The evolution and genomic landscape of CGB1 and CGB2 genes

Pille Hallast a, Kristiina Rull a,b, Maris Laan a,∗

a Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Riia 23, 51010 Tartu, Estonia
b Department of Obstetrics and Gynecology, University of Tartu, Estonia

Received 12 August 2005; accepted 28 November 2005

Abstract

The origin of completely novel proteins is a significant question in evolution. The luteinizing hormone (LHB)/chorionic gonadotropin (CGB) gene cluster in humans contains a candidate example of this process. Two genes in this cluster (CGB1 and CGB2) exhibit nucleotide sequence similarity with the other LHB/CGB genes, but as a result of frameshifting are predicted to encode a completely novel protein. Our analysis of these genes from humans and related primates indicates a recent origin in the lineage specific to humans and African great apes. While the function of these genes is not yet known, they are strongly conserved between human and chimpanzee and exhibit three-fold lower diversity than LHB across human populations with no mutations that would disrupt the coding sequence. The 5′-upstream region of CGB1/2 contains most of the promoter sequence of hCGβ plus a novel region proximal to the putative transcription start site. In silico prediction of putative transcription factor binding sites supports the hypothesis that CGB1 and CGB2 gene products are expressed in, and may contribute to, implantation and placental development.

© 2006 Elsevier Ireland Ltd. Open access under CC BY license.

Keywords: Chorionic gonadotropin beta 1 and 2; Gene evolution; Polymorphism patterns; Putative promoter region; In silico transcription factor binding site prediction

1. Introduction

The human luteinizing hormone/chorionic gonadotropin beta (LHB/CGB) gene cluster on chromosome 19q13.3 consists of one LHB gene and six CGB genes (Fiddes and Goodman, 1980; Talmadge et al., 1984a; Maston and Ruvolo, 2002; Fig. 1A). These seven genes are highly conserved at the nucleotide level (85–99% DNA sequence identity) and appear to have originated from an ancestral LHB gene as a result of duplication during primate evolution. Four of the genes (CGB, CGB5, CGB7 and CGB8) encode the beta subunit of human chorionic gonadotropin, a 163 amino acid protein that is produced by the implanting conceptus and is essential for alternations to the maternal reproductive system in support of pregnancy. The other CGB genes, CGB1 and CGB2, encode a hypothetical protein of 132 amino acids that is completely different from the hCGβ-subunit and lacks similarity to any known protein (Bo and Boime, 1992). These genes appear to have evolved by insertion of a DNA fragment (736 bp for CGB1, 724 bp for CGB2) that replaces 52 bp of the proximal end of the promoter and the entire 5′-UTR of an ancestral hCGβ-subunit coding gene (Bo and Boime, 1992; Hollenberg et al., 1994; Fig. 1B). This insertion creates a CGB1/CGB2 specific putative promoter fragment, an alternative 5′untranslated region (5′-UTR) and a novel exon 1, leading to a one basepair frameshift in the open reading frame (ORF) for exons 2 and 3.

Although a protein product corresponding to CGB1 and CGB2 has not yet been isolated, mRNA from these genes has been detected in the placenta (Bo and Boime, 1992; Rull and Laan, 2005) as well in the testis (Berger et al., 1994), pituitary (Dirnhofer et al., 1996), and in breast cancer tissue (Giovagnardi et al., 2001). The repeated observations of expression suggest that these genes are functional. In transgenic mice carrying a 36-kb cosmid insert with all the six CGB genes, the CGB1 and CGB2 transcripts were also observed in brain at levels comparable with placenta, the expression site for all the CGB genes (Strauss et al., 1993).

As the next step toward understanding the evolution and functional relevance of CGB1 and CGB2 we sequenced and analyzed the genes from three human populations as well as from the closest living relatives of humans. As a reference for considering relative conservation of CGB1 and CGB2 we used LHB, the founding member of this gene cluster and a gene that has a well-established, essential and conserved function in mam-
Fig. 1. Genomic context of CGB1 and CGB2. (A) Schematic presentation of the structure of the LHB/CGB gene cluster (covering 39.76 kb from LHB to CGB7) drawn to an approximate scale. Individual LHB/CGB genes (white boxes) cover 1.11–1.466 kb. Arrows indicate the direction of transcription either from a sense or an antisense strand. Experimentally identified hCG/H9252 drawn to an approximate scale. Individual sequence are indicated with (*) and based on CGB1/CGB1 arrows above the aligned sequences of CGB1 does the variation pattern of human CGB8, CGB2 anti-sense strand. Experimentally identified hCG/H9252 putative CGB1/2 protein with no amino acid similarity to hCG 2003). *CCAAT box has been identified by Matinspector and Alibaba TFBS prediction softwares. (D) Prediction of transcription factor binding sites (TFBS) onto hCG black brackets), trophoblast-specific element TSE from −305 to −279 (Steger et al., 1993; dotted brackets). Other experimentally proven regulatory elements of hCG promoter include activating protein 2 (AP2) and selective promoter factor 1 (Sp1) (Johnson and Jameson, 1999) as well as Ets-2 binding sites (Ghosh et al., 2003). *CCAAT box has been identified by Matinspector and Alibaba TFBS prediction softwares. (D) Prediction of transcription factor binding sites (TFBS) onto the 5′-upstream segment unique to CGB1 and CGB2 created by the insertion (B). TFBSs predicted by both MatInspector and Alibaba methods are marked with solid arrows above the aligned sequences of CGB1 and CGB2; TFBSs recognized by MatInspector alone are marked by broken arrows. TFBSs predicted solely based on CGB1 sequence are indicated with (*) and based on CGB2 (**). ATF: activating transcription factor; AP2: activating protein 2; Cdx2: Caudal-related transcription factor; CREB: cAMP responsive element binding protein; ERE: Estrogen response element; HIF: Hypoxia-inducible factor 1; NFkappaB: nuclear factor κB; GATA2: GATA-biding protein 2; SF1: steroidogenic factor 1; Sp1: selective promoter factor 1. Transcription start site has been indicated based on NCBI GenBank locus no NG_000019 information.

2. Materials and methods

2.1. Experimental subjects

The study was approved by the Ethics Committee of the University of Tartu, Estonia (protocol no. 117/9, 16 June 2003). CGB1 and CGB2 genes were resequenced for 47 Estonian (Europe), 23 Mandenka (Africa) and 25 Chinese Han (Asia) individuals. Estonian DNA samples originate from the DNA bank of Department of Biotechnology, IMCB, University of Tartu. Mandenka and Han DNA samples were obtained from HGDP-CEPH Human Genome Diversity Cell Line Panel (http://www.ceph.fr/HGDP-CEPH-Panel/). Common chimpanzee (Pan troglodytes) DNA was extracted from sperm material obtained from Tallinn Zoo, Estonia; sources of orangutan (Pongo pygmaeus) and gorilla (Gorilla gorilla) DNAs were primary cell lines AG12256 and AG05251B, purchased from ECACC.

2.2. Sequencing of human and great ape CGB1 and CGB2 genes

The structure of the CGB1/CGB2 (#MIM, 608823; 608824) genomic region has been reconstructed by web-based global alignment (http://www.ebi.ac.uk/clustalw/;CLUSTALW) and BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) tools. For the analysis we used the sequence obtained from NCBI GenBank database (http://www.ncbi.nlm.nih.gov; locus no NG_000019; 26 June 2002 release).

PCR and sequencing primers for CGB1, CGB2 and a reference gene LHB were designed based on human sequence using the web-based version of the

malian reproduction. We used the resulting data to explore the following questions: (i) what is the origin of CGB1 and CGB2? (ii) how conserved are CGB1 and CGB2 among primates? (iii) does the variation pattern of human CGB1 and CGB2 support constraints on variation consistent with functionality? (iv) does the upstream region of human CGB1/2 have expected features of a functional promoter and what transcription factor binding sites are present that could direct expression of these genes to specific tissues?
Additional PCR primers used to amplify CGB1 and CGB2 of great apes

**CGB1**
- Forest-rich in-f: TCTCAAGGAGATGGACGAGTG
- Forest-rich in-R: CAGAAGAGTCTACACGGTACGA
- Forest-rich R: ATCCCAGTAATTTTCTGTATTTTA

**CGB2**
- Forest-rich in-f: AGGAGAGGCTTACACGGTACGA
- Forest-rich R: CACGTACCTCCTGTACCTTCTT

**LHB**
- LHB-f: ATAGCAGTACACGGTGCTAC
- LHB-R: GTGAGTTGTCCAGGTTACCC

Additional sequencing primers for great apes

**CGB1**
- Chimpr-CGB1-3R2-R: GAAATGTGATCCACGTATCTACCT
- Chimpr-CGB1-1F1R-F: ATCAAGGAGATGGACGAGTG

**CGB2**
- Gor-CGB2-2F2-R: GTCCACGCAAACGTCTCTTT
- Gor-CGB2-2R2-R: CTGTGAAAGCCTGGTGAGA

**LHB**
- Chimpr-LHB-1F1R-R: AGCTGAGACCTACCTTTAGT
- Gor-LHB-1F1R-F: CACCACGATCTTACACGTACGA
- Orang-LHB-1R2-R: CTGACGTGCTACACGGTGCTAC

2.3. Sequence diversity parameters and neutrality tests

Sequence diversity parameters were calculated by DnaSP software (Version 4.0) (Rozen and Rosas, 1999). The direct estimate of per-site heterozygosity (π) was derived from the average pairwise sequence diversity, and Watterson’s θ (Watterson, 1975) represents an estimate of the expected per-site heterozygosity based on the number of segregating sites (S). Tajima’s D (Tajima, 1989) statistic (Tajima, 1989) was performed to determine if the observed patterns of human CGB1 and CGB2 were consistent with the standard neutral model. The basis of θ and D values is the difference between the π and θ estimates; under neutral expectation π = θ = D = 0. Significant positive D values indicate an excess of intermediate-frequency alleles in a population consistent with either balancing selection or population bottleneck, whereas significant negative D values indicate an excess of rare SNPs consistent with either recent directional selection or an increase in population size.

A simple neutral model (Kimura, 1983) predicts that drift and mutation rate determine the level of nucleotide variation accumulating within and between species. Therefore, the relative amount of within-species polymorphism should reflect the amount of between-species fixation under neutrality. Genetic diversity of human CGB1 and CGB2 was compared with fixation between human and chimpanzee, as well as human and gorilla sequences to test neutrality. We applied the Hudson, Kreitman and Aguade (HKA) test (Hudson et al., 1987) to estimate whether there was a significant difference in the ratio of polymorphism to divergence of across CGB1 and CGB2 using LHB as a reference locus.

2.4. In silico prediction of TFBS to human CGB1 and CGB2 5′-upstream region

Prediction of transcription factor binding sites (TFBS) was performed using the MatInspector 2.2 (http://www.genomatix.de/products/MatInspector; Cartharius et al., 2005) and Alabiba 2.1 (http://www.gene-regulation.com/pub/programs/alabiba2/index.html; Grabe, 2002) programs. Both approaches
3. Results and discussion

3.1. CGB1 and CGB2 have possibly arisen in the common ancestor of African great apes

First, we addressed the question of conservation of CGB1 and CGB2 genes among the species. Human-specific primers were used to amplify a unique gene product of CGB1 for chimpanzee (primers CGB1_2F and CGB1_6R, predicted length based on human sequence 2312 bp; Genbank accession no. DQ238547) and of CGB2 for gorilla (CGB2_1F and CGB2_3R, 1812 bp; DQ238550). Chimpanzee CGB2 was inferred from the jointly amplified CGB1/2 products (primers CGB2_1F and CGB2_5R, 2269 bp; DQ238549) using the chimpanzee CGB1 sequence as a reference. With a similar approach, the gorilla CGB2 was used as a reference to derive CGB1 from a common CGB1/2 product amplified from gorilla DNA (primers CGB1_1F and CGB1_2R, 1600 bp; DQ238548). None of the human-specific CGB1 and CGB2 primer combinations were capable to amplify the expected products from orangutan genomic DNA. Therefore, either the orangutan CGB1 and CGB2 sequences are highly divergent from other studied primates; or this species lacks CGB1/2 insertion region (target of one of the primers), and consequently CGB1/2 genes. The latter scenario is also supported by a recent study suggesting the total copy number of orangutan CGB genes to be four (Maston and Ruvolo, 2002). Consequently, we raise the hypothesis of the origin of CGB1 and CGB2 in the common ancestor of African great apes.

Amplification of the reference gene LHB, a functional ancestral member of the same gene cluster was successful with human-specific primers in all four studied species: human, chimpanzee (Genbank accession no. DQ238551), gorilla (DQ238552) and orangutan (DQ238553).

3.2. CGB1 and CGB2 are conserved between human and chimpanzee

Divergence of chimpanzee (C) and gorilla (G) CGB1, CGB2 and LHB from human (H) sequences (Table 2; H/C: across the genes 1.35–2.19%, exons 0.5–1.42%, introns 1.53–2.68; H/G: across the genes 1.44–3.00%, exons 1.42–4.04%, introns 1.36–3.31%) somewhat exceeds previous estimations. The average divergence across 53 autosomal intergenic regions has been reported 1.24 ± 0.07% for H/C and 1.62 ± 0.08% for H/G (Chen and Li, 2001). Human/chimpanzee comparison of

| Gene | Region | Length (bp) | Nucleotide diversity (%) |
|------|--------|-------------|--------------------------|
|      |        |             | H/C | H/G | H/O | C/G | C/O | G/O |
| LHB  | mRNA   | 1111        | 1.35 | 1.44 | 5.39 | 1.88 | 6.29 | 6.38 |
|      | Exons  | 423         | 1.42 | 1.42 | 4.02 | 2.84 | 5.91 | 5.20 |
|      | Introns| 588         | 1.53 | 1.36 | 6.12 | 1.19 | 6.47 | 6.81 |
|      | 5’-UTR | 9           | 0    | 0    | 11.11| 0    | 11.11| 11.11|
|      | 3’-UTR | 91          | 0    | 2.20 | 6.45 | 2.15 | 6.45 | 8.60 |
| CGB1 | mRNA   | 1366        | 2.19 | 2.34 | –    | 2.26 | –    | –    |
|      | Exons  | 396         | 0.50 | 1.51 | –    | 1.49 | –    | –    |
|      | Introns| 634         | 2.68 | 2.21 | –    | 2.20 | –    | –    |
|      | 5’-UTR | 174         | 3.40 | 4.60 | –    | 3.41 | –    | –    |
|      | 3’-UTR | 162         | 3.08 | 2.47 | –    | 3.08 | –    | –    |
| CGB2 | mRNA   | 1366        | 1.46 | 3.00 | –    | 3.80 | –    | –    |
|      | Exons  | 396         | 1.26 | 4.04 | –    | 5.30 | –    | –    |
|      | Introns| 634         | 1.89 | 3.31 | –    | 3.47 | –    | –    |
|      | 5’-UTR | 174         | 1.70 | 2.30 | –    | 3.97 | –    | –    |
|      | 3’-UTR | 162         | 0    | 1.23 | –    | 1.23 | –    | –    |

H: human; C: common chimpanzee; G: gorilla; O: orangutan.

*a Length based on human sequence.*
127 genes mapped to human chr. 21 resulted in estimates of overall divergence for coding sequences 0.75 ± 0.01% (range 0.53–2.05%), for exons 0.51% ± 0.02% (range 0.08–2.52%), for exon/intron junction 0.85 ± 0.02% (range 0.41–2.78%) for 5′-UTR 1.00% ± 0.10 and for 3′-UTR 0.93% ± 0.09 (Shi et al., 2003). Relatively high divergence (across the gene 5.39% compared to 3.08% reported for intergenic regions; Chen and Li, 2001) was also estimated between human and orangutan (O) for LHB including 11 non-synonymous changes. Higher interspecific divergence could result from the intraspecific gene conversion among highly homologous genes in the LHB/CGB cluster (Maston and Ruvolo, 2002; Hallast et al., 2005). For gorilla CGB2 gene approximately two fold higher sequence divergence for H/G compared to H/C largely arises from two gorilla-specific deletions (2 and 12 bp) increasing substantially the number of fixed nucleotide differences between species (Fig. 2; supplementary figure).

Divergence patterns between human and chimpanzee CGB1 and CGB2 resemble the reference gene LHB, characterized by higher conservation in exons compared to introns (Table 2). We identified only a few fixed differences among species causing non-synonymous changes in chimpanzee CGB1 (3), CGB2 (5) and LHB (1) relative to human sequence (Fig. 2; supplementary figure). None of the sequence differences alter the ORF nor create a preliminary stop-codon. The evolution of 5′- and 3′-UTR sequences is variable among the genes, from 0 differences to 3.4% divergence. In gorilla the overall number of non-synonymous changes is even higher for LHB (8) than for

![Fig. 2. SNP patterns and fixed differences between human and great apes in CGB1 (A), CGB2 (B) and LHB (C) genes. Human polymorphic positions (vertical black bars) are marked as long bars for common SNPs (minor allele frequency >10%) and short bars for rare SNPs (<10%). For human and great ape comparison fixed differences (black arrows), non-synonymous changes (black arrows with an asterisk), SNPs found in apes (black triangle) and protein altering insertions/deletions are shown (exclamation mark).]
CGB1 (1) and CGB2 (4). However, in gorilla we identified for both CGB1 (1 bp insertion in exon 2) and CGB2 (12 bp deletion at the beginning of exon 2 removing an Ala-Val-Ala-Ala motif) a change presumably leading to the disruption of a predicted protein (Fig. 2). Whether these represent consensus sequences for gorilla CGB1 and CGB2, or only mutations in the genome of the sequenced individual will be solved when additional gorilla sequences are available for comparison.

In summary, the interspecific analysis of CGB1 and CGB2 indicates that the level of conservation between human and chimpanzee is as high as for LHB, thus supporting the functional importance of these genes in these species. However, in gorilla the functionality of CGB1 and CGB2 is less likely as disrupted ORF was identified for CGB1 and a large deletion in exon 2 for CGB2.

3.3. Resequencing of human CGB1 and CGB2 genes revealed low variation and no nonsense mutations

As a next step we studied the polymorphism patterns of human CGB1 and CGB2 in comparison to LHB gene. Resequencing of total 190 chromosomes from three human populations (Estonians n = 94, Chinese Han n = 50 and Mandenka n = 46) identified 22 single nucleotide polymorphisms (SNPs) in CGB1 (ss48399944–ss48399963), 30 in CGB2 (ss48399964–ss48399977) and 24 in LHB (ss48399828–ss48399908) (Fig. 2; supplementary table). Interestingly, the LHB gene exhibited even three-fold higher variation than CGB1 and CGB2 (Table 3; average across populations: \( \pi_{CGB1}/kb = 1.39 \); \( \pi_{CGB2}/kb = 1.26 \)). Only one polymorphism in CGB1, five in CGB2 and four in LHB were identified leading to a non-synonymous change. None of the polymorphisms found in coding regions created a preliminary stop codon. Thus, the diversity patterns of human CGB1 and CGB2 comparable with a typical variation of human genes (African Americans \( \pi(74\ genes)/kb = 1.00 \); European Americans \( \pi(74\ genes)/kb = 0.80 \); Crawford et al., 2004) and rare non-synonymous substitutions give support to the functionality of these genes. Identification of ancestral alleles of human SNPs in comparison with other great ape sequences revealed that for most of the SNPs the major allele in human is also the ancestral variant (supplementary table).

We performed two alternative analyses to test whether the human LHB, CGB1 and CGB2 have evolved under standard neutral model. Tajima’s test examines whether the average number of pairwise nucleotide differences between sequences (\( \pi \)) is larger than expected from the observed number of polymorphic sites (\( \theta \)). The expected difference (Tajima D) between \( \pi \) and \( \theta \) is roughly zero under the standard neutral model. As differences between \( \pi \) and \( \theta \) for the studied genes were small and Tajima D values were close to zero (Table 3), the hypothesis of neutral evolution of these genes was not rejected. The HKA test was performed to test the neutral evolution of CGB1 and CGB2 among the studied species with LHB as a reference (Table 4). The test is based on prediction from the Neutral Theory of Molecular Evolution (Kimura, 1983) that the amount of within-species diversity should be correlated with levels of between-species divergence, due to the dependence of both on the neutral muta-

| Gene       | Pop   | mRNA | Exons | 3′-UTR | 5′-UTR | Ed | Md | All |
|------------|-------|------|-------|--------|--------|----|----|-----|
| LHB        | E     |       | 3.55  | 2.82   | 0.74   | 4.37| 3.65| 4.37|
|            | H     |       | 3.09  | 2.61   | 0.54   | 3.69| 3.58| 3.77|
|            | M     |       | 3.04  | 3.28   | 0.23   | 3.35| 3.77| 3.28|
| CGB1       | E     | All   | 3.92  | 2.31   | 0.13   | 3.62| 3.48| 3.48|
|            | H     |       | 1.08  | 1.00   | 0.18   | 0.82| 0.78| 0.82|
|            | M     |       | 1.31  | 1.96   | 0.98   | 0.19| 0.96| 0.19|
| CGB2       | E     | All   | 1.39  | 2.18   | 0.62   | 1.76| 1.76| 1.76|
|            | H     |       | 0.92  | 0.86   | 0.16   | 0.96| 0.96| 0.96|
|            | M     |       | 0.90  | 1.80   | −1.45  | 0.69| 0.69| 0.69|

\( a \) Estimate of nucleotide diversity per site from average pairwise difference among individuals.

\( b \) Estimate of nucleotide diversity per site from number of segregating sites (S).

\( c \) Value of Tajima’s \( D \) statistics and significance level: \( * \) \( p < 0.05; \) NA: not applicable.

\( d \) E: Estonians \( (n = 47); \) H: Chinese Han \( (n = 53); \) M: Mandenka \( (n = 23) \).
tion rate. Consistently with Tajima’s test, we could not reject the null hypothesis of neutral evolution. The exception was the Estonian sample set, where the significant result could be spurious outcome of the population history in Europe shaped by bottlenecks (Barbujani and Goldstein, 2004) and thus capturing the least the human intrapopulation variation applied in HKA test. However, as neither Tajima’s nor HKA test takes into account genes within a species (Maston and Ruvolo, 2002; Hallast et al., 2005), we should interpret the overall test results with caution and could not entirely exclude selection. Innan (2003) has shown that statistical tests of neutrality based on the standard coalescent theory for a single-copy gene may not be appropriate for duplicated genes.

3.4. CGB1 and CGB2 genes possess almost complete hCGβ promoter sequence

In order to predict the regulatory elements and patterns putatively involved in driving the expression of CGB1 and CGB2, we investigated the upstream regions of these genes. Alignment of the experimentally identified hCGβ promoter (−311 bp from hCGβ 5′-CAP; Otani et al., 1988; Hollenberg et al., 1994) with the 5′-upstream region of CGB1 and CGB2 genes revealed a more proximal location of an almost complete hCGβ promoter sequence (−757 to −481 for CGB1 and −745 to −469 for CGB2 from predicted transcription start site), lacking only 52 bp of proximal promoter segment of hCGβ (Fig. 1B and C). Despite the absence of the minimal promoter region (MPR; −37 to +104; Hollenberg et al., 1994) including two Ets-2 binding sites (Ghosh et al., 2003), the other sequence motifs playing a crucial role in regulating hCGβ expression are conserved among the genes (Fig. 1C). These include cAMP-dependent transcription element mapped to −311 to −202 (Albanese et al., 1991), trophoblast-specific element (TSE) between −305 and −279 maintaining basal expression (Steger et al., 1993), as well as binding sites for AP-2 and Sp1 transcription factors required for the full activity of the promoter (Johnson et al., 1997; Johnson and Jameson, 1999; Knöfler et al., 2004). Hollenberg et al. (1994) has suggested that the individual domains of the hCGB promoter act in an additive or combinatory manner. Thus, the absence of hCGβ MPR from CGB1/2 putative promoter region could possibly be compensated by the sequences within CGB1/2-specific insertion. However, whether this segment indeed has a regulatory function for CGB1/2 needs to be proven in wet-lab experiments.

3.5. CGB1- and CGB2-specific upstream region is predicted to harbor binding sites for transcription factors related to early placental development and implantation

In addition to alternative 5′-UTR and exon 1 for CGB1 and CGB2 genes, the CGB1/2-specific insert (736 and 724 bp, respectively) provides a novel putative proximal promoter segment (481 and 469 bp, respectively) upstream the transcription start site (Fig. 1B). We evaluated this fragment as a potential additional promoter component. In silico analysis predicted several regulatory elements in CGB1/2 insert experimentally determined to be essential for gonadotrope expression (Fig. 1D): two copies of CRE sites binding cAMP-responsive element binding protein (CREB) and activating transcription factor (ATF), binding sites for CG-β and Cβ transcription inducer AP-2 as well as for repressor c-Jun (Pestell et al., 1994; Johnson et al., 1997), binding sites for GATA-2 regulating CGβα-subunit expression in placenta (Steger et al., 1994). Interestingly, the CGB1/2 specific upstream region is predicted to harbor interaction sites for several transcription factors regulating implantation and placental development (Fig. 1D): NF-kB, Cdx-2, ERR-β, HIF1 and SF-1. Although NF-kB is a transcriptional factor involved mainly in inflammatory and immune responses, it regulates also several genes responsible for immunological adaptation at the feto-maternal interface and early embryonic development (Chen et al., 1999; Muggia et al.,
1999). Both, in human (Page et al., 2002) and mouse (Nakamura et al., 2004) NF-kB is activated in the pregnant uterus during preimplantation period and is highly expressed during the implantation window.

Cdx-2 and ERR-β exhibit highly specific expression pattern during embryogenesis. Besides its main role in driving embryonic axial elongation and anterior–posterior patterning, Cdx2 is also essential for trophoblastic development (Chawengsaksophak et al., 2004). Consistently, aberrant expression of bovine Cdx2 in the preimplantation cloned embryo has been reported to lead to the failure of implantation (Hall et al., 2005). ERE (reviewed by Gruber et al., 2004) is a binding site (consensus sequence 5′-GGTCAANNTGACC-3′) not only for the estrogen receptor ligand complex, but also for ERR-β, an orphan member of the superfamily of nuclear hormone receptors (Pettersson et al., 1996). Studies on mice have shown that ERR-β is expressed during embryogenesis by ectodermally derived regions of the amniotic fold, forming chorion. Homozygous mutant mouse embryos generated by targeted disruption of the Estrrb gene have severely impaired placental development, and die 10.5 days post-coitum (Luo et al., 1997).

Hypoxia-inducible factors (HIFs) mediating oxygen homeostasis have been suggested to regulate uterine vascular permeability and angiogenesis (Daikoku et al., 2003). Transcription factor SF-1 is a key regulator of the transcription of many genes involved in sexual differentiation, steroidogenesis and reproduction (GnRHR, α-GSUr, FSHB and LHB; reviewed by Parker and Schimmer, 1997).

In silico prediction of putative transcription factor binding sites allows to postulate the hypothesis of the involvement of CGB1 and CGB2 gene products in implantation and placental development. The hypothesis is supported by the detection of CGB1 and CGB2 transcripts in the placenta, although at much lower level compared to hCGβ coding genes (Bo and Boime, 1992; Rull and Laan, 2005).

4. Conclusions

This report aimed to explore the evolution, variation and putative regulatory regions of CGB1 and CGB2 in order to seek indirect evidence for the functionality of these genes, originally considered as pseudogenes (Talmadge et al., 1984b).

As both of the genes were amplified additionally to human also in chimpanzee and gorilla but not in orangutan, we suggest that they have arisen among the common ancestor of African great apes. Gene duplication was accompanied by the replacement of the hCGβ 5′-UTR with a non-coding sequence providing novel putative promoter segment, 5′-UTR and exon 1.

In human, CGB1 and CGB2 exhibit three times lower diversity than for LHB and no ORF disturbing mutations for a sample representing three continents. Both genes are conserved between human and chimpanzee, exhibiting the same level of interspecific divergence as LHB. Especially CGB1 stands out with a strong exonic conservation with only 0.5% divergence between human and chimpanzee, whereas the respective number for LHB is 1.42%. In contrast, for gorilla both CGB1 and CGB2 harbor insertion/deletion changes, which disrupt the predicted protein and thus there is little support for the functionality of these genes. We hypothesize that the fate of duplicated CGB1 and CGB2 genes has split for human–chimpanzee and gorilla lineages evolving towards a novel functional gene for the former and pseudogenization for the latter.

Upstream CGB1 and CGB2 is preserved almost full and well conserved (among genes) sequence of the promoter for hCGβ coding genes. Additionally, CGB1/2 possess a novel putative proximal promoter segment created by the CGB1/2-specific insertion. Analysis of this segment in silico for TFBSs highlighted several elements shown to regulate gene expression during implantation and placental development. However, as TFBS prediction programs can only infer the binding potential, and not the functionality of the site, only succeeding wet-lab experiments are able to uncover whether the predictions and postulated hypothesis hold true.

Acknowledgements

We thank Mrs. Liina Nagirnaja, Imbi Taniel and Tallinn Zoo for the chimpanzee DNA sample, Tõnu Margus for the help in DNA sequence assembly, Viljo Soo for the technical assistance and dr. Robert K. Campbell for the editing of the English language. The study has been supported by Wellcome Trust International Senior Research Fellowship in Biomedical Science in Central Europe (grant no. 070191/Z/03/Z for M.L.), Howard Hughes Medical Institute International Scholarship (grant #55005617 for M.L.), Estonian Science Foundation (grant no. 5796 for M.L.) and Estonian Ministry of Education and Science (Core grants no. 0182721s06 for M.L. and no. 0182641s04 for K.R.); as well as personal scholarships from the World Federation of Scientists’ (K.R. and P.H.) and Kristjan Jaak Stipend Program (travel to ICGR for K.R.); from Artur Lind Gene Technology Fund and Estonian-Revelia Harald Raudsepp Academic Fund (for P.H.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mce.2005.11.049.

References

Albanese, C., Kay, T.W., Troccoli, N.M., Jameson, J.L., 1991. Novel cyclic adenosine 3′,5′-monophosphate response element in the human chorionic gonadotropin beta-subunit gene. Mol. Endocrinol. 5, 693–702.
Barbujani, G., Goldstein, D.B., 2004. Africans and Asians abroad: genetic diversity in Europe. Annu. Rev. Genomics Hum. Genet. 5, 119–150.
Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploviz: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.
Bo, M., Boime, I., 1992. Identification of the transcriptionally active genes of the chorionic gonadotropin beta subunit gene in vivo. J. Biol. Chem. 267, 3179–3184.
Berger, P., Kranenwitter, W., Madersbacher, S., Gerth, R., Geley, S., Dinnhöfer, S., 1994. Eutopic production of human chorionic gonadotropin beta (hCG beta) and lutetising hormone beta (hLH beta) in the human testis. FEBS Lett. 343, 229–233.
Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., Werner, T., 2005. MatInspector and beyond:
promoter analysis based on transcription factor binding sites. Bioinformatics 21, 2933–2942.

Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J., Beck, F., 2004. Cdx2 is essential for axial elongation in mouse development. Proc. Natl. Acad. Sci. U.S.A. 101, 7641–7645.

Chen, F., Castranova, V., Shi, X., Demers, L.M., 1999. New insights into the role of nuclear factor-kappaB, a ubiquitous transcription factor in the initiation of diseases. Clin. Chem. 45, 7–17.

Chen, F.-C., Li, W.H., 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestors of humans and chimpanzees. Am. J. Hum. Genet. 68, 444–456.

Crawford, D.C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M.J., Nickerson, D.A., Stephens, M., 2004. Evidence for substantial fine-scale variation in recombination rates across the human genome. Nat. Genet. 36, 700–706.

Daikoku, T., Matsumoto, H., Gupta, R.A., Das, S.K., Gassmann, M., DuBois, R.N., Dey, S.K., 2003. Expression of hypoxia-inducible factors in the peri-implantation mouse uterus is regulated in a cell-specific and ovarian steroid hormone-dependent manner. Evidence for differential function of HIFs during early pregnancy. J. Biol. Chem. 278, 7683–7691.

Dimhofer, S., Hermann, M., Hittmair, K., Kapelari, K., Berger, P., 1996. Expression of the human chorionic gonadotropin-beta gene cluster in human placentas and alternate use of exons 1, 2. J. Clin. Endocrinol. Metab. 81, 4212–4217.

Fiddes, J.C., Goodman, H.M., 1980. The cDNA for the beta-subunit of human chorionic gonadotropin suggests evolution of a gene by readthrough into the 3′-untranslated region. Nature 286, 684–687.

Ghosh, D., Ezathi, S., Ostrowski, M.C., Roberts, R.M., 2003. A central role for Ets-2 in the transcriptional regulation and cyclic adenosine 5′-monophosphate responsiveness of the human chorionic gonadotropin-beta subunit gene. Mol. Endocrinol. 17, 11–26.

Giovagnardi, Y., Parfait, B., Asher, M., Olivi, M., Lidereau, R., Vidaud, M., Bieche, I., 2001. Analysis of the human CGB/LHB gene cluster in breast tumors by real-time quantitative RT-PCR assays. Cancer Lett. 168, 93–100.

Grabe, N., 2002. AiiBaba2: context specific identification of transcription factor binding sites. In Silico Biol. 2, S1–S15.

Gruber, C.J., Gruber, D.M., Gruber, I.M., Wieser, F., Huber, J.C., 2004. Anatomy of the estrogen response element. Trends Endocrinol. Metab. 15, 73–78.

Hallast, P., Nagirnaja, L., Margus, T., Laan, M., 2005. Segmental duplications and gene conversion: human luteinizing hormone/chorionic gonadotropin genes during the normal and failed pregnancy. Hum. Reprod. 20 (12), 3360–3368.

Rozas, J., Rozas, R., 1999. DnaSP Version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15, 174–175.

Shi, J., Xu, H., Wang, Y., Zhang, C., Jiang, Z., Zhang, K., Shen, Y., Jin, L., Zhang, K., Yuan, W., Wang, Y., Lin, J., Hua, Q., Wang, F., Xu, S., Ren, X., Su, X., Zhao, G., Chen, Z., Jin, L., Huang, W., 2003. Divergence of the genes on human chromosome 21 between human and other hominoids and variation of substitution rates among transcription units. Proc. Natl. Acad. Sci. U.S.A. 100, 8331–8336.

Stegger, D., Schrader, A., 1996. Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis. Mech. Dev. 54, 211–223.

Rozas, J., Rozas, R., 1999. DnaSP Version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15, 174–175.

Steeg, D.J., Buscher, M., Hecht, J.H., Mellon, P.L., 1993. Coordination of the alpha- and beta-subunit genes of human chorionic gonadotropin by trophoblast-specific element-binding protein. Mol. Endocrinol. 7, 1579–1588.

Steeg, D.J., Hecht, J.H., Mellon, P.L., 1994. GATA-binding proteins regulate the human chorionic gonadotropin alpha-subunit gene in the placenta and pituitary gland. Mol. Cell. Biol. 14, 5592–5602.

Stormo, G.D., 2000. DNA binding sites: representation and discovery. Bioinformatics 16, 16–23.

Strauss, B.L., Pettman, R., Pixley, M., Nilson, J.H., Boime, H.I., 1993. Expression of the beta subunit of chorionic gonadotropin in transgenic mice. J. Biol. Chem. 269, 4968–4973.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595.

Talmadge, K., Vamvakopoulos, N.C., Fiddes, J.C., 1984a. Evolution of the genes for the beta subunits of human chorionic gonadotropin and luteinizing hormone. Nature 307, 37–40.

Pettersson, K., Svensson, K., Mattsson, R., Carlsson, B., Ohlsson, R., Berkenstam, A., 1996. Expression of the estrogen response element. Trends Endocrinol. Metab. 15, 73–78.