Sussex Research Online

The chimpanzee GH locus: composition, organization, and evolution

Article (Accepted Version)

Pérez-Maya, Antonio A, Rodríguez-Sánchez, Iram P, de Jong, Pieter, Wallis, Michael and Barrera-Saldaña, Hugo A (2012) The chimpanzee GH locus: composition, organization, and evolution. Mammalian Genome, 23. pp. 387-398. ISSN 0938-8990

This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/71916/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

http://sro.sussex.ac.uk

The Chimpanzee *GH* locus*:* composition, organization and evolution.

Antonio A. Pérez-Maya[1], Irám P. Rodríguez-Sánchez[1], Pieter de Jong[2], Michael Wallis[3], Hugo A. Barrera-Saldaña[1*].

[1]Departamento de Bioquímica y Medicina Molecular, Facultad de Medicina, Universidad Autónoma de Nuevo León, Monterrey, Nuevo León 64460, México.

[2]Children's Hospital Oakland Research Institute, Oakland, California.

[3]Biochemistry Department, School of Life Sciences, University of Sussex, Falmer, Brighton, Sussex BN1 9QG, UK.

Keywords: Growth hormone, chorionic somatomammotropin, evolution, *Pan troglodytes*, hominids, comparative genomics, gene conversion.

Short title: Chimpanzee *GH* locus

*Correspondence:

Dr. Hugo A. Barrera-Saldaña.

Departamento de Bioquímica y Medicina Molecular.

Facultad de Medicina de la Universidad Autónoma de Nuevo León.

Av. Madero y Dr. Aguirre Pequeño, Col. Mitras Centro.

Monterrey, Nuevo León, México 64460.

Tel.: +52 81 83294173 x2591; fax: +52 81 83337747.

E-mail: hbarrera@fm.uanl.mx

Abstract

In most mammals the growth hormone (*GH*) locus comprises a single gene expressed primarily in the anterior pituitary gland. However, in higher primates multiple duplications of the *GH* gene gave rise to a complex locus containing several genes. In man this locus comprises 5 genes including *GH-N* (expressed in pituitary) and four genes expressed in the placenta, but in other species the number and organization of these genes varies. The situation in chimpanzee has been unclear, with suggestions of up to seven *GH*-like genes. We have re-examined the *GH* locus in chimpanzee and have deduced the complete sequence. The locus includes five genes apparently organized in a similar fashion to those in human, with two of these genes encoding GH-like proteins, and three encoding chorionic somatomammotropins/placental lactogens (CSHs/PLs). There are notable differences between the human and chimpanzee loci with regard to the expressed proteins, gene regulation and gene conversion events. In particular, one human gene (*hCSH-L*) has changed substantially since the chimpanzee/human split, potentially becoming a pseudogene, while the corresponding chimpanzee gene (*CSH-A1*) has been conserved, giving a product almost identical to the adjacent *CSH-A2*. Chimpanzee appears to produce two CSHs, with potentially differing biological properties, whereas human produces a single CSH. The pattern of gene conversion in human has been quite different from that in chimpanzee. The region around the *GH-N* gene in chimpanzee is remarkable polymorphic, unlike the corresponding region in human. The results shed new light on the complex evolution of the *GH*-locus in higher primates.

Introduction

Pituitary growth hormone (GH, somatotropin) is a protein hormone synthesized in and secreted from somatotroph cells of the anterior pituitary gland in all vertebrates. It regulates growth and has a variety of metabolic actions. In most mammals, including prosimians, the *GH* locus comprises a single gene. However, in higher primates a series of tandem gene duplications followed by divergent evolution has given rise to a complex gene locus in which the gene encoding pituitary GH, at the 5' end of the locus, is associated with a number of similar genes expressed in the placenta (Chen et al. 1989). The gene duplications that initiated the expansion of *GH*-like genes in higher primates appear to have occurred independently on lineages leading to new-world monkeys and to old-world monkeys and apes (Li et al. 2005; Papper et al. 2009; Wallis and Wallis 2002, 2006), and followed an episode of rapid evolution of the *GH* gene (Forsyth and Wallis 2002; Liu et al. 2001; Wallis 1996); subsequent diversification was rapid, resulting in variable numbers and organization of *GH*-like genes (González Alvarez et al. 2006; Revol de Mendoza et al. 2001, 2004; Rodríguez-Sánchez et al 2010; Ye et al. 2005,).

The best-characterized *GH* gene cluster is that of man (Chen et al. 1989). Here the locus comprises 5 genes in the same transcriptional orientation, with that encoding pituitary GH (*GH-N*) at the 5' end, followed by 4 genes that are expressed in the placenta. Two of these (*hCSH-A* and *hCSH-B*) encode chorionic somatomammotropin (hCSH; also known as placental lactogen, PL) which is expressed at very high levels during pregnancy. The mature proteins encoded by these genes are identical, though their signal peptides differ (Barrera-Saldaña et al., 1983; Chen etal. 1989). A third gene (*hCSH-L*) was initially considered a pseudogene (Reséndez-Pérez et al. 1990), but it is expressed in the placenta at low levels (Misra-Press et al. 1994; Männik et al. 2010). This gene is subject to an altered splicing pattern that suggests that the majority of the transcripts could encode a placental-lactogen-like protein, but that this would not be secreted owing to

lack of a functional signal peptide (Misra-Press et al. 1994). The fourth gene expressed in the placenta (*hGH-V*) is a variant form of GH that is expressed at moderate levels and appears to take over from pituitary GH during pregnancy (Frankenne et al. 1988; Lacroix et al. 2002). The rapid evolution of the genes in the *GH*-cluster during primate evolution thus involved not only gene duplication and divergence, but also changes in regulatory mechanisms so that these genes are expressed in either the pituitary or the syncytiotrophoblast (the outer fetal part of the placenta), although it should be noted that low-level expression in various other tissues also occurs.

The pattern of *GH*-like genes in other higher primates varies. In rhesus macaque (*Macaca mulatta*; Golos et al. 1993; González Alvarez et al. 2006) and baboon (*Papio hamadryas*; Rodríguez-Sánchez et al. 2010) there are 6 'genes', including at least one pseudogene, while gibbon (*Hylobates leucogenys*) has 7 *GH*-like genes and pseudogenes (Ye et al. 2005). The organization of such genes in new-world monkeys differs from that in old-world monkeys and the gene clusters seem to have arisen independently (Li et al. 2005; Papper et al. 2009; Wallis and Wallis 2002), with 8 genes and pseudogenes in marmoset (*Callithrix jacchus*; Wallis and Wallis 2002, 2006) and at least 40, mostly pseudogenes, in white fronted capuchin (*Cebus albifrons*; Wallis and Wallis 2006).

The organization of the *GH* gene cluster in chimpanzee (*Pan troglodytes*) has been rather unclear. Revol de Mendoza et al. (2004) identified 6 *GH*-like genes, while the draft chimpanzee genome (http://www.ensembl.org/Pan_troglodytes; Mikkelsen et al. 2005) shows seven, three of which are incomplete. Given the overall very close similarity between human and chimpanzee genomes (Mikkelsen et al. 2005; Kehrer-Sawatski & Cooper 2007) the possibility of a significant difference at the *GH* locus is clearly of interest and we have therefore re-examined the organization of the

chimpanzee *GH* cluster in detail. The results reported here show that the chimpanzee locus is similar to that of human in possessing 5 *GH*-like genes, but that there are some notable differences with regard to the expressed proteins, gene regulation and gene conversion events, which may have functional significance.

Materials and methods

Characterization of the chimpanzee *GH* locus

The re-examination of the chimpanzee *GH* locus involved two approaches: 1) a bacterial artificial chromosome (BAC) containing the whole locus was analysed by amplifying specific regions and by direct sequencing, and 2) the data available from whole genome sequencing of the chimpanzee was examined in detail.

Amplification of the chimpanzee *GH* locus

A BAC (CHORI-251-61D11) containing the chimpanzee *GH* (*cGH*) locus was obtained by screening a chimpanzee BAC library (CHORI-251; constructed from male chimpanzee "Clint"; average insert size 164 Kbp) from the BACPAC Resource Centre (Children's Hospital Oakland Research Institute, Oakland, California) with a radiolabeled probe derived from the human *GH-N (hGH-N)* gene. Extrachromosomal DNA from a bacterial stock carrying this BAC was extracted with the BACMAX DNA Purification Kit, following the manufacturer's instructions (EPICENTRE, Madison, WI, USA). The insert size of this BAC was about 215 Kbp.

Genes and inter-genic regions (IGRs) of the *cGH* locus were amplified from CHORI-251-61D11 using polymerase chain reaction (PCR) and the primers listed in Table 1. For *GH*-like genes a PCR reaction used 10 pmol of primers, 10 ng of BAC DNA, 5 μl of 10X buffer (provided with the enzyme), 0.2 mM of each dNTP, 1.5 mM $MgCl_2$, one unit of Gold Taq DNA Polymerase (EPICENTRE, Madison, WI, USA) and Milli-Q water to a final volume of 50 μl per reaction. After an initial denaturation step at 95ºC for 5 min, PCR was carried out for 30 cycles using a denaturation temperature of 94ºC for 30 sec,

an annealing temperature of 64ºC for 90 secs and an extension temperature of 72ºC for 2 min. The reaction was stopped after a final extension time of 10 min at 72ºC.

The IGRs of the *cGH* locus were amplified using the MasterAmp Extra-long PCR kit (Epicenter), following the manufacturer's instructions. The reaction mixture of 10 pmol of each primer, 40 ng of BAC DNA, 25 μl of 2X Pre-mix 4 (provided with the enzyme) and 2.5 U of MasterAmp Extra DNA Polymerase, was adjusted with Milli-Q water to a final volume of 50 μl. After an initial denaturation step at 94ºC for 5 min, amplifications were carried out for 12 cycles, comprising incubation at 94ºC for 30 sec (denaturation), 55-57ºC for 40 sec (annealing) and 72ºC for 7-14 min (elongation), followed by 23 similar cycles, but with increments of 15 sec per cycle in the extension time. Cycling concluded with a final extension step at 72ºC for 20 min. Amplification was carried out on an ABI Veriti thermocycler (Applied Biosystems, Foster City, CA). PCR products were analysed by electrophoresis in 1% agarose gel followed by ethidium bromide staining.

Molecular cloning and sequencing

The amplified products of the *cGH* locus (amplicons) were cloned in the commercial vector pTOPO-XL-3.5, included in the TOPO® XL PCR Cloning Kit (Invitrogen, La Jolla, CA, USA), according to the manufacturer's instructions, and transformed into electrocompetent *Escherichia coli* cells supplied with the kit. Cloned amplicons were sequenced using BigDye® terminator (Applied Biosystems, Foster City, CA) and universal M13 or specific primers and a DNA analyzer, ABI PRISM® 3100 Genetic Analyzer and software (Applied Biosystems). Additionally, the chimpanzee BAC DNA was sequenced by the McGill University and Génome Québec Innovation Centre, Montreal, Canada using

a Roche Genome Sequencer FLX 454, with a depth of coverage of about 25X. Chromatograms were imported into CodonCode Aligner Version 3.7.1 (CodonCode Corporation).

Re-evaluation of genomic assembly

The sequence of the *GH* locus available from the chimpanzee genome project (Mikkelsen et al. 2005) was re-evaluated using the approach used previously (Wallis 2008; Rodriguez-Sanchez et al. 2010). The genome assembly available at http://www.ensembl.org/Pan_troglodytes (assembly of March 2006, Pan_troglodytes_2.1), and sequence traces identified in the WGS Traces database for Pan troglodytes (http://www.ncbi.nlm.nih.gov/Traces) using the BLAST and BLAT search method (Altschul et al. 1990; Kent 2002), were integrated with data for gene and intergene sequences derived in this study and the BAC 454 sequencing, to give a finalized assembly. The WGS Traces originated at the Washington University Genome Sequencing centre and the Broad Institute. Sequences from Traces were assembled using the Staden Package (https://sourceforge.net/projects/staden/).

The assembled chimpanzee *GH* locus sequence has been deposited in GenBank/EMBL/DDBJ database, with accession number JN622009. A version of this sequence annotated to show established and potential polymorphic sites is given in the supplementary material, as are sequence alignments used in the analysis.

Sequence analysis

The nucleotide sequences obtained in this work were aligned using the ClustalW method (Higgins and Sharp 1988). Protein sequences were derived by conceptual translation of the coding sequences. The significance

of amino acid differences was assessed by molecular modelling using Rasmol and the 3D structure of the human GH:receptor model of deVos et al. (1992). Identification of repetitive elements was made with the RepeatMasker program (Smit et al. 1996-2010), using those of the human *GH* locus as a reference . Gene conversion was investigated using the GENECONV program (Sawyer 1999). Phylogenetic analysis was carried out using parsimony, neighbor-joining and maximum likelihood methods in PAUP* (Swofford 1998). Rates of evolution for non-synonymous and synonymous sites in coding sequences (dN and dS) were analysed using the codeml program in the paml package (Yang, 2007).

Results and Discussion

Organization of the chimpanzee *GH* locus

Previous studies on the *cGH* locus suggested that it contains 6 (Revol de Mendoza et al. 2004) or 7 (chimpanzee genome assembly 2.1; http://www.ensembl.org/Pan_troglodytes/Info/Index) *GH*-like genes, in contrast with the 5 genes in the well-characterized human locus (Chen et al. 1989; Fig.1). Given the overall similarity between the human and chimpanzee genomes such a difference seemed potentially important and worthy of further investigation. However, it is clear that the very repetitive nature of this region, arising as a consequence of several rounds of gene duplication (see below), leads to problems in achieving an accurate sequence assembly.

In order to re-evaluate the gene cluster, the genes and intergenic regions were amplified from a BAC containing the entire *cGH* gene cluster, using primers based on the corresponding human sequences. The BAC was also subjected to complete sequencing using a Roche Genome Sequencer FLX 454. De novo automatic assembly of the reads from this did not give a reliable assembly, but detailed analysis of these reads did help resolve a number of problem regions, and confirm many parts of the assembled locus. The results from these two approaches identified 5 *GH*-related genes, corresponding closely to the 5 genes in the human locus, and five of the six genes in the assembly proposed by Revol de Mendoza et al. (2004). No evidence was seen to support the *PLA* gene reported there, from either of these approaches (or from re-evaluation of the genome assembly - see below), and corrections are proposed in several of the other gene sequences. The possibility that the 6th gene was a feature of the particular genomic DNA sample used by Revol de Mendoza et al. (2004) cannot be ruled out.

The data available for the GH locus in the genomic assembly (http://www.ensembl.org/Pan_troglodytes/) were also examined in detail, using data available from the WGS Trace database (http://www.ncbi.nlm.nih.gov/Traces). The 4 complete genes there (Fig. 1) were well supported. The incomplete gene at the 3' end of this assembly was well supported and could be extended, but not with good reliability. Its sequence was completed on the basis of data derived from the BAC sequencing. The other two incomplete genes were not supported by this re-evaluation of the data. The first of these (Fig. 1B) is very similar to the pituitary *GH* gene (*GH-N*); its sequence is supported by the data, but these are consistent with the occurrence of a single gene showing considerable polymorphic variation, and this conclusion was supported by the number of sequence reads obtained and intergenic distances based on sequence 'mate-pairs'. This also accords with the data from the BAC, in which just a single gene was detected - polymorphism would of course have been lost on cloning of the BAC. The second incomplete gene (Fig.1B) was supported by just a single trace, of very poor quality, which clearly should be discounted; the assembly can then be reliably revised with the first and second *CSH* genes as neighbors. Detailed evaluation allowed assembly of the data from the genomic sequencing to give a complete sequence for the chimpanzee *GH* locus (Supplementary Fig. 1), with gaps being filled using information from the BAC sequencing.

On the basis of this, it is concluded that the overall organization of the chimpanzee *GH* locus is similar to that of human, with 5 genes, encoding pituitary GH (*cGH-N*), variant GH (*cGH-V*) and three placental lactogens (*cCSH-A1*, *cCSH-A2* and *cCSH-B*, equivalent to *hCSH-L*, *hCSH-A* and *hCSH-B* respectively; Fig. 1 and Table 2). Overall similarity between these 5 genes is about 92%, reflecting their recent generation by multiple rounds of gene duplication (Chen et al 1989), and some extended sections show much greater similarity than this, partly due to gene conversion (see below). The genes

flanking the locus are similar in human and chimpanzee. The chimpanzee and human *GH* gene clusters differ in detail in a number of respects, and these are discussed below.

Encoded proteins

The chimpanzee *GH-N* gene encodes a protein differing from the corresponding human gene at two positions (Fig. 2), one in the signal peptide (residue -23, Pro in chimpanzee, Thr in man) and one in the mature protein (residue 132, Glu in chimpanzee, Gly in man). This last substitution is distant (>5 Å) from both receptor-binding sites and thus unlikely to directly affect hormone-receptor interaction. The chimpanzee *GH-V* gene encodes a protein identical to human GH-V except for substitution of Pro for Ser at residue -8 in the signal peptide.

The human *CSH-L* gene potentially encodes a protein that is substantially different from placental lactogen, because of a substitution in the splice donor site in intron 2. mRNA from this gene is expressed at low level, but it is not clear that this gives rise to significant levels of functional protein. In the corresponding chimpanzee gene (*CSH-A1*) the donor splice site in intron 2 is intact, and production of a protein equivalent to CSH-A/B in human and CSH-A2 and CSH-B in chimpanzee is likely. In human the *CSH-A* and *CSH-B* genes encode the same mature protein (hCSH/hPL), although the signal peptides differ at a single base. In chimpanzee the *CSH-A1* and *CSH-A2* genes encode the same mature protein (again with signal peptides differing at one residue - residue -24 is Ala in CSH-A1, Pro in CSH-A2). The protein encoded by *CSH-B* differs from CSH-A1/A2 at 3 residues all in the mature protein, Gln-69, Met-123, Gly-126 in CSH-A1/A2 replaced by His, Thr, Arg respectively in CSH-B. Thr at residue 123 is also seen in human GH, and is within the second binding site in the human GH:receptor model (de Vos et al. 1992), so this could affect interaction with the receptor. Thus

chimpanzee potentially has two different CSHs in contrast to human, and these may differ in biological activity.

The differences between the chimpanzee CSH proteins and human CSH are indicated in Fig.2. hCSH-A/B and cCSH-A1/A2 (mature proteins) differ at three residues, 34, 123 and 164 (respectively Thr, Thr, Tyr in human, Ala, Met, His in chimpanzee). Residue 164 is within the first binding site of the human GH:receptor model (de Vos et al 1992) and residue 123 is within the second binding site (see above), so these substitutions may affect interaction with the receptor.

Regulatory elements

Many regulatory elements have been identified in the *hGH* gene cluster, including those associated with the promoter regions immediately upstream of each gene, a glucocorticoid response element in intron 1 of the *GH-N* gene, an enhancer of placental expression downstream of *hCSH-B*, "P-elements" upstream of each gene except *GH-N*, and a locus control region far upstream of *GH-N* (Ho et al. 2004). The last of these falls outside the *GH* locus considered here; the other elements are considered in detail below.

5' promoter region.

A number of response elements have been identified in the promoter region of *hGH-N*, and are compared with the corresponding regions of other genes in the *hGH* and *cGH* loci in the alignment of Fig. 3. The TATA box is retained in all genes. The *Sp-1* element (Lemaigre et al. 1989) is completely conserved, possibly reflecting its role in both placental and pituitary expression.

The distal and proximal *Pit-1* elements (Theill and Karin 1993; Krawczak et al. 1999) are generally well conserved, despite observations that Pit-1 plays an important role in controlling expression of *GH-N* in the pituitary, but not *GH/CSH* genes in the placenta (Jiang et al. 1995). For the chimpanzee *CSH-A1* gene, unlike any of the human *CSH* genes, the sequences of both distal and proximal *Pit-1* elements are identical to those for *GH-N*. A striking exception to this conservation of *Pit-1* elements is seen in the proximal element for the *GH-V* genes which show a number of substitutions, largely shared in human and chimpanzee. For this element the substitution rate on the line leading to *GH-V*, since divergence of *GH-N* and *GH-V* genes (6/18 for human, 5/18 for chimpanzee) is significantly greater than that for the divergence of the introns in the *GH-N* and *GH-V* gene sequences (for human: 6/18 v 50/811, P = 0.00072; for chimpanzee, 5/18 v 50/814, P=0.0046; Fisher's exact test). This strongly suggests that this element has been subject to adaptive evolution since the duplication giving rise to *GH-N* and *GH-V*, reflecting their differential expression in pituitary and placenta.

The thyroid hormone response element (*TRE*; Glass et al. 1987) is reported to function in placenta but not pituitary (Leidig et al. 1992), and accordingly this element shows considerable sequence variation. For the human - chimpanzee comparison, this variation differs from expectation in terms of strict orthology. Thus, the element preceding chimpanzee *CSH-A2* resembles that before chimpanzee (or human) *GH-N* rather than before other *CSH* sequences. Gene conversion events have probably influenced variation in this region, but a consequence may be that the placentally-expressed genes in chimpanzee show a pattern of responsiveness to thyroid hormones different from that seen in human.

The initiator binding site (*InrE*), which is required for efficient activity of the promoter and maximum activity of the enhancer (Jiang et al. 1995), is generally well conserved. The cyclic AMP response

element (*CRE*, Eberhardt et al. 1996) is highly conserved, contrasting with the situation in non-primate mammals (Wallis et al. 2001).

The placental enhancer

The placental enhancer is located ~2.2 kb downstream from *CSH-B* in human, and markedly increases expression of *CSH-B* in placental cells but not other tissues (Rogers et al. 1986; Walker et al. 1990). Very similar sequences are found downstream of *hCSH-L* and *hCSH-A* (but not *GH-N* or *GH-V*), but these have very little enhancer activity. An alignment of the region including the enhancer is given in Fig. 4; this shows 4 regions identified by Jacquemin et al. (1994) by footprinting, of which *DF3* and *DF4* appear to be most important. The low enhancer activity of the sequences following *hCSH-L* and *hCSH-A* has been explained by a few substitutions (Lytras et al. 1996; Jacquemin et al. 1996) which are indicated on Fig.4. Only one of these was detected in both of these studies; interestingly this substitution (position 208 in Fig.4, in the DF-3 domain - A in *hCSH-B*, G in *hCSH-A* and *hCSH-L*) is found in all three chimpanzee *CSH* genes, suggesting that the enhancer here is much less active than in human.

P-elements and GRE

P-elements are found upstream of *GH-V* and the three *CSH* genes in human, and are thought to inhibit expression of these genes in the pituitary (Norquay et al. 2006) and/or activate their expression in the placenta (Elefant et al. 2000; Ho et al. 2004). Equivalent sequences are found in the chimpanzee, and may well serve the same function. It should be noted, however, that although there is no P-element upstream of the *GH-N* gene in human or chimpanzee, such sequences are found in this position in the

marmoset and dog, rendering a simple interpretation in terms of pituitary suppression less tenable (Wallis and Wallis 2006).

A GRE has been reported in the first intron of *hGH-N* and *hGH-V* (Slater et al., 1985). This is conserved completely in the corresponding chimpanzee genes.

Polymorphism in the *GH* locus

Analysis of data available from the chimpanzee genome project suggests that the GH locus shows considerable polymorphism, particularly in the region of the *GH-N* gene (Supplementary Fig. 1). Overall the full sequence of the locus shows at least 53 clear-cut polymorphisms, mostly snps, with two or more instances of each nucleotide (nt) represented in the traces. About 47 additional potential polymorphisms were detected, in which there was only one instance of one of the two nt - some of these may represent sequencing errors, but many may be real polymorphisms, given the low coverage at many sites. Given the ~66,000 nt of sequence encompassing the locus, such polymorphisms represent 0.086-0.152% of sites, according with estimates for the overall heterozygosity of the chimpanzee genome of 0.080-0.176% (Mikkelesen et al. 2005). In the sequence of the *GH-N* gene, including 500 nt upstream and downstream of the coding sequence (~2500 nt total), there are 14 clear-cut polymorphisms and 10 potential ones, representing 0.56-0.96% of all sites, a much higher proportion than in the locus as a whole ($P<10^{-6}$, Fisher's exact test). The biological significance of this high level of polymorphism - in effect 2 haplotypes - for the region around the *GH-N* gene is not clear.

Six of these polymorphisms fall in the coding region of the *GH-N* gene, compared with a total of five in the other four genes of the cluster. Two of the six are nonsynonymous leading to single amino acid changes in mature GH at residues 92 (Phe/Leu) and 153 (Gln/His). Both of these sites (like residue 132 where chimpanzee and human sequences differ) are located on the side of the hormone facing away from the membrane in the hormone receptor complex, well away (>5 Å) from either receptor-binding site. They would not be expected to influence hormone-receptor interactions directly, although this region does show accelerated adaptive evolution in some species (Wallis 2008), suggesting that it has a specific biological role.

Gene conversion within the *GH* locus

The close similarity between the human *CSH-A* and *CSH-B* genes is probably a consequence of a gene conversion event (Chen et al. 1989). The marked differences between the equivalent chimpanzee genes (*CSH-A1* and *CSH-B*) suggests that this gene conversion event was specific to the lineage leading to man, after separation from that leading to chimpanzee. To assess the possibility of gene conversion within the chimpanzee *GH* locus, an alignment of the five gene sequences (from about 900 bp upstream of the start codon to about 250 bp downstream of the stop codon; Supplementary Fig. 2) was analysed using the GENECONV programme (Sawyer 1999).

Several potential gene conversion events were identified (Table 3). Conversions in the 5' upstream region, extending into the promoter region, were observed for *CSH-A1:GH-N* and *CSH-A2:GH-N*. Whether these represent independent events is not clear, but they may have important implications for gene regulation. Conversions involving *CSH-A1*, *CSH-A2* and/or *CSH-B* were observed within the section between intron 1 and exon 4, which may have led to some homogenization of these gene

sequences as proposed for the human *GH* locus. Finally, evidence for conversion between *CSH-A1*, *CSH-A2* and *CSH-B* in the section intron 4 to 3' downstream was seen, though statistical support for this was relatively weak.

The pattern of gene conversion observed for the chimpanzee *GH* locus contrasts with that seen for the human sequence (Table 3). Here the expected extensive conversion between *CSH-A* and *CSH-B* was confirmed, together with a possible conversion in the 5' sequence of *CSH-A* and *GH-V*, well upstream of the promoter region. The results indicate that conversion events have occurred independently in the chimpanzee and human *GH* gene clusters, since the divergence of the two species, and may have affected both expressed genes and regulatory sequences.

Repetitive elements and indels

Chen et al. (1989) identified 48 *Alu* (SINE) repetitive elements in the human *GH* locus, some of them truncated. Use of RepeatMasker confirmed that 48 equivalent *Alu* sequences are also present in the chimpanzee genome, together with a number of additional SINE- and LINE-related elements. This analysis, together with visual inspection, showed that the only clear difference in repetitive-element content between human and chimpanzee loci is a LINE 1 related element (224 nt, corresponding to the 3' end of LINE 1) upstream of *CSH-A2*/*CSH-A*, present in chimpanzee, but not in human. A very similar sequence is found upstream of *CSH-B*/*CSH-B* in both chimpanzee and human, suggesting that the difference arose by deletion of the element in human.

This LINE 1 element upstream of *CSH-A2*/*CSH-A* represents the largest insertion/deletion event (indel) seen in a comparison of the human and chimpanzee loci. 4 additional indels of more than 15 nt are seen, 3 extensions/reductions of simple AG rich sequences and a 112 nt sequence in intron 3. This last is present in human but not chimpanzee, and is not identified as a repetitive element by RepeatMasker, despite the presence of a long poly T tract.

Evolutionary aspects

Gene duplication

The evolution of the human *GH* locus was considered in detail by Chen et al. (1989), who proposed that an initial gene duplication gave rise to the ancestors of the *GH*-like and *CSH*-like genes. A second duplication gave rise to a cluster of four genes, ancestors of *GH-N, CSH-L, GH-V* and *CSH-B*. Finally, a further duplication of the *CSH-L* gene gave rise to *CSH-A*. The close similarity between *CSH-A* and *CSH-B* is then explained by gene conversion rather than phylogeny. The similarity betwen the chimpanzee and human *GH* loci implies that they shared the same evolutionary origin. In chimpanzee, as for the human locus, phylogenetic analysis (Fig. 5) indicates that *CSH-A2* (*CSH-A*) and *CSH-B* are more similar than either is to *CSH-A1 (CSH-L*). The tree of Fig. 5a is based on an alignment of all the sequence (excluding repetitive elements) that is repeated 5 times within the locus (the 5 genes, plus about 815 nt of 5' and about 218 nt of 3' flanking sequence, plus about 721 sequence upstream of each gene, 2532 nt upstream for the *GH-N* gene; Supplementary Fig. 3). Here the human *CSH-A* gene forms a clade with human *CSH-B* and chimpanzee *CSH-B*, rather than its presumed ortholog chimpanzee *CSH-A2*, possibly reflecting the long (1803 nt) gene conversion involving the human *CSHA*/*CSHB* genes (Table 3). This explanation is supported by the tree of Fig. 5b, based on an alignment excluding this 1803 nt region, where all the chimpanzee and human genes group according to the expected orthology. But both these trees indicate that the human *CSH-A* and *CSH-B*

genes and the chimpanzee *CSH-A2* and *CSH-B* genes are more closely related to each other than to the *CSH-L/CSH-A1* genes, suggesting that the final gene duplication event in the evolution of the gene cluster was not tandem duplication of the *CSH -L/CSH-A1* gene, as previously proposed (Chen et al 1989).

Rate variation in the evolution of coding sequences

Previous studies have identified variable evolutionary rates during the diversification of the GH-like proteins of higher primates; in at least some cases this reflects adaptive evolution, based on dN/dS values greater than 1 (Wallis 1996; Rodríguez-Sánchez et al. 2010). Phylogenetic analysis was carried out here using the codeml method in paml, with all branches unconstrained, and an alignment of coding sequences for mature proteins (Supplementary Fig. 4). The tree obtained (Fig.6) showed no evidence for rate variation since the divergence of human and chimpanzee, except for the rapid evolution of *hCSH-L* (dN/dS 0.82), compared with chimpanzee *CSH-A1* (dN/dS 0), which would be consistent with *hCSH-L* effectively becoming a pseudogene shortly after the divergence, and evolving at an unconstrained (neutral) rate. The increased dN/dS for the branch leading to *hCSH-A* is based on only three nt substitutions, and could also reflect gene conversion. Accelerated evolution at earlier stages in the evolution of the locus is confirmed, including rapid divergence of *CSH* prior to further duplications (dN/dS 1.35) and on the lineage leading to *GH-V* after the duplication giving *GH-N* and *GH-V* (dN/dS 0.48 for *GH-V*, 0 for *GH-N*). The availability of the chimpanzee sequences shows that dN/dS values mostly fall to very low levels after these episodes of rapid change, supporting the idea that these episodes reflect adaptive evolution, despite dN/dS levels that do not significantly exceed 1.0. The slow evolution of *GH/CSH* genes since human chimpanzee divergence contrasts with that observed for old-world monkeys (Rodríguez-Sánchez et al. 2010).

Conclusions

The main conclusion from this work is that despite previous suggestions, the *GH* gene loci of human and chimpanzee are similar, with 5 genes each showing complete orthology. Some significant differences are seen, however. In particular, the *hCSH-L* gene has changed substantially since the chimpanzee/human split, potentially becoming a pseudogene, while the corresponding *CSH-A1* gene in chimpanzee has been conserved, giving an identical product to *CSH-A2*. Chimpanzee produces two CSHs, with potentially differing biological properties, whereas human produces a single CSH. The pattern of gene conversion in human is quite different from that in chimpanzee. The region around the *GH-N* gene in chimpanzee is remarkable polymorphic, unlike the corresponding region in human.

Acknowledgements We thank Sergio Lozano-Rodriguez, M.D. and Hugo R. Gonzalez Cardenas for their critical reading of the manuscript. This research was possible thanks to grants contributed by CONACYT (Q-43987).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Barrera-Saldaña HA, Seeburg PH, Saunders GF (1983) Two structurally different genes produce the same secreted human placental lactogen hormone. J Biol Chem 258: 3787-3793

Chen EY, Liao YC, Smith DH, Barrera-Saldana HA, Gelinas RE, Seeburg PH (1989) The human growth hormone locus: nucleotide sequence, biology, and evolution. Genomics 4: 479–497

De Vos AM, Ultsch M, Kossiakoff AA (1992) Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science 255: 306-312

Eberhardt NL, Jiang SW, Shepard AR, Arnold AM, Trujillo MA (1996) Hormonal and cell-specific regulation of the human growth hormone and chorionic somatomammotropin genes. Prog Nucleic Acid Res Mol Biol 54: 127–163

Elefant F, Su Y, Liebhaber SA, Cooke NE (2000) Patterns of histone acetylation suggest dual pathways for gene activation by a bifunctional locus control region. EMBO J 19: 6814–6822

Forsyth IA, Wallis M (2002) Growth hormone and prolactin - molecular and functional evolution. J Mammary Gland Biol Neoplasia 7: 291-312

Frankenne F, Closset J, Gomez F, Scippo ML, Smal J, Hennen G (1988) The physiology of growth hormones (GHs) in pregnant women and partial characterization of the placental GH variant. J Clin Endocrinol Metab 66: 1171-1180

Glass CK, Franco C, Weinberger C, Albert VR, Evans RM, Rosenfeld MG (1987) A *c-erb-A* binding-site in rat growth hormone gene mediates *trans*-activation by thyroid-hormone. Nature 329: 738-741

Golos TG, Durning M, Fisher JM, Fowler PD (1993) Cloning of four GH/chorionic

somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta. Endocrinology 133: 1744-1752

González Alvarez R, Revol de Mendoza A, Esquivel Escobedo D, Corrales Félix G, Rodríguez Sánchez I, González V, Dávila G, Cao Q, de Jong P, Fu YX, Barrera Saldaña HA (2006) Growth hormone locus expands and diverges after the separation of New and Old World Monkeys. Gene 380: 38-45

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. Gene 73: 237–244

Ho Y, Liebhaber SA, Cooke NE (2004) Activation of the human GH gene cluster: roles for targeted chromatin modification. Trends Endocrinol Metab 15: 40-45

Jacquemin P, Oury C, Peers B, Morin A, Belayew A, Martial JA (1994) Characterization of a single strong tissue-specific enhancer downstream from the three human genes encoding placental lactogen. Mol Cell Biol 14: 93–103

Jacquemin P, Alsat E, Oury C, Belayew A, Muller M, Evain-Brion D, Martial JA (1996) The enhancers of the human placental lactogen B, A, and L genes: progressive activation during *in vitro* trophoblast differentiation and importance of the DF-3 element in determining their respective activities. DNA Cell Biol 15: 845-54

Jiang SW, Shepard AR, Eberhardt NL (1995) An initiator element is required for maximal human chorionic somatomammotropin gene promoter and enhancer function. J Biol Chem 270: 3683-3692

Kehrer-Sawatzki H, Cooper DN (2007) Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. Human Mutat 28: 99-130

Kent WJ (2002) BLAT – the BLAST-like alignment tool. Genome Res 12: 656–664

Krawczak M, Chuzhanova NA, Cooper DN (1999) Evolution of the proximal promoter region of the mammalian growth hormone gene. Gene 237: 143– 151

Lacroix MC, Guibourdenche J, Frendo JL, Muller F, Evain-Brion D (2002) Human placental growth hormone - a review. Placenta 23 (Suppl A, Trophoblast Research): S87-S94

Leidig F, Shepard AR, Zhang W, Stelter A, Cattini PA, Baxter JD, Eberhardt NL (1992) Thyroid hormone responsiveness in human growth hormone-related genes. Possible correlation with receptor-induced DNA conformational changes. J Biol Chem 267: 913-921

Lemaigre FP, Courtois SJ, Lafontaine DA, Rousseau GG (1989) Evidence that the upstream stimulatory factor and the Spl transcription factor bind *in vitro* to the promoter of the human-growth-hormone gene. Eur J Biochem 181: 555-561

Li Y, Ye C, Shi P, Zou X-J, Xiao R, Gong Y-Y, Zhang Y-P (2005) Independent origin of the growth hormone gene family in New World monkeys and Old World monkeys/hominoids. J Mol Endocrinol 35: 399-409

Liu J-C, Makova KD, Adkins RM, Gibson S, Li W-H (2001) Episodic evolution of growth hormone in primates and emergence of the species specificity of human growth hormone receptor. Mol Biol Evo 18: 945-953

Lytras A, Surabhi RM, Zhang JF, Jin Y, Cattini PA (1996) 'Repair' of the chorionic somatomammotropin-A 'enhancer' region reveals a novel functional element in the chorionic somatomammotropin-B enhancer. Mol Cell Endocrinol 119: 1-10

Männik J, Vaas P, Rull K, Teesalu P, Rebane T, Laan M (2010) Differential expression profile of growth hormone/chorionic somatomammotropin genes in placenta of small- and large-for-gestational-age newborns. J Clin Endocrinol Metab 95: 2433–2442

Mikkelsen TS, and 66 other authors (2005) Initial sequence of the chimpanzee genome and comparison with

the human genome. Nature 437: 69-87

Misra-Press A, Cooke NE, Liebhaber SA (1994) Complex alternative splicing partially inactivates the human chorionic somatomammotropin-like (hCS-L) gene. J Biol Chem 269: 23220–23229

Norquay LD, Yang X, Jin Y, Detillieux KA, Cattini PA (2006) Hepatocyte nuclear factor-3α binding at P sequences of the human growth hormone locus is associated with pituitary repressor function. Mol Endocrinol 20: 598-607

Papper Z, Jameson NM, Romero R, Weckle AL, Mittal P, Benirschke K, Santolaya-Forgas J, Uddin M, Haig D, Goodman M, Wildman DE (2009) Ancient origin of placental expression in the growth hormone genes of anthropoid primates. Proc Natl Acad Sci USA 106: 17083–17088

Reséndez-Pérez D, Ramirez-Solís R, Varela-Echavarría A, Martínez-Rodríguez HG, Barrera-Saldaña HA (1990) Coding potential of transfected human placental lactogen genes. Nucl Acids Res 18: 4665-4670

Revol de Mendoza A, Esquivel Escobedo D, Santiago Alarcón D, Barrera Saldaña H (2001) Independent duplication of the growth hormone gene in three anthropoidean lineages. J Endocr Gen 2: 151-159

Revol de Mendoza A, Esquivel Escobedo D, Martínez Dávila I, Barrera Saldaña H (2004) Expansion and divergence of the GH locus between spider monkey and chimpanzee. Gene 336: 185-193

Rodríguez-Sánchez IP, Tejero ME, Cole SA, Comuzzie AG, Nathanielsz PW, Wallis M, Barrera-Saldaña HA (2010) Growth hormone-related genes from baboon (*Papio hamadryas*): characterization, placental expression and evolutionary aspects. Gene 450: 1-7

Rogers BL, Sobnosky MG, Saunders GF (1986) Transcriptional enhancer within the human placental-lactogen and growth-hormone multigene cluster. Nucl Acids Res 14: 7647-7659

Sawyer SA (1999) GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University, St. Louis, MO; available at: http://www.math.wustl.edu/_sawyer

Smit AFA, Hubley R, Green P (1996-2010) RepeatMasker Open-3.0. http://www.repeatmasker.org

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer, Sunderland, MA

Theill LE, Karin M (1993) Transcriptional control of GH expression and anterior pituitary development. Endocrine Reviews 14: 670–689

Walker WH, Fitzpatrick SL, Saunders GF (1990) Human placental-lactogen transcriptional enhancer - tissue-specificity and binding with specific proteins. J Biol Chem 265: 12940-12948

Wallis M (1996) The molecular evolution of vertebrate growth hormones: a pattern of near-stasis interrupted by sustained bursts of rapid change. J Mol Evol 43: 93-100

Wallis M (2008) Mammalian genome projects reveal new growth hormone (GH) sequences. Characterization of the GH-encoding genes of armadillo (*Dasypus novemcinctus*), hedgehog (*Erinaceus europaeus*), bat (*Myotis lucifugus*), hyrax (*Procavia capensis*), shrew (*Sorex araneus*), ground squirrel (*Spermophilus tridecemlineatus*), elephant (*Loxodonta africana*), cat (*Felis catus*) and opossum (*Monodelphis domestica*). Gen Comp Endocrinol 155: 271–279

Wallis OC, Wallis M (2002) Characterisation of the GH gene cluster in a new-world monkey, the marmoset (*Callithrix jacchus*). J Mol Endocrinol 29: 89-97

Wallis OC, Wallis M (2006) Evolution of growth hormone in primates: the GH gene clusters of the New World monkeys marmoset (*Callithrix jacchus*) and white-fronted capuchin (*Cebus albifrons*). J Mol Evo 63: 591–601

Wallis OC, Zhang YP, Wallis M (2001) Molecular evolution of GH in primates: characterization of the GH

genes from slow loris and marmoset defines an episode of rapid evolutionary change. J Mol Endocrinol 26: 249–258

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586–1591

Ye C, Li Y, Shi P, Zhang YP (2005) Molecular evolution of growth hormone gene family in old world monkeys and hominoids. Gene 350: 183-192

Table 1. Primers used in this study.

Name	Sequence	Orientation	Target
GHPRIMB	TTAAGGATCCCAAGGCCCAACTCC	sense	genes
GHPRIMK	ATTAAGGGTACCGTGGACAGCTCACCTAG	sense	genes
GHPRIMEX	CCTCGAGATATCCTAGAAGCCACAGCT	antisense	genes
GHPRIMF	CATCTCCCTGCTGCTCATC	sense	IGRs
GHPRIMR	GAAAACAACCCTGAGCTCC	antisense	IGRs
IGLEcoF	TTTGCAGATACAGGATATCTACAGCCCTGATG	sense	IGRs
IGLKpnR	CCCCACAGTTTGGTACCCTAAGATTTAGGACTAC	antisense	IGRs
CSH378F	GGCTTTTTGACCACGCTATGCTCCA	sense	IGRs
CSH348R	CGGAACGGTTTGGACGGCACC	antisense	IGRs
GH348R	GGGAATGGTTGGGAAGGCACTG	antisense	IGRs

Table 2. Nomenclature used for genes of the *GH* loci.

Position in the locus	**Gene**	**Abbreviations**		**Tissue of expression**
		Human	**Chimpanzee**	
1	'Normal' Growth Hormone	*hGH-N*	*cGH-N*	pituitary
2	Chorionic somatomammotropin	*hCSH-L*	*cCSH-A1*	placenta
3	Chorionic somatomammotropin	*hCSH-A*	*cCSH-A2*	placenta
4	Variant Growth Hormone	*hGH-V*	*cGH-V*	placenta
5	Chorionic somatomammotropin	*hCSH-B*	*cCSH-B*	placenta

hCSH-L and *cCSH-A1* are at equivalent locations in the locus, but *hCSH-L* is spliced differently from other *CSH* genes and may effectively be pseudogene (see text), whereas *cCSH-A1* appears to undergo normal splicing, equivalent to that of *cCSH-A2* and *hCSH-A*. Note that hPL (placental lactogen) is frequently used instead of hCSH (chorionic somatomammotropin) to designate both the genes and their hormonal products in human.

Table 3. Gene conversion between GH/CSH genes

Sequences	nt positions	Length	Location	p value
Chimpanzee				
CSH-A1:GH-N	368-807	440	5' upstream-promoter	<0.0001
CSH-B:CSH-A1	1198-1960	763	intron 1 - exon 4	0.0001
CSH-A2:GH-N	743-923	181	5' upstream - promoter	0.0001
CSH-A2:CSH-A1	1240-1987	748	exon 2 - exon 4	0.0005
CSH-A2:CSH-B	1048-1965	918	intron 1 - exon 4	0.0120
CSH-A2:CSH-A1	2145-2632	487	intron 4- 3' downstream	0.0177
CSH-B:CSH-A1	2203-2665	463	intron 4- 3' downstream	0.0192
Human				
CSH-A:GH-V	1-172	172	5' upstream	0.0004
CSH-A:CSH-B	182-2012	1831	promoter-exon 4	0.001

For chimpanzee: ATG start codon at 928-930, TAG stop codon at 2425-2427. For human: ATG start codon at 946-948, TAG stop codon at 2440-2442. GENECONV was run with mismatches allowed (setting g1). p values are corrected for multiple hypothesis testing.

Figure legends

Fig. 1 The genomic organization of the *GH* locus in man and chimpanzee. **A.** The human *GH* locus (based on Chen et al. 1989). **B.** The chimpanzee *GH* locus based on the assembly available at http://www.ensembl.org/Pan_troglodytes (released in March 2006). **C.** the chimpanzee *GH* locus based on the present work. The boxes in gray indicate the genes of GH-like sequences, white regions on the horizontal bar represent gaps in the chimpanzee *GH* locus sequences. *CD79B* and *TCAM* (*TCAM1* homolog pseudogene) are the genes immediately flanking the GH locus in human and chimpanzee assemblies (http://www.ensembl.org). In man and chimpanzee the *GH* locus is found on chromosome 17 (17q24.2 in man) and is shown here in reverse orientation. Arrows indicate direction of transcription. Additional genes identified in the BAC but not shown are *SCN4A* and *ICAM2* (upstream of *CD79B*) and *SMARCD2* (downstream of *TCAM1* pseudogene).

Fig. 2 Alignment of amino acid sequences. hGH-N was used as a reference; identities to this in other sequences are shown as . and deletions as -. The sequence given for hCSH-L is a conceptual translation, assuming splicing equivalent to that seen in the other genes; in practice alternatively spliced forms predominate for *hCSH-L* and it is not certain whether these are translated.

Fig. 3 Alignment of proximal promoters. The *hGH-N* gene was used as a reference; identities to this in other sequences are shown as . and deletions as -. Regulatory elements that have been identified in the *hGH-N* gene are shown in boxes.

Fig. 4 Alignment of the placental enhancers. The sequences of the three putative placental enhancers of the chimpanzee locus were aligned with their counterparts in human. The *hGH-N* gene was used as a reference;

identities to this in other sequences are shown as . and deletions as -. The rectangles show four DF domains of the placental enhancer. Nucleotides that when mutated cause a reduction in activity of the enhancer are shown as * (Lytras et al. 1996) and $ (Jacquemin et al. 1996).

Fig. 5 Phylogenetic trees for the five genes in the *GH* gene locus of human and chimpanzee. **A.** Tree based on alignment of all sequence (excluding repetitive elements) repeated five times within the locus (see text for details). **B.** Tree based on the same alignment but with exclusion of the region showing gene conversion between *hCSH-A* and *hCSH-B* (Table 3). The trees were constructed using the parsimony method in PAUP*; Neighbor Joining and maximum likelihood methods gave similar results. Branch lengths are proportional to number of substitutions, indicated by the scale bars. Numbers at nodes are percentages of at least 100 bootstrap replications supporting that clade.

Fig. 6. Phylogenetic tree based on coding sequences of human and chimpanzee *GH* locus genes. The tree was constructed using an alignment of coding sequences for mature proteins (no signal peptide) and codeml, with a defined tree and no constraint on dN/dS values (model 1). Branch lengths represent total substitutions (dN + dS; the scale bar shows 0.1 substitution/site) and numbers on branches are dN/dS values. The sequence encoding slow loris (*Nycticebus pygmaeus*) GH (Wallis et al. 2001) was used as an outgroup.

Fig. 1

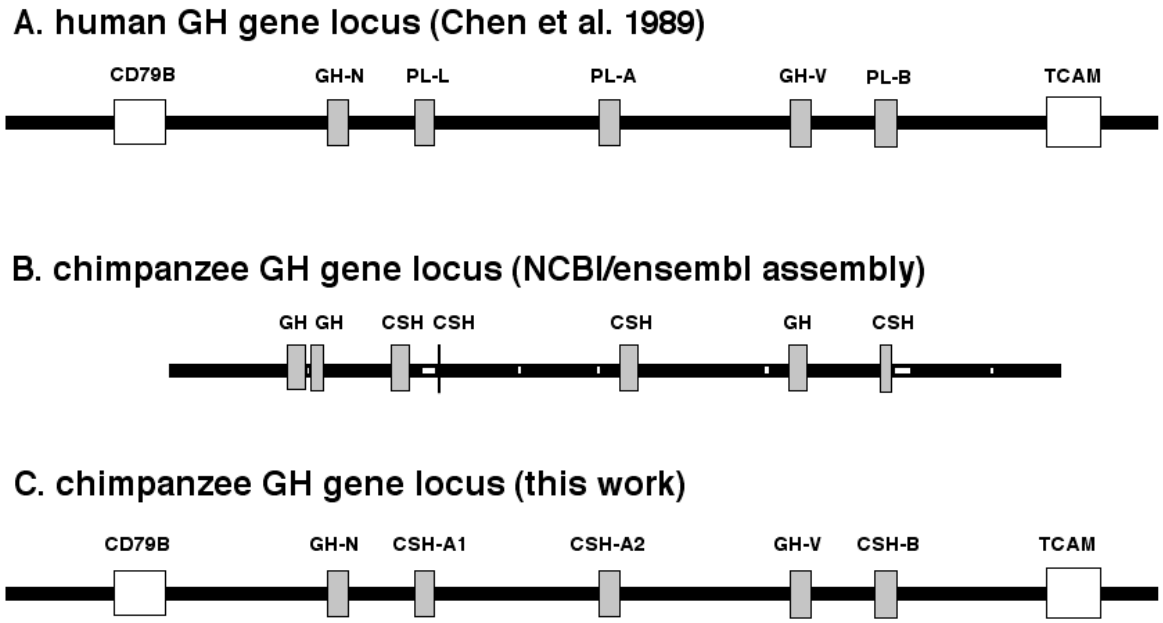


Fig. 2

```
                        20                  40                  60                  80                 100
hGH-N    FPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANS
cGH-N    ....................................................................................................

hGH-V    ................R..Y...Y..........L.........................VK.......................L........
cGH-V    ................R..Y...Y..........L.........................VK.......................L........

hCSH-L   VQ.V......KE...Q...A....I..........T........HDS...F...D....S..M...........H......E.R....R....T.T.N
hCSH-A   VQ.V.......H...Q...A....I........T....D......HDS...F...D.......M.................E......R....M...N
hCSH-B   VQ.V.......H...Q...A....I........T....D......HDS...F...D.......M.................E......R....M...N

cCSH-A1  VQ.V.......H...Q...A....I............D......HDS...F...D.......M.................E......R....M...N
cCSH-A2  VQ.V.......H...Q...A....I............D......HDS...F...D.......M.................E......R....M...N
cCSH-B   VQ.V.......H...Q...A....I............D......HDS...F...D.......M....H............E......R....M...N


                       120                 140                 160                 180         192
hGH-N    LVYGASDSNVYDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMDKVETFLRIVQC-RSVEGSCGF
cGH-N    ........................E...........................................................

hGH-V    ...........RH............W.............N.S......K.....................................
cGH-V    ...........RH............W.............N.S......K.....................................

hCSH-L   ...DT...DD.H..........M.........HL...TL.............H..........H.............M.............
hCSH-A   ...DT...DD.H.....................R.....L.............H..........................M.............
hCSH-B   ...DT...DD.H.....................R.....L.............H..........................M.............

cCSH-A1  ...DT...DD.H..........M.........R.....L.............H..........H.............M.............
cCSH-A2  ...DT...DD.H..........M.........R.....L.............H..........H.............M.............
cCSH-B   ...DT...DD.H.............R......R.....L.............H..........H.............M.............
```

Fig.3

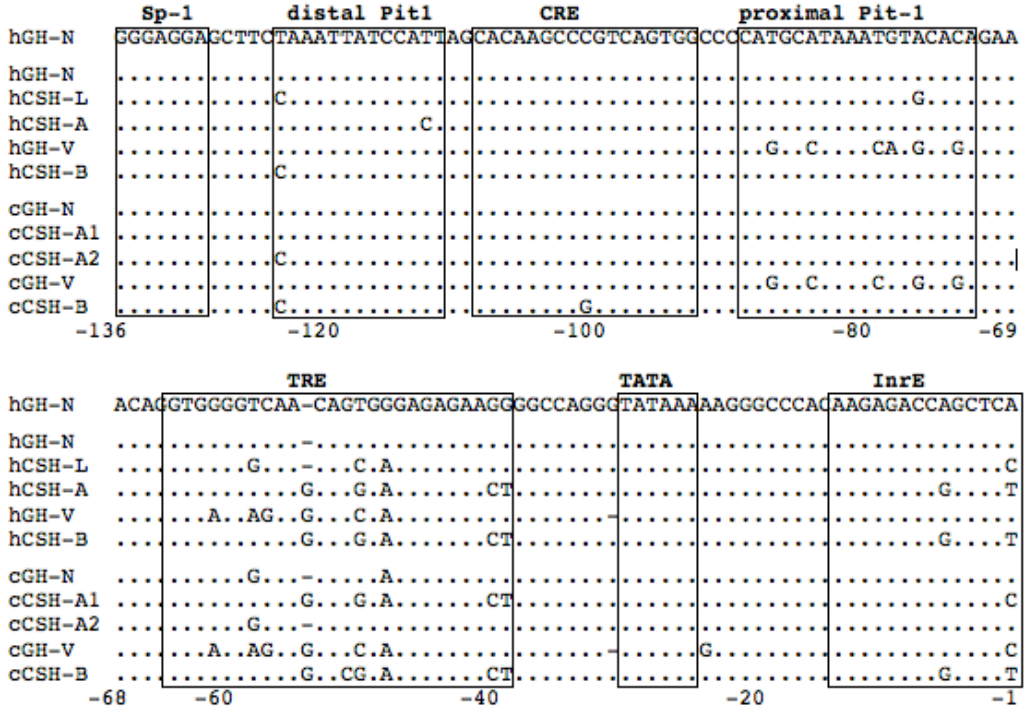

Sp-1
distal Pit1
CRE
proximal Pit-1
hGH-N GGGAGGAGCTTCTAAATTATCCATTAGCACAAGCCCGTCAGTGGCCCCATGCATAAATGTACACAGAA
hGH-N
hCSH-L
hCSH-A
hGH-V
hCSH-B
cGH-N
cCSH-A1
cCSH-A2
cGH-V
cCSH-B
-136
-120
-100
-80
-69
TRE
TATA
InrE
hGH-N ACAGGTGGGGTCAA-CAGTGGGAGAGAAGGGGCCAGGGTATAAAAAGGGCCCACAAGAGACCAGCTCA
hCSH-L
hCSH-A
hGH-V
hCSH-B
cGH-N
cCSH-A1
cCSH-A2
cGH-V
cCSH-B
-68
-60
-40
-20
-1

Fig.4

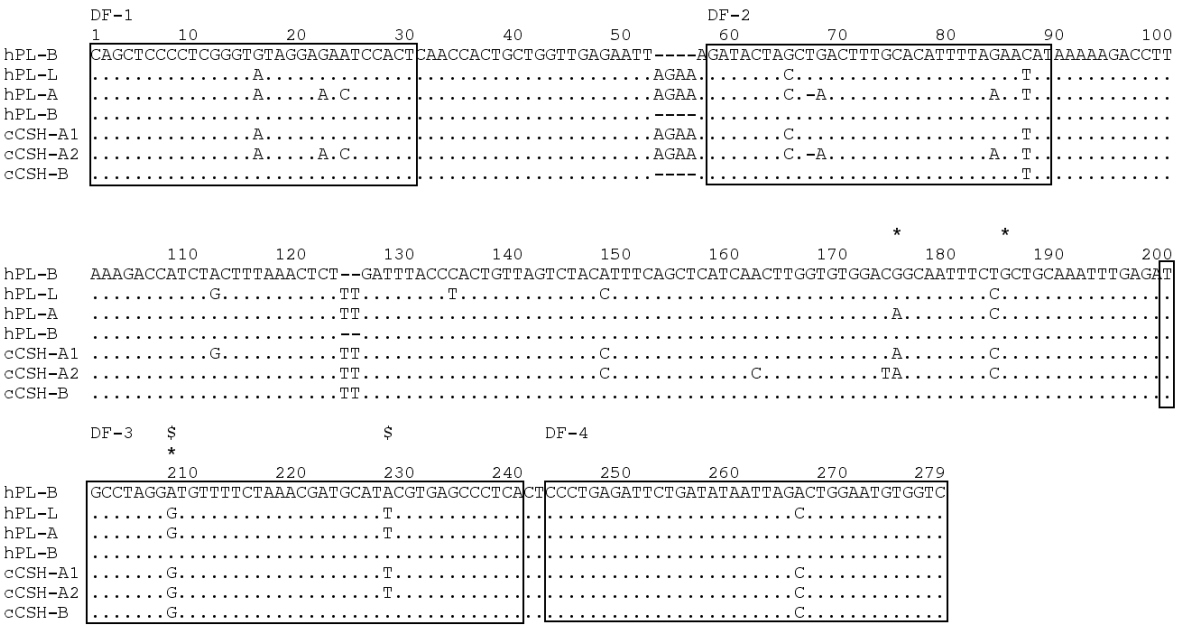

DF-1
DF-2
hPL-B CAGCTCCCCTCGGGTGTAGGAGAATCCACTCAACCACTGCTGGTTGAGAATT----AGATACTAGCTGACTTTGCACATTTTAGAACATAAAAAGACCTT
hPL-L
hPL-A
hPL-B
cCSH-A1
cCSH-A2
cCSH-B
hPL-B AAAGACCATCTACTTTAAACTCT--GATTTACCCACTGTTAGTCTACATTTCAGCTCATCAACTTGGTGTGGACGGCAATTTCTGCTGCAAATTTGAGAT
hPL-L
hPL-A
hPL-B
cCSH-A1
cCSH-A2
cCSH-B
DF-3
DF-4
hPL-B GCCTAGGATGTTTTCTAAACGATGCATACGTGAGCCCTCACTCCCTGAGATTCTGATATAATTAGACTGGAATGTGGTC
hPL-L
hPL-A
hPL-B
cCSH-A1
cCSH-A2
cCSH-B

Fig.5

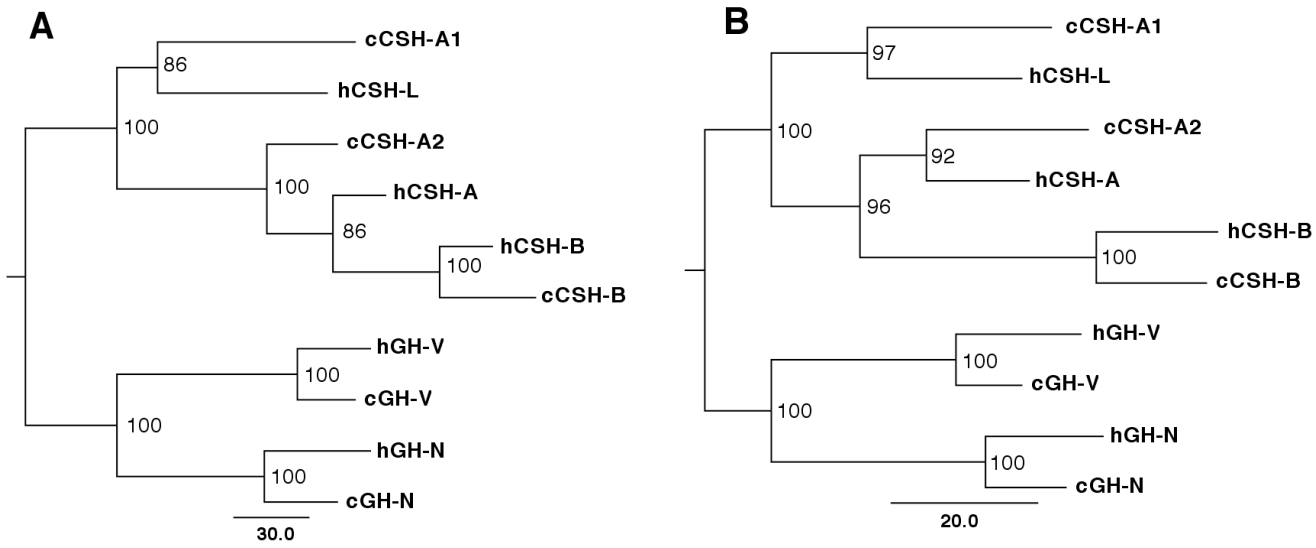
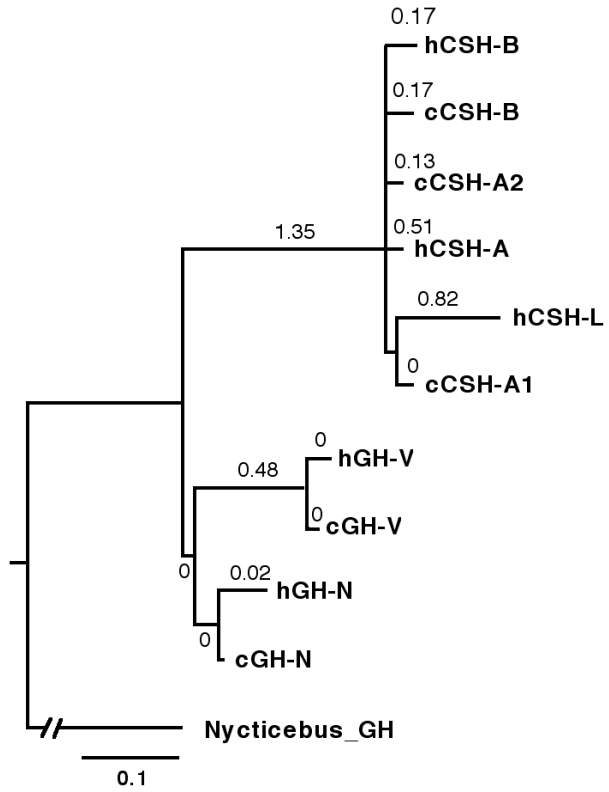

A
cCSH-A1
86
hCSH-L
100
cCSH-A2
100
hCSH-A
86
hCSH-B
100
cCSH-B
hGH-V
100
cGH-V
100
hGH-N
100
cGH-N
30.0
B
cCSH-A1
97
hCSH-L
100
cCSH-A2
92
hCSH-A
96
hCSH-B
100
cCSH-B
hGH-V
100
cGH-V
100
hGH-N
100
cGH-N
20.0

Fig. 6


0.17
hCSH-B
0.17
cCSH-B
0.13
cCSH-A2
1.35
0.51
hCSH-A
0.82
hCSH-L
0
cCSH-A1
0
hGH-V
0.48
0
cGH-V
0
0.02
hGH-N
0
cGH-N
Nycticebus_GH
0.1