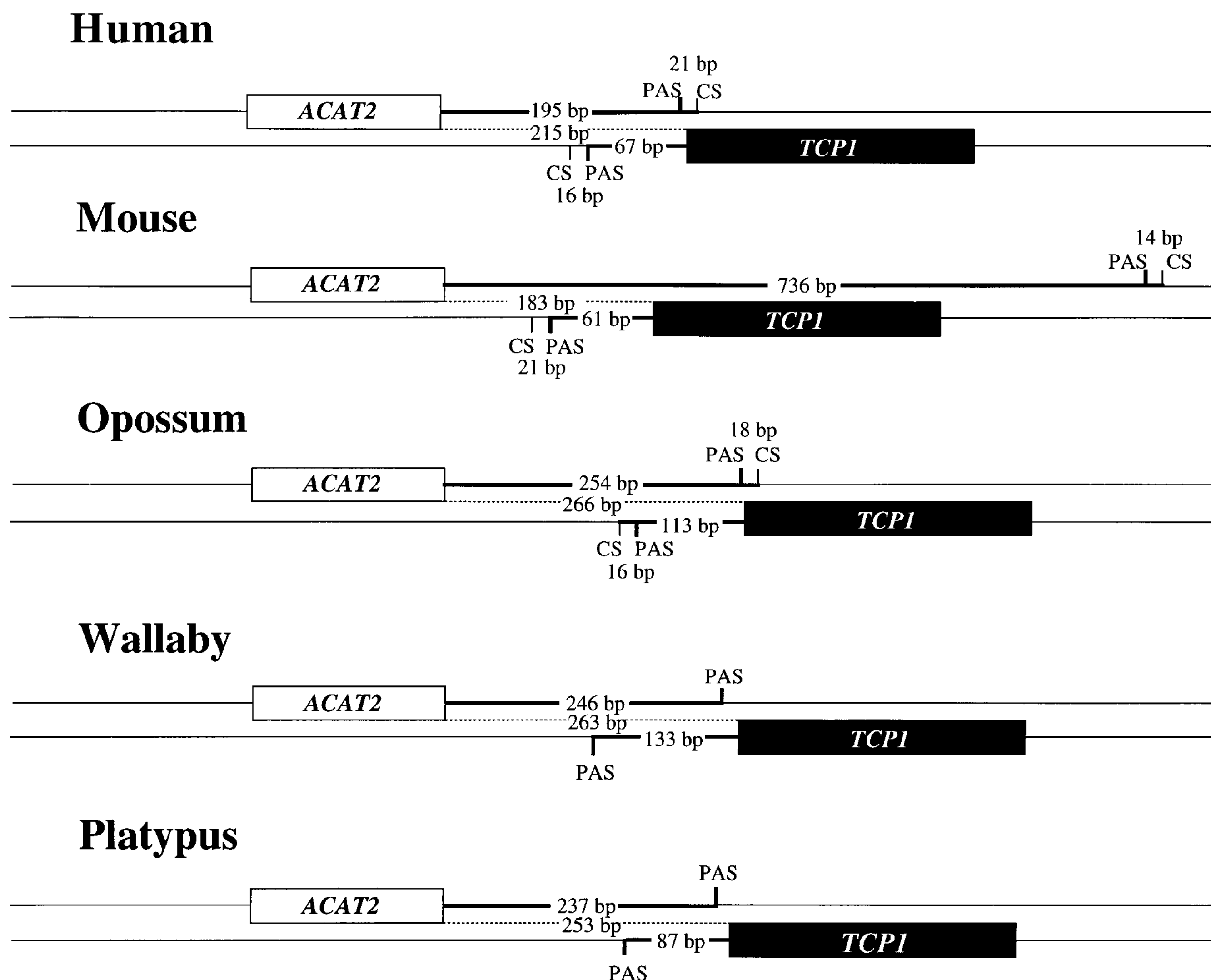

Figure 1.—Diagram depicting the *ACAT2-TCP1* overlapping region in different vertebrate species. The human and mouse diagrams are based on the data of Ashworth (1993), Song *et al.* (1994), Dudley *et al.* (1991), and Kubota *et al.* (1994). The diagrams for the remaining species are deduced from the data described in the present study. For each species, the two strands of genomic DNA are depicted, the last exon of the coding strand is shown as a rectangle, the 3′-untranslated region as a thick line, and the rest of the strand as a thin line. The distances between the stop codons of the two genes (dotted line), between the stop codon and the conserved polyadenylation signal (PAS), as well as between the PAS and the cleavage site (CS) of the transcript are indicated in base pairs. Alternative PAS, which are known, for example, in humans (Willison *et al.* 1987), are not indicated.

the annealing temperature, and 2 min extension at 72°. The final extension was for 10 min at 72°. Long PCR was carried out with the help of the GeneAmp XL PCR Kit (Perkin Elmer Applied Biosystems, Freiburg, Germany) in the GeneAmp PCR system 9600 (Perkin Elmer-Cetus, Norwalk, CT) and consisted of one cycle at 94° for 30 sec, followed by 12 cycles, each for 30 sec at 94° and 10 min at 64°. In the next 24 cycles, the reaction time at 64° was extended by 15 sec in every cycle; the reaction was completed by a final primer extension for 10 min at 72°.

Cloning and sequencing: Twenty microliters of the PCR amplification product was purified by electrophoresis in 1.5% low-melting-point agarose (GIBCO BRL, Eggenstein, Germany) and the band was identified by ethidium bromide staining, excised, and isolated from the gel using the QIAEX extraction kit (Hilden, Germany). The isolated DNA was blunt ended, phosphorylated, and ligated to *Sma*I-digested pUC18 plasmid vector with the SureClone ligation kit (Pharmacia Biotech). The reaction products were transformed into competent *E. coli* XL-1 blue bacteria by standard methods and plated on LB agar containing ampicillin (50 μl/ml). Transformants were grown overnight in LB broth containing ampicillin, and minipreps were prepared according to the standard Qiagen protocol. Two to five micrograms of DNA were used in the dideoxy sequencing reactions with the AutoRead Sequencing kit (Pharmacia Biotech). The reactions were processed by the automated laser fluorescent sequencer (Pharmacia Biotech). The GenBank accession nos. of the sequences are AF143488–AF143500.

Data analysis: The nucleotide sequences were aligned with the aid of the SeqPup computer program (Gilbert 1995). Sequence similarities were evaluated with the aid of the DottyPlot computer program (Gilbert 1989). Substitution rates were estimated by the method of Li (1993) from sequences

TABLE 1

Oligonucleotide primers used in this study

Designation	Sequence	Orientation	Location
Acat 4	5′-ACTGGTACAATCTCTTTGTCAAA-3′	A	E3
Acat 7	5′-CCATCATTTATTCCTGAAGCATT-3′	A	E4/E3
Acat C1	5′-TATTGCCCTTGGACACCCTCT-3′	S	E7
Acat F4	5′-TTTGACAAGGAAATTGTGCCAGT-3′	S	E3
Acat F8	5′-GCGATCTCTCTTGGTCATCCTCT-3′	S	E7
Acat F12	5′-CACACATAGCGATGCCCATTC-3′	A	E7
Acat F14	5′-GGTACTGTGACAGCTGGTAATG-3′	S	E5
Acat G1	5′-AGCCATTGCCYTKGGTCATC-3′	S	E7
Acat G3	5′-TGTGCATYGGCGGAGGGAT-3′	S	E7
Acat O4	5′-CTGGCYATTTTGACAARGAGATTGT-3′	S	E3
Acat X1	5′-CACATGGGAATAACAGCTGAAAA-3′	S	E3
Acat Xa1	5′-CCCTCTGGAATGTCTGGTTGC-3′	S	E7
Tcp C1	5′-ATTGGGCTTGACTTAATAAATGGAA-3′	S	E12
Tcp F4	5′-ATTACCATCCTTCGAATCGATCA-3′	S	E12
Tcp F5	5′-GATTGGACGGCGTCCTGATA-3′	A	E12
Tcp F7	5′-GGTTTGCTCTTCTCTGTCCAA-3′	S	E9
Tcp G1	5′-TGTGCATYGGCGGAGGGAT-3′	S	E12
Tcp G3	5′-AAGCAAGCAGGGGTGTTTGAA-3′	S	E12
T cp G8	5′-AGCTGCTATTGCAGAGTTTGC-3′	S	E11
Tcp Xa 1	5′-TAACAAGCAGGCTGGGGTCTTC-3′	S	E12

A, antisense; S, sense; E, exon.

aligned using the CLUSTAL W program (Thompson *et al.* 1994).

RESULTS

Human *ACAT2* gene organization: As a prelude to the main study, we determined the exon-intron borders of the human *ACAT2* gene to better plan experiments concerned with the overlap of the *ACAT2* and *TCP1* genes and interpret their results. To this end, we used the published cDNA sequences (Song *et al.* 1994) to synthesize a series of primer pairs corresponding to stretches spaced at short intervals along the sequence. In each pair, the sense and antisense primers corresponded to cDNA sequence stretches at distances ranging from 49 to 290 bp from each other (Figure 2). The PCR amplification products obtained by the application of these primer pairs to human genomic DNA were cloned and sequenced. The comparison of the genomic with the cDNA sequences revealed the positions of the exon-intron borders in the human *ACAT2* gene (Figure 2). The gene consists of at least seven exons interrupted by six introns; the number could be higher should the 5′- and 3′-UTRs turn out to contain additional introns. Because the *ACAT2* gene is conserved, it can be assumed that it has a similar exon-intron organization in other vertebrates.

Strategy: Before the start of this study, the *ACAT2-TCP1* gene overlap had been known to exist only in representative species of the mammalian orders Primates (humans, Ashworth 1993) and Rodentia (house mouse, Dudley *et al.* 1991). Because primates and rodents were among the first orders to diverge from each other during the adaptive radiation of modern eutherian mammals (Novacek 1992), it can be assumed that the overlap was established before the emergence of extant Eutheria. To trace its evolutionary origin, we therefore turned to representatives of two noneutherian orders, Marsupialia (opossum and wallaby) and Monotremata (duck-billed platypus), as well as representatives of other classes of jawed vertebrates, the Reptilia (caiman), Amphibia (clawed toad), and Osteichthyes/Actinopterygii (zebrafish). In each instance, we first cloned and sequenced the 3′ part of the two genes (including a segment of the translated region) to make sure that we identified the homologs of the human/mouse *ACAT2* and *TCP1* genes rather than some other members of the two gene families. We then carried out Southern blot analysis of genomic DNA from the individual species, using the identified gene segment as a probe to determine whether multiple copies of the gene were present in the genome. Gene pairs thus identified were then tested for an overlap of their 3′ ends by PCR amplification with primers complementary to a sequence stretch of the last exon in each of the two genes, using genomic DNA as a template. A failure to amplify a product, even under long-PCR conditions, which should allow an amplification with primers up to 15 kb apart, was taken as evidence that *ACAT2* and *TCP1* homologs were not overlapping in the species under investigation. Description of the results obtained in the study of the individual taxa follows.

Zebrafish homologs: The zebrafish *TCP1* homolog

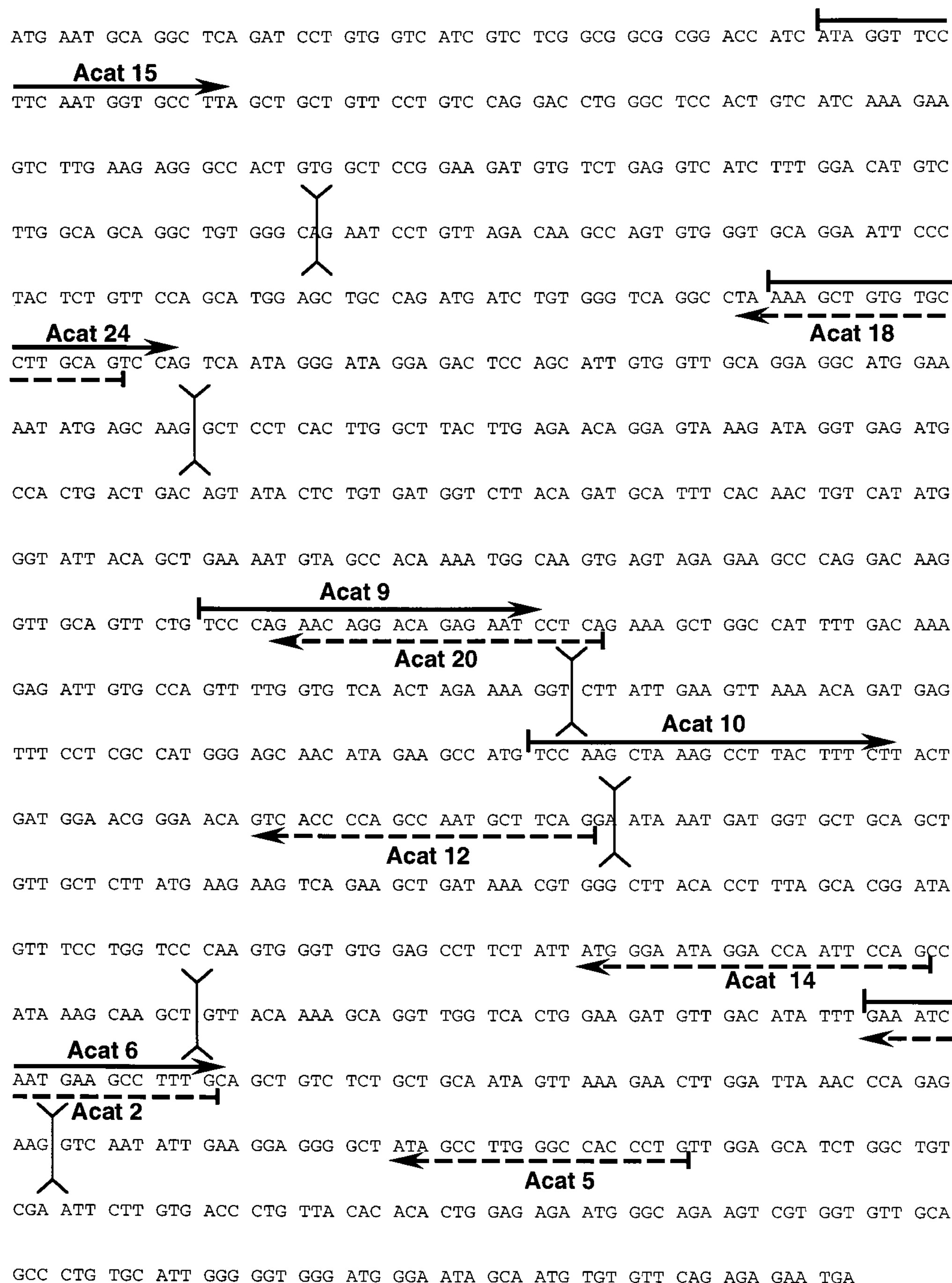

Figure 2.—Exon-intron borders (⟩⟨) of the human *ACAT2* gene as determined by PCR amplification of genomic DNA with the indicated primer pairs (solid and broken arrows show the positions of the sense and antisense primers, respectively). The sequence is from Song *et al.* (1994).

was cloned, and the entire coding sequence was determined in a separate study that had been initiated for a different purpose (K. Takami, F. Figueroa, and J. Klein, unpublished results; see Figures 3 and 4). The entire zebrafish *ACAT2* coding sequence was determined in two steps using the liver cDNA library. In the first step, the Acat 7 primer spanning the exon 3/4 border was used in conjunction with the vector primer to amplify and sequence the 5′ half of the coding region. (Here we assumed that the exon-intron organization of the zebrafish *ACAT2* gene is the same as that of the human gene. The Acat 7 primer was based on the human sequence, which we correctly assumed to be very similar to the zebrafish sequence in this part of the gene.) In the second step, the primer Acat F4, which was based on the zebrafish sequence obtained in the first step, was used in combination with another vector primer to amplify and sequence the 3′ half of the zebrafish *ACAT2* coding region (Figure 4). Comparison of the zebrafish *ACAT2* and *TCP1* sequences failed to identify a homology region that would be indicative of an overlap between the two genes. Similarly, both standard and long-PCR experiments using the Acat F8 and Tcp F4 primers failed to yield a product that would be expected

TCP1

```
               1          11         21         31         41         51         61         71         81         91
CONSENSUS ==>  MAAMEGPLAV LGERSTGEVV RSQNVMAAAS IANIVKSSLG PVGLDKMLVD DIGDVTITND GATILKLLEV EHPAAKVLCE LADLQDNEVG DGTTSVVIIA
Zebrafish      ..MIDS--S- --Q-T---S- ---------- ---------- ---------- ---------- ---------- ---------- --E---K--- ----------
Toad A         ...------- F--------- ---------- ---------- ---------- G--------- ---------- ---------- ---------- ----------
Toad B         ...--A---- F--------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Caiman         ---L------ -----S-DT- -N---T---T ---------- ---------- ---------- ---------- ---------- --E---K--- ----------

               101        111        121        131        141        151        161        171        181        191
CONSENSUS ==>  AELLKNADEL VKQKIHPTSV IGGYRLACKE AVRYINENLT INTDELGKDC LLNAAKTSMS SKIIGVDGDF FAAMVVDAAL AVKFVDPKGQ ARYPINSVNV
Zebrafish      -----S---- ---------I -S-------- ---------- -G--D--RE- ---------- -------AE- --N------V ------G--V ----------
Toad A         ---------- ---------- ---------- ---------- --------E- ---------- -----I---- SS-------- ---Y------ ----------
Toad B         ---------- ---------- ---------- ---------- ---------- ---------- -------S-- -ST------- -------R-- -------I--
Caiman         ------S--- ---------I ---------- ---------I -------R-- -I-S------ -----I---- --S------S ---YT-Q--- -------I--

               201        211        221        231        241        251        261        271        281        291
CONSENSUS ==>  LKAHGRSQKE SILVNGYALN CIVGSQGMTK RIVNAKIACL DFSLQKTKMK LGVQVVISDP TKLDQIRQRE SDITKERIQK ILATGANVIL TTGGIDDMCL
Zebrafish      ---------- -F-------- -T------V- -VA------- ---------- -------N-- E--------- -------V-- ---S----V- ----------
Toad A         --------M- ---------- ------S-N- ---------- ---------- -----I---- ---------- ---------- ---------- ----------
Toad B         ---------- ---------- -▼----S--- --L------- ---------- -----I---- ---------- ---------- ---------- ----------
Caiman         ---------- ---------- -V-------- ---------- ---------- -------T-- E--------- ---------- ---------- ----------

               301        311        321        331        341        351        361        371        381        391
CONSENSUS ==>  KYFVDAGAMA VRRVLKKDLK RIAKATGATV CSTLANLEGE ESFEASMLGQ AEEVVQERVC DDELILVKNT KARTCASIIL RGANDFMCDE MERSVHDALC
Zebrafish      -----V---- ---------- ---------- --S-S----- -T--P----- ---------- ------I--- ----S----- ---------- ----L-----
Toad A         ---------- ---------- ---------M ---------- ---------- --------I- ---------- ---------- ---------- ----------
Toad B         ---------- ---------- C------G-L L--------- ---------- --------I- ---------- -------V-- ---------- ----------
Caiman         ---------- ---------- -----S---- ---------- ---------- ---------- ------I--- ----S----- ---------- ----I-----

               401        411        421        431        441        451        461        471        481        491
CONSENSUS ==>  VVKRVLESKS VVPGGGAVEA ALSIYLENYA TSMGSREQLA IAEFARALLV IPKTLAVNAA QDSTDLVAKL RAFHNEAQVN PERKNLKWIG LDLLNGKPRD
Zebrafish      ---------- ---------- ---------- ---------- ------S--- ---------- ---------- ---------- -D-------- ---V------
Toad A         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Toad B         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---R------ ----------
Caiman         ---------- ---------- ---------- ---------- ------S--I --N------- --A------- ---------- ---------- ---I------

               501        511        521        531        541        551
CONSENSUS ==>  NKQAGVFEPT LVKTKSLKFA TEAAITILRI DDLIKLHPEA KEDEGGKYHD AVQSGSLEN
Zebrafish      ------Y--- M--------- ---------- ------F-DQ --*G-PS-Q- --------G
Toad A         ---------- ---------- ---------- --V------- --*-K----E ------I--
Toad B         ---------- ---------- ---------- ------L--- --*-K----E ------I--
Caiman         ---------- ---------- ---------- ---------S -D-K--C-E- --R--A--E
```

ACAT2

```
               1          11         21         31         41         51         61         71         81         91
CONSENSUS ==>  MSGEHMSSGG DAVVIVSAAR TPIGSFNGAL STLPAHTLGS TVIKEVLKRA TIKPEEVSEV IFGQVLTAGA GQNPARQASV AAGVPYSVPA WSCQMICGSG
Zebrafish      .......MNT --I------- ----A---T- -SV-L-E--T L---D----V NV-------- -M-H-----H ---------- G--------- ----------
Toad A         .....---SQ ET-------- ---------- ---------- ---------- ---------- ---------- ---------- -------I-- ----------
Toad B         .....-N--E ET---I---- ---------- ---------- ---------- ---------- ---------- ---------- -----#-I-- ----------
Caiman         -----GDGD- ----V----- --V------- --MQ--E--- V--------- N-N------- ---------- ----V----- G--I--A--- ----------

               101        111        121        131        141        151        161        171        181        191
CONSENSUS ==>  LKAVCLGAQS IMTGEADIVV AGGMESMSQA PHLVHMRAGV KAGDVSLQDS IICDGLNDAF YKYHMGITAE NVAKQWQITR EEQDQLAVQS QNRTEAAQKA
Zebrafish      --S------- --MK-ST--- --------R- --I-Q--S-- -M--AT---T -LT---T--- -N--S----- --XS--GVSC -A------T- ----------
Toad A         ----S----- -K-------- -----N---- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Toad B         ----S----- -K-------- ----GN---- ---------- -------#-- ---------- ---------- ----#----- ---------- ----------
Caiman         ---------- L--EDSS--- --------K- ---I------ -M-ET----- ------R-V- -Q-------- ------K-S- A--------- -T-A-G----

               201        211        221        231        241        251        261        271        281        291
CONSENSUS ==>  GYFDKEIVPV SVPSRKGPVE VKVDEFPRHG SNVDAMSKLK PCFLKDGSGT VTAANASGIN DGAAAVVLIK ESEARRRGLT PMARIVSWAQ VGLDPSIMGV
Zebrafish      ---------- T--------- ---------- --I------- --SV------ ---G------ -----T--MS Q---Q----K -----T---- A-----V--T
Toad A         ---------- ---------- ---------- --F------- -Y-------- --P------- ---------- -----H---- -------SG- ----------
Toad B         ---------- ---------- ---------- ---E------ -Y-------- --P------- ---------- ---------- ------AS-- ----------
Caiman         --------S- L--------- --T--H---- --LET-A--- ----T--T-- ---------- ------I-M- K---A----L -L-Q------ ----------

               301        311        321        331        341        351        361        371        381        391
CONSENSUS ==>  GPIAAIRKAV EKAGWSLDEV DLFEINEAFA AQALAVVKDL GLNPEKVNCQ GGAVALGHPL GMSGCRILVT LLYALERTGG KKGVAALCIG GGMGIAMCVE
Zebrafish      ---P-----I ---R-K---- ---------- S-SI---Q-- ----D---VC ---IS----- ------V--- --H--Q---- R----S---- ----------
Toad A         ---------- ---------I ---------- ---------- ---------- ---------- ---------- ---------- -----R---- --------I-
Toad B         ---------- --------DI ---------- ---------- ---------- ---------- ----S----- ---------- ----V----- ----------
Caiman         ------K--- N----M-EQ- ---------- SV-A-IA-E- K-------IE ---I------ -A-------- --H------- -R-------- --------I-

               401
CONSENSUS ==>  RTS
Zebrafish      -*E
Toad A         --N
Toad B         ---
Caiman         -**
```

Figure 3.—Amino acid sequence of zebrafish, clawed toad, and caiman TCP1 and ACAT2 deduced from the corresponding cDNA sequences. Simple majority consensus is shown at the top. Identity with the consensus is indicated by dashes (—), absence of information by dots (·), and gaps introduced to optimize the alignment by asterisks (*). The ▼ and # symbols indicate the beginning of a frame shift and premature stop codons, respectively. Amino acids are given in the single-letter international code.

Zebrafish

ACAT2

...GCTATGTGTGTAGAGAGAGAGTAA GGGCAAATGAACACCCTGTAATTTTGAGCTTTTAACCAGACTTTGAATCCAGCTTTTCAC
ATTCAAAACAGATATGAAAACATATGGTAAAACTTAAGTATATATGCAGCGTTAATAGTAAGGAAACAGCCTGCAGTAAAATGCACTC
AATTCTGCCTTCTTATCTGAACTCATCTCTAATGTTTCATCATATACAGTGGATCTGTTTTGCACTTAACATGTATTGATTCTCTTCG
GGTTTATGGTACATGAAAAGGACTTGTTGAGGTATACCAGCTTTTAAGACATTCCCTTTCTGTATAAATTGTTTACAGTTGCTGAGAC
ACTCATGTAGTAGAGCTTTTATTGGTACTAACAGTAGGCCTGTCCACATGCGTGTTAGTTATGACCTGTAAGTCATATTGTGGTATTT
ATAAATATTAAATTCATCAGTTGAATGGAGTCGGTCCTGTTTCCAGCACCTTTTCCAGACCTGTCTACACTGTGCAATAGAAAACTGC
CTTAGAGAATGTCAAATGTGCCAAAAAGTAAAATTAGCCTCTATCCTGTGCTTCCATTAGGCTGCTTTTAACTCTCTCTTCTCCTAAT
TAATACCTCAGGGGAAATGCTGTTTACTTATATAGCCCGTGTTGTACA

TCP1

...GTCCAATCAGGCTCTCTAGAGGGCTAA GGTCTCTCATTGGCTGATTCTGGTGTCACTCCGCTCAGCTTTGTATTTATATGTGGA
TCCAACTGATGAGCTGTTAGTAGAGATTGTGCACTTTTTTGTTTGCGTTACTTATATTTTTGCTATTAGCATGGCTGTCTGTCATGTT
TAATGTAAAAAAAGTTCCTTTTTCTGCCATTTCAAAGAAAATAAACCAGATTGGTGGTGTGGAAAAGAGAAAAAAAA

African clawed toad

ACAT2

...GCCATGTGCATTGAAAGAACTAATTGA ATTTTTCCCATAACTAAACGAGTGTAGGTGTTTTCTTGTTCATCTGCACCTTTTGTA
CGTGGACACAAAGCACTACATACAAATGAAATGGGAATTGAAGACATTCCAGTTAAGACGCCATTGATACAGTTTTTATTAGATCACT
GTGATTTAAATCTTGCATTAAAAGCATTTACGTTTTATTTAACA

TCP1

...GTGCAGTCTGGGTCGATTGAAAATTAA CTGGTCTTCCTGCATTTCTTTTCATTACTGTTATTTATAACACATCTGTTTTCATTG
TCTATCTTGCACTTGTTGGGTCTTGCACCCAACCCCACTCCTAACAGAAATAAAAGCAAAGA

Caiman

ACAT2

...GCTATGTGTATTGAAAGA******TGA AAATGAATGAAGCTTGTTTATATGGTGTTTTAGTGGTTTCAGTTTTCCAATAAAGTT
TATCAAGCTCAAAAAAAAAAAAAAAA

TCP1

...GTTCGTTCTGGAGCACTTGAAGAATAA GGGGCATTTTATTTGTTTGAGTTATGTAATGTAGATTTTCCTGTCTGTAGTTGGGCA
GTGTGTGTTTAAAAAAAAACCACGTGCACAGTAAGAAAGGATTGTCTATATTCAGCTTTGACTTACAAACTTCTCTTGAATTTAGTGT
TTAGAATATCTCTTCCTGTTAGGAGAGCAGTGCAAGTGTAAGGCAGTGTTGTATAACTATTAAAAATA

Figure 4.—Nucleotide sequences of a portion of the last exon (rectangle) and the 3′-untranslated region of zebrafish, clawed toad, and caiman *ACAT2* and *TCP1* genes. Potential polyadenylation signals for a gene in the opposite orientation on the complementary strand (see text) are underlined.

if the two genes were overlapping. Finally, screening of the zebrafish PAC library by hybridization with zebrafish *ACAT2* and *TCP1* coding sequence probes and testing of the positive clones by PCR amplification with specific primers revealed the *ACAT2* and *TCP1* genes to be located on different PAC clones (not shown). Taken together, these three pieces of evidence indicate that the *ACAT2* and *TCP1* genes are not closely linked in the zebrafish and, hence, not overlapping.

Clawed toad homologs: To obtain the Xenopus homolog of the mammalian *TCP1* gene, the cDNA library prepared from the jaws of adult frogs (Toyosawa *et al.* 1998) was PCR amplified using Tcp G8 (a degenerate sense primer based on the comparison of available sequences) in combination with the vector primer. The amplification product of ∼469 bp covered enough of the translated region sequence to identify it as the *TCP1* homolog. Sequencing of the multiple clones obtained

from the single amplification band revealed, however, the existence of two different genes in the Xenopus, *TCP1A* and *TCP1B* (Figures 3 and 4). The *TCP1B* gene has an insertion of 4 bp in exon 7 that is responsible for a frame shift in the rest of the translated sequence (Figure 3). We assume, therefore, that *TCP1B* is a pseudogene and that *TCP1A*, which has no identifiable defect in the part sequenced, is the toad's functional gene. The two genes may be located at distinct loci, possibly in different chromosomes as a result of the genomic tetraploidization that *X. laevis* is believed to have undergone in its evolutionary history (Kobel and Du Pasquier 1986). To obtain the Xenopus *ACAT2* homolog, we first PCR amplified a product from the cDNA library using the Acat 4 antisense human sequence-based primer in conjunction with the vector primer. In the second step, we used the Acat X1 primer on the basis of the sequence of the product from the first step, in conjunction with the vector primer, to obtain by PCR the 3′-UTR of the clone. The amplification yielded a 750-bp-long fragment containing a large portion of the translated region. Sequencing of several clones isolated from the 908-bp band again revealed the existence of two distinct genes, *ACAT2A* and *ACAT2B* (Figure 3). The former appears to be an intact gene, whereas the latter is apparently a pseudogene on account of the presence of at least three premature stop codons (one in exon 2 and two in exon 3) in its sequence (Figure 3). The assumptions made above about the two copies of the *TCP1* gene, therefore, also apply to the two copies of the *ACAT2* gene. To test the possible overlap between the toad *TCP1* and *ACAT2* genes, we obtained two primers specific for different sequence stretches of each of the four genes and used them in standard and long PCR in all possible pairwise combinations, always matching a sense *ACAT2* primer with an antisense *TCP1* primer or vice versa. In none of these combinations did the PCR yield a detectable band. We conclude, therefore, that in the clawed toad, as in the zebrafish, neither of the two *ACAT2* genes is overlapping with either of the two *TCP1* genes.

Caiman homologs: Reptilian *TCP1* and *ACAT2* sequences were cloned from a cDNA library prepared from the jaws of a 3-day-old caiman (Toyosawa *et al.* 1999). The *TCP1* clones were obtained from the library by PCR amplification using the Tcp G8 sense and the vector antisense primers. The cloned and sequenced amplification product was 581 bp long and contained a large part of the translated *TCP1* region as well as the entire 3′-UTR (Figures 3 and 4). An *ACAT2* clone obtained in a similar manner using the Acat X1 (Xenopus-specific) primer was 780 bp long and also encompassed both the 3′-translated and -untranslated regions. Efforts to PCR amplify a product with one primer (Acat C1) located in the translated region of caiman *ACAT2* and another primer (Tcp C1) located in the translated region of caiman *TCP1* failed, leading us to the conclusion that the two genes are nonoverlapping in this species.

Platypus homologs: Because a platypus cDNA library was not available to us, we resorted to the use of genomic DNA. Under the assumption that the *TCP1* and *ACAT2* genes either overlap or are closely linked, we used the Tcp G3 and Acat G1 combination of primers annealing to the translated sequences of the corresponding eutherian genes. The PCR amplification of the genomic DNA with these two primers yielded a 493-bp product that upon cloning and sequencing proved to represent the overlapping *ACAT2* and *TCP1* sequences (Figure 5). The platypus sequence from the *ACAT2* stop codon on one DNA strand to the *TCP1* stop codon on the complementary strand encompasses 253 bp compared to 215 and 183 bp of the human and mouse sequences, respectively. The unavailability of platypus cDNA precluded a definitive identification of the polyadenylation signals. A putative *ACAT2* polyadenylation signal is, however, present at a distance of 237 bp from the *ACAT2* stop codon and at a distance of 11 bp from the *TCP1* stop codon. Assuming that the polyadenylation site is >15 bp downstream of the signal, the cleavage site of the *ACAT2* transcript probably overlaps with the *TCP1* translated region on the complementary DNA strand. Similarly, a putative polyadenylation is present on the *TCP1* coding strand 87 bp downstream of the *TCP1* stop codon and, therefore, the *TCP1* transcript cleavage site probably does not overlap with the translated region of the *ACAT2* gene on this strand.

Opossum and wallaby homologs: In our study, marsupials were represented by two species, the opossum and the wallaby. A cDNA library, however, was available to us only from the former species. Sequencing of PCR products obtained from both genomic DNA and cDNA (primers Acat G3 and Tcp G1) revealed an overlap of the *ACAT2* and *TCP1* genes in the opossum, and sequencing of a product of genomic DNA amplification similarly revealed an overlap of the two genes in the wallaby. In the opossum, the distance between the stop codons of the two genes was 266 bp (Figures 1 and 5).

Figure 5.—Nucleotide sequence alignment of a portion of the last exon (boxed) and the overlapping 3′ untranslated regions of the *ACAT2* and *TCP1* genes in various mammalian species. The human and mouse sequences are from Song *et al.* (1994), Kubota *et al.* (1994), and Dudley *et al.* (1991); the remaining sequences are from the present study. The known extent of 3′ untranslated regions is indicated by arrows. The conserved segment is underlined, and putative polyadenylation signals are indicated by an overhead line. (—) Identity with the simple majority consensus sequence at the top of the alignment, (*) gaps introduced to optimize the alignment, (·) nonavailability of sequence information.

```
               1          11         21         31         41         51         61         71         81         91
CONSENSUS ==> GTCAACATTG AAGGAGGAGC CATAGCCTTG GGCCATCCTG TGGGAGCATC TGGCTGCCGA ATTCTTGTGA CCCTGCTACA CACACTGGAG AGAATGGGCG
Human         -----T---- -------G-- T--------- -----C---- -T-------- ------T--- ---------- -----T---- ---------- ---------A
Mouse         --------C- -T-------- ---T------ ---------- ---------- -------A-G -----A---- --T----G-- ---C------ ---G------
Opossum       ---------C ----T----- ------T--- --A-----CA -T--GT-C-- A--A--T--- --------T- -T---T---- T--------- C-G-----T-
Wallaby       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Platypus      .......... .......... .......... .......-GC -------G-- C--T-----C --C--G--C- -T--C--CT- -G-C------ C-G-C-----

               101        111        121        131        141        151        161        171        181        191
CONSENSUS ==> GAACTCGTGG TGTTGCAGCC CTGTGCATTG GAGGAGGGAT GGGAATAGCA ATGTGTGTTC AGAGAGAATG A AATGAAAA CCTTCATCTT AAAAATTGTG
Human         ---G------ ---------- ---------- -G--T----- ---------- ---------- ---------- - -T--**** *******--- ---CT--***
Mouse         -G--C----- ---------- ---------- -G--T----- ----G-C--- ---------- ------GG-- - -C--**** *******-A- G-C-******
Opossum       ---------- A--G-----T --C------- ---------- ---------- ---------A -A-------- - --G----- ---------C -----**---
Wallaby       .......... .......... .......... .......... ---------- ---------- -A----G--- - --AA---- ******---* --T-------
Platypus      ---AG--C-- C--C-----T ------G-C- ---------- --CCG----C --------GG -----***-A G -******* ********** **********

               201        211        221        231        241        251        261        271        281        291
CONSENSUS ==> GGTCTCATGG AACAAACTGT AAAATGTAAC ATTTAATTC TTTTTAAACC GTTACGGCCT CGTCTCCGGG GTAATCGGAA CGGGGGGGGG GGTGAGGAGA
Human         *********- -----C--CA ********** *****-T--- --------** ********** ********** ********** ********** **********
Mouse         ********-C C--C-C--** ****--*G-- *****-G--- --G-----** ********** ********** ********** ********** **********
Opossum       ******--T- ---------- --------G* ***------* ***----*** ********** ********** ********** ********** **********
Wallaby       ------**** ***--***-- -----T-C-- ---------- **-----*** ********** ********** ********** ********** **********
Platypus      ********-C CGACG--*** ********** ********-- -----***-- ---------- ---------- ---------- ---------- ----------

                                                                                                        ◄—— Mouse TCP1
                                                                                                         ◄—— Opossum TCP1
               301        311        321        331        341        351        361        371        381        391
CONSENSUS ==> CGTCCCCTCA TTGAGTCTTA AAAAAGAAAC TAAGTTAACT AAACGACCTG TGAAAACATG TTTCAGGGGC CTAAACCAAA AGAACAGAGC AGAGATGGAA
Human         ********** ******-*-- -T----T-*- --G--***** ********** --C--**TAT *****-T-** **---T--G- G--C--A--T -C--------
Mouse         ********** ********-C -GTGCA-C-- ----CA**** ********** *----***** *****----** **------C- G----C---G T*********
Opossum       ******---- ---------- -----C---- -T-T------ ---TT----- ---------- --------** ****------ -A****---- ---***----
Wallaby       ********** *****----- ---TG-**-- -*****---- *****----- ---------- --------** ****------ --****---- ----**----
Platypus      ---------C ********** *********- -C-AG----C TTC--GG--T G-TCCCGGCA --C------- --CCC----T -ATT****** ****G--**-

                                                 ◄—— Human TCP1
               401        411        421        431        441        451        461        471        481        491
CONSENSUS ==> AGAAGCCGAT TTCCTACATC ACAAAAACCC AAGTTGAAAA AGTTTGCTGT ACTTTAATTT GATATCATCA CCCTTATAAA AACTCAAGGC ACAAGAAAAG
Human         -****-*-- ---------- ---------- -----T-C-G CT-G-A--T- --------G- -TA--AC--- A-T******* ****--C--* **********
Mouse         ********** ********** ********** *****GTC- GC-------- --------G- --G-****-- ********** ********** **********
Opossum       ---T-***** ********** ********** ********** ********** TT----T--- *C-T-----T ---------- --------T- -A-GC--C--
Wallaby       ---*-***** ********** ********** ********** ********** -T----T--- *C-T-----T -T-------- -------AT- -GC-TG-C--
Platypus      ----C----- --******** ********** *****T--C --C-----*** **----T--- A-CCC-G-A- --GC-CGGTG T-AGACG-AA G-GG******

               501        511        521        531        541        551        561        571                   591
CONSENSUS ==> GCTGTTCTAT TTTTTATGCC TAATAAGTGCA AGACCAACTG GTACAAGGAT AAAACATAAG CAGTAACATT GTTATAAATA AAAGGAAAAT CAGA  T
Human         ********** ********** *********** ********** *------ACA -TTG---T** ***------- ---------- -------C-- ----  -
Mouse         ********** ********** *********** ********** *CC-----C- -CGG---TC- ACC-G----- ---G------ -G---G---- -CA-  -
Opossum       -T-------- ------**** **********- ---------- ----*--A-- T----*---* ------T--- ---------- -******--- -T--  -
Wallaby       AT----GC-- --------T- ----------- ----A--A-- ----*----- G----*---T --T-----C- ---------- -*--A------ ----  -
Platypus      ********** ********** *********** ********** *G-G*---GC -CT--*-G-- -C-----C-- ---------- ---A-G-*-- -GC-  -

Opossum ACAT2 ——►
Human ACAT2 ——►
               601        611        621        631        641        651        661        671        681        691
CONSENSUS ==> CAGTCTTCAA TGGCTCCAGA GTGAACAGCA TCTTCATAAC TTCCTCCTCC TCCTCTATGT TTATCATCTT TGCTTTCTGG ATGTAATTTA ATGAGGTCAT
Human         --A--A-T-- G--------- ---------- ---------- ----****** ******---- ---ATCCGAA G-A------- ---------- --A--A----
Mouse         -----A---- G--------- ---------- -T-------- ----****** *****G--- -----G---- ---------- G--------T --C--A--G-
Opossum       -------- -A-------- ---T--T--- ---------G ----****** ------***- --------C- -A-------- ------C--- ----------
Wallaby       -------- -A--C------ ---------- ---------- ---------- ------***- ---------- -A-------- ------C--- ----------
Platypus      -------G- ------C-- ---T--G--G -----G---- A---****** -----G***C --C--G---- ----C--C-- G--C------ --------G-

               701        711        721        731        741        751        761        771        781        791
CONSENSUS ==> CAATTCGAAG AATGGTAATT GCAGCTTCTG TTGCAAACTT CAAGCTCTTA ACTTTAACCA TGGTAGGTTC AAACACACCA GCCTGCCTGT TGTCTCGAGG
Human         ---------- ------G--- ---------- -------T-- ---A------ --------T- ----T----- ------C--T --T--TT--- ----------
Mouse         ----C----- G-----G--- -----C---- ----G----- --G------C --------T- ----T----- ------C--T --T---T--- -------T--
Opossum       ---------- ------T--- ---------- -A-----T-- ---A------ ---------- ---------- G--------- -T-------- -A--C-----
Wallaby       ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- CT-------- ----------
Platypus      -G-----C-- G--------G --G-----G- -G--G----- ---------G -----G---- G------... .......... .......... ..........

               801        811        821
CONSENSUS ==> TTTACCATTG ACCAAATCAA GACCAATCCT GTAGTGAAAG TCGGATTCAG TGAGCCTCTT CCAGGCTGAG GTGGGGTAGG TATTCCAGCT GTTCTAGTCA
Human         ---------- CT-------- -------... .......... .......... .......... .......... .......... .......... ..........
Mouse         ---C----G- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Opossum       ---------- --T--G---- ---------- .......... .......... .......... .......... .......... .......... ..........
Wallaby       ---T...... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Platypus      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

               901        911        921        931        941        951        961        991        981        991
CONSENSUS ==> GAAGACTGGC AGTGCCAACA AGTGTGCCTC TGGATTGTTG AAAGGTGGGC AGCAAGCAGC ATCCACGTAC TTATTTCTCT GCTTGACGGG ATGGGATGTT
HUMAN         .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Mouse         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
Opossum       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Wallaby       .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........
Platypus      .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........

                                                                         Mouse ACAT2 (TCP-1x) ——►
              1001       1011       1021       1031       1041       1051       1061       1071       1081
CONSENSUS ==> ACCTGCTTCA GATATCTACC TTGACTTCCC CAAAATGATG ACTGGAACTG GAACTGACAG TCAAATAAAC TCTTTCCTCC CTT
HUMAN         .......... .......... .......... .......... .......... .......... .......... .......... ...
Mouse         ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---
Opossum       .......... .......... .......... .......... .......... .......... .......... .......... ...
Wallaby       .......... .......... .......... .......... .......... .......... .......... .......... ...
Platypus      .......... .......... .......... .......... .......... .......... .......... .......... ...
```

The *ACAT2* polyadenylation signal was located 254 bp downstream from the gene's stop codon and 6 bp from the *TCP1* stop codon on the complementary strand, so the transcript cleavage site overlapped with the *TCP1* translated sequence. The *TCP1* polyadenylation signal was located 113 bp downstream from the stop codon and at a distance of 147 bp from the *ACAT2* stop codon on the complementary strand (Figures 1 and 5). The transcript cleavage site was located 16 bp downstream of the polyadenylation signal and, hence, did not overlap with the *ACAT2* translated signal. In the wallaby, the arrangement of the polyadenylation signals of the *ACAT2* and *TCP1* genes was similar to that in the opossum, except that the distance between the two signal sites was somewhat longer (116 bp in the wallaby compared to 101 bp in the opossum).

DISCUSSION

The results of the present study suggest that for most of vertebrate evolutionary history, the *ACAT2* and *TCP1* genes have been independent entities, as indeed they still are in modern bony fish and presumably all tetrapods except mammals. In the extant nonmammalian gnathostomes, the two genes are either not linked or, if they are on the same chromosome, they are at a distance from each other that precludes the formation of an overlap between them. The overlap observed in all mammals tested, including the monotremes and the marsupials, may have therefore arisen during the transition from therapsid reptiles to mammals (Carroll 1998), presumably by a chromosomal rearrangement that brought the two genes together. The arrangement may have been a translocation, a deletion if the *ACAT2* and *TCP1* genes had already been located on the same chromosome, or a more complex event that may have involved several successive steps. The overlap may have arisen in one of two ways. First, the rearrangement may have been accompanied by the loss of a part of the 3′-UTR, including the polyadenylation signal from, say, the *TCP1* gene. By chance, however, the 3′-UTR of the new neighbor, the *ACAT2* gene, contained on the noncoding strand all the signals necessary for the termination of transcription and processing of the transcript so that the *TCP1* gene could continue to function normally. Second and perhaps more likely, the two genes became neighbors through the rearrangement but at first did not overlap. Only later, when one of the genes lost its original polyadenylation signal and began to use a signal that happened to be present on the noncoding strand of the other gene, did the pair become locked in. The noncoding strands of the *ACAT2* genes in the zebrafish, toad, and caiman do indeed contain one or more potential, correctly oriented and spaced polyadenylation signals in their 3′-UTRs that could be used by the *TCP1* gene if the two genes were to come together now (Figure 4). The abundance of potential polyadenylation signals in the noncoding strand is undoubtedly caused by the relatively high AT content (60–70%) that characterizes the 3′-UTRs of the *ACAT2* and *TCP1* genes.

The rearrangement thus generated a genetic odd couple that has henceforth been inherited as a unit. The restriction of the overlap to a single phylogenetic lineage, the mammals, suggests that the link-up of the two genes occurred only once and then persisted for more than 200 million years. This conclusion is further supported by the observation of sequence conservation in the overlapping 3′-UTRs of the *ACAT2* and *TCP1* genes (Figure 5). The sequence similarity among the mammalian sequences is poor in the part of the overlap flanking the *ACAT2* translated region, but is rather striking in the part flanking the *TCP1* translated region, in which a whole sequence block has been conserved during evolution of monotremes, marsupials, and eutherians from their common ancestors (Figure 5). For 53 bp of the alignment up to the *ACAT2* stop codon, the comparisons of human with mouse, opossum, wallaby, and platypus sequences, respectively, give 71, 78, 76, and 71% identity, with only seven indel events postulated for alignment. The presence of the postulated *ACAT2* polyadenylation signal in this segment may account for some of the conservation, although the extent of conservation appears to exceed simple polyadenylation signal requirements. To examine whether such conservation is common in eutherian to noneutherian comparisons, we have collated the 3′-UTRs of 10 sequences available from eutherian and metatherian sources and selected at random from DNA databases (sperm protein Sp17, α-tumor necrosis factor, preprolactin, occludin, transthyretin, β-casein, preprouroguanylin, pyruvate dehydrogenase E1-α, β-actin, and protamine P1). We examined the sequences by dot plot, using sensitivities of 70% matches in windows of 10-bp size, and we assessed the degree of homology. Although crude, this method avoids the problems of changes in 3′-UTR length, gap penalties, and base composition corrections in the alignment of such poorly conserved sequences. Among the 10 sequences selected, only 1 (β-actin) exceeds the homology found in *TCP1-ACAT2* 3′-UTR comparisons, and an additional 3 (α-tumor necrosis factor, transthyretin, and protamine P1) show comparable degrees of conservation. The other 6 sequences show little or no sequence similarity between metatherian and eutherian 3′-UTRs.

Human-mouse comparisons show that the *ACAT2* and *TCP1* genes evolve at moderately high synonymous substitution rates that are comparable to the average rates observed in large-scale surveys of mammalian genes (O'hUigin and Li 1992). Hence, relative to the synonymous rate of the rest of the *ACAT2* and *TCP1* genes, the rate has slowed in parts of the overlap region. We assume, therefore, that the sequence similarity observed in the overlap region is retained by selection for reasons

related to the overlap. Miyata and Yasunaga (1978) have argued that the rate of evolution can be expected to slow down in the overlapping stretches, but their argument applies to the overlap of translated regions only: they reason that the proportion of nondegenerate sites is higher in overlapping genes than in nonoverlapping ones, thus reducing the proportion of synonymous substitutions relative to the total number of substitutions. In the 3′-UTR, the slowdown in overlapping stretches might be related to the retention of signals or tertiary structures necessary for processing of the transcript. The overlap might restrict the number of permissible substitutions in certain parts of the 3′-UTR.

Finally, the conclusion that the observed overlaps derive from the same event is also supported by the fact that the relative arrangement of the polyadenylation signals is very similar in the platypus, wallaby, opossum, and human (Figure 1). Only in the mouse does the *ACAT2* 3′-UTR appear to extend all the way into the last intron of the *TCP1* gene (Ashworth 1993). This use of an alternative *ACAT2* polyadenylation signal appears to predate the duplication of the gene because both *ACAT2* copies have a 3′-UTR extending into the last intron of *TCP1.* In all other known mammalian sequences, the distances of the polyadenylation signals from the stop codons are comparable.

The reasons the *ACAT2* and *TCP1* genes have remained coupled together for some 200 million years are unclear. In fact, we cannot exclude the possibility that in some of the 4629 living species of mammals (Wilson and Reeder 1993) they have not. But even if a secondary separation has occurred in some species, it is probably safe to assume that in the great majority of mammals, the two genes have stayed together. There is no evidence that *ACAT2* and *TCP1* are in any way related to each other evolutionarily, structurally, or functionally. The same is true for the other documented cases of gene overlap in vertebrates (Williams and Fried 1986; Adelman *et al.* 1987; Emi *et al.* 1988; Morel *et al.* 1989; Cawthon *et al.* 1991; Shayiq and Avadhani 1992 and other references cited earlier). It is also difficult to imagine any potential advantages that the gene overlap might have for the participating loci. In fact, it may have certain disadvantages manifested when the couple duplicates and then begins to undergo cycles of expansions and contractions, which has been well documented especially for the mammalian CYP21-C4 pair (Morel *et al.* 1989; Kawaguchi *et al.* 1991). In such cases, deletions and other more complex changes may lead to deficiencies with clinical consequences, which in this case would affect two genes instead of one, as they might in the case of uncoupled genes. In the absence of any known advantage, the persistence of gene overlap may be a consequence either of the conservative nature of the evolutionary process or of the difficulties associated with the separation. A divorce of the two genes would require that one or both of them would find, after their separation, all the signals necessary for the termination of transcription and the processing of the transcript in their immediate vicinity. The signals would have to be on the right strand, in the right order, and at the appropriate distances from one another. The probability of this happening might be quite low.

We thank Ms. Jane Kraushaar for editorial assistance, as well as Prof. W. H. Stone and Dr. R. W. Slade for tissue samples.

LITERATURE CITED

Aaronson, J. S., B. Eckman, R. A. Blevins, J. A. Borkowski, J. Myerson *et al.*, 1996 Toward the development of a gene index to the human genome: an assessment of the nature of high-throughput EST sequence data. Genome Res. **6:** 829–845.

Abe, H., A. Ohtake, S. Yamamoto, Y. Satoh, M. Takayanagi *et al.*, 1993 Cloning and sequence analysis of a full length cDNA encoding human mitochondrial 3-oxoacyl-CoA thiolase. Biochim. Biophys. Acta **1216:** 304–306.

Adelman, J. P., C. T. Bond, J. Douglass and E. Herbert, 1987 Two mammalian genes transcribed from opposite strands of the same DNA locus. Science **235:** 1514–1517.

Ashworth, A., 1993 Two acetyl-CoA acetyltransferase genes located in the *t*-complex region of mouse chromosome 17 partially overlap the *Tcp-1* and *Tcp-1x* genes. Genomics **18:** 195–198.

Barrell, B. G., G. M. Air and C. A. Hutchinson III, 1976 Overlapping genes in bacteriophage φX174. Nature **264:** 34–41.

Carroll, R. L., 1988 *Vertebrate Paleontology and Evolution.* W. H. Freeman, New York.

Cawthon, R. M., L. B. Andersed, A. M. Buchberg, G. Xu, P. O'Connell *et al.*, 1991 cDNA sequence and genomic structure of EV12B, a gene lying within an intron of the neurofibromatosis type 1 gene. Genomics **9:** 446–460.

Clinkenbeard, K. D., T. Sugiyama, J. Moss, W. G. Reed and M. D. Lane, 1973 Molecular and catalytic properties of cytosolic acetoacetyl coenzyme A thiolase from avian liver. J. Biol. Chem. **248:** 2275–2284.

Cooper, P. R., N. J. Smilinich, C. D. Day, N. J. Nowak, L. H. Reid *et al.*, 1998 Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. Genomics **49:** 38–51.

Dudley, K., F. Shanahan, M. Burtenshaw, E. P. Evans, S. Ruddy *et al.*, 1991 Isolation and characterization of a cDNA clone corresponding to the mouse *t*-complex gene *Tcp-1x.* Genet. Res. **57:** 147–152.

Ellis, R. J., and S. M. van der Vies, 1991 Molecular chaperones. Annu. Rev. Biochem. **60:** 321–347.

Emi, M., A. Horii, N. Tomita, T. Nishide, M. Ogawa *et al.*, 1988 Overlapping two genes in human DNA: a salivary amylase gene overlaps with a gamma-actin pseudogene that carries an integrated human endogenous retroviral DNA. Gene **62:** 229–235.

Fairbairn, L. J., and M. J. A. Tanner, 1989 Complete cDNA sequence of human foetal liver peroxisomal 3-oxoacyl-CoA thiolase. Nucleic Acids Res. **17:** 3588.

Fukao, T., S. Yamaguchi, M. Kano, T. Orii, Y. Fujiki *et al.*, 1990 Molecular cloning and sequence of the complementary DNA encoding human mitochondrial acetoacetyl-coenzyme A thiolase and study of the variant enzymes in cultured fibroblasts from patients with 3-ketothiolase deficiency. J. Clin. Invest. **86:** 2086–2092.

Gilbert, D. G., 1989 DottyPlotter Version 1.0c, http://iubio.bio.indiana.edu/IUBio-Software+Data/molbio/mac/.

Gilbert, D. G., 1995 SeqPup Version 0.4j: a biosequence editor and analysis application, http://iubio.bio.indiana.edu/soft/molbio/.

Grima, B., D. Zelenika and B. Pessac, 1992 A novel transcript overlapping the myelin basic protein gene. J. Neurochem. **59:** 2318–2323.

Hadano, S., Y. Ishida, H. Tomiyasu, K. Yamamoto, G. P. Bates *et al.*, 1996 Transcript map of the human chromosome 4p16.3 consisting of 627 cDNA clones derived from 1 Mb of the Huntington's disease locus. DNA Res. **3:** 239–255.

Horwich, A. L., and K. R. Willison, 1993 Protein folding in the cell: function of two families of molecular chaperone, hsp60 and TF55-TCP1. Phil. Trans. R. Soc. Lond. **339:** 313–326.

Igual, J. C., C. Gonzalez-Bosch, J. Dopazo and J. E. Perez-Ortin, 1992 Phylogenetic analysis of the thiolase family. Implications for the evolutionary origin of peroxisomes. J. Mol. Evol. **35:** 147–155.

Joseph, D. R., 1998 The rat androgen-binding protein (ABP/SHBG) gene contains triplet repeats similar to unstable triplets: evidence that the ABP/SHBG and the fragile X-related 2 genes overlap. Steroids **63:** 2–4.

Kamijo, T., T. Aoyama, A. Komiyama and T. Hashimoto, 1994 Structural analysis of cDNAs for subunits of human mitochondrial fatty acid beta-oxidation trifunctional protein. Biochem. Biophys. Res. Commun. **199:** 818–825.

Kawaguchi, H., C. O'hUigin and J. Klein, 1991 Evolution of primate C4 and CYP21 genes, pp. 357–381 in *Molecular Evolution of the Major Histocompatibility Complex*, edited by J. Klein and D. Klein. Springer-Verlag, Heidelberg.

Kobel, H., and L. Du Pasquier, 1986 Genetics of polyploid *Xenopus.* Trends Genet. **2:** 310–315.

Kubota, H., G. Hynes, A. Carne, A. Ashworth and K. Willison, 1994 Identification of six *Tcp-1*-related genes encoding divergent subunits of the TCP-1-containing chaperonin. Curr. Biol. **4:** 89–99.

Laudet, V., A. Begue, C. Henry-Duthoit, A. Joubel, P. Martin *et al.*, 1991 Genomic organization of the human thyroid hormone receptor α (*c-erbA-1*) gene. Nucleic Acids Res. **19:** 1105–1112.

Li, W., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

Masuno, M., T. Fukao, X.-Q. Song, S. Yamaguchi, T. Orii *et al.*, 1996 Assignment of the human cytosolic acetoacetyl-coenzyme A thiolase (ACAT2) gene to chromosome 6q25.3-q26. Genomics **36:** 217–218.

Middleton, B., 1973 The oxoacyl-coenzyme A thiolases of animal tissue. Biochem. J. **132:** 717–730.

Middleton, B., 1974 The kinetic mechanism and properties of the cytoplasmic acetoacetyl-coenzyme A thiolase from rat liver. Biochem. J. **139:** 109–121.

Middleton, B., 1975 3-Ketoacyl-CoA thiolases of mammalian tissues. Methods Enzymol. **35:** 128–136.

Miyata, T., and T. Yasunaga, 1978 Evolution of overlapping genes. Nature **272:** 532–535.

Morel, Y., J. Bristow, S. E. Gitelman and W. L. Miller, 1989 Transcript encoded on the opposite strand of the human steroid 21-hydroxylase/complement component C4 gene locus. Proc. Natl. Acad. Sci. USA **86:** 6582–6586.

Nicoloso, M., M. Caizergues-Ferrer, B. Michot, M. C. Azum and J. P. Bachellerie, 1994 U20, a novel small nucleolar RNA, is encoded in an intron of the nucleolin gene in mammals. Mol. Cell. Biol. **14:** 5766–5776.

Normark, S., S. Bergström, T. Edlund, T. Grundström, B. Jaurin *et al.*, 1983 Overlapping genes. Annu. Rev. Genet. **17:** 499–525.

Novacek, M. J., 1992 Mammalian phylogeny: shaking the tree. Nature **356:** 121–125.

O'hUigin, C., and W.-H. Li, 1992 The molecular clock ticks regularly in muroid rodents and hamsters. J. Mol. Evol. **35:** 377–384.

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson *et al.*, 1977 Nucleotide sequence of bacteriophage φX174 DNA. Nature **265:** 687–695.

Shayiq, R. M., and N. G. Avadhani, 1992 Sequence complementarity between the 5′ terminal regions of mRNAs for rat mitochondrial cytochrome P-450c27/25 and a growth hormone-inducible serine protease inhibitor. J. Biol. Chem. **267:** 2421–2428.

Song, X.-Q., T. Fukao, S. Yamaguchi, S. Miyazawa, T. Hashimoto *et al.*, 1994 Molecular cloning and nucleotide sequence of complementary DNA for human hepatic cytosolic acetoacetyl-coenzyme A thiolase. Biochem. Biophys. Res. Commun. **201:** 478–485.

Swalla, B. J., and W. R. Jeffery, 1996 PCNA mRNA has a 3′UTR antisense to yellow crescent RNA and is localized in ascidian eggs and embryos. Dev. Biol. **178:** 23–34.

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

Toyosawa, S., C. O'hUigin, F. Figueroa, H. Tichy and J. Klein, 1998 Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. Proc. Natl. Acad. Sci. USA **95:** 13056–13061.

Toyosawa, S., C. O'hUigin, H. Tichy and J. Klein, 1999 Characterization of dentin matrix protein 1 gene in crocodilia: implications for the evolution of mineralized tissues. Gene (in press).

Van Bokhoven, H., R. B. Rawson, G. F. Merkz, F. P. Cremers and M. C. Seabra, 1996 cDNA cloning and chromosomal localization of the genes encoding the alpha- and beta-subunits of human Rab geranylgeranyl transferase: the 3′ end of the alpha-subunit gene overlaps with the transglutaminase 1 gene promoter. Genomics **38:** 133–140.

Williams, T., and M. Fried, 1986 A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3′ ends. Nature **322:** 275–279.

Willison, K. R., K. Dudley and J. Potter, 1986 Molecular cloning and sequence analysis of a haploid expressed gene encoding t complex polypeptide 1. Cell **44:** 727–738.

Willison, K., A. Kelly, K. Dudley, P. Goodfellow, N. Spurr *et al.*, 1987 The human homologue of the mouse t-complex gene, TCP1, is located on chromosome 6 but is not near the HLA region. EMBO J. **6:** 1967–1974.

Wilson, D. E., and D.-A. M. Reeder, 1993 *Mammal Species of the World. A Taxonomic and Geographic Reference.* Smithsonian Institute Press, Washington, D.C.

Yaffe, M. B., G. W. Farr, D. Miklos, A. L. Horwich, M. L. Sternlicht *et al.*, 1992 TCP1 complex is a molecular chaperone in tubulin biogenesis. Nature **358:** 245–248.

Communicating editor: N. Takahata