Duplication of the southern white rhinoceros (*Ceratotherium simum simum*) luteinizing hormone β subunit gene

L A Lund and G B Sherman

Department of Veterinary Biosciences, College of Veterinary Medicine, University of Illinois, 2001 South Lincoln Avenue, Urbana, Illinois 61802, USA

(Requests for offprints should be addressed to G B Sherman who is now at University of Nebraska-Lincoln, Great Plains Veterinary Educational Center, PO Box 187, Clay Center, Nebraska 68933, USA)

ABSTRACT

Luteotropic glycoprotein hormones (LGH) include luteinizing hormone (LH) and chorionic gonadotropin (CG). The order Primates is the only phylogenetic clad known to exhibit more than one LGHβ subunit gene per haploid genome. In the present study, we report the discovery of a second case of LGHβ gene replication, in the white (w) rhinoceros (r or rhino). The presence of more than one gene was strongly suggested by a complex banding pattern observed on Southern blots of DNA prepared from two unrelated white rhinos. The existence of two LGHβ genes per haploid genome was estimated by genomic equivalence assay. However, genomic restriction-site mapping studies, together with other findings, suggested that the replicates are probably not tandemly arranged as occurs in primates. A simple band pattern was observed in Southern blots of four other perissodactyl species, indicating that a single-copy LHβ gene is the consensus condition. Two distinct white rhino LHβ genomic clones (wrLHβ1 and wrLHβ2) were isolated. The nucleotide sequence of wrLHβ1 was identical with that of wrLHβ2, except that the latter lacked the consensus mammalian LGHβ second intron. Sequences of the TATA-containing proximal 5′-flanking regions of the two genes were homologous to at least − 57 relative to the site of pituitary transcriptional initiation. We conclude that wrLHβ1 is the extant form of the ancestral perissodactyl LHβ gene, whereas wrLHβ2 is a randomly integrated cDNA element (processed gene) reverse transcribed from a partially spliced ancestral wrLHβ1 mRNA. That wrLHβ2 was heritable demonstrates that wrLHβ1 was transcribed in gametes or early conceptus cells contributing to the germline at some point in time since the divergence of white rhinos from other members of the family Rhinocerotidae. Furthermore, because homologous proximal (pituitary) promoter sequence is present in wrLHβ2, it can be concluded that the wrLHβ1 mRNA template from which wrLHβ2 is derived was transcribed from a secondary promoter located upstream of the consensus TATA-regulated pituitary promoter.

Journal of Molecular Endocrinology (1998) **21**, 19–30

INTRODUCTION

There are two types of mammalian luteotropic glycoprotein hormones (LGH), luteinizing hormone (LH), produced by gonadotropes in the anterior pituitary gland, and chorionic gonadotropin (CG), produced by trophoblasts within the placenta. Pituitary expression of LH is ubiquitous among mammals, while placental CG expression is well described only in equids and primates (Pierce & Parsons 1981). LH plays an essential role in the endocrine regulation of gamete development in mammalian testes and ovaries, whereas CG is postulated to support ovarian function during pregnancy. Despite differences in their tissue sources, secretory patterns and the reproductive periods during which they function, LH and CG are classified together as luteotropic gonadotropins because they bind to the same LH/CG gonadal receptor and induce similar luteotropic responses. Moreover, these two gonadotropins are biochemically related, each containing a common α subunit

non-covalently bound to a hormone-specific β subunit, a heterodimeric structure required for high-affinity receptor binding and induction of biological activity (Bousfield *et al*. 1994). Similarity is high (>95%) between LHβ and CGβ gene sequences, reflecting the fact that they evolved from a common ancestral LHβ gene (Talmadge *et al*. 1984).

In the haploid genomes of most mammals, an LHβ gene is present as a single copy, directs pituitary-limited TATA-driven transcription of an mRNA with a short (6–11 nucleotide) 5′-untranslated region (UTR), and encodes a mature peptide of 121 amino acids (Bousfield *et al*. 1994). These consensus LHβ gene and locus attributes are postulated to represent the ancestral mammalian condition. By contrast, different structural and functional features are evident at LGHβ loci within the order Primates and the family Equidae, the two lineages known to produce both LH and CG.

In equids, there is an LH/CGβ gene that is present as a single copy (Sherman *et al*. 1992) which has gained the capacity for placenta-specific as well as pituitary-specific expression (Murphy & Martinuk 1991, Sherman *et al*. 1992). This is in contrast with primates in which a single pituitary-specific LHβ gene is linked to a cluster of six CGβ genes. While some primate CGβ replicates may be pseudogenes, those that are functional are expressed in a placenta-specific manner (Bo & Boime 1992). In the horse, LH/CGβ gene transcription is TATA-driven in both pituitary and placenta (endometrial cup). However, the presence of tandemly repeated functional TATA elements within a distinctive 51 bp insert in the horse leads to synthesis in both tissues of mRNAs with elongated 5′-UTRs (Sherman *et al*. 1992). By comparison, pituitary LHβ gene expression in primates is directed by a consensus single-TATA promoter, whereas functional CGβ genes adopt an alternative TATA-less upstream promoter to drive placental-specific expression. Thus, in both equids and primates, modifications of proximal promoter usage and 5′-UTR length and/or sequence structure correlate with acquisition of the capacity for placental expression.

CG expression is widespread throughout the order Primates (Crawford *et al*. 1986, Gwenda *et al*. 1990, Hearn *et al*. 1991, Steinetz *et al*. 1992, Seshagiri & Hearn 1993, Summers *et al*. 1993). However, it is not known whether placental CG expression is similarly widespread within the order Perissodactyla, which is composed of three extant families, Equidae, Rhinocerotidae and Tapiridae. By virtue of their relatively close phylogenetic relationship to equids, non-equid perissodactyl

species represent a potentially valuable comparative genetic model for the study of the evolutionary pathways and molecular events associated with transition from pituitary-limited LHβ gene expression (consensus mammalian condition) to the dual pituitary/placental expression capacity of the equid LH/CGβ gene. We have previously found preliminary evidence that some rhinoceros species exhibit pituitary LHβ cDNA/mRNA features similar to the consensus mammalian pituitary transcript (Sherman *et al*. 1997). In the present study, the objective was to further describe the perissodactyl LGHβ locus by determining gene copy number for several perissodactyl species and by isolating and characterizing the nucleotide sequence of the first LGHβ genomic clones from a rhinoceros species.

MATERIALS AND METHODS

Animals

Tissue samples were collected from five perissodactyl species including Sumatran rhino (*Dicerorhinus sumatrensis*), Indian rhino (*Rhinoceros unicornis*), black rhino (*Diceros bicornis*), southern white rhino (*Ceratotherium simum simum*), and Baird's tapir (*Tapirus bairdii*). Samples were obtained from two unrelated southern white rhinos; one was a captive 31-year-old female in residence at the Fossil Rim Wildlife Refuge, Texas (white rhino 1), and the other was a 5-year-old male from Matobo National Park in Zimbabwe (white rhino 2).

PCR amplification and sequence analysis

White rhino LHβ (wrLHβ) sequence amplifications were performed in a total volume of 100 µl using the Perkin–Elmer DNA Thermal Cycler (Norwalk, CT, USA) and 1 µg genomic DNA as target. (Primer locations are shown in Fig. 2.) Except where noted, biochemicals were purchased from Gibco–BRL (Grand Island, NY, USA). White rhino 2 LHβ1 and LHβ2 sequences were amplified using a consensus mammalian LGHβ forward primer P-3 (5′-GGCCTTGCCGCCCCACA) and a white rhino-based (Sherman *et al*. 1997) reverse primer P-2 (5′-ACGGTTCGAGATG CCTTTAT) with the Perkin–Elmer GeneAmp XL PCR Kit in reactions containing 2 U/100 µl r*Tth* polymerase, 200 µM deoxynucleotide triphosphates, 0·25 µM primers and 1·5 mM magnesium acetate. PCR was carried out for 33 cycles (denaturation at 95 °C for 1 min; annealing at 64 °C for 1 min; extension at 72 °C for 2 min). An Indian rhino-based primer (P-1; 5′-GCCTGCAGGTATAAG ACCAGGTAA   ACA)   located   immediately

downstream from P-3 was used with P-2 to amplify white rhino 1 LHβ2 sequence. Final PCR reagent concentrations were: 200 µM deoxynucleoside triphosphates, 0·25 µM primers, 10 mM Tris, pH 8·3, 50 mM KCl, 1·5 mM MgCl₂, 0·001% gelatin and 2·5 U/100 µl *Taq* polymerase (Gibco–BRL). The reaction was performed as above with the exception of a 59 °C annealing temperature. Another PCR was conducted using as target BamHI-digested white rhino 2 genomic DNA gel, isolated from an approximately 2500 bp band that hybridized with wrLHβ cDNA probe. PCR was carried out for 33 cycles (denaturation at 95 °C for 1 min; annealing and extension at 72 °C for 2·5 min) using the reagents as specified above for the *Taq* polymerase and white rhino-based primers P-6 (5′-TGCCCCGTCTGCATCACCTTCAC) and P-7 (5′-CGGGAGCCGGATGGAGGCAA). The resulting 450 bp product was subsequently used as template in a semi-nested PCR employing a degenerate consensus mammalian primer (5′-ACCA(C/G)CATCTG(C/T)GCCGGCTACTG(C/T)CC(C/T)A(C/G)ATG) complementary to the final 33 bp of LGHβ exon 2 paired with P-7. Sequence analysis of a 425 bp product revealed the first 66 bp of wrLHβ exon 3 at the 3′ end of the fragment, indicating that the remaining sequence contained LHβ intron 2 (data not shown). PCR was carried out for 33 cycles (denaturation at 95 °C for 1 min; annealing at 63 °C for 1 min; extension at 72 °C for 2 min) using the reagents as specified above for the *Taq* polymerase.

After cloning into pCR-Script (Stratagene, La Jolla, CA, USA), amplification products were sequenced in both directions using an Applied Biosystems 373A Automated DNA Sequencer (PE Applied Biosystems, Foster City, CA, USA). (Sequencing primer locations are shown in Fig. 2 for the principal full-length cloned PCR products). Vector-based primers include P-F (5′-GTAAAACGACGGCCAGT), P-R (5′-AGCGGATAACAATTTCACACAGGA) and P-KS (5′-TCGAGGTCGACGGTATC). Primers based on LGHβ sequence include P-4 (5′-GTGTGCACCTACCATGA), P-21 (5′-ACCACCAGCATCTGTGCCGGCTACTG) and P-5 (5′-TGGGCCGGCACAGTGGCCGCAGTGG).

Genomic Southern blots

Tissues used for genomic DNA isolation included liver (Sumatran rhino and white rhino 1), testes (Indian rhino and tapir) and pituitary (black rhino, white rhino 1 and white rhino 2). DNA was prepared by proteinase K digestion and organic extraction using standard techniques (Strauss 1995).

Biochemicals and restriction enzymes were purchased from Gibco–BRL. A 10 µg sample of restriction enzyme-digested DNA was electrophoresed in a 0·7% agarose gel prepared using TBE buffer (89 mM Tris, 89 mM borate, 2·76 mM EDTA, pH 8·3) at 2·5 V/cm for 8-11 h. Gels were prepared for blotting on to Duralon-UV membrane (Stratagene) according to the manufacturer's specifications and DNA was transferred by standard capillary blotting (Brown 1995). Blots were hybridized to ³²P-labeled probe prepared by random-primed labeling (Multiprime DNA Labeling System; Amersham, Arlington Heights, IL, USA) for 2·5 h at 65 °C in QuikHyb solution (Stratagene). Two washes in 2 × SSC/0·1% SDS at 25 °C for 15 min were followed by a final wash in 0·1 × SSC/1% SDS at 60 °C for 30 min. (Single-strength SSC is 0·15 M NaCl and 0·015 M sodium citrate.) Membranes were exposed to autoradiography film (X-OMAT AR; Eastman Kodak, Rochester, NY, USA) with an intensifying screen at −70 °C for 30–48 h.

DNA used for probe synthesis included wrLHβ cDNA sequence corresponding to codons −20 to +121 (Sherman *et al*. 1997), and a fragment comprising the first 220 bp of wrLHβ1 intron 2. The latter was a restriction digested product of a larger amplification product that spanned intron 2 and included the first 66 bp of exon 3 (see PCR methods above). The intron 2 fragment was used as a probe to identify bands containing only wrLHβ1 sequence, while wrLHβ cDNA probe, which could hybridize to both wrLHβ1 and wrLHβ2, was used for all other Southern blots.

To perform genomic titration analysis, plasmid containing wrLHβ1 PCR product was digested to release a 3450 bp fragment containing full-length wrLHβ1 sequence flanked by plasmid sequence. Aliquots of the digest containing 1, 2 or 3 genomic equivalents were added to 10 µg white rhino genomic DNA digested with either HindIII or BamHI and then subjected to standard Southern blotting procedures. The number of genomic equivalents present in 10 µg white rhino genomic DNA was calculated from an estimated value of $3·0 × 10^9$ bp per haploid mammalian genome (Lewin 1997*a*). Densitometric analysis (total optical density in a constant area) was performed using the PDI Discovery Series densitometry system running RFLPrint software (PDI, Huntington Station, NY, USA).

RESULTS

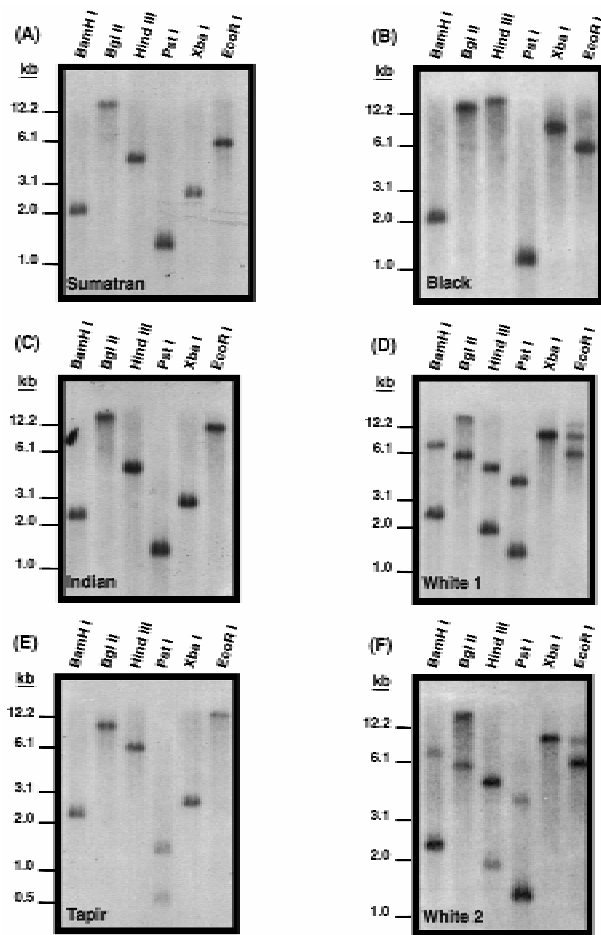To determine whether there are one or multiple wrLGHβ genes per haploid genome, Southern blot

FIGURE 1. Southern blots of genomic DNA isolated from four rhinoceros species and one tapir species. Blots were hybridized with full-length wrLHβ cDNA probe. The simple banding pattern observed for Sumatran, black and Indian rhinos (A–C) and Baird's tapir (E) strongly suggests a single-copy LHβ gene per haploid genome. By contrast, the predominance of lanes with multiple bands for two unrelated white rhinos (D and F) indicates the presence of two or more LGHβ genes in this species.

analyses of restriction enzyme-digested white rhino, black rhino, Sumatran rhino, Indian rhino and Baird's tapir genomic DNA were performed (Javan rhino tissue was unavailable). All species except the white rhino exhibited a simple banding pattern (generally a single band per lane; Fig. 1), strongly suggesting a single-copy gene. By contrast, two bands were observed in most lanes for the white rhino (Fig. 1D). To confirm that this more complex banding pattern was not unique to the original white rhino subject being evaluated (white rhino 1), genomic DNA was prepared from an unrelated white rhino (white rhino 2) and assayed in the same

manner. The complex Southern banding pattern was again observed (Fig. 1F). The presence of multiple bands in a majority of the white rhino lanes supports the contention that there are two or more LGHβ gene copies per haploid genome in this species.

Homologous primers based on wrLHβ cDNA sequence were used with consensus mammalian LGHβ primers to amplify full-length wrLHβ gene fragments. Two major amplification products of different size were generated (Fig. 2). One of the amplified fragments (wrLHβ1) was of the expected size (~ 1260 bp), whereas the other amplified fragment (wrLHβ2) was smaller (~ 960 bp) than predicted for a typical mammalian LGHβ gene (Fig. 2A). Putative coding sequence regions of wrLHβ1 and wrLHβ2 clones (Fig. 2B) were identical, differing at only two nucleotide positions from pituitary-derived wrLHβ cDNA sequence previously characterized in our laboratory (Sherman *et al.* 1997). One substitution constituted a silent mutation (Arg[112]), while the other encoded Arg[2] in the gene and Lys[2] in the cDNA. As in the cDNA, (1) all cysteine residues were found to be in register with the 14 highly conserved cysteines present in the consensus mammalian LHβ subunit (Bousfield *et al.* 1994), (2) a single consensus glycosylation attachment site was present at Asn[13], which is the typical location for oligosaccharide attachment in other mammalian LHβ subunits (Bousfield *et al.* 1994), and (3) the in-frame translational stop codons for wrLHβ1 and wrLHβ2 defined amino acid position +121 as the C-terminus, demonstrating that neither wrLHβ gene encodes the C-terminal peptide extension (CTP) which is present on both equid LH/CGβ and human CGβ. Thus, the common wrLHβ1/wrLHβ2 coding sequence suggests a consensus LHβ peptide structure (Bousfield *et al.* 1994). (Although the use of the term wrLGHβ would acknowledge uncertainty about the tissue-specific expression patterns, these genes are designated wrLHβ here to adhere to convention.)

Unlike the coding sequence regions, there was a salient intronic sequence difference between the two wrLHβ genes. While wrLHβ1 exhibited an exon–intron structure typical of the consensus mammalian LHβ condition (three exons and two introns), wrLHβ2 was missing the 289 bp consensus second intron (Fig. 2). The resulting wrLHβ2 gene was structured such that exons 2 and 3 were perfectly conjoined, preserving the correct reading frame. Accordingly, if expressed, the wrLHβ2 gene would contain two exons and one intron. By contrast, consensus first introns were present in both wrLHβ genes and their nucleotide sequences were identical. To ensure that the absence of the

(A) wrLHβ1

wrLHβ2

~ 0.2 kb

■ Coding Sequence	▭ Flanking Sequence
▨ Intronic Sequence	▶ Oligonucleotide Primer

(B)
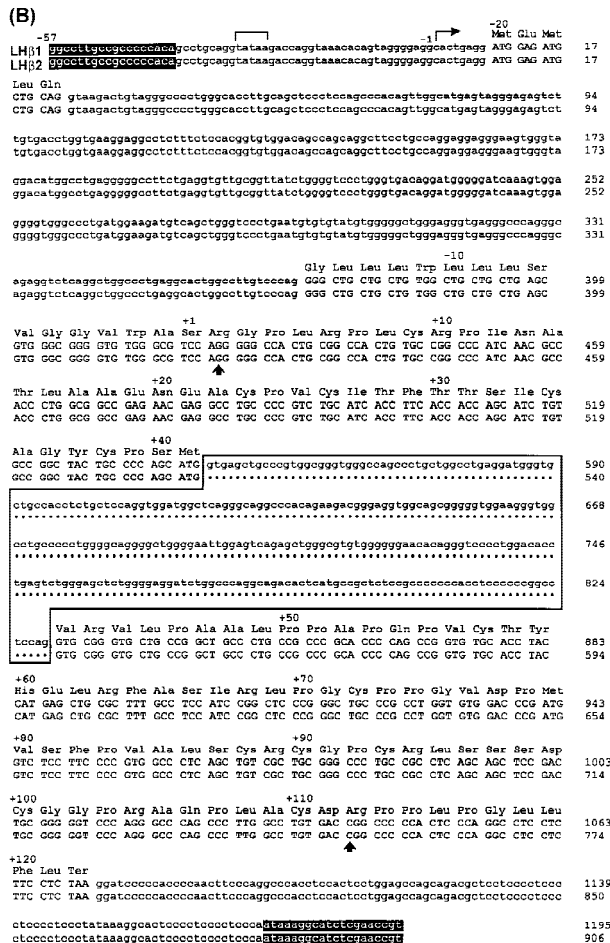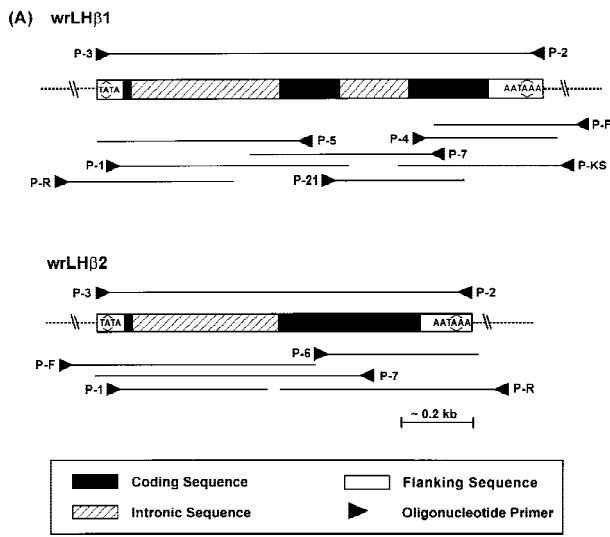
second intron was not an artefact, additional amplification reactions were carried out using white rhino genomic DNA as template. Fragments containing and lacking second intron sequence were routinely amplified, confirming the existence of both intron-containing and intronless wrLHβ target (data not shown). An amplified Indian rhino LHβ genomic fragment corresponding to amino acid position − 20 to codon +70 of the third exon was also isolated and sequenced to determine whether a second intron was present in a rhino species known to exhibit a single-copy LHβ gene. Like wrLHβ1, this fragment contained the consensus second intron (data not shown).

Homology between the two genes included 65 bp of 5′-flanking sequence (Fig. 2B) extending from the beginning of the cloned inserts to the putative translational start codon. Similarly, 100% sequence identity was observed over the 103 bp span of 3′-UTR sequence included in the genomic clones (Fig. 2B). There was, however, a 25 bp deletion in the 3′-UTR of the cDNA compared with the genes. This deletion is presumably a consequence of a cDNA amplification artefact, alternative splicing during transcriptional termination, or the existence of an alternative wrLHβ allele. Other features of the two genes, including relative locations of TATA elements, polyadenylation signals, and intron–exon structure in the region of the first intron, were

FIGURE 2.   (A) White rhinoceros LHβ1 and LHβ2 gene amplification and sequencing strategies. Triangles represent primers used for PCR or sequencing. Lines extending from triangles depict full-length PCR-amplified fragments generated from pairs of primers (above gene diagram) or sequence obtained using a given primer (below gene diagram). (B) Nucleotide and deduced amino acid sequences of two wrLHβ genes (LHβ1 and LHβ2). The sequence of wrLHβ1 reflects consensus LHβ gene features, whereas wrLHβ2 lacks the second intron (dots within boxed area) but otherwise shares perfect sequence identity with wrLHβ1. Numbers directly above the amino acid sequence indicate codon position relative to the first residue of the mature peptide. Numbers directly above the nucleotide sequence (first line) and in the right margin indicate nucleotide position relative to pituitary transcriptional initiation (right-angled arrow). Arrowheads denote base differences relative to pituitary-derived wrLHβ cDNA which contained adenine at both positions. Brackets indicate consensus TATA elements and polyadenylation signal sequences. Reverse-type designates homologous, but not necessarily wild-type, wrLHβ sequence corresponding to consensus ungulate amplification primers. GenBank accession numbers for wrLHβ1 and wrLHβ2 are AF024520 and AF024521 respectively.
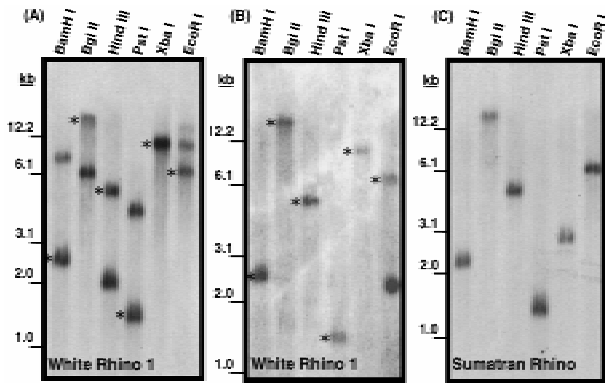
FIGURE 3.   Gene-specific Southern blots. The complex white rhino Southern banding pattern observed with full-length wrLHβ cDNA probe (A) was reduced to a single band-per-lane pattern with wrLHβ1-specific second intron probe (B). Asterisks indicate corresponding wrLHβ1 bands. The high degree of similarity between probe-positive restriction fragment lengths observed in the side-by-side comparison of the intron-hybridized blot (B) and the blot of a rhino species known to carry a single-copy LHβ gene (C; hybridized with the same probe) suggests that wrLHβ1 corresponds to the ancestral single-copy perissodactyl LHβ gene.
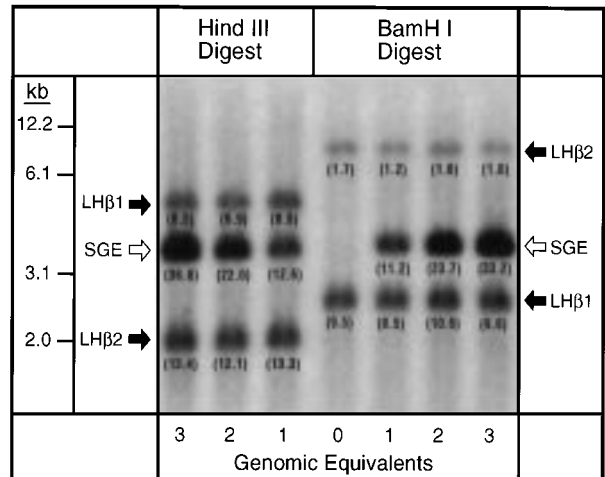

FIGURE 4.   Estimation of wrLHβ gene copy number by genomic titration analysis suggests the presence of only two genes per haploid genome. Genomic blots were hybridized with full-length wrLHβ cDNA probe. Open arrows indicate bands containing standard genomic equivalents (SGE; per haploid genome) of cloned full-length wrLHβ1 gene sequence, and filled arrows indicate bands containing wrLHβ genes of genomic origin. Values from densitometric quantification are indicated in parentheses under the bands.

consistent with consensus mammalian LHβ gene structure.

Second intron sequence was isolated, labeled, and used to probe white rhino genomic Southern blots to identify which of the bands in each lane corresponded to wrLHβ1 (Fig. 3). Use of this gene-specific intronic probe reduced the complex banding pattern to the more simple form observed for other rhino species. Side-by-side comparison of the wrLHβ1 band positions with those of the Sumatran rhino, which has a single-copy LHβ gene (Fig. 3B and C), reveals a high degree of conservation of restriction enzyme site locations within and flanking these two genes, indicating that they are homologous loci.

The preponderance of Southern blot lanes with only two bands for the white rhino (Fig. 1D and F) suggests the presence of only two genes; if there were three or more genes, there would be a high probability of observing more than two bands in several of the lanes. To obtain a more reliable estimate of gene-copy number, genomic titration assays were performed. Figure 4 shows the results of genomic titration Southern blot analyses in which lanes of genomic DNA digested with selected restriction enzymes were spiked with genomic equivalent standards of linearized wrLHβ1-containing plasmid. Sequence and mapping analyses (see Fig. 5 below) were used to select enzymes

(HindIII and BamHI) for genomic digestion such that cleavage would not occur within the span of genomic sequence to which the homologous cDNA probe would hybridize. Total optical density unit values of each of the two HindIII bands were closest to the value obtained for the single genomic equivalent standard band, indicating two LHβ gene copies per haploid genome. For the BamHI lanes, the total optical density unit value for the lower band (wrLHβ1) was closest to the single genomic equivalent standard, while the value for the upper genomic band (known to contain at least one full-length wrLHβ2 copy) anomalously indicated the presence of a small fraction of one genomic equivalent. However, this upper BamHI band consistently exhibits low intensity (see also Fig. 1D and F), presumably a consequence of microenviron-mental blot conditions (e.g. local background DNA milieu) that are not conducive to hybridization at that lane location. Nevertheless, because the closest non-zero whole integer standard value interpolated for this band is one genomic equivalent, the BamHI genomic titration data are most consistent with the interpretation that there are probably two wrLHβ genes. While these data are persuasive that there are more than one and fewer than four genes, the possibility of a third partial or full-length LHβ

FIGURE 5.   White rhino LHβ gene loci restriction-site mapping by Southern blot analysis. (A) Southern blots used to generate restriction maps (see Table 1 for estimated sizes of restriction fragments). Genomic blots were hybridized with full-length wrLHβ cDNA probe. (B) Deduced restriction-site maps. Restriction-site mapping data show that wrLHβ1 and wrLHβ2 reside on two non-overlapping ∼ 9 kb DNA fragments and are therefore not closely linked.

sequence cannot be entirely dismissed because of the semi-quantitative nature of this type of assay.

The extra-dark single band in the XbaI lane of the Southern blots in Fig. 1 (D and E) suggested that wrLHβ1 and wrLHβ2 may be closely linked within a single ∼ 9 kb restriction fragment, analogous to the tandem arrangement of LHβ and CGβ genes in primates. To explore this possibility further, genomic restriction-mapping studies were undertaken (Fig. 5 and Table 1). Comparisons among the three Southern blot lanes containing DNA digested with XbaI, EcoRI and XbaI+EcoRI (Fig. 5A, lanes 5, 7 and 6) ruled out the possibility

TABLE 1 Lengths (kb) of genomic restriction fragments mapped to specific white rhino LHβ genes

Length of restriction fragment (kb)

Gene	BamHI	HindIII	PstI	XbaI	EcoRI	XbaI + BamHI	PstI + BamHI	XbaI* + EcoRI	SacI	SacI + BamHI	SacI + XbaI	SacI + HindIII	XbaI + HindIII	BamHI + HindIII	PstI + HindIII	PstI + XbaI	SacI + PstI
LHβ1	2·49	5·37	1·40	8·92	6·72	2·55	1·13	5·88	1·02 0·69	0·99	0·98 0·67	1·00 0·69	3·68	1·38	1·43	1·30	0·64
LHβ2	7·63	2·04	4·20	8·92	8·88	2·76	0·87	6·36	2·39	0·99	2·28	1·86	1·96	1·10	1·53	3·98	1·92

*XbaI–EcoRI fragments cannot be definitively mapped to a specific LHβ gene.

that both gene copies were contained within the single XbaI fragment. This conclusion is based on the finding that the sum of the fragment lengths resulting from the XbaI+EcoRI co-digestion (5·88 kb+6·36 kb=12·24 kb) is significantly greater than the 8·92 kb length of the entire XbaI fragment (Table 1). The experiment was repeated with the same outcome, indicating that the two genes are contained within two different XbaI fragments that are coincidentally the same size. Reflecting this finding, restriction-site maps are proposed for each of two distinct wrLHβ-containing DNA fragments (Fig. 5B). Analysis of genomic blot restriction fragment length data (Fig. 5A and Table 1), in conjunction with gene-specific probe analysis findings (Fig. 3) and restriction-site location data derived directly from nucleotide sequencing data (Fig. 2B), revealed only one plausible linear order of restriction sites for the two LHβ loci (Fig. 5B).

The blots generated for restriction mapping provided further data supporting the contention that there are two, rather than three, genes. Sequence analysis of wrLHβ1 and wrLHβ2 genomic clones identified conserved PstI and BamHI sites within exon 1 and 3′-UTR respectively. Two bands of precisely the predicted sizes (differing only by the 289 bp intron length) and of similar signal intensity were detected in the lane where DNA was co-digested with PstI and BamHI (Fig. 5, lane 4). Because only two bands were observed in this lane, if there were a third gene-containing fragment, it would be co-migrating with one of the other two genes. This in turn would lead to an increase (approximate doubling) in the relative density of one band. However, the intensities of the two closely migrating bands in the BamHI+PstI lane were consistently observed to be comparable. Bands of similar intensity were also observed in each of the other two instances where bands within the same lane migrated close to each other (Fig. 5; lane 3, XbaI+BamHI and lane 6, XbaI+EcoRI).

DISCUSSION

Previous studies in our laboratory revealed that the haploid genome of the Indian rhino carries a single-copy LHβ gene (Sherman *et al*. 1997). Because this matched the equid and consensus mammalian conditions (Bousfield *et al*. 1994), it was not considered likely that more than a single LHβ gene would be found in other perissodactyl species. While this prediction was accurate for Baird's tapir and Sumatran and black rhinos, evidence for multiple gene copies was discovered for the white

rhino (Fig. 1), suggesting that LHβ locus structure, and possibly function, have diverged along a distinctive evolutionary pathway in this species. This is the first case of LGHβ gene replication reported in a mammalian species outside the order Primates.

Although results from any one gene-copy number analysis may be questionable, evidence supporting the same estimated value was derived from several semi-quantitative experiments (Figs. 1, 4 and 5), thereby providing a greater measure of confidence that the estimate of two wrLHβ gene copies is correct. Despite the preponderance of evidence pointing to two genes, however, the existence of other partial or full-length wrLHβ gene sequence elements, or alternative allelic forms, cannot be entirely ruled out. Nevertheless, the assumption is made for the purposes of the following discussion that there are two LHβ genes, LHβ1 and LHβ2, per haploid genome in the white rhino.

Two lines of evidence indicate that gene duplication was a relatively recent mutational event. First, because only single LHβ genes are present in the other three rhino species examined, duplication did not occur until after divergence of the white rhino (genus *Ceratotherium*) from other rhino lineages. Indeed, morphological and molecular analysis of the family Rhinocerotidae indicate that *Ceratotherium* was the most recent genus to diverge (7–2·5 million years ago) during the radiation of extant rhinos which began an estimated 22 million years ago (Prothero *et al*. 1989, Morales & Melnick 1994). Second, wrLHβ1 and wrLHβ2 exhibit identical coding sequences. A high degree of nucleotide sequence similarity is consistent with relatively recent divergence from a common ancestral gene. However, conservation of coding sequence can also result from concerted evolution, wherein there is a selective advantage for the co-expression of two or more homologous genes (Lewin 1997*b*). Either or both circumstance(s) may be operational in this case.

The consensus second intron was present in wrLHβ1 but absent in wrLHβ2 (Fig. 2). The absence of introns is a defining characteristic of a class of genomic repetitive elements called processed pseudogenes. They arise as a consequence of unintended re-integration into the eucaryotic genome of a cDNA reverse-transcribed from a host cell mRNA during a retroviral infection (Lewin 1997*c*). Because mRNA processing is the only known mechanism by which precise excision of introns occurs (Lewin 1997*c*), it is concluded that wrLHβ2 is a processed gene, with the qualification that a partially spliced ancestral

wrLHβ1 mRNA served as template for reverse transcription.

The two wrLHβ loci would not be expected to be in close proximity because processed pseudogenes integrate randomly into the genome. Consistent with this expectation, genomic restriction-mapping analyses indicate that wrLHβ1 and wrLHβ2 reside in distinct non-overlapping ~ 9 kb XbaI fragments (Fig. 5). This proposed genesis and structure of white rhino LHβ loci contrasts with the primate LHβ–CGβ gene cluster-locus wherein the repeated members of the gene family are tandemly linked (Talmadge *et al*. 1983, Policastro *et al*. 1983, Graham *et al*. 1987, Jameson & Lindell 1988), having evolved by a different replicative process involving unequal crossing over of adjacent DNA regions (Stark & Wahl 1995).

A mutation must be introduced into the germline in order to be inherited. Therefore, the transcriptional event creating wrLHβ1 mRNA, and hence wrLHβ2, must have occurred within a gamete or early conceptus. While there appears to be no evidence in extant species for CGβ synthesis in gametes, expression of CGβ and LH/CGβ genes by conceptus trophoblast cells is well documented in primates and equids respectively (Baird *et al*. 1991, Woodward *et al*. 1993, Hoppen 1994). Furthermore, the inner cell mass, which gives rise to germline cells, is supplemented with cells dividing from the trophoblast layer before germline differentiation (Pederson *et al*. 1986, Fleming 1987). Such an exchange thus constitutes a potential mechanism for CGβ-producing trophoblast layer cells to mediate transfer of mRNA and/or cDNA (i.e. processed gene template) into the germline genome, a process analogous to that used in the production of stable transgenic mice using contemporary stem cell methodologies.

The consensus ungulate LHβ PCR primer used to amplify wrLHβ2 corresponds to sequence upstream of the TATA box, which in turn suggests that an LHβ-like promoter sequence in wrLHβ2 extends an undetermined distance upstream of the TATA element. We have previously demonstrated that pituitary transcription of LHβ genes in the white and black rhino initiates 26 bp downstream (nucleotide +1 in Fig. 2B) of the consensus LHβ TATA box (Sherman *et al*. 1997). Therefore, the homologous sequence observed between nucleotides -1 and -57 (Fig. 2B) in wrLHβ2 would not be included in cDNA derived from an mRNA transcribed under the direction of either the consensus TATA-regulated pituitary LHβ promoter or a TATA-regulated common placental–pituitary equid LH/CGβ-like promoter. Taken together, the data support the contention that

transcription of the wrLHβ1 mRNA that served as template for wrLHβ2 synthesis was driven by an alternative upstream promoter, thereby generating an mRNA with a lengthened TATA-containing 5′-UTR.

What is the significance of a secondary upstream promoter site? One possible scenario is that the same retroviral infection that introduced the enzyme reverse transcriptase also induced, directly or indirectly, transcription of the wrLHβ1 gene from an anomalous upstream initiation site. Another possibility is that there were cryptic or promiscuous (i.e. non-physiological) upstream promoters that occasionally drove wrLHβ1 transcription. Finally, wrLHβ1 transcription could have been directed by a second physiologically relevant promoter located upstream of the consensus TATA-regulated promoter. Among these options, the notion of a physiological upstream promoter is particularly noteworthy because of its resemblance to the transcriptional mechanism employed by primate CGβ genes expressed in trophoblast cells of the early conceptus and placenta. In humans, CGβ gene transcription is driven by a TATA-less upstream promoter which transcribes a 350 bp long 5′-UTR that includes a TATA element located in a position homologous to the functional TATA box of consensus LHβ genes. Thus full-length reverse transcripts of human CGβ mRNA would include TATA-containing LHβ-like proximal promoter sequence. By analogy, the existence of an alternative upstream wrLHβ1 promoter would account for the presence of TATA-containing pituitary proximal promoter sequence in the 5′-UTR of wrLHβ1 mRNA, and hence also in the corresponding wrLHβ2 processed gene. It is not known how much further upstream of position -57 wrLHβ1-like flanking sequence may be retained in wrLHβ2. Evidence that promoter homology may extend as far as the corresponding human CGβ transcriptional start site is the finding that both white rhino genes contain similarly positioned HindIII and SacI sites at approximately -400 (Fig. 5). Overall, it can be deduced that at some point in the evolutionary past an upstream promoter directed synthesis of wrLHβ1 mRNA in germline tissue, and that reverse transcription of an incompletely spliced form of this transcript produced a cDNA that re-integrated randomly into the genome, thereby creating a novel wrLHβ2 locus. It is less certain in which germline cell type(s) (gamete or early conceptus cells) upstream wrLHβ1 transcriptional initiation occurred and whether upstream transcriptional initiation was physiological or anomalous.

Although characterization of additional promoter sequence was beyond the scope of this study, it will

be of interest to determine how much wrLHβ1 promoter sequence transmitted to wrLHβ2. Transient transfection assays using primary cultures of pituitary cells have shown that as little as 75 bp of rat LHβ 5′-flanking sequence was sufficient to support 38% of the transcriptional activity exhibited by a 1·7 kb rat LHβ promoter (Kim *et al.* 1990). If the same is true for wrLHβ1, then only a relatively small amount of 5′-flanking sequence carried along during the creation of wrLHβ2 would be sufficient to constitute a partially or even fully active pituitary promoter.

Re-integrated cDNA elements are typically transcriptionally and/or translationally incompetent (i.e. pseudogenes) as a consequence of the absence of promoter sequence and/or the presence of readily identifiable disruptions of regulatory or coding nucleotide sequence structure (Lewin 1997*c*). As revealed in the present studies, however, an undetermined length of intact proximal promoter is present in wrLHβ2, and no other structural features known to induce transcriptional or translational dysfunction have been identified. Indeed, wrLHβ2 processed gene structure is consistent with what would be predicted to constitute a functional artificial transgene construct. Hence, use of the term pseudogene to describe wrLHβ2 would be presumptuous. Accordingly, we propose the less qualified term 'processed gene' to acknowledge the unconfirmed functional status of wrLHβ2, and 'processed eugene' to designate a putative re-integrated cDNA element shown to be functional. While evidence to date indicates that wrLHβ2 may be a processed eugene, evaluation of this possibility must await experimental confirmation.

The data we present herein necessitate consideration of several options regarding extant and ancestral structure–function relationships for wrLGHβ genes (summarized in Fig. 6). The existence of a potentially active wrLHβ2 leads to three possibilities with regard to contemporary pituitary expression patterns: (1) both wrLHβ1 and wrLHβ2 are functional, (2) wrLHβ1 is functional and wrLHβ2 is a processed pseudogene, or (3) wrLHβ2 is functional and wrLHβ1 has become a pseudogene. In addition, wrLHβ1 may at present be expressed in both pituitary and germline (wrLH/CGβ), germline only (wrCGβ), or not at all (pseudogene). With regard to ancestral promoter function, it can be concluded that wrLHβ1 was transcriptionally active in the pituitary because this homologue corresponds to the consensus mammalian single-copy LHβ gene. It is also known that wrLHβ1 exhibited ancestral germline transcriptional activity utilizing an alternative upstream
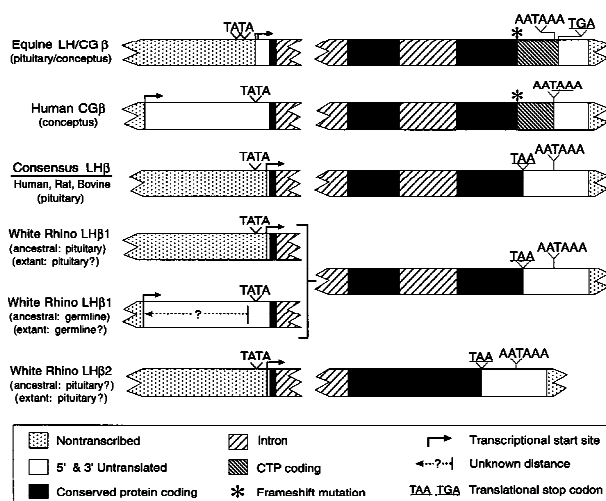

FIGURE 6.   Diagram summarizing and comparing structure–function relationships of the two wrLHβ genes and LGHβ genes of other mammals. For equine, human and consensus mammalian LGHβ genes, established sites of tissue-specific expression are shown in parentheses beneath the construct label. For wrLHβ gene/promoter combinations, tissue-specific expression is further qualified in parentheses as follows: ancestral, evolutionary past; extant, present-day; ?, uncertain promoter activity for the tissue/time-frame combination indicated. See the Discussion for further details.

promoter, and that one such transcript served as template during the genesis of wrLHβ2. Finally, ancestral wrLHβ2 pituitary expression (i.e. transcriptional activation immediately after stable integration into the genome) cannot be ruled out. Neither extant nor ancestral germline expression of wrLHβ2 is postulated because of the low probability of random re-insertion downstream of sequence capable of directing germline-specific transcription.

In conclusion, the presence of two distinct and potentially functional LHβ loci in the white rhinoceros, but not other perissodactyl species, indicates recent evolutionary divergence of wrLHβ gene structure, and possibly function, relative to the consensus mammalian condition. Inclusion of homologous pituitary proximal promoter sequence in the conserved processed wrLHβ2 gene demonstrates that the ancestral wrLHβ1 gene was expressed from an alternative upstream promoter site in the gamete or early conceptus (i.e. germline cells). While germline expression of wrLHβ1 from an upstream transcriptional start site could have been spurious, it is also possible that this alternative site was (and perhaps still is) a physiologically relevant conceptus-specific promoter that functions

in a manner analogous to hCGβ genes. It is anticipated that the findings presented in this report will provide a basis for novel comparative analyses of mammalian LHβ, CGβ and LH/CGβ gene structure and promoter function. Furthermore, it is hoped that these studies will underpin efforts to develop a better understanding of the reproductive endocrinology of the many critically endangered species within the order Perissodactyla.

REFERENCES

Baird DD, Weinberg CR, Wilcox AJ, McConnaughey DR, Musey OI & Collins DC 1991 Hormonal profiles of natural conception cycles ending in early unrecognized pregnancy loss. *Journal of Clinical Endocrinology and Metabolism* **72** 793–800.

Bo M & Boime I 1992 Identification of the transcriptionally active genes of the chorionic gonadotropin gene cluster *in vivo*. *Journal of Biological Chemistry* **267** 3172–3184.

Bousfield GR, Perry WM & Ward DN 1994 Gonadotropins: chemistry and biosynthesis. In *The Physiology of Reproduction,* edn 2, pp 1749–1792. Eds E Knobil & JD Neill. New York: Raven Press.

Brown T 1995 Analysis of DNA sequences by blotting and hybridization. In *Current Protocols in Molecular Biology,* pp 2·9·1–2·10·16. Eds FM Ausubel, R Brent, RE Kingston,

DD Moore, JG Seidman, JA Smith & K Struhl. New York: John Wiley & Sons.

Crawford RJ, Tregear GW & Niall HD 1986 The nucleotide sequences of baboon chorionic gonadotropin β-subunit genes have diverged from the human. *Gene* **46** 161–169.

Fleming TP 1987 Quantitative analysis of cell allocation to trophectoderm and inner cell mass in the mouse embryo. *Developmental Biology* **119** 520–531.

Graham MY, Otani T, Boime I, Olson MV, Carle GF & Chaplin DD 1987 Cosmid mapping of the human chorionic gonadotropin subunit genes by field-inversion gel electrophoresis. *Nucleic Acids Research* **15** 4437–4448.

Gwenda JM, Bambra CS & Tarara RP 1990 Studies with human, baboon, and vervet monkey chorionic gonadotrophins. *Journal of Medical Primatology* **19** 715–724.

Hearn JP, Webley GE & Gidley-Baird AA 1991 Chorionic gonadotrophin and embryo–maternal recognition during the peri-implantation period in primates. *Journal of Reproduction and Fertility* **92** 497–509.

Hoppen HO 1994 The equine placenta and equine chorionic gonadotrophin: an overview. *Experimental and Clinical Endocrinology* **102** 235–243.

Jameson JL & Lindell CM 1988 Isolation and characterization of the human chorionic gonadotropin beta subunit (CG beta) gene cluster: regulation of transcriptionally active CG beta gene by cyclic AMP. *Molecular and Cellular Biology* **8** 5100–5107.

Kim KE, Day KH, Howard P, Salton SRJ, Roberts JL & Maurer RA 1990 DNA sequences required for expression of the LHβ promoter in primary cultures of rat pituitary cells. *Molecular and Cellular Endocrinology* **74** 101–107.

Lewin B 1997*a* Genomes. In *Genes,* edn 6, p 646. Ed B Lewin. New York: Oxford University Press.

Lewin B 1997*b* Gene numbers. In *Genes,* edn 6, pp 708–709. Ed B Lewin. New York: Oxford University Press.

Lewin B 1997*c* Retroviruses and retroposons. In *Genes,* edn 6, pp 615–616. Ed B Lewin. New York: Oxford University Press.

Morales JC & Melnick DJ 1994 Molecular systematics of the living rhinoceros. *Molecular Phylogenetics and Evolution* **3** 128–134.

Murphy BD & Martinuk SD 1991 Equine chorionic gonadotropin. *Endocrine Reviews* **12** 27–44.

Pederson RA, Wu K & Batakier H 1986 Origin of the inner cell mass in mouse embryos: cell lineage analysis by microinjection. *Developmental Biology* **117** 581–595.

Pierce JG & Parsons TF 1981 Glycoprotein hormones: structure and function. *Annual Review of Biochemistry* **50** 465–495.

Policastro P, Ovitt CE, Hoshina M, Fukuoka H, Boothby MR & Boime I 1983 The β-subunit of human chorionic gonadotropin is encoded by multiple genes. *Journal of Biological Chemistry* **258** 11492–11499.

Prothero DR, Guerin C & Manning E 1989 The history of the rhinocerotoidea. In *The Evolution of Perissodactyls,* p 336. Eds DR Prothero & RM Schoch. New York: Clarendon Press.

Seshagiri PB & Hearn JP 1993 *In vitro* development of *in vivo* produced rhesus monkey morulae and blastocysts to hatched, attached, and post-attached blastocyst stages: morphology and early secretion of chorionic gonadotrophin. *Human Reproduction* **8** 279–287.

Sherman GB, Wolfe MW, Farmerie TA, Clay CM, Threadgill DS, Sharp DC & Nilson JH 1992 A single gene encodes the β-subunits of equine luteinizing hormone and chorionic gonadotropin. *Molecular Endocrinology* **6** 951–959.

Sherman GB, Lund LA, Bunick D & Winn RJ 1997 Characterization and phylogenetic significance of rhinoceros

luteinizing hormone beta (LHβ) subunit mRNA structure, cDNA sequence, and gene copy number. *Gene* **195** 131–139.

Stark GR & Wahl GM 1995 Gene amplification. *Annual Review of Biochemistry* **53** 447–491.

Steinetz BG, Randolph C & Mahoney CJ 1992 Serum concentrations of relaxin, chorionic gonadotropin, estradiol-17β, and progesterone during the reproductive cycle of the chimpanzee (*Pan troglodytes*). *Endocrinology* **130** 3601–3607.

Strauss WM 1995 Preparation of genomic DNA from mammalian tissue. In *Current Protocols in Molecular Biology,* pp 2·2·1–2·2·3. Eds FM Ausubel, R Brent, RE Kingston, DD Moore, JG Seidman, JA Smith, & K Struhl. New York: John Wiley & Sons.

Summers PM, Taylor CT & Miller MW 1993 Requirement of inner cell mass for efficient chorionic gonadotrophin

secretion by blastocysts of common marmosets (*Callithrix jacchus*). *Journal of Reproduction and Fertility* **97** 321–327.

Talmadge K, Boorstein WR & Fiddes JC 1983 The human genome contains seven genes for the β-subunit of chorionic gonadotropin but only one gene for the β-subunit of luteinizing hormone. *DNA* **2** 281–289.

Talmadge K, Vamvakopoulos NC & Fiddes JC 1984 Evolution of the genes for the β-subunits of human chorionic gonadotropin and luteinizing hormone. *Nature* **307** 37–40.

Woodward BJ, Lenton EA & Turner K 1993 Human chorionic gonadotrophin: embryonic secretion is a time-dependent phenomenon. *Human Reproduction* **8** 1463–1468.