Chorionic Gonadotropin Has a Recent Origin Within Primates and an Evolutionary History of Selection

Glenn A. Maston and Maryellen Ruvolo

Department of Anthropology, Harvard University, Cambridge, Massachusetts

Chorionic gonadotropin (CG) is a critical signal in establishing pregnancy in humans and some other primates, but this placentally expressed hormone has not been found in other mammalian orders. The gene for one of its two subunits (CG β subunit [CGβ]) arose by duplication from the luteinizing hormone β subunit gene (LHβ), present in all mammals tested. In this study, 14 primate and related mammalian species were examined by Southern blotting and DNA sequencing to determine where in mammalian phylogeny the CGβ gene originated. Bats (order Chiroptera), flying lemur (order Dermoptera), strepsirrhine primates, and tarsiers do not have a CGβ gene, although they possess one copy of the LHβ gene. The CGβ gene first arose in the common ancestor of the anthropoid primates (New World monkeys, Old World monkeys, apes, and humans), after the anthropoids diverged from tarsiers. At least two subsequent duplication events occurred in the catarrhine primates, all of which possess multiple CGβ copies. The LHβ-CGβ family of genes has undergone frequent gene conversion among the catarrhines, as well as periods of strong positive selection in the New World monkeys (platyrrhines). In addition, newly generated DNA sequences from the promoter of the CG alpha subunit gene indicate that platyrrhine monkeys use a different mechanism of alpha gene expression control than that found in catarrhines.

Introduction

One of the more intriguing molecular evolutionary questions under investigation concerns the origin of new genes and their acquisition of new functional activities. Recent studies have demonstrated that new genes can evolve by complex combinations of gene duplication, exon shuffling, and transposition (Nurminsky et al. 1998; Wang et al. 2000; Courseaux and Nahon 2001; Malik and Henikoff 2001). The evolution of chorionic gonadotropin (CG) in primates presents an opportunity to study the molecular evolution of a gene family from its origin throughout its evolutionary history, including changes in gene expression and functional properties. The well-characterized function of CG allows us to understand the connections between the molecular evolution of the new gene and the morphological evolution of the tissue in which it is expressed.

CG is a glycoprotein hormone expressed in the human placenta which acts as a signal to the maternal physiology to establish pregnancy. Specifically, the binding of CG molecules to LH/CG receptors on the corpus luteum prevents regression of the corpus luteum at menstruation and stimulates continued progesterone production which maintains the uterine lining in a specialized state receptive to implantation and placental development. CG is a member of a larger family of glycoprotein hormones which includes luteinizing hormone (LH), follicle stimulating hormone (FSH), and thyroid stimulating hormone (TSH). Each of these hormones is composed of two protein subunits. The alpha subunit (here labeled GPHα) is shared by all four glycoprotein hormones, whereas each of the four hormones has a unique beta subunit which confers biological specificity. Of the four glycoprotein hormones, only CG is expressed in the placenta; the other three are expressed in the anterior pituitary gland.

The first nucleotide sequence of a human gene encoding the beta subunit of CG (CGβ) suggested that CGβ evolved from a duplicate copy of the beta subunit of the related glycoprotein hormone LH (Fiddes and Goodman 1980). Subsequent nuclear mapping has shown that humans possess six copies of the CGβ gene, found together with the single copy of the human LH beta subunit (LHβ) gene on chromosome 19p13.33 (Policastro et al. 1986; Graham et al. 1987). Human CGβ and LHβ genes share a high degree of sequence similarity (94%), and previous analyses suggest that CGβ genes may be evolving under a regime of positive selection (Talmadge, Vamvakopoulos, and Fiddes 1984).

From a recent phylogenetic analysis of genes from the entire glycoprotein hormone family, Li and Ford (1998) have proposed that the CGβ gene first arose around 94 MYA. If CGβ were indeed that old, it would predate the origin of eutherian mammals and should therefore be widespread in living mammalian taxa. Yet, genomic analyses have clearly shown that CGβ genes do not exist in rats (Jameson et al. 1984; Tepper and Roberts 1984; Carr and Chin 1985), mice (Kumar and Matzuk 1995), cows (Virgin et al. 1985), pigs (Ezashi et al. 1990), sheep (Brown et al. 1993), and rhinoceri (Lund and Sherman 1998). Within Primates, biological and immunological assays have found CG in every species tested, although tests have been almost exclusively limited to anthropoid primates (Tullner 1974; Hobson and Wide 1981). The only exception is a report of CG from the term placenta of a lemur (Hobson and Wide 1981), but the amount of CG reported is very low, and the study reported no negative controls, so this result may represent a spurious nonspecific immunological cross-reaction. Thus, the origin of the gene encoding the

CGβ subunit in primates is unknown, but it would appear to fall between the common ancestor of eutherian mammals and the common ancestor of anthropoid primates.

There are a minimum of four molecular evolutionary events that must have occurred in order to evolve a functional CGβ gene (as found in humans) from an LHβ gene ancestor. These include (in no specific order) (1) the original duplication of the ancestral LHβ gene, (2) a frameshift mutation in the third exon of the duplicated gene, (3) expression gain of the duplicated gene (or its prototype) in the placenta and expression reduction in the pituitary, and (4) expression gain of the GPHα gene in the placenta with expression retention in the pituitary (for use in LH, FSH, and TSH). Here we report on experiments designed to determine when these molecular changes occurred.

Materials and Methods
Tissues and Genomic DNA Preparation

Frozen placenta, kidney, or liver tissue samples were collected from the Duke University Primate Center (Philippine tarsier, *Tarsius syrichta*; Horsfield's tarsier, *Tarsius bancanus*; bushbaby, *Galago senegalensis*; ring-tailed lemur, *Lemur catta*; black and white lemur, *Varecia variegata*; slender loris, *Loris tardigradus*; aye-aye, *Daubentonia madagascariensis*), Yerkes Regional Primate Center (orangutan, *Pongo pygmaeus*), Dr. C.-B. Stewart, University at Albany-SUNY (dusky leaf monkey, *Presbytis obscura*; guereza monkey, *Colobus guereza*), Dr. George Amato, Bronx Zoo (owl monkey, *Aotus trivirgatus*), the Museum at Texas Tech University (colugo or so-called flying lemur, *Cynocephalus variegatus*; vampire bat, *Desmodus rotundus*; flying fox, *Pteropus lylei*), and UC Davis (dusky titi monkey, *Callicebus moloch*). Purified DNA from rhesus macaque (*Macaca mulatta*) and rat (*Rattus norvegicus*, Sprague-Dawley strain) were purchased from Clonetech Inc. Purified human DNA was obtained from cell line GM 3043 (!Kung), Human Genetic Mutant Cell Repository, Camden, NJ. Total genomic DNA was extracted from homogenized frozen tissue by proteinase K digestion, phenol-chloroform extraction, and dialysis.

Southern Blotting

The genomes of 11 extant primates and 3 related mammalian species were surveyed by Southern blotting to determine the number of CGβ plus LHβ genes present in each species. All restriction enzymes and reaction buffers were from Life Technologies (Gaithersburg, Md). Eleven micrograms of digested DNA was electrophoresed in a 0.6% agarose gel, transferred onto nylon membranes using the Turboblotter kit (Schleicher & Schuell, Keene, NH), and immobilized by UV cross-linking according to the manufacturer's protocol.

Hybridization probes radiolabeled with ^{32}P-α-dCTP (NEN Life Sciences) were synthesized using the NEBlot random-primed labeling kit (New England Biolabs, Beverly, Mass.). Probes were synthesized from five different templates: *Homo* CGβ5, *Aotus* CGβ, *Tarsius*

Table 1
Primers Used in this Study

Name	Sequence (5′ → 3′)
M69F	GGTGGCCTTGCCGCCCCCACN
P06F	ATGCTCCAGGTAAGWCTGTAG
1096R	GATTGAGAAGCCTTTATTGTGGGA
1105R	GCGGATTGAGAAGCCTTTATTG
CGA-M205F . . .	TTTCATAGGCTGACCTTGTGG
CGA-P63R	TCAGATCCACAGTCAACTGCC
1080F	TCCCACAATAAAGGCTTCTCAA
P25R	GCAGTCTTACCTGGARCATCTC
M13R	CAGGAAACAGCTATGAC
T7	TAATACGACTCACTATAGGG
T3	ATTAACCCTCACTAAAGGGA
179F	TGGGGTATCTCAGGTCCTCTGG
353F	AGTGTCCTTATCCCAGGGGCT
NWM-330F	AGTCTCAGGCTCGGCTGAAGC
Callic-330F.	CAGGTCCTAAAGAAGACATGC
495F	CCGTCAACACCACCATCTGTGC
501F	CACCACCATCTGTGCCGGCTACTG
835F	GTGTGCAACTACCGCGAGCTGCG
390R	AGCAGCAGCCCCTGGGACAAGGAC
NWM-390R	AGCAGCAGCCCCTGGGACAGGAC
500R	CAGATGGTGGTGTTGACGGTG
NWM-500R	CAGATGGTGGTGTTGAAGGCG
757R	CGTGTGGGTAGCGGAGTGAGCAT
855R	CGCACATCGCGGTAGTTGCACAC
OWM-855R	CGCACCTCGCGGTAGGTGCATAC
NWM-828R	CGTGGTAGTTGCACACCAACTG
Callic-828R.	CGTGGTAGTTGCACACCGGCTG
Tars-483F	CACCTTCACCACCAGTATT
Tars-649R	CCTTCCCAGCCCCTACCCA
Gal-679R	TGCCCCTACAGACCTCAGAC
Daub-613R	ACCAGACCACCCTTCCAACC
Daub-459F	TCACCTTCACCACCAGCATC
Bat-593R.	CGCTGCCACCTCCTGTCTTCT
Col-599R	TGCTGCCACCTCCTGATTTCT

LHβ, *Galago* LHβ, and *Cynocephalus* LHβ. Template DNA for the labeling reaction was prepared by polymerase chain reaction (PCR) amplification (see later). Hybridization was carried out using PerfectHyb Plus Buffer (Sigma Chemical Co.), supplemented with 100 μg/ml yeast tRNA. Washed membranes were exposed to X-ray film for 2–14 days. Autoradiographs were scanned, and the program NIH Image (Rasband and Bright 1995) was used for densitometric analyses of hybrid-band intensity.

PCR, Cloning, and DNA Sequencing

LHβ and CGβ genes were amplified using either the *Taq* Mastermix kit (Qiagen Inc., Valencia, Calif.) or the Platinum *Pfx* Polymerase kit (Life Technologies, Gaithersburg, Md); the latter is preferable for cloning because of its lower error rate. PCR reactions were amplified in an Eppendorf MasterCycler Gradient thermocycler. Reactions used one of two different upstream primers (M69F or P06F, table 1) and one of two different downstream primers (1096R or 1105R). With primer M69F, the expected product was only weakly amplified, so the product was gel-purified and reamplified using primer P06F in the place of M69F. The proximal promoter of the GPHα gene was amplified using the Qiagen *Taq* Mastermix kit with primers CGA-M205F and

CGA-P63R which were designed to match regions conserved among human, rat, mouse, cow, and horse (Steger et al. 1991). PCR products were sequenced directly using the same primers. PCR was also used in some species to amplify the noncoding space between adjacent CGβ-LHβ gene copies in order to test for the number of gene copies. For these reactions, the Failsafe PCR kit (Epicentre Technologies, Inc.) was used with primers 1080F and P25R; buffer G was found to be the optimal reaction buffer.

LHβ and CGβ PCR products were cloned using the TOPO TA Cloning kit (*Taq*) or the ZERO BLUNT TOPO Cloning kit (*Pfx*) (both kits from Invitrogen Corp, Carlsbad, Calif.). Colonies were tested for the expected insert by *Taq* Mastermix PCR, using a small amount of colony material directly as template. Primers for these colony-testing PCR reactions were the vector-matching primers T7 and M13R (table 1). Purified colony-test PCR products were sequenced directly.

Sequencing reactions used the Prism Dye-Terminator Cycle Sequencing kit (Applied Biosystems, Inc.). All clones were sequenced in both directions, starting with the vector primers and then using four to six additional internal primers (2–3 in each direction) designed to match the sequence differences found in each species' clones. The complete list of universal and species-specific primers used in sequencing is given in table 1. This sequencing strategy guaranteed that every base was covered by at least two sequencing reactions, and most were covered by four.

DNA Sequence Analysis

The full-length nucleotide sequence for each clone was assembled from the individual sequencing reactions using the program AutoAssembler (Applied Biosystems, Inc.). Clones were then aligned to each other using the program Clustal X (Thompson et al. 1997), first for all of the clones from one species only, then later to align sequences from different species. Treating LHβ and CGβ clones separately, all the clones from a single species were first aligned together, and any single nucleotide variants that occurred uniquely in single clones were assumed to be PCR errors and changed to match the consensus nucleotide found in every other clone at that site. The same was done for uniquely occurring clones which appeared to be PCR mosaics—artifactual recombinant clone sequences composed of pieces amplified from two different CGβ gene copies. In cases where PCR recombination was suspected to occur uniquely in a single clone, the mosaic clone was removed from consideration. Duplicate clones were then removed from the data set until a minimum number of clones remained which represented the total amount of nonunique sequence diversity found in the clones of a given species. Although this strategy to remove PCR errors and PCR mosaics may also remove some allelic variation from the data set, we were more interested in retaining sequences from each different locus in a species, rather than all polymorphic variants. Alignments were then inspected by eye to improve the alignment.

Phylogenetic analyses on the sequences were performed using PAUP* (Swofford 1998).

Maximum likelihood analyses of substitution rates were performed using the program codeml in the PAML software package (Yang 2000), which uses a codon-based substitution model (Goldman and Yang 1994). The following parameter settings were used: codon frequencies were estimated from the average nucleotide frequency at each codon position; the transition-transversion ratio was estimated from the data; rates were assumed to be equal for all sites (no gamma correction); the correlation coefficient was assumed to be zero; and a molecular clock was not assumed. The user-input tree for each analysis was previously determined by maximum likelihood search using PAUP*; these trees always agreed with well-established phylogenetic relationships of primates.

Statistical tests for gene conversion were performed using the program GENECONV 1.81 (Sawyer 2000). Only the sequences from species for which multiple sequences were cloned (human, orangutan, rhesus, leaf monkey, and guereza) were input as data. By default, the program uses all polymorphic sites in an alignment in scoring the likelihood of conversion between two sequences in a given stretch of DNA. As the sequences in this alignment come from different species, some polymorphic sites vary between species but not within species; using the between-species polymorphic sites would artificially increase the likelihood of finding proposed conversion events between sequences within a species. To avoid this problem, all the sequences from each species were defined as a group for the sake of the analysis, which limits the program to using only sites which are polymorphic within a group (species) when testing for conversion events between sequences of the same group. Analyses were repeated where mismatches were either not allowed, or allowed but given a relative penalty of 1, 2, or 5. These four settings tended to produce similar numbers of significant fragments (likely gene conversions), but they differed in the estimated lengths of the converted fragments.

Results and Discussion
Origin of the CGβ Gene

Given the high level of sequence similarity between human LHβ and CGβ genes, any human CGβ or LHβ probe is expected to hybridize to both loci at low stringency, so the observed number of hybridizing bands on Southern blots represents the total number of LHβ plus CGβ genes in a species' genome. Human DNA digested with *Kpn*I and *Bam*HI produces six fragments which hybridize to a human CGβ probe (not shown) and to a *Galago* LHβ probe (fig. 1*A*, lane 1). One of the bands (at 7.1 kb) is twice as dark as the others, giving a total of seven gene copies in the human genome. As this replicates previous results (Policastro et al. 1986), human DNA was used throughout as a positive control. It was assumed initially that the above enzyme combination would produce a pattern of one DNA fragment per gene copy for all of the species
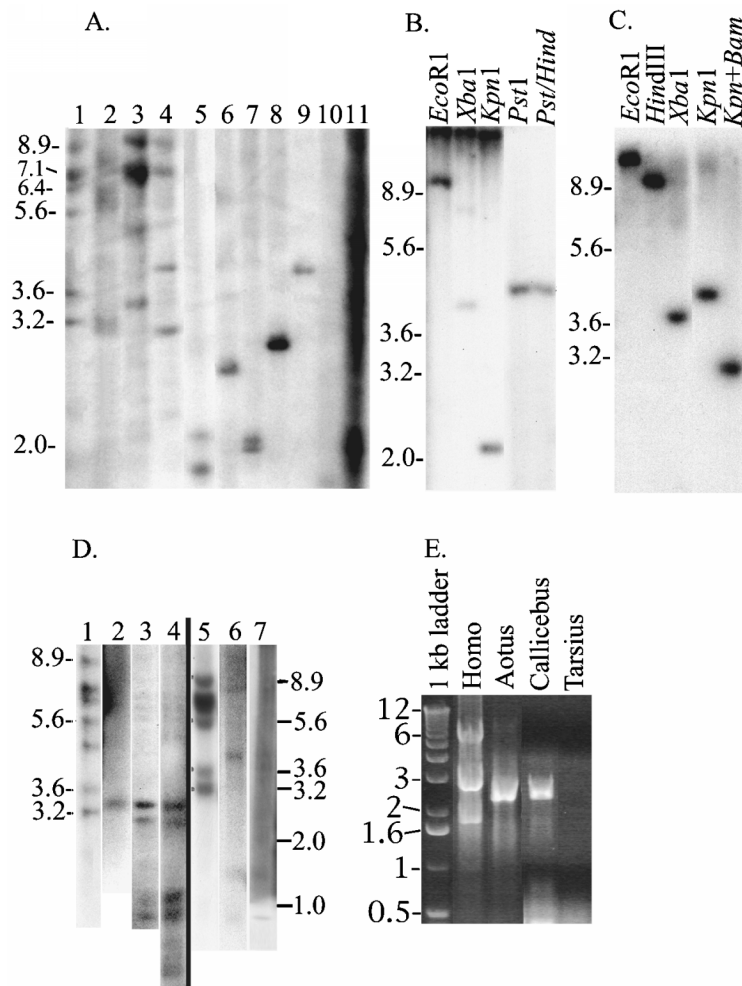
FIG. 1.—Autoradiographs of CGβ-LHβ Southern blots. *A,* Genomic DNAs digested with *Kpn*I and *Bam*HI and hybridized to radiolabeled *Galago* LHβ probe at low stringency (35% sequence mismatch allowed). Lane 1, human; lane 2, orangutan; lane 3, leaf monkey; lane 4, rhesus macaque; lane 5, tarsier; lane 6, lemur; lane 7, aye-aye; lane 8, loris; lane 9, bushbaby; lane 10, flying fox; lane 11, rat. Single band in flying fox (lane 10) is at the very bottom of lane. *B,* Tarsier genomic DNA hybridized with tarsier LHβ probe at high stringency (20% mismatch allowed). The enzymes used in each digestion are indicated above the lane. *C,* Loris genomic DNA hybridized to *Galago* LHβ at moderate stringency (25% mismatch allowed). *D,* Genomic *Aotus* (lanes 2–4) and *Callicebus* (lanes 6–7) DNAs digested with *Kpn*I and *Bam*HI and hybridized to radiolabeled *Homo* CGβ probe at low stringency (35% mismatch, lanes 2 and 6), *Galago* LHβ probe at low stringency (lane 3), or *Aotus* CGβ probe at high stringency (15% mismatch allowed, lanes 4 and 7). Lanes 1 and 5 are digested human DNA used as control and size marker. *E,* Ethidium bromide–stained agarose gel showing inter-gene PCR used to test New World monkeys for the presence of multiple CGβ-LHβ genes. DNA fragment sizes are indicated in kilobases at left in each panel.

analyzed; this assumption was tested by checking the DNA sequences cloned from each species (below) for the presence of any *Kpn*I and *Bam*HI restriction sites within the genes.

Hybridization to a *Galago* LHβ probe found five equally dark fragments in orangutan (fig. 1*A,* lane 2), four fragments in dusky leaf monkey (fig. 1*A,* lane 3), two of which are twice as dark as the other two, and four equal fragments in rhesus macaque (fig. 1*A,* lane 4). For each species, the same pattern was found using an *Aotus* CGβ probe (not shown). These results indicate that there are a total of five gene copies in orangutan, six in dusky leaf monkey, and four in rhesus macaque. Guereza monkey was also found to have six gene copies (not shown). Whereas the exact identity of each of the hybridizing genes is not known, we inferred that each

species has one LHβ gene as do humans, given that there was no case of a species having two different DNA sequences among their LHβ gene clones.

Tarsier has two hybridizing fragments (fig. 1*A,* lane 5). The DNA sequence of the tarsier gene has a single *Bam*HI cut site 222 nucleotides (nt) from the 3′ end of the gene, thus it would appear that tarsier has just one hybridizing gene sequence, represented by two fragments in the *Kpn*I-*Bam*HI double digest. To confirm that tarsier does not have any CGβ genes and has only one LHβ gene, a more detailed genomic analysis was conducted using an array of enzyme digests and using tarsier LHβ as the probe (fig. 1*B*). All of the digests produce either one or two hybridizing fragments, consistent with the number of enzyme cut sites predicted from the tarsier DNA sequence. Thus, tarsier has just one cross-

hybridizing sequence in its genome, and DNA sequencing indicates it is an LHβ gene, not a CGβ gene.

For all the strepsirrhine primate species analyzed (lemur, aye-aye, loris, and bushbaby), a single gene copy hybridized to the *Galago* LHβ probe (fig. 1*A,* lanes 6–9) and to a human CGβ probe (not shown). To further confirm these results, a more detailed genomic analysis was conducted for loris. Five different enzyme digests all produce just one hybridizing fragment (fig. 1*C*), indicating that there is just one gene copy present in the loris genome. DNA sequence analysis shows that the single gene in all of these strepsirrhine species is an LHβ gene. Therefore, all of the strepsirrhine primates lack CGβ genes. Both Southern blotting and DNA sequencing also found a single LHβ gene and no CGβ genes for the flying fox (fig. 1*A,* lane 10), the vampire bat (not shown), and the colugo (not shown).

The New World monkey *Aotus* has from one to five hybridizing copies, depending on the hybridization probe used (fig. 1*D*). A human CGβ probe only hybridizes to one fragment (fig. 1*D,* lane 2), whereas a *Galago* LHβ probe hybridizes with 3–4 fragments (fig. 1*D,* lane 3), and with an *Aotus* CGβ probe (self-hybridization), five dark bands hybridize, as well as five or more lightly hybridizing bands (fig. 1*D,* lane 4). A single *Kpn*I cut site is predicted from the *Aotus* CGβ DNA sequence (subsequently) at ~100 bp downstream from the start of the CGβ coding sequence, suggesting a single *Aotus* gene should be represented by two fragments, one about 10 times darker than the other. One way to interpret these results is that *Aotus* has five LHβ-like genes in its genome, some of which are recently evolved pseudogenes of either LHβ or CGβ. This would explain why probes of different phylogenetic distance from *Aotus* gave different hybridization patterns. Presumably, at least one of the five genes in *Aotus* is a functioning LHβ gene, but from the blot results alone, it is not clear if any of the other four genes are CGβ and, if so, whether any are functional.

In *Callicebus*, probes from both human (fig. 1*D,* lane 6) and *Aotus* (fig. 1*D,* lane 7) CGβ hybridize to three fragments. The *Callicebus* CGβ sequence predicts a *Kpn*I site at 82 bp from the 5′ end of the PCR fragment, as well as a *Bam*HI site 910 bp from the 5′ end. This indicates that in *Callicebus* three fragments (one of which would be ~800 bp and the other two of unknown length) represent a single gene. The Southern blot results are thus consistent with the presence of only one gene copy in *Callicebus*, yet PCR amplification and cloning finds only CGβ-like sequences for both *Callicebus* and *Aotus*.

In order to further characterize the genomes of these New World monkey species, a different experimental approach was used to estimate the number of tandemly arrayed gene copies present. PCR amplification was performed with primers designed to amplify between gene copies if and only if more than one gene is present. In humans, seven linked genes are present; so six PCR products are expected. These six fragments are predicted to form three pairs of similarly sized intergenic spaces based on the previously determined map

of the cluster (Policastro et al. 1986). Three amplified products of the expected sizes are clearly obtained using human DNA (fig. 1*E,* lane 1). Tarsier, with only one gene, should not amplify any product, and none is seen (fig. 1*E,* lane 4). Both *Aotus* (lane 2) and *Callicebus* (lane 3) amplify a single band; from this we can infer that these New World monkeys both have at least two linked LHβ-CGβ genes; they may possibly have more if the additional genes are either (1) spaced exactly the same distance apart from the first two, or if (2) the additional copies are too far away to allow PCR amplification between them (the upper limit on amplifiable fragments is around 20 kb; Epicentre Technologies technical staff, personal communication). Both the *Aotus* and *Callicebus* sequences found here from cloned PCR amplifications have a fairly low degree of sequence divergence (7.9% and 9.4%, respectively) from the previously published CGβ cDNA sequence for marmoset (*Callithrix jacchus,* Simula et al. 1995), and there are no nonsense mutations in the *Aotus* and *Callicebus* CGβ sequences to suggest that either is a pseudogene. Comparisons among our *Aotus* CGβ sequences find zero sequence variation in 30 clones from 3 amplifications using different primers, all of which bind in coding regions. As a further measure, all three New World monkey CGβ sequences are roughly equidistant from human CGβ (19.4%–20.4%) and from tarsier LHβ (23.1%–25.0%). These results suggest that both *Aotus* and *Callicebus* CGβ sequences generated here are likely to be functional genes.

To summarize, all of the anthropoid primate species tested here show evidence of one or more CGβ genes in their genome. The tarsier, all of the strepsirrhine primates, and the three nonprimate outgroup species have all been shown to have exactly one gene copy (later shown by sequencing to be LHβ genes). Therefore, the most parsimonious reconstruction places the first gene duplication event in the common ancestor of the anthropoid primates, after the divergence of the tarsier lineage (between nodes D and C, fig. 2). From fossil and molecular phylogenetic studies, this places the origin of the CGβ gene between 50 and 34 MYA (Bailey et al. 1991). This date is considerably more recent than the 94-MYA origin date proposed previously from calibrated molecular clock calculations (Li and Ford 1998). If we take the average DNA sequence divergence between the introns of the tarsier LHβ gene and the anthropoid CGβ genes, corrected by the Hasegawa, Kishino, and Yano (1985) model, and divide by twice the estimated age of the tarsier-anthropoid common ancestor (50 Myr), we get a mutation rate of 3.37×10^{-9} substitutions per site per year, which is in close agreement with previous estimates for noncoding DNA in primates (Bailey et al. 1991). This supports our estimate of the age of the origin of the first CGβ gene.

In order to explain the number of gene copies found in each species, additional duplications are inferred to have occurred as shown in figure 2. The New World monkeys each have evidence for only one functional CGβ gene, so there was most likely only one CGβ gene in the last common ancestor of the anthropoids.
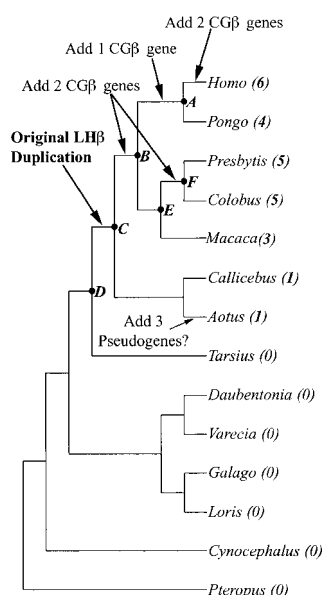
FIG. 2.—Phylogenetic tree of the species used in this study showing the most parsimonious reconstruction of the CGβ gene duplication events. The number of CGβ genes present in each species is indicated in parentheses after the species name. Node *A,* the large-bodied hominoid common ancestor; *B,* the catarrhine common ancestor; *C,* the anthropoid common ancestor; *D,* the haplorhine common ancestor; *E,* the cercopithecoid common ancestor; and *F,* the colobine common ancestor.

Table 2
CGβ/LHβ Sequencing Results

Taxon	Number of Clones Sequenced	Number of Unique Clones
Homo	26 (11 LHβ, 15 CGβ)	7
Pongo	17 (7 LHβ, 10 CGβ)	5
Macaca	18 (4 LHβ, 14 CGβ)	4
Colobus	25 (1 LHβ, 24 CGβ)	5
Presbytis	16 (4 LHβ, 12 CGβ)	6
Aotus	30 (all CGβ)	1
Callicebus	20 (all CGβ)	1
Tarsius	9 (all LHβ)	1
Varecia	6 (all LHβ)	1
Daubentonia	9 (all LHβ)	1
Loris	4 (all LHβ)	1
Galago	4 (all LHβ)	1
Cynocephalus	9 (all LHβ)	1
Pteropus	5 (all LHβ)	1

(The additional hybridizing genes in *Aotus,* proposed to be pseudogenes of either LHβ or CGβ, probably arose in the *Aotus* lineage after its divergence from *Callicebus.*) Assuming gene loss is less likely than gene duplication, it is then most parsimonious to reconstruct the addition of two gene copies in the catarrhine common ancestor (branch C-B), one copy along the branch between the catarrhine ancestor and the common ancestor of human and orangutan (branch B-A), two copies sometime after the orangutan divergence from the lineage leading to human-chimp-gorilla (after node A), and two copies in the colobine monkey common ancestor (node E-F). Crawford, Tregear, and Niall (1986) report that baboons have more than five CGβ gene copies; however, visual inspection of their figure does not suggest more than four copies, and one of these is likely to be baboon LHβ. If there were five (CGβ plus LHβ) genes in baboon, this could change the above reconstructed history of duplications in Old World monkeys, in having either one additional duplication unique to the baboon (*Papio*) lineage (which is sister to *Macaca*) or an additional duplication in the catarrhine common ancestor (branch C-B) with three fewer later gene additions and a loss of a gene copy in rhesus macaque. The relative likelihoods of the different scenarios depend on the relative probabilities of gene copy gains versus losses.

Evolution of CGβ-Specific Sequence Characteristics

The primary differences between LHβ and CGβ are in their gene expression patterns (LHβ in pituitary,

CGβ in placenta), and in the lengths of their coding sequence. Human CGβ genes have a single-base deletion relative to the human LHβ gene at position +988 (counting from the first translated base of exon 1), which is eight codons before the LHβ termination codon. This causes a frameshift which incorporates much of what is the 3′-untranslated region in LHβ into the third exon of CGβ, in turn adding 24 amino acids to the length of the CGβ peptide. A two-base insertion in human CGβ (relative to human LHβ) at position 1060 again adjusts the reading frame in human CGβ, which produces a termination signal eight codons later. Knowing that the CGβ gene family first arose in the anthropoid common ancestor, the next step was to reconstruct when these CGβ-specific characteristics evolved.

CGβ and LHβ genes were cloned and sequenced from 13 previously uncharacterized species, and also from human, in which only two of six CGβ loci had been completely sequenced at the time this study began. The species sequenced and the number of sequences reported for each species is shown in table 2. Genomic DNA sequences produced here start at the 13th base of the first intron and include the rest of intron A (340 nt), exon 2 (168 nt), intron B (234 nt), all of the coding portion of exon 3 for LHβ genes, and all of the coding region, except the last amino acid of exon three for the CGβ genes. The first exon includes only 15 translated nucleotides, and the amino acids encoded in this region are not part of the mature protein, so their absence does not hinder analyses of the evolution of the protein's function. A total of 36 new LHβ and CGβ sequences were produced in this study. The aligned sequences are available in GenBank (accession numbers AF397576–AF397611). Given the experimental approach used here, we cannot determine whether clones from a given species are alleles of the same physical locus or whether they are representatives of different loci. Furthermore, we cannot say definitively whether all (or only some) of the LHβ and CGβ copies present in a given species were amplified and represented among the clones sampled for each species.

In an attempt to assess whether a random sampling of clones from a given species might represent all of the LHβ and CGβ loci present, the human clones sequenced in this study were compared to the CGβ gene sequences produced by the Human Genome Project (http://genome.cse.ucsc.edu/). In four cases, the assignment of a clone to a specific physical locus was clear: clone Hsa11 is a copy of human locus CGβ3, clone Hsa13 is from locus CGβ5, clone Hsa04 is locus CGβ7, and clone Hsa17 is from the LHβ locus (data not shown). Of the remaining three human clones, each shows evidence of gene conversion between it and the human LHβ gene (see subsequently), which makes assignment to a physical locus difficult. No human clones produced in this study show a clear similarity to loci CGβ1, CGβ2, or CGβ8. This brief analysis suggests that the set of species-clones produced in this study might not contain representatives of every different CGβ locus present in each species. Nevertheless, the DNA sequences presented here all show some sequence differences and therefore represent a conservative minimum estimate of the range of sequence diversity that exists among the CGβ and LHβ genes in each species.

Inspection of the aligned sequences reveals that the second frameshift, the two-base deletion at sites 1060–1061 (1211–1212 in the alignment), is found only in the human LHβ gene; all other sequences have the same bases at these sites as in the human CGβ genes. Thus, this event is most parsimoniously reconstructed as a deletion mutation which occurred uniquely in the 3′-untranslated region of the LHβ gene recently, after the orangutan lineage diverged from that leading to humans, chimps, and gorillas. Therefore, this two-base frameshift is not an important event in the deeper phylogenetic evolution of the CGβ gene family.

At least one cloned sequence from each anthropoid species studied was found to contain the single-base deletion at position 988 (site 1137 of the aligned sequences). In the absence of data on the actual expression patterns of each gene sequence produced from each species, the only way to define a sequence as being either an LHβ or a CGβ gene is based on the presence or absence of this single-base deletion. Using this criterion, all of the anthropoid species tested here possess a CGβ gene. All of the species possessing only one gene lack the deletion at this site, identifying these unique genes as LHβ genes. In all the catarrhines studied here, a single gene lacking this deletion (i.e., LHβ) was also found. In both New World monkeys, even though two gene copies were found by intergene PCR, no LHβ sequence was found; all sequences had the deletion and thus were CGβ using the above criteria. LH has been found in every vertebrate species tested to date, and it plays a vital role in the regulation of reproduction, so it is unlikely that New World monkeys lack an LHβ gene. Rather, it is more likely that a sequence mismatch between one of the PCR primers and the New World monkey LHβ sequences caused a PCR amplification failure of the LHβ gene in these species, thus explaining its absence among the sequences generated in this study. Overall, it is most parsimonious to reconstruct the oc-
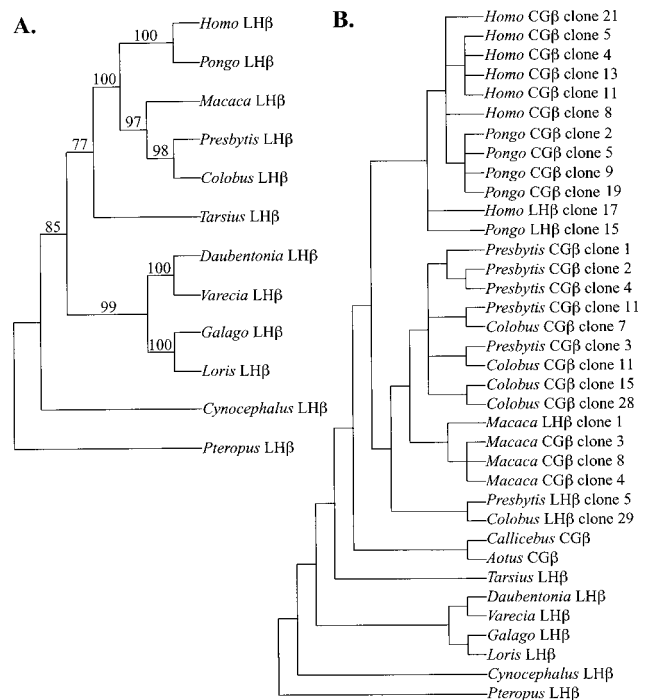

FIG. 3.—Gene trees of the LHβ and CGβ sequences. *A,* maximum parsimony gene tree of only the LHβ clone sequences produced in this study, based on the entire coding and noncoding (intron) sequence produced for each species (transversions weighted twice transitions; tree rooted using the *Pteropus* LHβ sequence as an outgroup). Numbers above branches are estimates of branch support from 500 bootstrapping replicates. *B,* Gene tree reconstructed for the entire set of 36 CGβ and LHβ clones sequenced in this study. Shown is the semistrict consensus of the maximum likelihood, maximum parsimony, and minimum evolution (distance) trees for the data set. The tree was rooted with the *Pteropus* LHβ sequence.

currence of the CGβ-specific deletion event after the initial LHβ gene duplication, but before the divergence of the catarrhines from the platyrrhines, placing it very early in the evolutionary history of the CGβ genes.

Gene Conversion in CGβ Genes

Given that the DNA sequences of the (linked) human LHβ and CGβ genes are more similar to each other than is human CGβ to other species' CGβ genes (Crawford, Tregear, and Niall 1986; Simula et al. 1995, and this study), it is clear that the CGβ and LHβ genes are not evolving independently. One of the prominent mechanisms of concerted evolution that could affect the evolution of these genes is gene conversion, in which a nonreciprocal transfer of DNA sequence information occurs, with a portion of one gene acting as the parent, copying its sequence into the daughter or recombinant gene. There are numerous ways to test for gene conversion. If we first reconstruct the phylogenetic tree for LHβ sequences alone (fig. 3*A*), the tree topology matches the well-established phylogenetic relationships of primates. However, when the CGβ sequences are added to the LHβ set, the phylogenetic results clearly deviate from the pattern expected if the loci were to have evolved independently (fig. 3*B*). Instead of observing a

Table 3
CGβ/LHβ Fragments Identified as Likely Gene Conversions

Sequence Names	Begin	End	Length	# Poly	# Diff	TD	P
Presbytis clone 11CG; *Presbytis* clone 04CG . . .	669	1,210	542	167	0	46	0.0001
Homo clone 21CG; *Homo* clone 17LH	1	722	722	192	2	50	0.0001
Colobus 29LH; *Colobus* clone 07CG	587	851	265	77	1	70	0.022
Presbytis clone 03CG; *Presbytis* clone 04CG . . .	972	1,210	239	67	0	70	0.0003
Colobus clone 29LH; *Colobus* 11CG	1	228	228	57	0	81	0.0003
Presbytis clone 11 CG; *Presbytis* clone 03CG . .	972	1,210	239	67	0	69	0.0003
Macaca clone 03CG; *Macaca* clone 01LH	1	910	910	257	9	34	0.0238
Presbytis clone 01CG; *Presbytis* clone 03CG . . .	1,012	1,210	199	57	0	72	0.0004
Presbytis clone 03CG; *Presbytis* clone 02CG . . .	1,013	1,210	198	56	0	73	0.0004
Colobus clone 07CG; *Colobus* clone 11CG.	702	1,178	477	150	4	52	0.0027
Homo clone 08CG; *Homo* clone 17LH	1	885	885	249	8	43	0.0036
Homo clone 13CG; *Homo* clone 17LH	111	722	612	168	10	65	0.0001
Homo clone 11CG; *Homo* clone 17LH	170	722	553	154	8	64	0.0001
Homo clone 04CG; *Homo* clone 08CG	821	1,210	390	125	1	41	0.0011
Homo clone 21CG; *Homo* clone 04CG	597	1,118	522	160	1	29	0.0045
Colobus clone 29LH; *Colobus* clone 28CG.	654	831	178	50	0	64	0.0316
Pongo clone 19CG; *Pongo* clone 15LH.	108	505	398	119	4	64	0.0003
Presbytis clone 01CG; *Presbytis* clone 11CG . . .	725	1,210	486	154	10	56	0.0339
Colobus clone 29LH; *Colobus* clone 11CG.	702	831	130	38	0	81	0.0472
Presbytis clone 11CG; *Presbytis* clone 02CG . . .	1,013	1,210	198	56	0	55	0.0499
Homo clone 13CG; *Homo* clone 08CG	111	443	333	100	0	30	0.0473
Homo clone 05CG; *Homo* clone 17LH	1	722	722	192	9	58	0.0001
Pongo clone 02CG; *Pongo* clone 15LH.	257	655	399	110	5	66	0.0011
Colobus clone 28CG; *Colobus* clone 11CG.	702	1,004	303	99	3	64	0.0027
Pongo clone 05CG; *Pongo* clone 15LH.	1	668	668	181	11	65	0.0078
Colobus clone 07CG; *Colobus* clone 28CG.	654	931	278	88	1	51	0.0081
Pongo clone 09CG; *Pongo* clone 15LH.	108	655	548	156	9	61	0.0029
Presbytis clone 11CG; *Presbytis* clone 03CG . . .	391	826	436	114	8	69	0.0112
Homo clone 05CG; *Homo* clone 13CG	182	1,149	968	187	2	11	0.0314

NOTE.—''Begin'' and ''End'' are the estimated boundaries of the sequence fragment affected by gene conversion; positions given are relative to the first translated base. "# Poly" is the number of polymorphic sites within a converted fragment shared by the two sequences involved in a conversion event; "# Diff" is the number of differences in the converted region; "TD" is the total number of differences between the two compared sequences; P is the probability of getting the apportionment of shared polymorphisms and differences between the two sequences in the fragment proposed by chance alone. The probabilities are estimated by permutation (Sawyer 2000).

tree with two halves (one half with LHβ sequences, the other with CGβ sequences) in which each half reflects the same topology, we observe a tree of intermixed LHβ and CGβ sequences. The hominoid LHβ and CGβ sequences form a separate clade, as do the Old World monkey LHβ and CGβ sequences. Within each clade, the LHβ sequences generally group together, sister to the CGβ sequences. This pattern suggests significant gene conversion between the LHβ and CGβ genes in both the common ancestors of the large-bodied hominoid and cercopithecoid clades. The one exception, in which rhesus LHβ groups most closely with the rhesus CGβ sequences, represents a more recent gene conversion within the genus *Macaca*.

A recent review (Drouin et al. 1999) found that the statistical method of Sawyer (1989) gives the most consistent results in detecting conversion events. Using the program GENECONV 1.81 (Sawyer 2000) implementing Sawyer's (1989) method, 29 statistically likely cases of gene conversion were identified between the CGβ and LHβ sequences in this study (table 3). The boundaries of the converted regions vary depending on the degree of sequence mismatch allowed in the converted region; here we show only the longest estimates of the converted regions. It should be noted that some of the

regions identified as gene conversions (those hybrid sequences with one internal recombination boundary and not two) could be caused by PCR recombination and not real gene conversion. In an effort to minimize this possibility, a clone from a given species was removed from the data set if it was found to have a recombinant sequence which occurred uniquely in that clone and where there was only one internal recombination boundary. This strategy will not eliminate all PCR recombinants, as some might occur in more than one clone and thus may show up in the gene conversion analysis. However, the inclusion of some PCR-derived recombinants will not produce directional bias (in the sense that some regions might artifactually be observed to be free of gene conversion) because a PCR recombination event can occur anywhere along the amplified fragment.

Among all of these proposed cases of gene conversion (table 3), two general patterns emerge. First, when a gene conversion involves an LHβ gene, in all but one instance the direction of the conversion is from the LHβ gene into the CGβ gene; that is, in all but one case the LHβ gene remains unaffected, whereas the CGβ gene is converted. Second, no case identified here involves conversion between the third exon of a CGβ and an LHβ gene. This result suggests that selection acts

```
                        110       115        120       125      130       135      142
                         .         .          .         .        .         .        .
Homo clone 21 CG      H P L T C D D P R F Q A S S S S K A P P P S L P S P S R L P G P S D T P I
Homo clone 5 CG       . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . . . . . . .
Homo clone 4 CG       . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . . . . . . .
Homo clone 13 CG      . . . . . . . . . . . D . . . . . . . . . . . . C . . . . . . . . . . . .
Homo clone 11 CG      . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . . . . . . .
Homo clone 8 CG       . . . . . . . . . . . D . . . . . . . . . . . . . . . . . . . . . . . . .
Pongo clone 2 CG      . . . . . . . . . . . D P . . . . . . . . . . . . . . . . . E . . G . . F
Pongo clone 5 CG      . . . . . . . . . . . D P . . . . . . . . . . . . . . . . . E . . G . . F
Pongo clone 9 CG      . . . . . . . . . . . D P . . . . . . . . . . . . . . . . . E . . G . . F
Pongo clone 19 CG     . . . . . . . . . . . D P . . . . . . . . . . . . . . . . . E . . N . . F
Presbytis clone 1 CG  . . . . . . . . . H L . . . . . . . D L . . . P . . . . L E . A G . . F
Presbytis clone 11 CG . . . . . . . . . H L . . . . . . . D L . . . P . . . . L E . A G . . F
Presbytis clone 3 CG  . . . . . . . . . H L . . . . . . . D L . . . P . . . . L E . A G . . F
Presbytis clone 2 CG  . . . . . . . . . H L . . . . . . . D L . . . P . . . . L E . A G . . F
Presbytis clone 4 CG  . . . . . . . . . H L . . . . . . . D L . . . P . . . . L E . A G . . F
Colobus clone 7 CG    . . . . . . . . . H L . . . . . . . D L . . . P . . . . L E . A G . . F
Colobus clone 11 CG   . . . . . . . . . H L . . . . . . . D . . . . P . . . . L E . A G . . F
Colobus clone 15 CG   . . . . . . . . . H L . . . . . . . D . . . . . . . . . L E . A G . . F
Colobus clone 28 CG   . . . . . . . . . H L . . . . . . . D . . . . P . . . . L E . A G . . F
Macaca clone 3 CG     . . . . . . . . . H L . . . . . . . D . . . . P . . . G . L E . A . . . F
Macaca clone 8 CG     . . . . . . . . . H L . . . . . . . D . . . . P . . . . . L E . A . . . F
Macaca clone 4 CG     . . . A . . . . . H L . . . . . . . D . . . . P . . . . . L E . A . . . F
Callicebus CG         E . . G . . Y A I P K S . - . . . D . . R N . T . S . Q . . E . A . . A L
Aotus CG              E . . G . . Y H T S . D . - . . . D . . R N . T . . . Q . . E . A . A T L
Homo clone 17 LH      . . . . . . H . Q L S G L L F L *
Pongo clone 15 LH     . . . . . . H . Q L P G L L F L *
Presbytis clone 5 LH  . . . . . . H . Q L P G L L F L *
Colobus clone 29 LH   . . . . . . H . Q L P G L L F L *
Macaca clone 1 LH     . . . . . . H . Q L P G L L F L *
Tarsius LH            Q . . A . E R . H L P G L L F L *
Daubentonia LH        Q . F A . . H . H L P S L F L L *
Varecia LH            Q . F A . . H . H L P S L L L L *
Galago LH             Q . . A . . H . H L P S L F L L *
Loris LH              Q A . A . . H . H L P S L F L L *
Cynocephalus LH       Q . . A . . H . H L P G L L F F *
Pteropus LH           Q S . A . . R . S I P D L L F L *
```

FIG. 4.—Amino acid sequences predicted from the last 111 nt of the CGβ sequences and last 48 nt of the LHβ sequences. The sequences are aligned to the previously published human sequence; position numbers (above sequences) are relative to the first amino acid of the mature peptide. Amino acids matching the human sequence at each position are indicated by a period, gaps are indicated by a dash. The four serine residues which are glycosylated in humans are indicated by arrows at the bottom of the alignment.

against conversion between CGβ and LHβ third exons. The third exon of both LHβ and CGβ encodes over 60% of the mature peptide and includes several amino acids important for dimerization and receptor binding (Lapthorn, Harris, and Littlejohn 1994). It is unclear why gene conversion in the third exon would be selected against, given that the hormones CG and LH have similar structure, bind to the same receptor, and elicit the same signal. Researchers continue to debate how CG and LH differ in their molecular interactions. For example, it has been proposed that CG has an immunosuppressive function during pregnancy (Morse et al. 1976), but it is not known how CG performs this function or if it is specific to CG (Hearn and Gomme 2000). The molecular evolutionary patterns observed here support the idea that there is a difference between LH and CG in their molecular actions, and the difference is encoded in the third exon of these genes.

Functional Amino Acid Changes in the Evolution of CGβ

The only appreciable difference in the structure and molecular properties of these two hormones is in the number of sugar chains attached to each. Human LHβ has two N-linked glycosylations at amino acid sites 13 and 30 of the mature peptide, whereas human CGβ has two N-linked glycosylations at homologous sites and also four O-linked glycosylations at serine residues 121, 127, 132, and 138. The effect of these glycosylations is to slow the clearance of CG molecules from the maternal bloodstream; LH has a circulatory half-life of about 30 min (de Kretser, Atkins, and Paulsen 1973), whereas the half-life for CG is roughly 12 h (Braunstein, Vaitukaitis, and Ross 1972). This property is believed to make CG molecules more effective at establishing pregnancy than the equivalent dose of LH (Matzuk et al. 1990; Hearn and Gomme 2000). Thus, the evolution of these four serine residues in CGβ could be regarded as an adaptation, and it is of interest to reconstruct when and how these sites evolved.

A comparison of the predicted amino acid sequences (fig. 4) reveals that two of the serine residues (serines 121 and 132) which are glycosylated in humans are present in all of the anthropoid CGβ genes. The two New World monkeys have only these two serines and lack the other two (127, 138) which are glycosylated in humans. Therefore, it is likely that the serine residues at 121 and 132 were present in the ancestral CGβ gene around the time when the frameshift mutation was introduced, and this region of sequence was translated for the first time. It can be inferred from the amino acids sequences (fig. 4) that there was an asparagine to serine substitution at site 127 in the catarrhine common ancestor and an alanine to serine substitution at site 138 in the large-bodied hominoid ancestor, based on the shared presence or absence of serines in each species in the study. Assuming that the addition of each serine (and thus a sugar chain on that serine) has an incremental

effect on the metabolic clearance rate of the molecule, these results suggest that there may have been selection acting to increase the effectiveness of CG at resisting metabolic clearance throughout anthropoid evolution, particularly in the catarrhines.

Substitution Rates in the LHβ and CGβ Genes

In addition to selection favoring an increase in the number of serine residues in the evolution of the CGβ genes, there may also have been positive selection acting to remodel the amino acid sequence of CG throughout primate evolution. Positive selection in coding sequences is commonly tested for by comparing the rate of nonsynonymous substitutions per nonsynonymous site (d_N) to the rate of synonymous substitutions per synonymous site (d_S) (Yang 1998). The ratio d_N/d_S ($=\omega$), if significantly greater than 1.0, is taken to indicate positive selection. It has been proposed that positive selection may act in relatively short episodes in the evolution of a protein (Gillespie 1991, pp. 132–133); this would be indicated by finding $\omega > 1$ along a specific branch of a gene tree. An example of this type of selection has been found in primate lysozyme genes (Messier and Stewart 1997; Yang 1998). The analyses conducted on the LHβ and CGβ genes below are modeled after that of Yang (1998).

Looking first at just the LHβ genes, there is no evidence of positive selection along any branches (table 4A). If a null model assuming a uniform ω for the entire tree is assumed, the estimated ω for the data set is 0.22 (table 4A, line 1). Although a free-ratio model assuming an independent ω value for each branch (line 2) is a significantly better fit than the null model ($2\Delta L = 36.68$, $P < 0.025$, df $= 20$), none of the branches have ω values significantly greater than 1.0. Thus the LHβ genes appear to be evolving under a regime of generally negative or purifying selection, as is found in most functioning genes (Li 1997, pp. 179–182).

For the analysis of CGβ sequences, one representative sequence was chosen from each species. Included in this analysis were CGβ sequences from GenBank for baboon (*Papio hamadryas anubis*, accession number M14966) and marmoset (*C. jacchus*, accession number U04447). (Available sequences for CGβ from *Macaca fascicularis* were not included because of their close phylogenetic relationship to *Macaca mulatta*. Another group has also submitted a CGβ sequence for *M. mulatta* [accession number AY011015]; this sequence differs from all of our *M. mulatta* clone sequences, but it was not different enough to include in our analyses.) The following analyses also incorporated a maximum likelihood reconstruction of the nucleotide sequence of the ancestral haplorhine LHβ gene (labeled HapAncLH), based on the LHβ sequences from *Pteropus*, *Cynocephalus*, *Loris*, *Galago*, *Varecia*, *Daubentonia*, *Tarsius*, *Colobus*, and *Homo*. This sequence approximates the state of the proto-CGβ gene at the time when the the 3′-UTR first became coding sequence.

The likelihood analyses on these CGβ sequences (table 4B) show that the free-ratio model fits the CGβ

Table 4
Tests of Substitution Rate Differences Using Maximum Likelihood

Model	N	L	Parameter estimates
A. Twelve LHβ genes, using the coding sequence of the entire mature protein			
1. Uniform ω (estimate ω)........	23	−1767.53	$\omega_0 = 0.22$
2. Free-ratio (one ω per branch)........	43	−1749.19	ω estimated for 21 branches (0–0.763)
B. Ten CGβ genes, using the coding sequence of the entire mature protein (142 codons)			
1. Uniform ω	19	−1731.03	$\omega_0 = 0.48$
2. Free-ratio........	35	−1718.80	ω estimated for 17 branches (0.098–3.30)
3. 2 ω—*Aotus* versus rest........	20	−1727.49	$\omega_0 = 0.43$, $\omega_A = 3.04$
C. Ten CGβ genes, using only the carboxy-terminal 33 codons			
1. Uniform ω	19	−417.06	$\omega_0 = 1.66$
2. Free-ratio........	35	−405.41	(See Figure 5)
3. 7 ω (A–F, ω0)........	25	−409.68	$\omega_0 = 0.55$, $\omega_A = \infty$, $\omega_B = \infty$, $\omega_C = 87.8$, $\omega_D = 5.50$, $\omega_E = 2.45$, $\omega_F = 3.41$
4. 6 ω (A = ω0)........	24	−411.37	$\omega_0 = 0.68$, $\omega_B = \infty$, $\omega_C = \infty$, $\omega_D = 5.18$, $\omega_E = 2.45$, $\omega_F = 3.44$
5. 6 ω (B = ω0)........	24	−411.04	$\omega_0 = 0.65$, $\omega_A = \infty$, $\omega_C = \infty$, $\omega_D = 4.85$, $\omega_E = 2.45$, $\omega_F = 3.20$
6. 6 ω (C = ω0)........	24	−414.16	$\omega_0 = 0.99$, $\omega_A = \infty$, $\omega_B = \infty$, $\omega_D = 4.10$, $\omega_E = 2.45$, $\omega_F = 3.44$
7. 6 ω (D = ω0)........	24	−411.28	$\omega_0 = 0.82$, $\omega_A = \infty$, $\omega_B = \infty$, $\omega_C = \infty$, $\omega_E = 2.45$, $\omega_F = 3.31$
8. 6 ω (E = ω0)........	24	−410.76	$\omega_0 = 0.72$, $\omega_A = \infty$, $\omega_B = \infty$, $\omega_C = \infty$, $\omega_D = 29.1$, $\omega_F = 3.40$
9. 6 ω (F = ω0)........	24	−411.11	$\omega_0 = 0.78$, $\omega_A = \infty$, $\omega_B = \infty$, $\omega_C = \infty$, $\omega_D = 4.42$, $\omega_E = 2.39$,
10. 2 ω (C, ω0)........	20	−414.40	$\omega_0 = 1.30$, $\omega_C = 89.9$

sequences best, although it is not a significantly better fit than the single-ratio model ($2\Delta L = 24.46$, $0.05 < P < 0.10$, df = 16). The ω values estimated for each branch on the tree under the free-ratio model (not shown) are all less than or near to the neutral expectation of 1.0 except for the terminal *Aotus* branch, which has an ω of 3.3. To investigate the robusticity of this finding, a third model was constructed in which two distinct ω values were hypothesized: one for the *Aotus* branch and one for the rest of the tree (table 4*B,* line 3). This model is significantly better than the single-ratio model ($2\Delta L = 7.08$, $P < 0.01$, df = 1), supporting the conclusion that there has been positive selection in CGβ along the *Aotus* branch.

Selection can sometimes act on a specific domain or structural subportion of a protein in remodeling it for a new function (Li 1997, p. 186). In these cases, evidence of positive selection would not be detected if the entire protein sequence were analyzed, because the ω value is averaged across the entire sequence. In the case of the CGβ genes, there is reason to believe that different portions of the protein may have been subject to different selection pressures, in particular some portion of the third exon of CGβ, which is not involved in gene conversion events with LHβ. A logical, biologically defined region to test is the carboxyl-tail created by the frameshift which distinguishes CGβ from LHβ.

To test this possibility, selection analyses were conducted on just the carboxyl-terminal 33 codons. The free-ratio model (table 4*C,* line 2) is not a significantly better fit than the uniform-ratio null model, ($2\Delta L = 19.86$, $P = \sim 0.10$, df = 13), although this model is the best fit for the data overall (table 4*C*). Although not statistically rigorous, it is useful informally to look at the ω values calculated for each branch under the free-ratio model to survey the variability of the ω values and identify branches which appear to be subject to positive selection. Individual ω values and the reconstructed number of changes for each branch calculated under the free-ratio model are shown in figure 5.

Eleven branches on this tree have ω values greater than 1.0, but in five of the cases, the branches are so short (2.1 or less total reconstructed changes) that it is unlikely that the ratios are significant or meaningful. The other six branches are of interest because they are longer branches and may represent cases of significant positive selection. They are the terminal branch leading to *Aotus* (branch A in fig. 5), the terminal branch leading to *Callithrix* (branch B), the terminal branch leading to *Callicebus* (C), the branch leading to the common ancestor of the three New World monkeys (D), the terminal branch leading to *Homo* (E), and the branch leading to the common ancestor of human and orangutan (branch F). If we assume (as in table 4*C,* line 3) a seven-ω model that assigns a distinct ω to branches A–F and a uniform ω_0 for the rest of the tree, this model has the second best overall fit for the data and rejects the null model ($2\Delta L = 14.76$, $P < 0.0250$, df = 6). This shows that branches A through F all together are significantly different from the negative selection in the rest of the gene tree.
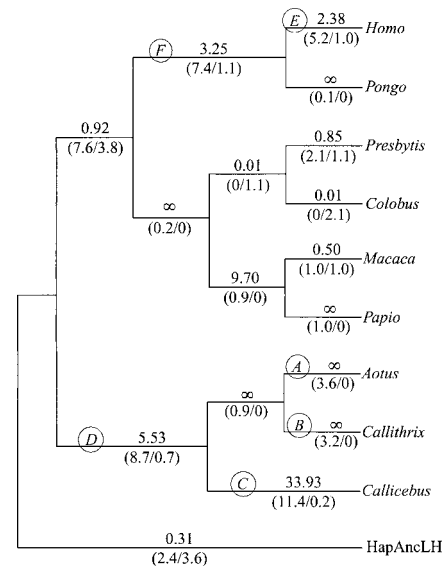

FIG. 5.—Phylogenetic tree showing the maximum likelihood analysis for selection in the carboxyl-tail portion of the CGβ genes mapped onto each branch of the tree. Estimates of ω (the nonsynonymous to synonymous substitution rate ratio) are given above each branch, and the estimated numbers of substitutions occurring along each branch are given below (replacement/silent). The ω values shown are calculated under the free-ratio model. HapAncLH is the maximum likelihood reconstructed sequence of the Haplorhine ancestor at the time of the initial primate LHβ gene duplication. Labeled branches A–E are discussed in the text.

Next, we can ask if each of these six branches individually contributes significantly to the robustness of the seven-ω model. To do this, we test a model which assumes only six ω values: ω_0 and only five branches which are different from the background. This model is tested six times (table 4*C*), in turn assuming the ω for one of the branches A–F is the same as ω_0. For branches A, B, and D–F, the seven-ω model is not significantly better than the six-ω model (comparing lines 4, 5, 7, 8, and 9 against line 3), yet the six-ω model is significantly better than the null model, so it appears that each of these branches alone does not contribute to the significance of the seven-ω model. Only branch C makes a significant contribution to the seven-ω model, for when branch C is set equal to ω_0, the resulting six-ω model (table 4*C,* line 6) does not reject the null model ($2\Delta L = 5.80$, $P > 0.10$, df = 5), and the seven-ω model is significantly better than this six-ω model ($2\Delta L = 8.96$, $P < 0.01$, df = 1).

These numbers indicate that whereas the tail region of the CGβ gene has been subject to varying degrees of negative (purifying) selection throughout most of its evolution in primates, there have been periods of positive selection acting on this portion of the CGβ gene in the platyrrhines, especially along the lineage leading to *Callicebus* after it diverged from the ancestor of *Callithrix* and *Aotus*. The analyses also find weak positive selection acting along the terminal *Aotus* and *Callicebus* lineages, the ancestral New World monkey lineage, and in the hominoid lineages both before and after the *Homo* lineage diverged from the lineage leading to *Pongo*;
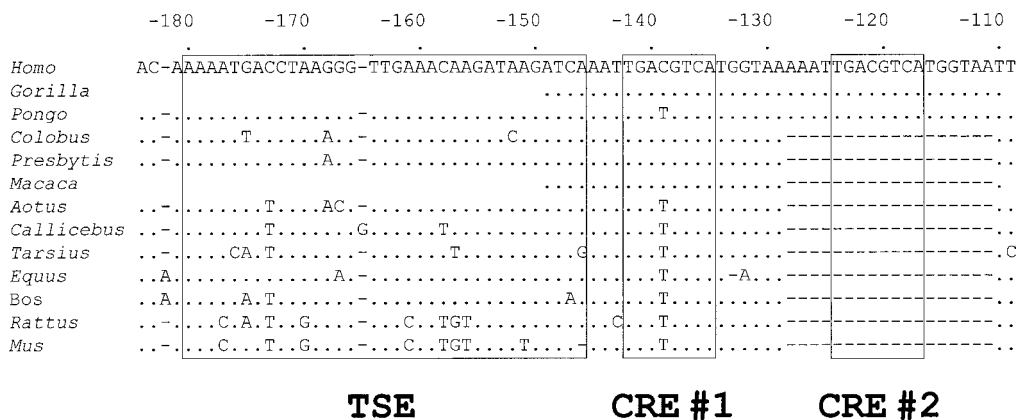
```
          -180      -170      -160      -150      -140      -130      -120      -110
           .    .    .    .    .    .    .    .    .    .    .    .    .    .    .
Homo       AC-AAAAATGACCTAAGGG-TTGAAACAAGATAAGATCAAATTGACGTCATGGTAAAAATTGACGTCATGGTAATT
Gorilla    ................................................
Pongo      ..-.............-................................T.....
Colobus    ..-......T......A.-...........C.....................---...........------..
Presbytis  ..-............A.-.................................---...........------..
Macaca     .............................................---...........------..
Aotus      ..-.......T....AC.-...............................T.....---...........------..
Callicebus ..-.......T....G......T...........................T.....---...........------..
Tarsius    ..-.....CA.T......-........T......G.....T.....---...........------..C
Equus      ..A.............A.-...............................T.....-A...---...........------~..
Bos        ..A......A.T......................A.....T.....---...........------..
Rattus     ..-....C.A.T...G....-...C..TGT..........C....T.....---...........------..
Mus        ..-....C...T..G....-...C..TGT....T.....-.....T.....---...........------..
                        TSE                        CRE #1      CRE #2
```

FIG. 6.—Nucleotide sequences of the promoter region of the GPHα gene including six from previously uncharacterized primate species. The sequences are aligned to the previously published human promoter; position numbers (above sequences) are relative to the first transcribed base of the gene. Bases matching the human sequence at each position are indicated by a period, gaps are indicated by a dash. The promoter elements required for expression in humans are boxed and labeled below the alignment; TSE: trophoblast-specific element, CRE: cyclic AMP response element. The human, horse, cow, mouse, and rat sequences are from Steger et al. (1991); the gorilla and rhesus macaque sequences are from Nilson et al. (1991).

however, the data are not robust. A total of 20.1 amino acid changes in this stretch of just 33 sites are inferred in the *Callicebus* CGβ gene since the common anthropoid ancestor, which translates to a rate of about 0.015 amino acid replacements per site per lineage per Myr, assuming the anthropoid ancestor lived 40 MYA. This rate is approximately five times faster than that seen in the interferons, which are some of the fastest evolving proteins known (Li 1997, pp. 180–181). Therefore, the new amino acids in the 3′-tail region may play an important role in the function of CG in the New World monkey clade represented here by *Callicebus*. It is noteworthy that the analysis of the CGβ-tail did not find strong positive selection in the *Aotus* lineage, yet the earlier analysis of the entire CGβ coding sequence did find positive selection in *Aotus*. This would suggest that the *Aotus* and *Callicebus* CGβ genes are evolving under different selection pressures.

Evolution of Placental Expression of the Alpha Subunit

Two of the significant molecular changes that must have occurred in the evolutionary history of primate CG are expression changes: CGβ had to gain expression in the placenta (and, at least in humans, greatly reduce expression in the pituitary), whereas GPHα had to gain placental expression, yet retain pituitary expression. The promoter of the CGβ genes is not understood well enough to identify specific nucleotide sites critical to placental expression. On the other hand, the expression of the GPHα gene is relatively well characterized and in the human placenta, depends on the presence of two promoter elements. The first, often called the trophoblast-specific element (TSE), spans nucleotides −180 to about −146 from the transcription initiation site (Roberts and Anthony 1994). The second promoter element is a perfect copy of the eight-base long cyclic AMP response element (termed CRE) (Bokar et al. 1989; Nilson et al. 1991). Humans have two copies of the CRE

(boxed in figure 6 and labeled below the alignment). It has been shown previously that the TSE is a required element for placental expression in humans, but it alone is not sufficient; at least one copy of the CRE is needed (Nilson et al. 1991). Therefore, this study focused on the CRE sequences. To investigate the evolution of these promoter elements, a 270-nt fragment of the GPHα gene was PCR amplified and sequenced from *Homo*, *Pongo*, *Presbytis*, *Colobus*, *Aotus*, *Callicebus*, and *Tarsius*. These sequences are available in GenBank (accession numbers AF401991–7), and a portion of them is shown in figure 6.

The first human CRE element has a homologous sequence in the promoter of all other species that have been studied, including the six primates for which new sequences are presented here (fig. 6). The second CRE, created by a duplication of an 18-nt stretch of the alpha promoter which was inserted just downstream of the first CRE element, is only found in humans, gorillas (Nilson et al. 1991), and orangutans (this study). It is most parsimonious to reconstruct this insertion event along the branch between the catarrhine common ancestor and the *Homo-Pongo* common ancestor. However, orangutan does not have two functional copies of the CRE. The second *Pongo* CRE is identical to the consensus sequence, but the first element differs from the consensus CRE sequence by a single nucleotide, a C → T transition at the fourth base of the eight-base element (site −139). It is known from in vitro assays that this single change is enough to decrease expression of the alpha gene more than 200-fold (Bokar et al. 1989; Steger et al. 1991). Therefore, orangutans probably have just one effectively functional copy of the CRE element, as do the Old World monkeys with their single CRE sequence.

Orangutans are not the only species with this substitution at −139 in the CRE element. The T base at this position appears to be the ancestral state for this base, as it is found in horses, cows, mice, rats, and rabbits (Steger et al. 1991), and also in tarsiers (this study;

fig. 6). This is not surprising, given that most of these species neither produce CG nor express GPHα in the placenta (with the exception of horses, discussed subsequently). In contrast, *Homo*, *Gorilla*, *Pongo*, *Macaca*, *Presbytis*, and *Colobus* all have a C at position −139, matching the consensus CRE sequence; this is consistent with the fact that each of these species would need to express the GPHα gene in the placenta in order to produce CG hormone. The most parsimonious reconstruction places the original (activating) T to C substitution at site −139 in the lineage leading to the common catarrhine ancestor. The mutation observed in the orangutan sequence reverting the consensus CRE back to the ancestral state (T) at this position presumably could not have occurred until after the second CRE element was inserted, because one functional CRE is essential for placental CG production and the maintenance of pregnancy.

Despite having at least one CGβ gene, both *Aotus* and *Callicebus* have a GPHα promoter with the ancestral T at position −139. They therefore do not have a consensus copy of the CRE known to be necessary for placental expression in humans. Nevertheless, placental expression of CG has clearly been shown in these and other New World monkey species (Hodgen et al. 1976; Hobson and Wide 1981; Crawford, Tregear, and Niall 1986; Einspanier et al. 1999). Therefore, the New World monkeys must have a mechanism of expression control for their GPHα gene that is different from the CRE-based mechanism in catarrhines.

Horses have evolved placental LH expression, which is functionally convergent upon the anthropoid CG, but with a different molecular basis (Sherman et al. 1992). GPHα is expressed in the horse placenta, but the promoter of the equine GPHα gene does not have a CRE: it differs by the same T nucleotide at position −139 found in the other non–CG-producing species (Steger et al. 1991; fig. 6). Rather, DNase-1 protection assays have shown that the regulatory protein α-ACT binds to the horse GPHα promoter (Steger et al. 1991). The *Aotus* sequence perfectly matches the horse sequence in the α-ACT binding site region (sites −161 to −142), and the *Callicebus* sequence differs by just one base. In fact, the human sequence also matches the horse in this region, yet binding of α-ACT alone to the human GPHα promoter does not stimulate expression (Steger et al. 1991), so from the sequence alone we cannot predict if α-ACT promotes GPHα expression in the New World monkeys. Whatever their mechanisms of expression control for the GPHα gene, this system may have evolved either within the platyrrhine lineage in parallel to the CRE-based control system of the catarrhines or it may have evolved before the anthropoid common ancestor and then later been replaced in the catarrhine lineage by the CRE system.

CG and Models of the Evolution of New Genes

The evolution of CG is a case study of how new genes with new functions arise from existing genes. The classical model (Ohno 1970, p. 71) posits that genes duplicate before new functions evolve; one gene copy retains the required ancestral function, whereas the fate of the second copy is likely to be nonfunctionalization, in which a mutation abolishes its ability to be expressed or to carry out the function of its progenitor. Only in rare cases does a duplicate gene evolve a new, selectively advantageous function under this model. A recent genome-wide survey shows that gene duplications occur at a high rate, and fewer pseudogenes exist than predicted by the classical model (Lynch and Conery 2000). This empirical observation prompted the proposal of an alternative model in which genes evolve multiple functions before duplication. Duplicate genes then evolve by subfunctionalization—the fixation of complementary degenerative mutations in the daughter copies, which requires the presence of both copies to maintain the ancestral functions (Force et al. 1999; Lynch and Force 2000). It predicts a much higher probability for the preservation of actively expressed duplicate genes than does the classical model.

The subfunctionalization model prompts a reexamination of CG evolution in ways that the classical model does not. It is possible that the ancestral primate LHβ gene was expressed in both the pituitary and the placenta, as is the case with horse LHβ. Additionally, primate LHβ and CGβ genes use different upstream regions for their proximal promoter elements, so that the promoters of these two genes do not overlap. It is not known whether the LHβ gene is expressed in the placentae of any extant species, but it has been shown that there is transcription of all of the human CGβ genes in the pituitary, although at very low levels (Dirnhofer et al. 1996). Finally, it is unclear from the reconstruction of events in this study when CGβ first gained placental expression, but it is possible that placental expression could have been gained before gene duplication. If the ancestral LH gene had acquired placental expression before the reconstructed gene duplication event (which could be tested by examining tarsier first-trimester placentae—a rare commodity), this would prove that CG evolved from LH by subfunctionalization.

The Origin of CG as a Functional Signal of Pregnancy

If placental expression of LH evolved before the origin of the CGβ gene family, then the origin of the functional activity of establishing pregnancy via a placentally expressed gonadotropic hormone precedes, and is independent of, the gene duplication event leading to the origin of the CGβ subunit gene. It may then be a different question to ask when this placental gene expression first evolved. Could placentally expressed LH have been functioning to establish pregnancy as long ago as the origin of placental mammals? The evolution of placental morphology suggests not. LH (or CG) has to move from the placenta into the maternal bloodstream and then be transported to the ovary in order to act on its target, the corpus luteum. Anthropoid primates all have a hemochorial placenta, in which placental tissue is directly bathed in maternal blood, making it easy for placentally derived molecules to enter the maternal

bloodstream (King 1993). Strepsirrhine primates and most other mammals have an epitheliochorial placenta, in which both the uterine epithelium and the maternal vascular endothelium remain present during pregnancy. These two additional tissue layers impede the flow of large macromolecules (such as the glycoprotein hormones LH and CG) from the placenta to the maternal bloodstream (Faber, Thornburg, and Binder 1992). Given the widespread distribution of mammals possessing an epitheliochorial placenta, this is thought to be the ancestral state of mammalian placentation (Luckett 1976).

We propose that LH could not have been efficiently delivered to the corpus luteum before hemochorial placentation evolved. In primates, hemochorial placentation first appears in tarsiers, making the haplorhine common ancestor the first primate that would have been capable of making efficient use of LH as a pregnancy-establishing signal. It should be noted that horses, which have evolved CG independently from primates, have also evolved specialized placental structures—endometrial cups—which help in the delivery of equine CG to the mare's bloodstream (Mossman 1987, p. 271). A recent report proposes that guinea pigs (*Cavia porcellus*) have also independently evolved CG (Sherman et al. 2001); however, the evidence is not convincing (primarily because the authors present Northern blots showing no CG expression in the guinea pig placenta). Nonetheless, guinea pigs have evolved a labyrinthine type of hemochorial placentation (Mossman 1987, p. 230), so if CG were to have evolved in this species, it could be efficiently delivered to the maternal bloodstream. This final discussion shows the power of combining the molecular evolutionary history of a gene with the anatomical evolution of a tissue expressing that gene, in order to build a more complete picture of the evolution of a higher primate adaptation.

Acknowledgments

LITERATURE CITED

BAILEY, W. J., D. H. A. FITCH, D. A. TAGLE, J. CZELUSNIAK, J. L. SLIGHTOM, and M. GOODMAN. 1991. Molecular evolution of the Ψη-globin gene locus: gibbon phylogeny and the hominoid slowdown. Mol. Biol. Evol. **8**:155–184.

BOKAR, J. A., R. A. KERI, T. A. FARMERIE, R. A. FENSTER-MAKER, B. ANDERSEN, D. L. HAMERNIK, J. YUN, T. WAGNER, and J. H. NILSON. 1989. Expression of glycoprotein hormone alpha-subunit gene in the placenta requires a functional cyclic AMP response element, whereas a different *cis*-acting element mediates pituitary-specific expression. Mol. Cell. Biol. **9**:5113–5122.

BRAUNSTEIN, G. D., J. L. VAITUKAITIS, and G. T. ROSS. 1972. The in vivo behavior of human chorionic gonadotropin after dissociation into subunits. Endocrinology **91**:1030–1036.

BROWN, P., J. R. MCNEILLY, R. M. WALLACE, A. S. MCNEILLY, and A. J. CLARK. 1993. Characterization of the ovine LH beta-subunit gene: the promoter directs gonadotrope-specific expression in transgenic mice. Mol. Cell. Endocrinol. **93**:157–165.

CARR, F. E., and W. W. CHIN. 1985. Absence of detectable chorionic gonadotropin subunit messenger ribonucleic acids in the rat placenta throughout gestation. Endocrinology **116**:1151–1157.

COURSEAUX, A., and J.-L. NAHON. 2001. Birth of two chimeric genes in the *Hominidae* lineage. Science **291**:1293–1297.

CRAWFORD, R. J., G. W. TREGEAR, and H. D. NIALL. 1986. The nucleotide sequences of baboon chorionic gonadotropin β-subunit genes have diverged from the human. Gene **46**:161–169.

DE KRETSER, D. M., R. C. ATKINS, and C. A. PAULSEN. 1973. Role of the kidney in the metabolism of luteinizing hormone. J. Endocrinol. **58**:425–434.

DIRNHOFER, S., M. HERMANN, and A. HITTMAIR. 1996. Expression of the human chorionic gonadotropin-β gene cluster in human pituitaries and alternate use of exon 1. J. Clin. Endocrinol. Metab. **81**:4212–4217.

DROUIN, G., F. PRAT, M. ELL, and G. D. P. CLARKE. 1999. Detecting and characterizing gene conversions between multigene family members. Mol. Biol. Evol. **16**:1369–1390.

EINSPANIER, A., R. NUBBEMEYER, S. SCHLOTE, M. SCHUMACHER, R. IVELL, K. FUHRMANN, and A. MARTEN. 1999. Relaxin in the marmoset monkey: secretion pattern in the ovarian cycle and early pregnancy. Biol. Reprod. **61**:512–520.

EZASHI, T., T. HIRAI, T. KATO, K. WAKABAYASHI, and Y. KATO. 1990. The gene for the beta subunit of porcine LH: clusters of GC boxes and CACCC elements. J. Mol. Endocrinol. **5**:137–146.

FABER, J. J., K. L. THORNBURG, and N. D. BINDER. 1992. Physiology of placental transfer in mammals. Am. Zool. **32**:343–354.

FIDDES, J. C., and H. M. GOODMAN. 1980. The cDNA for the β-subunit of human chorionic gonadotropin suggests evolution of a gene by readthrough into the 3′-untranslated region. Nature **286**:684–687.

FORCE, A., M. LYNCH, F. B. PICKETT, A. AMORES, Y. YAN, and J. POSTLETHWAIT. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**:1531–1545.

GILLESPIE, J. 1991. The causes of molecular evolution. Oxford University Press, New York.

GOLDMAN, N., and Z. YANG. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. **11**:725–736.

GRAHAM, M. Y., T. OTANI, I. BOIME, M. V. OLSON, G. F. CARLE, and D. D. CHAPLIN. 1987. Cosmid mapping of the human chorionic gonadotropin β genes by field-inversion gel electrophoresis. Nucleic Acids Res. **15**:4437–4448.

HASEGAWA, M., H. KISHINO, and T. A. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HEARN, M. T. W., and P. T. GOMME. 2000. Molecular architecture and biorecognition processes of the cystine knot protein superfamily: part I. The glycoprotein hormones. J. Mol. Recognit. **13**:223–278.

HOBSON, B. M., and L. WIDE. 1981. The similarity of chorionic gonadotrophin and its subunits in term placentae from man, apes, Old and New World monkeys and a prosimian. Folia Primatol. **35**:51–64.

HODGEN, G. D., L. G. WOLFE, J. D. OGDEN, M. R. ADAMS, C. C. DESCALZI, and D. F. HILDEBRAND. 1976. Diagnosis of pregnancy in marmosets: hemagglutination inhibition test and radioimmunoassay for urinary chorionic gonadotropin. Lab. Anim. Sci. **26**:224–229.

JAMESON, L., W. W. CHIN, A. N. HOLLENBERG, A. S. CHANG, and J. F. HABENER. 1984. The gene encoding the β-subunit of rat luteinizing hormone: analysis of gene structure and evolution of nucleotide sequence. J. Biol. Chem. **259**: 15,474–15,480.

KING, B. F. 1993. Development and structure of the placenta and fetal membranes of nonhuman primates. J. Exp. Zool. **266**:528–540.

KUMAR, T. R., and M. M. MATZUK. 1995. Cloning of the mouse gonadotropin β-subunit encoding genes, II. Structure of the luteinizing hormone β-subunit-encoding genes. Gene **166**:335–336.

LAPTHORN, A., D. HARRIS, and A. LITTLEJOHN. 1994. Crystal structure of human chorionic gonadotropin. Nature **369**: 455–461.

LI, W.-H. 1997. Molecular evolution. Sinauer Associates, Sunderland, Mass.

LI, M. D., and J. J. FORD. 1998. A comprehensive evolutionary analysis based on nucleotide and amino acid sequences of the α- and β-subunits of glycoprotein hormone gene family. J. Endocrinol. **156**:529–542.

LUCKETT, W. P. 1976. Cladistic relationships among primate higher categories: evidence of the fetal membranes and placenta. Folia Primatol. **25**:245–276.

LUND, L. A., and G. B. SHERMAN. 1998. Duplication of the southern white rhinoceros (*Ceratherium simum simum*) luteinizing hormone β subunit gene. J. Mol. Endocrinol. **21**: 19–30.

LYNCH, M., and J. S. CONERY. 2000. The evolutionary fate and consequences of duplicate genes. Science **290**:1151–1155.

LYNCH, M., and A. FORCE. 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics **154**: 459–473.

MALIK, H. S., and S. HENIKOFF. 2001. Adaptive evolution of Cid, a centromere-specific histone in drosophila. Genetics **157**:1293–1298.

MATZUK, M. M., A. J. W. HSUEH, P. LAPOLT, A. TSAFRIRI, J. L. KEENE, and I. BOIME. 1990. The biological role of the carboxyl-terminal extension of human chorionic gonadotropin β-subunit. Endocrinology **126**:376–383.

MESSIER, W., and C.-B. STEWART. 1997. Episodic adaptive evolution of primate lysozymes. Nature **385**:151–154.

MORSE, J. H., G. STEARNS, J. ARDEN, G. M. AGOSTA, and R. E. CANFIELD. 1976. The effects of crude and purified human gonadotropin on in vitro stimulated human lymphocyte cultures. Cell. Immunol. **25**:178–188.

MOSSMAN, H. W. 1987. Vertebrate fetal membranes. Rutgers University Press, New Brunswick, NJ.

NILSON, J. H., J. A. BOKAR, C. M. CLAY, T. A. FARMERIE, R. A. FENSTERMAKER, D. L. HAMERNIK, and R. A. KERI. 1991. Different combinations of regulatory elements may explain why placenta-specific expression of the glycoprotein hormone α-subunit gene occurs only in primates and horses. Biol. Reprod. **44**:231–237.

NURMINSKY, D. I., M. V. NURMINSKAYA, D. DE AGUIAR, and D. L. HARTL. 1998. Selective sweep of a newly evolved sperm-specific gene in *Drosophila*. Nature **396**:572–575.

OHNO, S. 1970. Evolution by gene duplication. Springer-Verlag, New York.

POLICASTRO, P. F., S. DANIELS-MCQUEEN, G. CARLE, and I. BOIME. 1986. A map of the hCGβ-LHβ gene cluster. J. Biol. Chem. **261**:5907–5916.

RASBAND, W. S., and D. S. BRIGHT. 1995. NIH image: a public domain image processing program for the macintosh. Microbeam Anal. Soc. J. **4**:137–149.

ROBERTS, R. M., and R. V. ANTHONY. 1994. Molecular biology of trophectoderm and placental hormones. Pp. 395–440 *in* J. K. FINDLAY, ed. Molecular biology of the female reproductive system. Academic Press, San Diego.

SAWYER, S. 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6**:526–538.

SAWYER, S. A. 2000. GENECONV: statistical tests for detecting gene conversion—version 1.81. Department of Mathematics, Washington University, St. Louis, Mo.

SHERMAN, G. B., D. F. HEILMAN, A. J. HOSS, D. BUNICK, and L. A. LUND. 2001. Messenger RNAs encoding the β subunits of guinea pig (*Cavia porcellus*) luteinizing hormone (gpLH) and putative chorionic gonadotropin (gpCG) are transcribed from a single-copy gpLH/CGβ gene. J. Mol. Endocrinol. **26**:267–280.

SHERMAN, G. B., M. WOLFE, T. FARMERIE, C. CLAY, D. THREAGILL, D. SHARP, and J. NILSON. 1992. A single gene encodes the β-subunits of equine luteinizing hormone and chorionic gonadotropin. Mol. Endocrinol. **6**:951–959.

SIMULA, A. P., F. AMATO, R. FAAST, A. LOPATA, J. BERKA, and R. J. NORMAN. 1995. Luteinizing hormone/chorionic gonadotropin bioactivity in the common marmoset (*Callithrix jacchus*) is due to a chorionic gonadotropin molecule with a structure intermediate between human chorionic gonadotropin and human luteinizing hormone. Biol. Reprod. **53**: 380–389.

STEGER, D. J., J. ALTSCHMIED, M. BUSCHER, and P. L. MELLON. 1991. Evolution of placenta-specific gene expression: comparison of the equine and human gonadotropin α-subunit genes. Mol. Endocrinol. **5**:243–255.

SWOFFORD, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0. Sinauer, Sunderland, Mass.

TALMADGE, K., N. C. VAMVAKOPOULOS, and J. C. FIDDES. 1984. Evolution of the genes for the β subunits of human chorionic gonadotropin and luteinizing hormone. Nature **307**:37–40.

TEPPER, M. A., and J. L. ROBERTS. 1984. Evidence for only one β-luteinizing hormone and no β-chorionic gonadotropin gene in the rat. Endocrinology **115**:385–391.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**:4876–4882.

TULLNER, W. W. 1974. Comparative aspects of primate chorionic gonadotropins. Contrib. Primatol. **3**:235–257.

VIRGIN, J. B., B. J. SILVER, A. R. THOMASON, and J. H. NILSON. 1985. The gene for the β subunit of bovine luteinizing hormone encodes a gonadotropin mRNA with an unusually short 5′-untranslated region. J. Biol. Chem. **260**:7072–7077.

WANG, W., J. ZHANG, C. ALVAREZ, A. LLOPART, and M. LONG. 2000. The origin of the *Jingwei* gene and the complex modular structure of its parental gene, *Yellow Emperor,* in *Drosophila melanogaster*. Mol. Biol. Evol. **17**:1294–1301.

YANG, Z. 1998. Likelihood ratio test for detecting positive selection and application to primate lysozyme evolution. Mol. Biol. Evol. **15**:568–573.

———. 2000. Phylogenetic analysis by maximum likelihood (PAML). Version 3.0. University College London, London, U.K.