

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FACULTAD DE MEDICINA



DISECCIÓN BIOINFORMÁTICA Y EXPERIMENTAL DE LA FASE  
DE EXPANSIÓN DEL *LOCUS* GH EN PRIMATES.

Presentado por

BIOQUÍMICO, M.C. ANTONIO ALÍ PÉREZ MAYA

Como requisito parcial para obtener el Grado de  
DOCTORADO EN CIENCIAS CON ORIENTACIÓN EN  
BIOLOGÍA MOLECULAR E INGENIERÍA GENÉTICA

DICIEMBRE, 2012

*“La emoción más hermosa y más profunda que podemos experimentar es la sensación de lo místico. Es el legado de toda ciencia verdadera.”*

**Albert Einstein**

***Dedicatoria:***

*A mi madre, por su amor incondicional y el apoyo que me ha brindado siempre y porque ha sabido educarme con buenos sentimientos, hábitos y valores, lo cual me ha ayudado a salir adelante en los momentos más difíciles. Madre, gracias por darme un futuro, todo esto te lo debo a ti.*

*A mis hermanos, por su cariño y comprensión, y porque siempre han creído en mí, esto también se lo debo a ustedes.*

*A los seres queridos que he perdido, pero que continúan a mi lado como guardianes invisibles cuidándome y guiándome.*

*A mi familia y amigos, por el cariño que me han brindado y por compartir conmigo buenos y malos momentos, haciendo de esta experiencia una de las más especiales.*

*A todos aquellos profesores que de una forma u otra, por su tiempo, apoyo y sabiduría contribuyeron a mi formación y desarrollo profesional.*

## AGRADECIMIENTOS

---

Desearía añadir una extensa lista con los nombres de las personas que, directa o indirectamente, han contribuido en alguna medida para que haya podido culminar este trabajo. Sin embargo, me limitaré a destacar los reconocimientos que considero más significativos.

En primer lugar agradezco al Dr. Hugo Barrera, por su dirección, apoyo, tiempo y entusiasmo, que han sido de inestimable valor. Él ha fomentado un ambiente de curiosidad científica y de búsqueda incesante del conocimiento, que además de hacer posible este trabajo, también lo ha hecho emocionante. Muchas gracias por la confianza que ha depositado en mí desde el primer momento y por haberme dado esta oportunidad!

Una mención especial para el PhD. Michael Wallis, por toda la colaboración brindada y enseñanzas, su apertura al intercambio de datos, métodos y conocimientos ha permitido que mi trabajo sea más detallado y completo. Me considero afortunado de haber sido capaz de aprovechar tal riqueza de experiencias.

Agradezco a los doctores que conforman mi comité de tesis, Lourdes Garza, Jorge Ascacio, Víctor Treviño y Roberto Montes de Oca, quienes revisaron todo el manuscrito y proporcionaron valiosas sugerencias.

A la Dra. Herminia Martínez, Jefe del Departamento de Bioquímica y Biología Molecular, por su confianza, comprensión y apoyo.

Al Dr. Francisco Velázquez por su valiosa ayuda y sus consejos, y al Dr. Mario Hernández por su invaluable apoyo.

Quiero agradecer a todos los que me han ayudado, especialmente a Raquel Cardiel, Norma Estrella, Julieta Hinojos, Ariadna López, Lourdes Mendoza y Daniel Rodríguez por la gran calidad humana que me han demostrado con su amistad y porque cada uno con sus aportaciones hicieron posible este proyecto.

Es un placer expresar un agradecimiento general a mis amigos y compañeros, con los que he compartido buenos y malos momentos en esta etapa de mi vida, a Oscar Fajardo, Ricardo Gutiérrez, Susana Becerril, Francisco Hernández, Rodolfo Müller, Roberto Cuellar, Juan Carlos Tamez, Esperanza Cuellar, Dulce González, Lídice Macola, Alejandro Belmonte y familia, Rafael León, Celia Sánchez, María Elena García, Lourdes Garza, Edith Flores, Clarissa Álvarez, Tufic Yutani, Emilio Soto, Dolores Esquivel, Blanca Alemán, Daniela Monsiváis, José Morales, Alan Márquez, Gissela Borrego, Sofía Bernal, Adriana Sánchez, Mario Simental, Luz Castellanos, Magda Vera, Sergio Cárdenas, Irám Rodríguez, Tanya Salas, Hugo González, Jorge Lara, Diana Pérez, Iván Marino, Alma Rosa, Ruth Cervantes, Mario Sánchez, Irazema Aguirre y a todos los que con las prisas haya olvidado mencionar, pero no por ello dejan de ser importantes para mí. A todos tengo algo que agradecer.

Pero entre ellos quiero agradecer especialmente a los amigos con los que he convivido más y que han sido para mí de inmenso valor, por su cariño, por su apoyo incondicional, su compañía, sus consejos, por demostrarme que siempre podré contar con ellos. A esos amigos que siempre están conmigo, estén lejos o cerca y que me han tendido la mano cuando más lo he necesitado. A ustedes los considero también parte de mi familia, esto también se lo debo a ustedes.

Finalmente, por encima de todo, quiero agradecer y destacar el apoyo y cariño incondicional de mi familia, especialmente de mis padres y mis hermanos, que han inspirado gran parte de este trabajo. A ellos quiero dedicarles este proyecto pues sé la ilusión que les hace y sé que ellos están orgullosos de la persona en la cual me he convertido. Este logro también es suyo!

**A todos mis más sinceros agradecimientos.**

<b>ÍNDICE GENERAL</b>	<b>PÁGINA</b>
<b>LISTA DE CUADROS</b>	i
<b>LISTA DE FIGURAS</b>	ii
<b>NOMENCLATURA</b>	v
<b>RESUMEN</b>	vii
<b>CAPÍTULO I. INTRODUCCIÓN</b>	
1.1. Anatomía y fisiología del <i>locus</i>	1
1.1.1. Estructura (composición y organización)	1
1.1.2. Los genes	2
1.1.3. Función de las hormonas codificadas	4
1.1.4. Regulación neuroendocrina de la secreción de GH en humanos	5
1.1.5. Mecanismos de acción moleculares de la GH	6
1.2. Regulación de la expresión de los genes del <i>locus</i> GH en el humano	8
1.2.1. Regulación por elementos cercanos	8
1.2.2. Promotor	9
1.2.3. Represor	10
1.2.4. Potenciador	11
1.2.5. Regulación por elementos distantes	12
1.3. Los primates. Clasificación	14
1.3.1. Secuenciación y genómica comparativa en primates	17
1.3.2. Comparaciones genómicas en homínidos	21
1.3.3. El proyecto ENCODE comienza a desvelar los enigmas del genoma humano	28
1.4. Elementos móviles	33
1.4.1. Papel de los elementos repetitivos en la evolución del <i>locus</i> GH y en la expresión de sus genes	36
1.5. Estado del <i>locus</i> GH de los primates al comenzar este trabajo	40
1.6. Justificación del trabajo	41
1.7. Hipótesis	43

<b>CAPÍTULO II. OBJETIVOS</b>	
2.1. Objetivo general	44
2.2. Objetivos específicos	44
<b>CAPÍTULO III. ESTRATEGIA GENERAL</b>	45
<b>CAPÍTULO IV. MATERIAL Y MÉTODOS</b>	
4.1. Material	47
4.1.1. Origen de los reactivos	47
4.1.2. Material biológico	48
4.2. Equipo	48
4.2.1. Programas computacionales	49
4.3. Métodos	50
4.3.1. Caracterización de los <i>loci</i> GH de las especies de primates de interés	50
4.3.2. Amplificación de los <i>loci</i> GH	50
4.3.3. Tamizaje de las clonas recombinantes portadoras de los genes GH/CSH	53
4.3.4. Extracción del ADN plasmídico	53
4.3.5. Selección de las clonas recombinantes portadoras de fragmentos de RIGs	54
4.3.6. Clonación molecular	54
4.3.7. Secuenciación	54
4.3.8. Re-evaluación del ensamblaje genómico	55
4.3.9. Identificación de las secuencias generadas	55
4.3.10. Análisis de las proteínas	55
4.3.11. Análisis del promotor proximal	56
4.3.12. Análisis de los elementos reguladores del control transcripcional	56
4.3.13. Identificación de los elementos repetitivos	56

## CAPÍTULO V. RESULTADOS

5.1. Disección experimental y resolución de conflictos de ensamblajes	57
5.1.1. Iniciando el análisis de los datos	57
5.1.2. Aclarando dudas sobre los genes	58
5.1.3. Las regiones intergénicas como claves en el arreglo de los genes	59
5.1.4. Un nuevo enfoque para verificar el ensamblaje del <i>locus</i>	62
5.1.5. Confirmación del arreglo de cinco genes para el <i>locus</i> GH del chimpancé y de cuatro para el <i>locus</i> GH del orangután	65
5.2. Disección bioinformática	66
5.2.1. Ensamblaje y anotación estructural de los <i>loci</i> GH de primates no homínidos	66
5.2.2. Ensamblaje y anotación estructural de los <i>loci</i> GH de homínidos. Genómica comparativa y evolutiva.	68
5.2.2.1. Análisis de las proteínas	71
5.2.2.2. Análisis filogenético de los genes GH/CSH de humano y chimpancé	76
5.2.2.3. Un vistazo a los reguladores de la transcripción	76
El promotor proximal	77
El inhibidor hipofisiario o elemento P	78
El potenciador placentario o enhancer	79
Elementos de respuesta a glucocorticoides	80
5.2.2.4. Comparaciones de las regiones intergénicas	81
Elementos repetitivos y eventos <i>indels</i>	81

## CAPÍTULO VI. DISCUSIÓN

Obtención de los genes GH/CSH y sus regiones intergénicas: chimpancé	91
Organización de los <i>loci</i> GH: chimpancé, gorila y orangután	91
Análisis de las proteínas	92
Análisis de los elementos reguladores	96
Análisis del promotor proximal	96
El potenciador placentario	97
El elemento P	100

Elemento de respuesta a glucocorticoides	102
Elementos repetitivos y eventos <i>indels</i>	102
Visión general (Reconstruyendo la historia evolutiva del <i>locus</i> GH humano)	104
El valor de la genómica comparativa	108
Perspectivas	108
<b>CAPÍTULO VII. CONCLUSIONES</b>	109
<b>CAPÍTULO VIII. BIBLIOGRAFÍA</b>	110
<b>CAPÍTULO IX. ANEXOS</b>	123

**LISTA DE CUADROS****PÁGINA**

<b>Cuadro 1.</b> Iniciadores utilizados en este estudio	51
<b>Cuadro 2.</b> Nomenclatura usada para los genes de los <i>loci</i> GH	68

<b>LISTA DE FIGURAS</b>	<b>PÁGINA</b>
<b>Figura 1.</b> Anatomía del <i>locus</i> hGH	2
<b>Figura 2.</b> Vías de señalización del receptor GH iniciadas por la activación de JAK2	6
<b>Figura 3.</b> Distribución de los sitios del promotor proximal que participan en la regulación de la expresión de los genes	9
<b>Figura 4.</b> Detalles de sitios de regulación de la expresión de los genes del <i>locus</i> GH	11
<b>Figura 5.</b> La región de control del <i>locus</i> hGH	13
<b>Figura 6.</b> Evolución de los primates	16
<b>Figura 7.</b> Un nuevo miembro de la familia humana.	27
<b>Figura 8.</b> Estrategia General	46
<b>Figura 9.</b> Iniciadores utilizados para amplificar los <i>loci</i> GH de chimpancé y gorila	52
<b>Figura 10.</b> Tamizaje mediante PCR en colonia de vectores recombinantes portadores de GH	59
<b>Figura 11.</b> Amplificación de las regiones intergénicas de los <i>loci</i> GH de chimpancé y gorila	60
<b>Figura 12.</b> Amplificación de las regiones intergénicas de los <i>loci</i> GH de chimpancé y gorila	60
<b>Figura 13.</b> Amplificación de las regiones intergénicas de los <i>loci</i> GH de chimpancé y gorila	61
<b>Figura 14.</b> Amplificación de las regiones intergénicas de los <i>loci</i> GH de chimpancé y gorila	61
<b>Figura 15.</b> Representación de las regiones amplificadas con tres pares de iniciadores en el arreglo de genes propuestos para el <i>locus</i> GH del chimpancé	62
<b>Figura 16.</b> Caracterización del producto amplificado de las intergénicas largas con flancos en genes	63
<b>Figura 17.</b> Caracterización del producto amplificado obtenido para R2	64
<b>Figura 18.</b> Caracterización del producto amplificado obtenido para R3	65

<b>Figura 19.</b> Representación de las regiones amplificadas en los arreglos de genes propuestos para los <i>loci</i> GH del chimpancé (Panel A) y del orangután (Panel B)	66
<b>Figura 20.</b> Anotación estructural de los <i>loci</i> GH de los primates no homínidos ensamblados	67
<b>Figura 21.</b> Organización genómica del <i>locus</i> GH en el humano, el chimpancé, el gorila y el orangután	70
<b>Figura 22.</b> Alineamiento de las secuencias aminoacídicas de las proteínas maduras	71
<b>Figura 23.</b> Alineamiento de las secuencias aminoacídicas correspondientes al péptido señal	72
<b>Figura 24.</b> Reconstrucción de la secuencia aminoacídica del ancestro común de los genes gCSH1, cCSHA1 y hCSH-L	73
<b>Figura 25.</b> Representación de los productos proteicos generados en los <i>loci</i> GH del gorila (A), del chimpancé (B) y del humano (C).	75
<b>Figura 26.</b> Árbol filogenético de los genes GH/CSH de humano, chimpancé, gorila y orangután	76
<b>Figura 27.</b> Alineamiento de los promotores proximales	77
<b>Figura 28.</b> Alineamiento de los inhibidores hipofisarios	78
<b>Figura 29.</b> Alineamiento de los potenciadores placentarios	80
<b>Figura 30.</b> Alineamiento del exón 1 y el comienzo del intrón 1 en los genes GH/CSH	81
<b>Figura 31.</b> Análisis comparativo de la distribución de los elementos repetitivos	82
<b>Figura 32.</b> Análisis comparativo de la distribución de elementos <i>ALU</i> e identificación de las diferencias encontradas	83
<b>Figura 33.</b> Análisis comparativo de la distribución de elementos <i>MIR</i> (panel superior) e identificación de las diferencias encontradas	84
<b>Figura 34.</b> Análisis comparativo de la distribución de elementos <i>LINE</i> e identificación de las diferencias encontradas	85
<b>Figura 35.</b> Análisis comparativo de la distribución de elementos hAT-Charlie	87
<b>Figura 36.</b> Contribución de los elementos hAT-Charlie al origen del potenciador	

placentario	88
<b>Figura 37.</b> Comparación de las GHs hipofisiarias de especies representativas de los diferentes subórdenes de primates	89
<b>Figura 38.</b> Evolución del <i>locus</i> GH	102
<b>Figura 39.</b> Detalle de la evolución del <i>locus</i> GH	103
<b>Figura 40.</b> Representación de las secuencias genómicas de los <i>loci</i> GHs	106
<b>Figura 41.</b> Representación de las secuencias genómicas contenidas en nuestra colección de BACs	107

## NOMENCLATURA

$\mu\text{g}$	Microgramo
$\mu\text{l}$	Microlitros
$\mu\text{M}$	Concentración micromolar
$^{\circ}\text{C}$	Grados centígrados
ADN	Ácido desoxirribonucleico
AMPc	Monofosfato de adenosina cíclico
BAC	Cromosoma artificial bacteriano
cbp	Cuanto basta para
CSH	Hormona somatomamotropina coriónica
dATP	Desoxiadenosina trifosfato
dNTPs	Desoxirribonucleósidos trifosfato
DO	Densidad óptica
EDTA	Ácido etilen-diamino-tetra-acético
GH	Hormona del crecimiento
GM	Grandes monos
h	Horas
IGF-1	Factor de crecimiento semejante a la insulina tipo I
kpb	Kilo pares de bases
kV	Kilo volts
LCR	Región de Control del <i>locus</i>
M	Concentración molar
mg	Miligramos
min	Minutos
mL	Mililitro
mM	Concentración milimolar
MNM	Monos del nuevo mundo
MVM	Monos del viejo mundo
MYA	Millones de años atrás

ng	Nanogramos
nm	Nanómetro
PA	Producto amplificado
pb	Pares de bases
PCR	Reacción en Cadena de la Polimerasa
pH	Potencial de hidrógeno
r.p.m.	Revoluciones por minuto
ARN	Ácido ribonucleico
ARNm	ARN mensajero
SNP	Polimorfismo de un solo nucleotido
Taq	ADN polimerasa de <i>Thermophilus aquaticus</i>
TE	Solución amortiguadora de Tris y EDTA
U	Unidades de enzima
UV	Ultravioleta
V	Voltios
X	Veces la concentración
$\mu\text{F}$	Micro Faradios
$\Omega$	Ohms

## RESUMEN

**Universidad Autónoma de Nuevo León**

**Facultad de Medicina**

**Departamento de Bioquímica y Medicina Molecular**

**M.C. Antonio Ali Pérez Maya**

Candidato al grado de Doctor en Ciencias con orientación en Biología Molecular e Ingeniería Genética.

Número de páginas: 155

### **DISECCIÓN BIOINFORMÁTICA Y EXPERIMENTAL DE LA FASE DE EXPANSIÓN DEL *LOCUS* GH EN PRIMATES.**

**INTRODUCCIÓN.** La historia evolutiva de la hormona del crecimiento (GH) en los primates es única por su evolución acelerada. El *locus* GH comenzó en este orden taxonómico con una estructura unigénica, que continuó sin grandes cambios hasta la aparición de los platirrinos. En un inicio se privilegiaron eventos de expansión génica a causa de entrecruzamientos desiguales, propiciados en principio por la presencia de elementos repetitivos alrededor y dentro del *locus* ancestral. Ya en los antropoides, los *loci* GH están integrados por múltiples genes no alélicos relacionados y específicamente en los catarrinos terminaron codificando para nuevas hormonas. La multiplicación génica del *locus* GH y la divergencia de sus miembros fueron acompañadas por diferencias en cuanto a su expresión en los tejidos, con al menos dos especificidades: hipófisis y placenta.

**OBJETIVOS.** Determinar la composición y organización génica del *locus* GH en especies representativas de primates y descubrir su historia evolutiva. **METODOLOGÍA.** Se secuenciaron y anotaron Cromosomas Artificiales Bacterianos portadores de los *loci* GH de chimpancé y gorila. Se compararon las secuencias nucleotídicas de ellos y del humano, así como las de proteínas codificadas por sus genes. **RESULTADOS Y DISCUSIÓN.** El *locus* del chimpancé incluye cinco genes organizados en una distribución similar a la del humano, con dos de estos codificando proteínas GH muy similares y tres codificando

somatomamotropinas coriónicas (CSHs). Por otro lado, el *locus* GH del gorila está formado por seis genes, dos GHs y cuatro CSHs. Existen notorias diferencias entre estos *loci* y el *locus* GH humano con respecto a las proteínas expresadas y su regulación. El gorila parece producir cuatro CSHs con propiedades biológicas potencialmente diferentes y el chimpancé dos, mientras que en el humano se produce una sola CSH. El análisis comparativo entre las proteínas GHs y CSHs ortólogos reveló que el humano y el chimpancé comparten la misma GH variante (GH-V) y que la GH normal (GH-N) de ambas especies difieren en tan sólo un aminoácido; por otra parte, el humano y el gorila comparten la misma GH-N y sus GH-V difieren en 3 aminoácidos, evidenciando que hay una presión selectiva para conservar las GHs. Las CSHs muestran un mayor número de cambios, algunos de los cuales son compartidos entre gorila y chimpancé. **CONCLUSIONES.** Los *loci* GH de humano y chimpancé son similares, con cinco genes cada uno mostrando una completa ortología. En contraste, el *locus* GH de gorila presenta seis genes y solo para los genes GHN, GHV y CSH1 se observa una clara ortología con respecto al humano y al chimpancé, para el resto de los genes se puede observar que se originaron de un ancestro común, pero en este punto es difícil establecer su ortología. Por otro lado, se observan algunas diferencias significativas. En particular, el gen hCSH-L ha cambiado sustancialmente desde la divergencia chimpancé / humano / gorila hasta convertirse potencialmente en un pseudogen, el correspondiente gen en el gorila (gCSH1), se ha conservado funcional y ha sufrido menos cambios que el del humano. Una situación similar ha experimentado en el chimpancé el gen CSH-A1 que da un producto idéntico al de CSH-A2. En los homínidos, las diferencias más importantes entre los *loci* GH se observan en los productos codificados por los genes CSHs, más que en el número de estos, notándose una intrigante reducción en las hormonas CSHs funcionales producidas (cuatro diferentes en el gorila, dos en chimpancé y únicamente una en el humano).

---

**Dr. Hugo A. Barrera Saldaña**

**Director de Tesis**

## CAPÍTULO I: INTRODUCCIÓN

### 1.1. Anatomía y fisiología del *locus* GH.

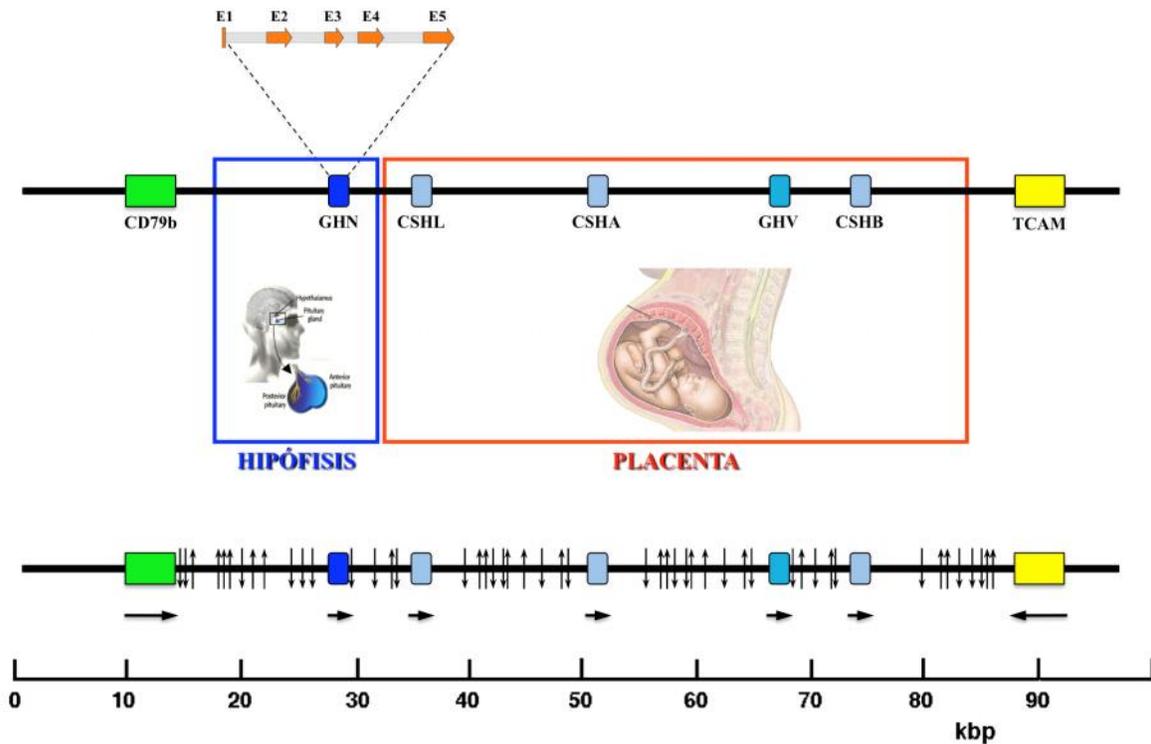
#### 1.1.1. Estructura (composición y organización).

El *locus* GH en el humano se localiza en la banda q24.2 del brazo largo del cromosoma 17 (George *et al.*, 1981; Harper *et al.*, 1982), abarca una extensión aproximada de 50 000 pares de bases (pb) y contiene cinco genes similares con la misma orientación transcripcional (figura 1) (Chen *et al.*, 1989). En su extremo 5' está flanqueado por el gen CD79b, activo en los linfocitos B (Cajiao *et al.*, 2004; Yoo *et al.*, 2006) y en su extremo 3' por el gen TCAM, activo en testículos (Sakatani *et al.*, 2000; Kimura *et al.*, 2004).

En su extremo 5', el *locus* inicia con el gen de la hormona del crecimiento hipofisiaria o hGH-N (referido como normal), cuya expresión es específica de los somatótrofos de la glándula pituitaria anterior. Le siguen los restantes cuatro genes, de expresión específica del sincitiotrofoblasto de la placenta. De estos, dos codifican para una CSH madura idéntica (hCSH-A y hCSH-B) (Barrera-Saldana *et al.*, 1983) y uno (también del tipo CSH y referido como hCSH-L) es considerado originalmente un pseudogen por poseer una mutación en su segundo intrón que evitaría el procesamiento correcto de sus transcritos (Resendez-Perez *et al.*, 1990), aunque ello está en duda pues ha sido reportada evidencia de su expresión en placenta (Misra-Perez *et al.*, 1994; MacLeod *et al.*, 1992; Hu *et al.*, 1999; Männik *et al.*, 2010). El gen restante es el hGH-V, el cual codifica para una variante de la GH que se postula sustituye a la GH hipofisiaria durante el embarazo (Chen *et al.*, 1989; Alsat *et al.*, 1997).

De la región que abarca el *locus* GH humano (~ 50 kpb) las unidades transcripcionales de sus cinco genes no ocupan más de 2 kpb cada una, estando separadas por cuatro regiones intergénicas: dos de 13 kpb, una de 6.5 kpb y una de 6 kpb (Chen *et al.*, 1989). En estas regiones se han identificado tanto secuencias repetitivas, como elementos que participan en la regulación de la expresión de los genes. De hecho, dentro del *locus* se encuentran 48 copias de secuencias repetitivas de la familia *Alu* y una de la familia *Kpn* (figura 1). Las inserciones *Alu* ocurren en ambas orientaciones y tres de ellas están

truncadas. Cabe señalar que más del 20% de las secuencias nucleotídicas del *locus* hGH corresponde a estos elementos repetitivos, cuyo patrón de distribución, semejante para los cinco genes, sugiere su posible implicación en la génesis del *locus* (Chen *et al.*, 1989). La composición y organización de éste se ilustra en la figura 1.



**Figura 1. Anatomía del *locus* hGH.** Se muestra la composición y organización genética dentro del *locus*, la anatomía canónica de los genes GH/CSH, los *loci* que lo flanquean, la orientación (flechas) y el tejido donde se expresan. En la parte inferior se representa la distribución de elementos Alu.

### 1.1.2. Los genes.

A pesar de sus diferencias en cuanto a especificidad tisular, niveles de expresión y funciones de sus proteínas, los genes GH / CSH presentan una elevada similitud en sus secuencias nucleotídicas (superior al 90%) y exhiben la misma anatomía consistente en cinco exones separados por cuatro intrones.

El gen hGH-N genera como su producto principal una proteína de 191 aminoácidos (HGH 22 kDa) de cadena única y con dos puentes disulfuro. Durante la maduración del transcrito, se generan otros dos ARNm que se traducen en sendas isoformas. Una de 20

kDa (176 aa) que constituye aproximadamente el 10% de la GH liberada por la glándula pituitaria, con la particularidad de carecer de los residuos del 32 al 46 (Baudet *et al.*, 2007; DeNoto *et al.*, 1981; Baumann, 2009) a consecuencia del uso en la génesis de su RNAm de un sitio alternativo de empalme en el exón 3 (Baumann, 2009; Barrera-Saldaña *et al.*, 1983, Lecomte *et al.*, 1987). Y una tercera de 17.5 kDa, caracterizada por carecer de los residuos 32 al 71 por haberse omitido el exón 3 en la génesis de su ARNm. Al parecer, esta isoforma está presente sólo en condiciones patológicas, cuando el gen GH1 ha mutado (deficiencia genética de GH tipo II) y no se ha visto expresada de manera significativa en condiciones normales (Lecomte *et al.*, 1987).

El producto más abundante del gen GH-V es también una proteína de cadena única de estructura muy similar a la GH-N. Está formada por 191 residuos (22 kDa) y posee dos puentes disulfuro, pero difiere de aquella en 13 posiciones aminoacídicas y en ser de un punto isoeléctrico mayor. Otras cuatro isoformas también se derivan del gen hGH-V (Frankenne *et al.*, 1987-A; Frankenne *et al.*, 1987-B). Una se origina de la retención en el ARNm del intrón D, resultando en un polipéptido de 26 kDa (230 aa) que presenta una divergencia notable en su carboxilo terminal con respecto a la isoforma principal (Lecomte *et al.*, 1987; Untergasser *et al.*, 1998). Dicha isoforma se cree que es una proteína de membrana. Otra difiere de la isoforma principal por el hecho de que se encuentra glicosilada en la asparagina 140, lo cual incrementa su peso a 25 kDa (Ray *et al.*, 1989). Una más, de 24 kDa (210 aa), resulta de un transcrito inmaduro que al carecer de cuatro nucleótidos (GTGG) hacia el inicio del intrón D, utiliza un segundo sitio donador de empalme y da lugar a un ARNm alternativo cuya traducción le origina. Esta isoforma también se ha encontrado en los testículos (Untergasser *et al.*, 1998; Boguszewski *et al.*, 1998). Una quinta isoforma resulta de otro transcrito inmaduro, que como en el caso de su contraparte proveniente del gen hGH-N, experimenta una delección de 45 pb en el exón 3, lo que genera una proteína de 20 kDa (Boguszewski *et al.*, 1998). El análisis estructural de las principales isoformas derivadas de estos genes, revela que las proteínas globulares que codifican poseen cuatro hélices alfa y contienen dos puentes disulfuro (Barrera-Saldana *et al.*, 1982, DeNoto *et al.*, 1981).

### 1.1.3. Función de las hormonas codificadas.

La principal función de la HGH-N consiste en promover el crecimiento lineal postnatal y regular diversas actividades metabólicas, dentro de las que destacan la homeostasis de la glucosa, la síntesis proteica en el tejido muscular, la lipólisis en el tejido adiposo, la reparación tisular y la regeneración ósea (Farreras-Rozma, 2000). Más recientemente se le ha involucrado en la neurogénesis y la neuroprotección, el aprendizaje y la memoria (Sonntag *et al.*, 2005).

Por su parte los niveles de la GH-V (principal isoforma) se asocian con índices de crecimiento intrauterino, peso al nacer y con niveles fetales del factor de crecimiento parecido a la insulina I (IGF-I) (Verhaeghe *et al.*, 2002; Chellakooty *et al.*, 2004; Rygaard *et al.*, 1998). Esta isoforma presenta actividad somatogénica (MacLeod *et al.*, 1991), lactogénica (Nickel *et al.*, 1990) y como se señaló arriba, sustituye a la GH-N durante la segunda mitad de la gestación (Frankenne *et al.*, 1987-A; Frankenne *et al.*, 1987-B); además se le ha implicado en el aumento de los niveles de IGF-I en el suero materno (Caufriez *et al.*, 1990).

Un papel similar se le adjudica a la CSH, cuyos niveles se han asociado con el buen funcionamiento placentario y también con los niveles fetales de IGF-1 (Hardwerger *et al.*, 2000; Karabulut *et al.*, 1972). Esta hormona, también de 22 kDa, se piensa que juega un importante papel como antagonista de la insulina en la regulación del metabolismo materno durante el embarazo (Josimovich *et al.*, 1964).

Los efectos contrarios a la insulina que tiene la CSH humana comprenden la inhibición de la síntesis de glucógeno, de la glucólisis, de la lipogénesis y del ingreso de glucosa a las células; además de la activación de la glucogenólisis, de la proteólisis en el músculo y de la gluconeogénesis a partir de los aminoácidos. Estas acciones tienen como propósito aumentar la concentración plasmática de nutrientes y energía para el feto (Ruiz-Parra *et al.*, 1999; Felber *et al.*, 1972; Williams y Coltart, 1978; Walker *et al.*, 1991; Kim y Felig, 1971). Además ejerce una actividad somatotrópica directa sobre los tejidos fetales (Felber *et al.*, 1972; Williams y Coltart, 1978). Durante el tercer trimestre del embarazo se

produce un incremento en la síntesis de esta hormona, llegando a representar el 10% de la producción total de las proteínas placentarias (Jacquemim *et al.*, 1994).

#### **1.1.4. Regulación neuroendocrina de la secreción de GH en humanos.**

La secreción de la GH-N está regulada bajo el control dual hipotalámico ejercido por un factor que la estimula: GHRH (del inglés growth hormone releasing hormone) y otro que la inhibe: SS (somatostatina). Cada episodio de secreción de GH esta dada por una liberación del neuropéptido GHRH al sistema portal hipotalámico-hipofisiario (sistema de doble capilarización, donde a nivel de las células neurosecretoras hipotalámicas se encuentra la primera capilarización a la cual se vierten las hormonas del hipotálamo, que viajan hacia la adenohipófisis o lóbulo anterior, donde se produce la segunda capilarización, que permite a dichas hormonas alcanzar las células glandulares estimulando o inhibiendo su secreción) y precedido por una disminución de SS. Este mecanismo trabaja en conjunto con una serie de factores como: neurotransmisores, otras hormonas y metabolitos intermediarios (aminoácidos, glucosa y ácidos grasos libres) (Frohman, 2009).

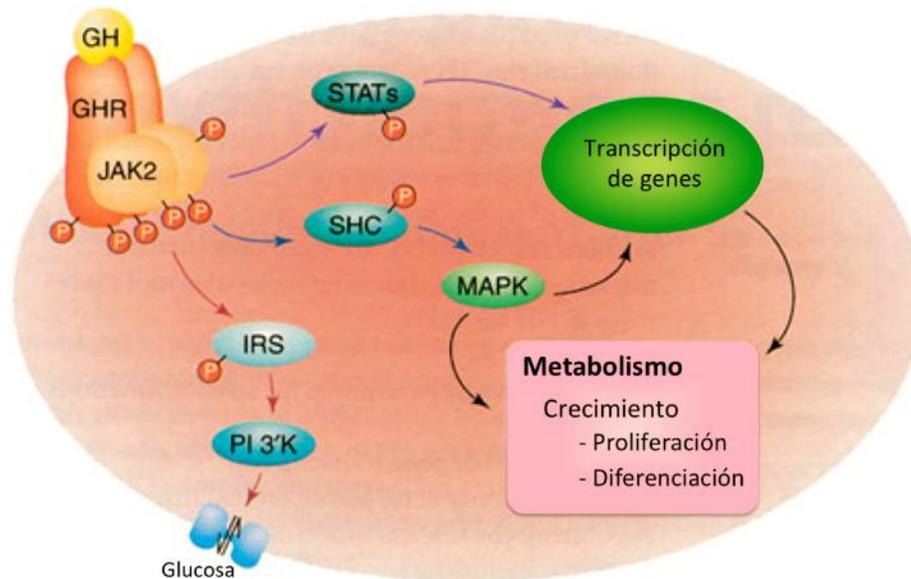
GHRH es un péptido de aproximadamente 44 aminoácidos (aa) que pertenece a la familia glucagón-secretina. Las neuronas que lo producen se encuentran localizadas en el núcleo arqueado del hipotálamo, cuyas proyecciones axonales alcanzan la eminencia media y liberan el péptido al sistema portal hipotalámico-hipofisiario. En la pituitaria, la GHRH se une a su receptor en los somatotrofos, el GHRHR (del inglés growth hormone releasing hormone receptor), para promover la liberación y la síntesis de GH, así como el crecimiento, proliferación y diferenciación de estas células (Frohman, 2009; Bhagavan y Ha, 2011).

La liberación de GHRH y SS está regulada a su vez principalmente por neurotransmisores: los  $\alpha_2$ -adrenérgicos y las dopaminas (adrenalina, noradrenalina, etc.) inhiben la liberación de SS y por lo tanto promueven la expresión y liberación de GH. La falta de hormonas tiroideas afecta la síntesis hipotalámica de GHRH. Los glucocorticoides estimulan la transcripción del gen GH-N y aumentan el número de receptores GHRH. Debido a todo esto, se puede apreciar que la regulación en la expresión y liberación de GH

no es un mecanismo simple, sino que resulta de un conjunto de vías que regulan de manera coordinada este proceso (Devesa *et al.*, 1992; Mayo *et al.*, 1995).

### 1.1.5. Mecanismos de acción moleculares de la GH.

La GH, una vez liberada al torrente sanguíneo, llega a su receptor GHR (del inglés growth hormone receptor), que es una proteína de aproximadamente 70 kDa integrante de la superfamilia de los receptores tipo citoquinas/hematopoyetina compuesta por dos subunidades. La unión de la GH al receptor dimérico induce un cambio conformacional que produce la activación de la tirosin-quinasa JAK2 (del inglés Janus activated kinase 2) (Gent *et al.*, 2002), lo que promueve la fosforilación del dominio citoplasmático del receptor y de la propia JAK2 en los residuos de tirosina (figura 2).



**Figure 2. Vías de señalización del receptor GH iniciadas por la activación de JAK2.** JAK2 fosforila: 1) a varios miembros de la familia de factores de transcripción STAT (flechas púrpuras) y 2) a las proteínas SHC, activando así la vía MAPK (flechas azules). De estas dos maneras la GH es capaz de regular la transcripción de genes. JAK2 también fosforila a proteínas IRS, activando de este modo a la PI3 quinasa (PI3'K: flechas rojas), que posee un papel fundamental en la estimulación del transporte de glucosa. Abreviaturas: GH, hormona del crecimiento; GHR, receptor de la hormona de crecimiento; IRS, sustratos del receptor de insulina; JAK2, Janus quinasa 2; MAPK, proteína quinasa activada por mitógenos; P, fosfato; PI3'K, fosfatidilinositol 3 quinasa; STAT, transductores de señal y activadores de la transcripción. Rediseñado de: Diamanti-Kandarakis *et al.*, 2004.

A continuación un sin número de otras proteínas son fosforiladas en los residuos de tirosina formando sitios de unión para proteínas de señalización que incluyen: 1) proteínas SHC, las cuáles inician la ruta Ras-MAP quinasa que posee un rol crítico en la regulación del crecimiento y la diferenciación (Herrington y Carter-Su, 2001; Diamanti-Kandaris *et al.*, 2004). También se induce la vía Akt, la cual promueve la supervivencia celular (Bhagavan y Ha, 2011).

La GH activa también a los activadores de la transcripción y transductores de señales (STATs), los cuáles regulan la transcripción (Diamanti-Kandarakis *et al.*, 2004; Herrington *et al.*, 2000). Estos factores citoplasmáticos viajan al núcleo, en donde se unen a sitios específicos en el DNA y activan la expresión de varios genes denominados de “expresión temprana” como *c-fos* y *c-jun*. Los productos proteicos de dichos genes son factores de transcripción implicados en los procesos de diferenciación y crecimiento celular. La GH también induce la expresión del gen *c-myc* que codifica para otro factor de transcripción involucrado en la proliferación celular y la de genes que participan en la diferenciación adipogénica, como los que codifican para: la glicerol fosfato deshidrogenasa (GDPH), la enzima lipogénica aP2, la proteína de unión a lípidos y la lipoproteína lipasa (LPL) (Bhagavan y Ha, 2011).

La GH responde de manera rápida a los efectos de la insulina sobre el transporte de glucosa a consecuencia del reclutamiento de los transportadores Glut1 y Glut4, que promueven el paso de la glucosa del plasma a la célula. La GH estimula la fosforilación en residuos de tirosina del receptor de insulina IRS-1, los cuáles se piensa que regulan eventos metabólicos en la célula (Diamanti-Kandaris *et al.*, 2004) lo que crea múltiples sitios de unión para proteínas con dominios SH2, incluyendo la subunidad de 85 kDa de la proteinquinasa PI3K. Esta última ha sido implicada en el transporte de glucosa, síntesis de DNA y la actividad de la enzima p70<sup>tsk</sup>, así como en la regulación del ciclo celular. La GH activa la vía del IP3. La hormona induce pequeñas cantidades de diacilglicerol (DAG), que es un activador de la protein quinasa C, lo cual estimula la lipogénesis, induce la expresión de *c-fos*, estimula la unión de una proteína nuclear al DNA en el sitio C/EBP e incrementa el calcio intracelular (Carter-Su *et al.*, 1996).

En particular, los mecanismos a través de los cuáles la GH disminuye la lipogénesis e incrementa la lipólisis, pueden involucrar tanto a una acción dependiente de STAT, como a una activación de una fosfodiesterasa específica dependiente de IP3 quinasa, mientras la estimulación de la síntesis de proteínas por parte de la GH se piensa que es mediada vía tirosinas específicas en el dominio citoplasmático del GHR (Herrington y Carter-Su, 2001; Diamanti-Kandaridis *et al.*, 2004).

En el hígado la GH regula la expresión del gen IGF-1, que es un mediador sérico de las acciones de esta hormona. IGF-1 regula funciones como: neurogénesis, crecimiento, supervivencia celular, formación de mielina, sinaptogénesis, neuroprotección y síntesis de proteínas, entre otras (Carter-Su *et al.*, 1996).

## **1.2. Regulación de la expresión de los genes del *locus* GH en el humano.**

En 1975, Marie-Claire King y Alan Wilson plantearon la paradoja de que a pesar de que los seres humanos y los chimpancés son tan similares a nivel de proteínas, su anatomía es muy diferente. King y Wilson hipotetizaron que se trataba de cambios en los "sistemas que controlan la expresión génica" que cuentan para las diferencias morfológicas (King y Wilson, 1975). Esta explicación estaba en desacuerdo con la postura bien establecida que enfatizaba una evolución a nivel de proteínas a través de duplicaciones génicas y de mutaciones en la secuencia aminoacídica (Ohno, 1970). Sin embargo, la idea de que los cambios en la regulación génica, en oposición a los cambios en los genes mismos, constituyen la principal fuerza subyacente para la diversidad morfológica, es una hipótesis que ha ido ganando apoyo experimental en la última década (Carroll, 2005).

### **1.2.1. Regulación por elementos cercanos.**

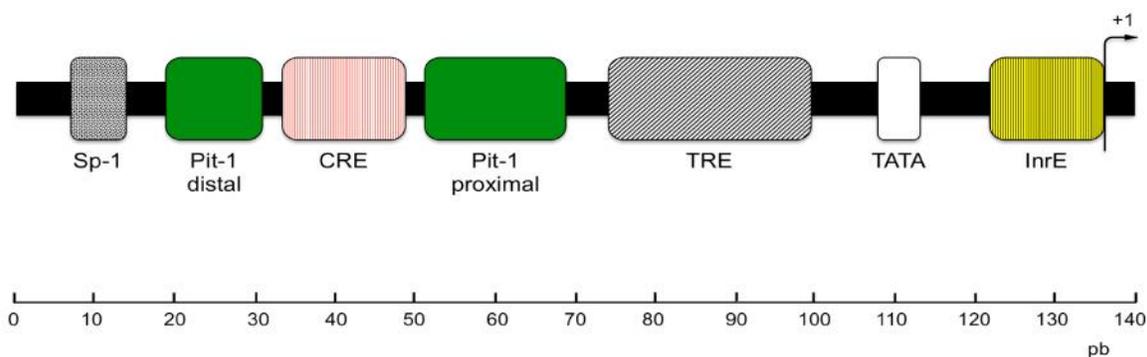
La GH se encuentra bajo una estricta regulación principalmente a nivel de la síntesis de su ARNm, por procesos de retroalimentación desde el hipotálamo, el hígado, las glándulas suprarrenales y las glándulas tiroideas (Giustina y Veldhuis, 1998). Entender la regulación de la GH, ayudará en la comprensión global del crecimiento y el metabolismo de los vertebrados.

Los principales reguladores del gen de la GH son el factor específico de hipófisis, Pit-1, glucocorticoides, hormonas tiroideas, la proteína específica 1 (SP1) y la proteína de unión al elemento de respuesta a AMPc (CREB). La expresión de los genes está estrechamente controlada por los mecanismos que operan a niveles de organización nuclear, estructura de los cromosomas y modificaciones del ADN (Kimura *et al.*, 2004).

Uno de los puntos claves de la regulación de la expresión de los genes, es el control transcripcional. En esta categoría está incluida la acción de los promotores (figura 3), la presencia de secuencias regulatorias accesorias (potenciadores y represores) y la interacción entre las múltiples proteínas activadoras o inhibidoras que se unen a dichas secuencias (figura 4). El control transcripcional no sólo se ocupa de "encender" o "apagar" genes, sino también de regular la intensidad de la transcripción de los genes encendidos.

#### **1.2.1.1. Promotor.**

El gen GH requiere una única combinación de factores de transcripción para su activación. Un grupo de factores *trans* activadores comunes - Sp1, USF/MLTF y NF-I - están involucrados en el establecimiento del nivel basal de la transcripción del gen hGH-N en la hipófisis (Jacquemin *et al.*, 1994). La expresión del gen hGH-N en los somatotrófos de la pituitaria anterior es dirigida por el factor transcripcional Pit-1 (figura 4), también conocido como GHF1 (Karin *et al.*, 1990). El factor Pit-1 se une al promotor del gen GH-N en dos sitios (figura 3): uno proximal (-92 a -65) y otro distal (-130 a -105) (Kimura *et al.*, 2004). Por ser Pit-1 un factor específico de tejido hipofisiario (Ingraham *et al.*, 1988), aparentemente no está involucrado en la inducción de la expresión de los genes placentarios. Este factor presenta dos dominios conservados: un dominio POU importante para la activación transcripcional y un dominio homeobox (HB) que interviene en su unión al ADN, ambos ubicados en el extremo carboxilo terminal de la proteína (Schanke *et al.*, 1997).



**Figura 3. Distribución de elementos regulatorios del promotor proximal de los genes del locus hGH.** Con una flecha se representa el sitio de inicio de la transcripción.

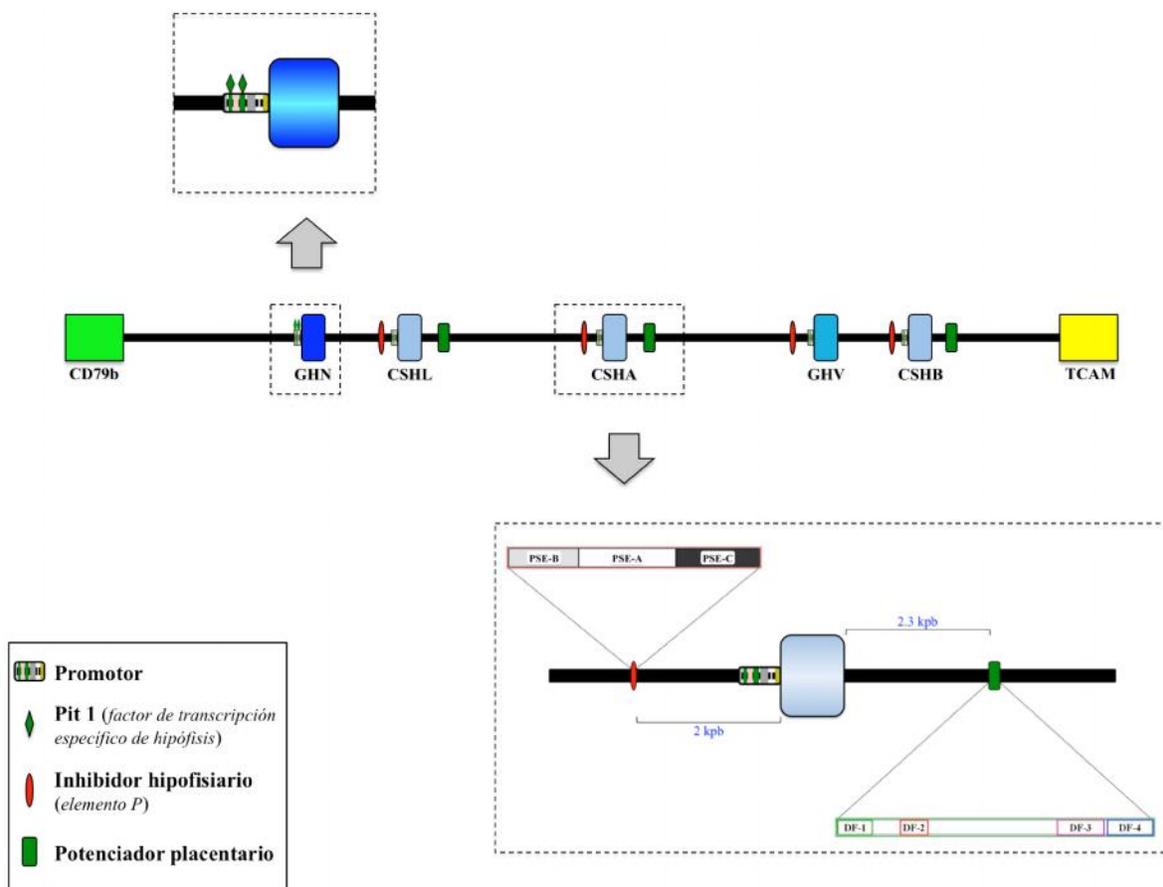
Aunque aún quedan muchas interrogantes sobre la regulación transcripcional en la placenta de los genes GH/CSH placentarios, ya se han realizado avances en este sentido. Para los genes hCSHs se ha identificado otro elemento regulador referido como elemento Iniciador o InrE, el cual se localiza entre la caja TATA y el sitio de iniciación de la transcripción (ver figura 3). Mutaciones en esta secuencia reducen en un 60% la expresión del promotor correspondiente (Jiang *et al.*, 1995).

### 1.2.1.2. Represor.

A pesar de que los promotores de los genes placentarios también presentan sitios de unión a Pit-1, estos están sujetos a una represión transcripcional mediada por una región conservada de 263 nucleótidos denominada elemento P o inhibidor hipofisiario, situada aproximadamente a 2 kpb enfrente del sitio de inicio de la transcripción (Norquay *et al.*, 2001), a la cual se unen en la hipófisis proteínas que luego interfieren con la unión de Pit-1 a los promotores de dichos genes, impidiendo su expresión en esta glándula (Nachtigal *et al.*, 1993; Barrera-Saldaña, 1998). El elemento P contiene tres dominios: PSE-A, PSE-B y PSE-C (figura 4). El primero interacciona con miembros de la familia del Factor Regulador X (RFX), el segundo lo hace con el Factor Nuclear-1 (NF-1, mientras que el tercero interactúa con HNF-3 y C/EBP) (Norquay *et al.*, 2001; Nachtigal *et al.*, 1993, Norquay *et al.*, 2003).

### 1.2.1.3. Potenciador.

Aproximadamente a 2.3 kpb después de cada gen hCSH se localiza un potenciador, que incrementa la expresión de estos en la placenta (Jacquemin *et al.*, 1994; Barrera-Saldaña, 1998; Rogers *et al.*, 1986). Ensayos de protección contra digestión con DNAsa I, han revelado en este potenciador la presencia de cuatro sitios de unión, denominados DF-1, DF-2, DF-3 y DF-4, para proteínas nucleares de placenta (figura 4) (Jacquemin *et al.*, 1994), habiéndose demostrado que la actividad potenciadora del enhancer en placenta está mediada por un sinergismo entre los sitios DF-3 y DF-4 (Jacquemin *et al.*, 1996).



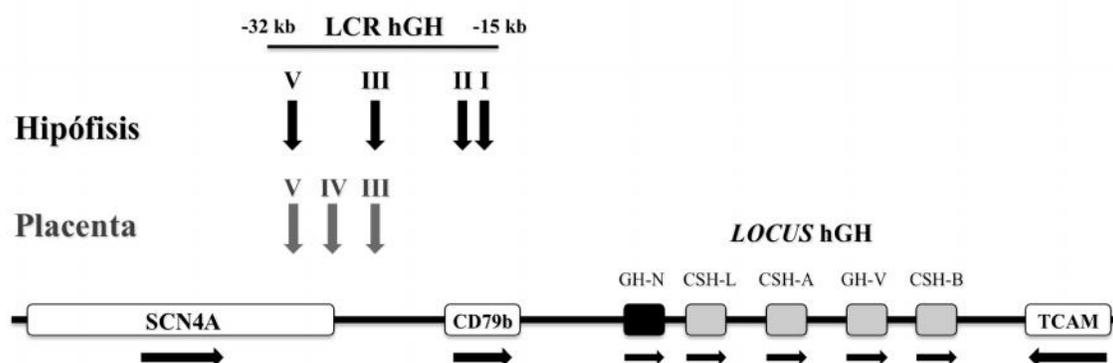
**Figura 4.** Detalles de sitios de regulación de la expresión de los genes del *locus* GH. Los dos sitios Pit-1 presentes en el promotor de hGH-N se esquematizan antes de dicho gen. El sitio del inhibidor hipofisiario o elemento P se representa antes de cada gen placentario y que ampliado destaca sus dominios PSE-A, PSE-B y PSE-C. El potenciador placentario se representa después de los genes CSHs y ampliado resalta los cuatro sitios de unión a proteínas nucleares de placenta que éste presenta: DF-1, DF-2, DF-3 y DF-4.

Se ha descrito que la función de este potenciador es independiente de su orientación y posición (Rogers *et al.*, 1986), así como que su actividad requiere de la unión cooperativa a varios sitios localizados dentro de este elemento, del factor potenciador de la transcripción-1 (TEF-1) y de un factor potenciador específico de la placenta denominado CSEF-1. Los potenciadores de los tres genes hCSHs presentan una alta similitud en su secuencia, sin embargo, las copias en genes CSHL y CSHA muestran una débil actividad de estimulación transcripcional en comparación a la de CSHB, debido a diferencias en sus respectivos sitios de unión a proteínas DF-3 (Jacquemin *et al.*, 1996).

### **1.2.2. Regulación por elementos distantes.**

El *locus* GH está bajo el control de un elemento regulatorio llamado región de control del *locus* (LCR) que contribuye al patrón de expresión tejido específico de los genes[8]. Este tipo de elemento tiene la capacidad de activar la expresión de genes blanco y de establecer, en la cromatina, dominios autónomos de regulación transcripcional. Estos se caracterizan por la presencia de sitios hipersensibles a la enzima DNAsa I (Hipersensitive Sites, HS), identificados como HSI a HSV y localizados entre 15 y 32 kb antes del *locus* (Figure 5). Esta hipersensibilidad está asociada con cambios en la estructura de la cromatina, que a su vez indican eventos de actividad transcripcional de genes blanco (Elefant *et al.*, 2000; Shewchuk *et al.*, 2002).

En general, una estructura compacta de la cromatina se asocia con una ausencia de actividad transcripcional y las histonas favorecen esta estructura particular. El proceso de activación transcripcional se asocia con el cambio a una estructura abierta de la cromatina que se logra mediante la acetilación de las histonas.



**Figura 5. La Región de Control del Locus hGH.** Se representa con un cuadrado negro el gen hGH-N y con cuadrados grises los cuatro genes de expresión placentaria que componen el locus hGH. Los rectángulos blancos representan los genes (SCN4A y CD79b) no relacionados que se encuentran río arriba del locus GH y el gen TCAM que se encuentra río abajo del mismo. Las flechas horizontales muestran la orientación transcripcional de los genes del locus hGH. La posición de cada sitio HS en la LCR del locus hGH está marcada con flechas verticales. Modificado de (Schewchuk *et al.*, 2006).

Hay evidencias de que la transcripción no codificante del gen cercano CD79b es esencial para la regulación por la LCR (Reséndez-Pérez *et al.*, 1990). La transcripción continua no codificante de CD79b, permite una conformación abierta de la cromatina y la presencia constante de la ARN polimerasa II (ARN pol II). La transcripción regulada del locus GH comienza con la unión de Pit-1 a HS-I y al promotor del gen GH-N, lo cual dispara la interacción entre estas dos regiones a través de un bucle que permite la yuxtaposición del promotor de GH-N en una región de alta actividad transcripcional, el dominio transcripcional LCR/CD79b (Schewchuk, 1999).

Pit-1 es esencial para el control coordinado de la transcripción del gen GH-N humano. La unión de Pit-1 a los sitios HS-I y HS-II es necesaria y suficiente para regular la expresión tejido específica, más que su unión al promotor del gen GH-N. El control específico de genes a través de la LCR es coordinado por la formación de dominios de histonas modificadas (hiperacetilación o metilación) y de bucles de cromatina (Schewchuk, 1999).

La expresión diferencial de los miembros del *locus* hGH en placenta e hipófisis depende en gran medida de un control ejercido por distintos subgrupos de la LCR. (Jones *et al.*, 1995; Elefant *et al.*, 2000, Kimura *et al.*, 2004). Regiones internas del *locus* multigénico son selectivamente acetiladas en la cromatina placentaria y una de estas regiones comprende precisamente a los elementos P situados frente a los genes placentarios e implicados en la represión de su expresión en hipófisis (Kimura *et al.*, 2004).

Los sitios HSI y HSII, son específicos para la expresión en la pituitaria y la acetilación de las histonas en una región centrada en estos sitios se asocia con la expresión del gen hGH-N en esta glándula. El sitio HSIV es específico de placenta y la acetilación de histonas asociada a la expresión de los genes placentarios está limitada a la región que abarca HSIII-HSV. Los sitios HSIII y HSV, localizados a 28 kpb y 32 kpb río arriba del gen hGH-N, se encuentra en ambos tejidos.

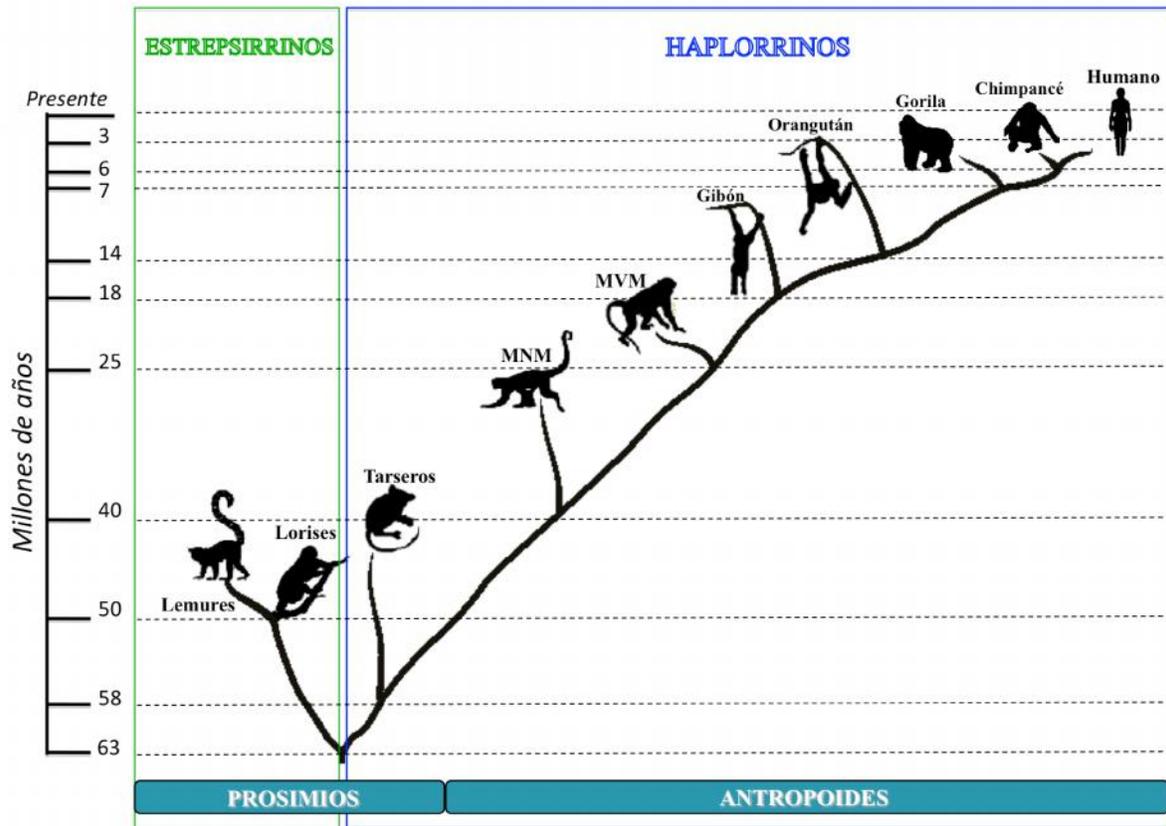
Los sitios HSI y HSII flanquean el extremo 5' del gen CD79b y los tres sitios restantes (HSIII-HSV) están localizados en la segunda mitad del gen SCN4A, mismo que codifica para un canal de sodio específico de músculo esquelético. La región HS I/II contiene tres sitios de unión a Pit-1 que actúan como un potenciador del promotor de la GH hipofisiaria. Tanto el promotor como la región HSI/II son requeridos para la expresión eficiente del gen GH-N (Reséndez-Pérez *et al.*, 1990).

### **1.3. Los primates. Clasificación.**

El proyecto del genoma humano ha revolucionado diversos campos como la investigación biomédica, la genética clínica y la biotecnología médica. Sin embargo, en los avances que se han logrado, se haya ausente la interpretación de los resultados en un contexto evolutivo, como es el caso de la jerarquía filogenética de las especies de primates. El número exacto de los géneros de primates es controvertido y el conteo de especies va desde 261 hasta 377 (Steiper y Young, 2009), de los cuales hasta ahora (Septiembre de 2012) se ha completado la secuenciación del genoma de doce de ellas. Para entender la evolución y estructura del genoma humano, lo que de paso redundaría en avances de la

biomedicina, es necesaria una representación más amplia del genoma de los primates más cercanos al hombre.

La taxonomía de los primates ha sido objeto de una considerable revisión, pero las opiniones actuales están de acuerdo (Steiper y Young, 2009) que entre 67 y 69 géneros de primates se originaron a partir de un ancestro común en el límite Cretácico/Paleoceno, hace aproximadamente 80 a 90 millones de años. Una expansión en el Eoceno originó los principales linajes existentes: 1) Strepsirrinos, que se compone de Lorisiformes (gálagos, potos y loris), Chiromyiformes (aye-aye de Madagascar) y Lemuriformes (lémures de Madagascar), 2) Tarsiiformes (tarseros) y 3) Simiiformes compuesto por Platiirrinos (monos del Nuevo Mundo) y catarrinos, que incluye Cercopithecoidea (monos del Viejo Mundo) y Hominoidea (gibones, grandes simios y humanos) (ver figura 6). La taxonomía de los primates es compleja e inicialmente fue realizada a partir de rasgos morfológicos, adaptativos, bio-geográficos, reproductivos y de comportamiento, con inferencias a partir de registros fósiles (Steiper y Young, 2009). La aplicación reciente de los datos genéticos moleculares para resolver la sistemática de los primates ha sido informativa, pero de alcance limitado y restringido a subconjuntos específicos de taxones. Los esfuerzos para superar esta deficiencia utilizando un enfoque de supermatriz (Freeman *et al.*, 2006, Birney *et al.*, 2007), con las secuencias publicadas en estudios previos, han sido infructuosos por la proporción excesivamente grande de datos faltantes para cada taxón (Birney *et al.*, 2007). Sin embargo, estudios recientes emplean secuenciación a gran escala y amplio muestreo de taxones para proporcionar una filogenia más resuelta que confirma, mejora y extiende las representaciones anteriores de la especiación en primates. A su vez, una filogenia más clara proporciona un marco sólido para nuevas representaciones de los patrones de evolución del genoma entre los linajes de primates. Esta visión es esencial en la genómica comparativa actual y en los futuros estudios de selección y adaptación en los primates, incluido el humano (Steiper y Young, 2009)



**Figura 6. Evolución de los primates.** La construcción de este árbol filogenético se apoyó en registros paleontológicos.

El Orden Primates se organiza en dos subórdenes: estrepsirrinos (lemures, lorises) y haplorrinos (tarseros, monos, simios y humanos). El “árbol del tiempo” (una representación filogenética calibrada a los tiempos de divergencia entre las especies, que usa una estructura jerárquica en correspondencia con el árbol de la vida), indica que estos dos subórdenes divergieron antes del límite Mesozoico-Cenozoico hace unos 78 millones de años (Ma), con la diversificación de familias individuales principalmente durante el Paleógeno tardío (~ 45 Ma) y Neógeno temprano (20 Ma).

La posición filogenética de los Tarseros es quizás la más enigmática dentro de los primates. Los tarseros se han relacionado con los estrepsirrinos basado en sus similitudes adaptativas (Lowe *et al.*, 2007). Sin embargo, rasgos morfológicos como un labio continuo, une a este linaje con los antropoides (monos y simios) en el Suborden Haplorrino (Birney *et*

*al.*, 2007). Aunque hay alguna discordancia entre los estudios moleculares, los trabajos recientes sobre los elementos nucleares pequeños interespaciados (SINEs), apoyan la inclusión de los tarseros en el taxón Haplorrino (Marques-Bonet *et al.*, 2009; Jones *et al.*, 1995).

Los Catarrinos, compuesto por los monos del Viejo Mundo (Cercopithecidae) y grandes simios (Hominoidea, que incluye a Hylobatidae y Hominidae), tienen fosas nasales estrechas, una reducción en los premolares y un tubo timpánico largo y continuo.

La divergencia entre platirrininos y catarrinos estimada en aproximadamente 44 Ma, así como la diversificación de las familias de platirrininos actualmente existentes en aproximadamente 21 Ma, indica que el antepasado de los monos del Nuevo Mundo emigró a América del Sur en algún momento durante este período (Steiper y Young, 2009).

### **1.3.1. Secuenciación y genómica comparativa en primates.**

Los proyectos genómicos de los primates que actualmente son secuenciados, no se llevan a cabo con el mismo estándar de calidad que el del genoma humano. Esta situación aumenta la frustración para diferenciar los artificios de las verdaderas diferencias en secuencia en el borrador de baja cobertura ensamblado y ha hecho que se reconozca la necesidad de ensamblar borradores genómicos de mayor calidad (Marques-Bonet *et al.*, 2005).

Hasta el momento han sido secuenciados los genomas de más de diez primates y alrededor de veintidos nuevos proyectos de secuenciación se encuentran en curso (Genbank Database, Noviembre, 2012).

¿Qué hemos aprendido de los genomas de los primates? Los genomas de los primates son entidades altamente dinámicas que han evolucionado rápidamente, pero no uniformemente a través del tiempo. Debido a la amplia información genómica disponible y a las nuevas herramientas bioinformáticas desarrolladas, la variación estructural y las

diferencias en el número de copias, han emergido como otro aspecto importante de la variación genética de los primates (Marques-Bonet *et al.*, 2005).

Con excepciones limitadas como el gibón, existen relativamente pocas diferencias entre los cariotipos de los primates. Una gran parte de las diferencias entre humanos y simios ha sido correctamente caracterizada y mayormente resuelta a nivel molecular. Sin embargo, la mayoría de los cambios nucleotídicos estructurales más pequeños no pueden ser detectados mediante las técnicas citogenéticas estándares, sino con la secuenciación de los genomas de los primates no humanos (Marques-Bonet *et al.*, 2005).

Al obtener la secuencia de un genoma se logra el acceso a la información que éste contiene, pero en este punto aún sería necesario hallar la manera de interpretar dicha información. El análisis de los genomas se refiere a la tarea de entender qué dice la secuencia de un genoma; básicamente qué genes contiene, dónde se encuentran y qué funciones realizan las proteínas que codifican (Rouze *et al.*, 1999; Stein, 2001; Rust *et al.*, 2002). A lo largo de la evolución, los cambios que sufren los genomas y sus genes, reflejan la acción de la selección natural. De las ventajas o desventajas que estos cambios acarreen, dependerá que se consoliden en las siguientes generaciones. La comparación de genomas, genes y proteínas de distintas especies, es la aproximación más intuitiva y directa para interpretar “el manual de instrucciones” que hay en los genomas (Koonin, 2001; Salzberg, 2003).

Cuando se comparan las secuencias de aminoácidos de proteínas que realizan una misma función en organismos distintos, se observa que son parecidas y que se puede establecer una correspondencia entre los residuos aminoacídicos en ciertas posiciones de unas y otras, existiendo algunas que son idénticas y otras donde hay diferencias. En el contexto de la evolución molecular, estas proteínas se parecen porque tienen un origen común y las diferencias se deben a que a lo largo del tiempo los genes han divergido por la acumulación de cambios o mutaciones en sus secuencias (Alberts, 2006).

La comparación de proteínas ortólogas (aquellas que tienen un origen común) proporciona mucha información acerca de cómo se ha producido la evolución de los organismos y las proteínas mismas, ya que el hecho de que unas posiciones aminoacídicas se conserven, está relacionado con cómo son clave para la estructura o la función de la proteína (Miller *et al.*, 2004), así como también con la importancia que tenga esta última para la sobrevivencia del organismo. En este sentido, los residuos más directamente implicados en el desempeño de la función, suelen presentar menor variación, ya que cambios en estos pueden provocar una pérdida de función (Alberts 2006, Khaitovich *et al.*, 2006).

La importancia de estos estudios comparativos de proteínas homólogas fue destacada por primera vez hace medio siglo (Zuckerlandl y Pauling, 1965) y desde entonces ha sido puesta de manifiesto una y otra vez en numerosos trabajos. Pero no siempre es fácil obtener información comparando las secuencias, ya que muchas veces lo que observamos no está ahí por ser óptimo, sino como consecuencia de un largo proceso evolutivo; como dijo en alguna ocasión el doctor Sydney Brenner, ganador del premio Nobel de medicina: "Si las matemáticas son el arte de lo perfecto y la física es el arte de lo óptimo, la biología no es más que el arte de lo satisfactorio: cualquier cosa sirve, siempre que funcione". Esto se puede expresar de forma más precisa utilizando los términos de selección positiva, negativa y neutra, que se refieren a cómo resulta un cambio para un organismo (Abascal, 2003).

Los eventos de duplicación génica han sido clave para moldear los genomas y familias génicas. En éstas, la generación de nuevas copias del gen ancestral permite que las copias queden libres para mutar y especializarse con el fin de llevar a cabo funciones diferentes, representando un cambio ventajoso para el individuo portador (Khaitovich *et al.*, 2006). La especialización se produce de forma gradual, permitiendo que series repetidas de duplicación y divergencia durante millones de años, conduzcan a que un gen de origen a una familia de genes dentro del mismo genoma.

Las diferencias nucleotídicas de relevancia entre los miembros de una misma familia, son aquellas que se concentran en sus regiones codificantes y regulatorias. Existen ejemplos de familias en las cuales la diferenciación aparece en ambas regiones, siendo más acentuada en genes con funciones diferenciadas, como lo son genes con especificidad tisular de expresión variada (Barsh et al., 1983). Tal es el caso de la familia GH, sobre todo en primates, donde también se incluye a los genes CSHs. Esta familia multigénica constituye un modelo ideal para estudiar los mecanismos moleculares mediante los cuales los genes se relacionan de manera muy cercana unos a otros, expresan proteínas muy similares, pero son activos en diferentes tejidos a muy diferentes niveles – arriba de miles de veces -, sin mencionar la ambigüedad funcional de algunos de ellos.

La duplicación génica está considerada como uno de los fenómenos más relevantes en la evolución de las proteínas y de sus nuevas funciones (Ohno *et al.*, 1968; Watts y Watts, 1968; Ohta, 1989; Lang *et al.*, 2000). Se ha postulado que al existir dos copias de un gen, la presión selectiva frente a cambios en residuos importantes es menor, ya que si uno de los genes pierde la función, aún queda la del otro. Esto facilita la exploración de nuevos nichos funcionales para las proteínas miembros de la familia, que empiezan a cambiar. Existen familias de proteínas en las que este fenómeno es especialmente notorio, por lo que en estas se puede observar gran variedad de funciones entre sus miembros (Alberts, 2006). Las proteínas con secuencias suficientemente parecidas suelen tener un origen común (Zuckerlandl y Pauling, 1965) y presentar una arquitectura tridimensional similar (Chothia y Lesk, 1986), pero frecuentemente realizan funciones distintas (Henikoff *et al.*, 1997); aunque también pueden traslaparse en otras. En muchas situaciones las familias de proteínas que tienen un origen común y una estructura tridimensional parecida, exhiben actividades enzimáticas similares. Por ello se podría considerar que estas proteínas realizan una misma función. Pero el caso es que esta actividad puede contribuir a distintos procesos celulares (Abascal, 2003).

Los investigadores han aprendido mucho sobre la evolución y función de los genes humanos examinando a sus contrapartes en modelos animales. Ello al comparar sus genomas para determinar la similitud de las secuencias, localización de los genes, la

longitud y número de exones en estos, las regiones genómicas altamente conservadas y la cantidad de ADN no codificante (Birney *et al.*, 2007). La secuenciación de los genoma de primates proporciona un medio para estudiar aspectos estructurales y funcionales de la evolución del genoma humano, desde su divergencia a partir de un antepasado común (Leakey *et al.*, 1998; Glazko y Nei, 2003; Kumar *et al.*, 2005).

Una manera obvia por la cual las variantes en el número de copias (VNCs) resulta en diversidad fenotípica en el humano, es mediante la alteración de los niveles de transcripción (y presumiblemente los subsecuentes niveles traduccionales) de los genes que están en un número variable de copias. Esta correlación ha sido ya demostrada para VNC de ciertos genes a niveles transcripcional (Hollox *et al.*, 2003; Aldred *et al.*, 2005; McCarroll *et al.*, 2006) y traduccional (González *et al.*, 2005; Linzmeier y Ganz, 2005). Los estudios que correlacionan los niveles de ARNm y de proteína con el número de copias genómicas de las VNC de los genes, necesitan tener en cuenta siempre que algunas VNCs pueden resultar en efectos fenotípicos que aparentemente se manifiestan sólo en determinados tejidos y/o etapas del desarrollo.

Otro aspecto de las VNC que debe abordarse es si los niveles y sus patrones de variación entre los seres humanos son similares a los de los primates no humanos y otros organismos. La duplicación de genes es conocida por ser una importante fuerza evolutiva a largo plazo (Freeman *et al.*, 2006)

### **1.3.2. Comparaciones genómicas en homínidos.**

¿Cuál es la utilidad de secuenciar y comparar los genomas de los grandes simios, nuestros parientes más cercanos en la escala evolutiva? Las comparaciones de sus genomas con el nuestro pueden ayudar a revelar las bases moleculares, así como las fuerzas evolutivas que han moldeado al genoma de nuestra especie, incluidos los procesos subyacentes de mutación y restricciones selectivas, entre otras cuestiones. Secuenciar un genoma de una especie cercana a la nuestra, nos proporciona una atalaya desde donde contemplar el del ser humano.

Por su proximidad evolutiva con los humanos, se secuenció en el año 2005 el genoma del chimpancé. El análisis de la secuencia genómica obtenida mostró que este primate comparte con los humanos el 98.8% del material genético. Los estudios comparativos de ambos genomas muestran patrones de evolución altamente correlacionados en los genes que codifican proteínas. Al alinearse los genomas del humano y del chimpancé, se estimó que éstos difieren en 1.23% de sus posiciones nucleotídicas (aproximadamente 35 millones) que incluyen alrededor de cinco millones de inserciones, deleciones, duplicaciones e inversiones (Mikkelsen *et al.*, 2005; Khaitovich *et al.*, 2006).

Teniendo en cuenta la pequeña diferencia entre las secuencias del humano y chimpancé, se puede presumir que en ellas están contenidos los cambios moleculares cruciales que dirigieron el rumbo evolutivo que marcó el surgimiento de nuestra especie (Khaitovich *et al.*, 2006; Kehrer-Sawatzki y Cooper, 2007). La mayor parte de las divergencias fenotípicas es resultado de cambios genómicos específicos que han dado lugar a diferencias en la expresión génica, así como a ganancias, pérdidas y cambios en la función (Mikkelsen *et al.*, 2005; King y Wilson, 1975; Carroll, 2003; Olson y Varki, 2003; Enard y Pääbo, 2004; Carroll, 2005; Goodman *et al.*, 2005; Varki y Altheide, 2005; Khaitovich *et al.*, 2006).

El reciente progreso en las comparaciones de los genomas de humano y chimpancé ha revelado que las inserciones y deleciones han alterado considerablemente el paisaje genómico de estas especies, después de la separación de ambos linajes (Khaitovich *et al.*, 2006; Kehrer-Sawatzki y Cooper, 2007). La secuenciación y comparación de ambos genomas han comenzado a revelar cambios genéticos de posible peso en la divergencia entre ambas especies. Sin embargo, la aún baja resolución del ensamblaje del genoma del chimpancé desafía la habilidad de los métodos actuales para identificar muchos de estos cambios estructurales (Mikkelsen *et al.*, 2005; Kehrer-Sawatzki y Cooper, 2007). Por ende, se requiere de investigaciones de genómica comparativa de regiones génicas específicas, de las que mientras tanto se puedan derivar lecciones.

Seis años después del chimpancé, le tocó al orangután la secuenciación de su genoma, siendo el más diferente de la familia de los primates. El orangután se sitúa a una distancia evolutiva mayor, por lo que se pudo conocer mejor las cuestiones que nos diferencian de otros grandes simios (Locke *et al.*, 2011).

El genoma del orangután aporta nuevos detalles a la evolución y proporciona nuevos conocimientos a los científicos sobre los aspectos únicos del ADN humano que separan al hombre de los grandes simios. La investigación demuestra que el genoma humano y el del orangután son similares en un 97%. Desde el punto de vista evolutivo, el genoma del orangután es bastante especial entre los grandes simios, porque ha permanecido extraordinariamente estable durante los últimos 15 millones de años. Esto le diferencia de los del chimpancé y humano, los cuales han experimentado grandes reordenamientos estructurales a gran escala, mismos que podrían haber acelerado su evolución (Locke *et al.*, 2011).

Cuando se observan los genomas de humanos y chimpancés, hay una aceleración de los cambios estructurales durante el transcurso de la historia evolutiva. Sin embargo, por alguna enigmática razón, los orangutanes no han participado en esa aceleración. Una posible clave para explicar la falta de dinamismo del genoma del orangután, es su escaso contenido de elementos repetitivos Alu. Estos representan alrededor del 10% del genoma humano y pueden aparecer en sitios “inesperados” para crear nuevas mutaciones o reordenamientos genéticos. El genoma humano posee unas 5,000 secuencias Alu específicas de los humanos, mientras que el chimpancé tiene unas 2,000 secuencias Alu específicas de su genoma. Pero en el genoma del orangután, sólo se han encontrado 250 copias de elementos Alu nuevas a lo largo de 15 millones de años de evolución. Esto es lo más parecido a una prueba irrefutable para poder explicar la estabilidad estructural del genoma del orangután (Locke *et al.*, 2011).

En marzo del presente año se descifró el genoma del gorila, que comparte un 98.25% del ADN con el humano, con quien tuvo un antepasado común hace unos ocho millones de años. Este gran simio comparte más material genético con los humanos de lo

que se creía. Se examinaron más de 11,000 genes en los humanos, chimpancés y gorilas, en busca de pistas evolutivas. Los resultados iniciales revelaron que aunque globalmente el genoma humano es más parecido al del chimpancé, casi en un tercio de éste se encuentra más cercano evolutivamente al gorila. El 15% del genoma del gorila es más parecido evolutivamente al del humano o al del chimpancé, que estos dos últimos entre ellos (Gibbs y Rogers, 2012; Scally *et al.*, 2012).

Los genes relacionados con la percepción sensorial, el oído y el desarrollo del cerebro, mostraron “una evolución acelerada” en los tres grupos (humanos, chimpancés y gorilas), pero particularmente en los humanos y en los gorilas. En varios casos, una variante de la proteína que causa enfermedad hereditaria en el ser humano es la única versión que se encuentra en gorilas. [Ej: la variante Arg432Cys del factor de crecimiento PGRN que está asociada a la demencia senil, enfermedad que los gorilas no sufren]. El estudio del genoma del gorila puede contribuir al desarrollo de nuevos tratamientos para enfermedades humanas causadas por variantes proteicas asociadas a un fenotipo normal en gorila (Scally *et al.*, 2012).

La secuenciación del genoma del bonobo en Junio de este mismo año pone la pieza que faltaba para trazar la relación evolutiva de los homínidos. La secuenciación del genoma del bonobo (*Pan paniscus*) completa la enciclopedia génica de los homínidos, que también incluye a chimpancés, gorilas, orangutanes y humanos. El bonobo es el "pariente" vivo más cercano al hombre junto al chimpancé, aunque más pacífico, juguetón y promiscuo sexualmente. La comparación de su genoma con el mapa genético de chimpancés y humanos, refleja que ambas especies de simios difieren con el hombre aproximadamente en el 1.3 % de su genoma, mientras que bonobos y chimpancés coinciden genéticamente en un 99.6%. La coincidencia genética es “bastante grande” porque los humanos se diferenciaron de los bonobos y de los chimpancés hace entre cinco y siete millones de años. En cambio, las dos especies del género *Pan* que se diferenciaron hace dos millones de años, solamente se distinguen en un 0.4% y aún así, sus comportamientos sociales y sexuales difieren mucho entre ellos (Prufer *et al.*, 2012).

Los neandertales son unos de los parientes evolutivos más cercanos de los seres humanos modernos. Surgieron hace cerca de 400,000 años, se extendieron a lo largo de Europa y Asia Occidental, y se extinguieron hace aproximadamente 30,000 años. El primer borrador del genoma neandertal, publicado en el 2010, demuestra que durante esas migraciones, los antepasados de los seres humanos modernos se entrecruzaron a su salida de África con los neandertales hace menos de 100,000 años. Este flujo génico entre los neandertales y el tronco común de todas las poblaciones no africanas, dio como resultado que el humano moderno no africano comparta con los extintos neandertales del 1 al 4% de los genes. Prácticamente todo nuestro ADN es muy similar a los neandertales, ya que estamos evolutivamente muy próximos. Lo que los humanos no africanos compartimos especialmente con ellos son variantes en rasgos no fundamentales de nuestra biología. El porcentaje "relativamente bajo" del ADN de neandertal en el genoma del humano moderno podría sugerir que el entrecruzamiento fue "bastante limitado" (Green *et al.*, 2010; Burbano *et al.*, 2010).

Los investigadores compararon el genoma del neandertal con los genomas de cinco humanos de la época moderna procedentes de diferentes lugares del mundo: un sudafricano del grupo San, un africano del grupo Yoruba, un chino Han, un francés, y un nativo de Papúa-Nueva Guinea. Los resultados revelan una variedad de genes que son únicos en los humanos actuales; incluso, un "puñado" de genes que se propagaron rápidamente entre el *Homo sapiens* después de que humanos y neandertales se separaran de un ancestro común (Green *et al.*, 2010). Los hallazgos también ofrecen una lista preliminar de regiones genómicas y genes claves para la identidad humana. Por primera vez se pudieron identificar características genéticas que diferencian a los humanos del resto de los organismos, incluyendo nuestros parientes evolutivos más cercanos. La investigación ha permitido identificar 83 genes específicos de *Homo sapiens*. Estos están relacionados con funciones cognitivas, con la anatomía y fisiología de la piel, y con el desarrollo esquelético del cráneo. Por ahora, los genes identificados corresponden a aspectos fisiológicos, metabólicos, morfológicos y cognitivos -estos últimos implicados en el síndrome de Down, la esquizofrenia y el autismo- que presentan diferencias entre los sapiens y los neandertales.

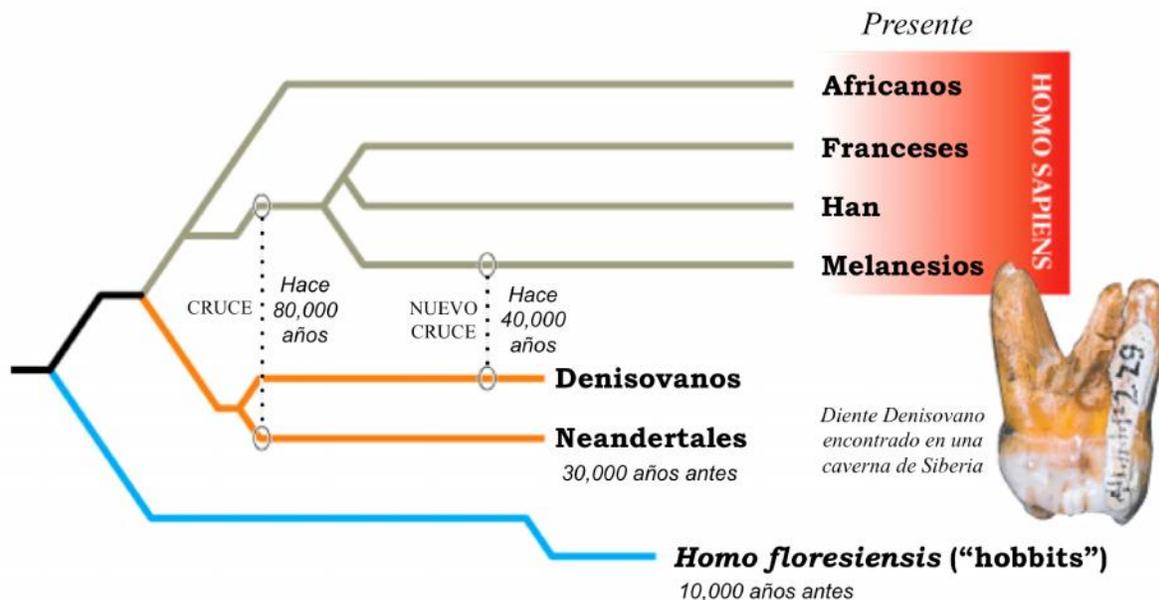
No obstante, aún se desconocen las funciones concretas que se relacionarían con los cambios genéticos (Green *et al.*, 2010; Burbano *et al.*, 2010).

Uno de estos genes, cuya evolución acelerada ha atraído un interés especial es FOXP2. Este gen está implicado en el lenguaje y en funciones cerebrales importantes. Lo poseen todos los vertebrados, pero particularmente en el humano existen dos variantes únicas, los investigadores sugieren que la capacidad del lenguaje estaría presente en al menos tres especies humanas (*sapiens*, neandertales y denisovanos) y obviamente en sus antepasados comunes hace entre 600,000 y 800,000 años. Sin embargo, no basta con poseer este gen. Para que haya comunicación entre individuos se requiere una estructuración neurológica y cognitiva, un aparato fonador para la emisión de un rango de sonidos, un aparato auditivo que permita entenderlo y por último, un entorno social en el que se desarrolle el lenguaje (Konopa *et al.*, 2009; Meyer *et al.*, 2012). Al extinguirse, los *Homo neanderthalensis* dejaron un legado genético, cultural y social que no siempre es reconocido, pero que posiblemente contribuyó en parte a lo que hoy es nuestra especie (Green *et al.*, 2010; Burbano *et al.*, 2010).

La reconstrucción en alta definición del genoma completo del homínido denisovano, descubierto en Asia en el 2010, arroja luz sobre las relaciones entre denisovanos y humanos modernos, y presenta una lista de cambios genéticos asociados al desarrollo de la función cerebral y el sistema nervioso, que permitieron diferenciarnos de los homínidos arcaicos (Meyer *et al.*, 2012). La secuenciación del ADN demuestra también que el antepasado de humanos modernos y denisovanos pudo haber vivido hace unos 800,000 años (Meyer *et al.*, 2012; Reich *et al.*, 2010; Reich *et al.*, 2011).

Con la comparación del genoma denisovano con el de los neandertales y el de 11 humanos modernos de todo el mundo, se confirmó que las poblaciones de grupos humanos aborígenes de las islas del sureste asiático, comparten más variantes genéticas con los denisovanos que otras poblaciones humanas (Meyer *et al.*, 2012; Reich *et al.*, 2010; Reich *et al.*, 2011). Los denisovanos y los neandertales descendían de la misma población ancestral que se había separado previamente de los ancestros de los humanos modernos. A

diferencia de los neandertales, los denisovanos no contribuyeron a todos los genes euroasiáticos de la actualidad. En cambio, evidencias de secuencias genéticas de los denisovanos en las poblaciones modernas de Melanesia —región de Oceanía que incluye la isla de Nueva Guinea—, sugieren que hubo mestizaje con los antepasados de los melanesios. Al igual que los neandertales, se mezclaron con los ancestros de los humanos no africanos actuales. Los melanesios actuales tienen entre un 4 y un 6% del material genético de los denisovanos extinguidos (Meyer *et al.*, 2012; Reich *et al.*, 2010; Reich *et al.*, 2011). Los genomas de las poblaciones del este de Asia y América del Sur incluyen ligeramente más genes procedentes de neandertales que los europeos. Se estima que la proporción de la ascendencia neandertal en Europa es más baja que en el este de Asia (Meyer *et al.*, 2012; Reich *et al.*, 2010; Reich *et al.*, 2011).



**Figura 7. Un nuevo miembro de la familia humana.** El ADN obtenido de un diente mostró que los Denisovanos estuvieron estrechamente relacionados a los Neandertales y consecuentemente relacionados a los humanos.

La secuencia del genoma ha permitido generar una lista con unos 100.000 cambios recientes en el genoma humano tras la separación de los denisovanos. Según el equipo de científicos, algunos de estos cambios afectan a genes asociados al desarrollo de la función

cerebral y el sistema nervioso y permite entender qué es lo que nos hace tan únicos. Otros afectan a la piel, los ojos y la morfología dental (Meyer *et al.*, 2012; Reich *et al.*, 2010; Reich *et al.*, 2011). Las regiones estructurales variantes del genoma son típicamente difíciles de caracterizar y están relacionadas con ciertas enfermedades humanas que indican, al mismo tiempo, que el genoma de los denisovanos, al menos por algunas regiones, es más arcaico que el de otros homínidos, ya que comparte algunas características con el genoma del chimpancé (Meyer *et al.*, 2012; Reich *et al.*, 2010; Reich *et al.*, 2011; Zhang *et al.*, 2011). Combinado con la secuencia del genoma neandertal, el genoma denisovano sugiere un cuadro complejo de interacciones genéticas entre nuestros ancestros y otros diferentes grupos de homínidos antiguos con los que convivían en ese momento (Meyer *et al.*, 2012; Reich *et al.*, 2010; Reich *et al.*, 2011).

### **1.3.3. El proyecto ENCODE comienza a desvelar enigmas del genoma humano.**

La secuencia del genoma humano proporciona el código subyacente de la biología humana. A pesar de un estudio intensivo, especialmente en la identificación de los genes que codifican proteínas, nuestro conocimiento del genoma está lejos de ser completo; en particular con respecto a los ARNs no codificantes, los transcritos empalmados alternativamente y las secuencias reguladoras. Los análisis sistemáticos de los transcritos y la información regulatoria son esenciales para la identificación de genes y regiones reguladoras, y son un recurso importante para el estudio de la biología humana y de las enfermedades. Estos análisis también pueden proporcionar un panorama sinóptico de la organización y la variabilidad de los genes y la información regulatoria a través de contextos celulares, individuales y de especies.

El genoma humano todavía alberga muchas incógnitas. Su secuenciación, concluida oficialmente en el 2003, fue un gran logro, pero tan sólo el primer gran paso. El mapa está listo, pero todavía falta mucho por tener del manual de instrucciones completo. El proyecto ENCODE (del inglés Encyclopedia of DNA Elements), que se inició entonces, supone el relevo del Proyecto Genoma Humano. El proyecto ENCODE tiene como objetivo delinear todos los elementos funcionales codificados en el genoma humano. Operacionalmente, se define un elemento funcional como un segmento discreto del genoma que codifica un

producto definido (por ejemplo, proteína o ARN no codificador) o muestra una firma bioquímica reproducible (por ejemplo, proteína de unión, o una estructura de la cromatina específica). Los estudios genómicos comparativos sugieren que 3-8% de las bases están bajo selección purificadora (negativa) y por lo tanto pueden ser funcionales, aunque otros análisis han sugerido estimados mucho mayores. Cubriendo el 1% del genoma, el proyecto ENCODE, en su fase piloto anotó el 60% de bases restringidas durante la evolución de los mamíferos, pero también identificó muchos elementos funcionales putativos adicionales sin evidencia de restricción (Birney *et al.*, 2007). El advenimiento de tecnologías de secuenciación de ADN más poderosas ahora permiten análisis de todo el genoma más precisos con un amplio repertorio de ensayos funcionales.

En el 2007 se publicaron los resultados preliminares del estudio piloto, el cual ya deparó algunas sorpresas (Birney *et al.*, 2007). Ahora se completan los datos con la totalidad del genoma, accesible a bajo costo gracias a las modernas técnicas de secuenciación masiva. Entre las principales aportaciones de ENCODE está la de echar por tierra la idea de que los genes se encuentran dispersos en un mar de ADN sin ninguna utilidad, lo que se conocía como “ADN basura”. Ahora se ha puesto de manifiesto que las tres cuartas partes del genoma contienen instrucciones importantes para la vida; tanto que constituyen un gran panel de control con millones de “interruptores” que regulan la actividad de nuestros genes. Alrededor del 80 % del genoma humano contiene elementos relacionados con algún tipo de función bioquímica, hasta un total de 120 funciones diferentes, lo que significa que casi todo nuestro genoma está activo en algún momento (Bernstein *et al.*, 2012).

La nueva información es tan exhaustiva y compleja que ha dado lugar a un nuevo modelo de publicación donde los documentos electrónicos y los conjuntos de datos están interconectados. Los resultados aparecieron en 30 artículos de acceso abierto conectados entre sí en las revistas Nature, Genome Biology y Genome Research, además de la revista Science. El proyecto está liderado por el National Genome Research Institute (NHGRI) en EEUU y el EMBL-European Bioinformatics Institute (EMBL-EBI) en Reino Unido. ENCODE presenta un mapa detallado de la función del genoma que identifica cuatro

millones de interruptores de genes. Esta referencia esencial ayudará a los investigadores a localizar áreas muy específicas de enfermedades humanas (Bernstein *et al.*, 2012).

El Proyecto ENCODE describe la producción y el análisis inicial de 1,640 conjuntos de datos diseñados para anotar elementos funcionales en el genoma humano. Además integra los resultados de diversos experimentos en tipos celulares, experimentos relacionados que involucran 147 tipos de células diferentes, y todos los datos ENCODE con otras fuentes, tales como las regiones candidatas a partir de estudios de asociación de todo el genoma (GWAS) y las regiones evolutivamente restringidas. Juntos, estos esfuerzos revelan características importantes acerca de la organización y función del genoma humano (Bernstein *et al.*, 2012)

El número sin precedentes de elementos funcionales identificados en el proyecto ENCODE, proporciona un recurso valioso para la comunidad científica y mejora significativamente nuestra comprensión del genoma humano. Los análisis han revelado muchos aspectos novedosos de la expresión génica y la regulación, así como la organización de la información. Sin embargo, hay todavía muchos detalles específicos que requieren experimentos adicionales para dilucidar en particular sobre los procesos que generan estos elementos y cómo y dónde funcionan (Bernstein *et al.*, 2012).

La gran dispersión de la cobertura presenta un espectro de elementos con diferentes propiedades funcionales descubiertos por ENCODE. Un total de 99% de las bases conocidas en el genoma están dentro de 1.7 kb de cualquier elemento ENCODE. Mientras que el 95% de las bases están dentro de 8 kb de un motivo de unión de un factor de transcripción o una “huella” de DNasa I. Curiosamente, incluso utilizando las estimaciones más conservadoras, la fracción de bases que pueden intervenir en la regulación directa de genes, aunque incompleta, es significativamente mayor que la atribuida a los exones que codifican proteínas (1.2%), aumentando la posibilidad de que más información en el genoma humano puede ser más importante para la regulación de genes que para la función bioquímica. Muchos de los elementos de regulación no están restringidos a través de la

evolución de los mamíferos, lo cual hasta ahora había sido una de las indicaciones más confiables de un evento bioquímico importante para el organismo (Bernstein *et al.*, 2012).

Es importante destacar que por primera vez tenemos suficiente poder estadístico para evaluar el impacto de la selección negativa en elementos específicos de los primates, de la cual el proyecto ENCODE muestra evidencia. Además, incluso con nuestra estimación más conservadora de elementos funcionales (8.5% de regiones putativas para proteínas específicas de unión a ADN) y suponiendo que ya se ha muestreado la mitad de los elementos de los factores de transcripción y de la diversidad de tipo celular, se podría estimar que, como mínimo, el 20% del genoma participa en estas funciones específicas, con una probabilidad significativamente mayor (17% de proteínas de unión a ADN y 2.9% de exones de genes que codifican proteína GENCODE) (Bernstein *et al.*, 2012).

La amplia cobertura de las anotaciones ENCODE mejora nuestra comprensión de las enfermedades comunes que tienen un componente genético, enfermedades genéticas raras y cáncer. El proyecto ENCODE y estudios similares proporcionan un primer paso para interpretar el resto del genoma más allá de los genes que codifican las proteínas. Tal información justifica la secuenciación del genoma completo en lugar de sólo el exoma (1.2% del genoma) para la investigación de enfermedades raras y variantes somáticas en elementos funcionales no codificantes. Además, así como los análisis GWAS de enfermedades típicamente se asocian a los SNPs en grandes regiones, la comparación con elementos funcionales no codificantes del proyecto ENCODE pueden ayudar a detectar supuestas variantes causales, además del refinamiento de la ubicación por técnicas de mapeo fino. La combinación de datos del proyecto ENCODE con información de alelos específicos derivada de las secuencias de genomas individuales, proporciona información específica sobre el impacto de una variante genética. En efecto, un objetivo importante sería el uso de datos funcionales, tales como los derivados de este proyecto, para asignar cada variante genómica a su posible impacto en los fenotipos humanos (Bernstein *et al.*, 2012).

Los datos aportados por el proyecto ENCODE indican que los genes son más complejos de lo que se pensaba hasta ahora: en vez de la visión tradicional, según la cual un gen da lugar a uno o varios transcritos alternativos que codifican una proteína en sus varias isoformas, parece claro que una región genómica puede codificar distintos productos proteicos y además dar lugar a otros transcritos (no necesariamente codificantes de proteínas) en ambas cadenas. Todo esto ha llevado a replantear el concepto de gen, que en la era post-ENCODE se definiría como “la unión de las secuencias genómicas que codifican un conjunto coherente de productos funcionales, potencialmente traslapados”. Esta definición hace hincapié en el producto funcional que se codifica (de ahí el uso de “coherente” para indicar que se trata de codificar una proteína o un ARN). Lo más novedoso de esta definición es que las regiones no traducidas (UTR) no formarían parte del gen, quedando incluidas –junto con los elementos reguladores- en la categoría de “regiones asociadas con genes”. La definición alternativa, más acorde con el pensamiento actual, de que un gen es “la región genómica que codifica un conjunto de transcritos alternativos traslapados”, aunque codifiquen distintos productos proteicos, es problemática a la luz de los datos aportados por el proyecto ENCODE. Si existe mucho solapamiento de transcritos, la aplicación de esta definición daría lugar a un número pequeño de genes muy extensos, los cuales además tendrían escaso significado biológico al codificar productos funcionales diversos (un mismo gen podría dar lugar a proteínas distintas y/o ARN no codificantes). La nueva definición probablemente aumentará el número total de genes del genoma, pero al estar centrada en el producto final, es más informativa de la función de cada gen concreto (Bernstein *et al.*, 2012).

Otra sorpresa del proyecto ENCODE ha sido comprobar que un alto porcentaje de los transcritos detectados no codifican proteínas, por lo que la categoría de "ARNs no codificantes" seguirá aumentando en el futuro. En concreto, los últimos años han sido testigos de la explosión de un nuevo tipo de ARNs no codificantes largos (en inglés lncRNAs), con funciones reguladoras importantes que se están empezando a conocer poco a poco. Se trata de ARNs con un tamaño superior a 200 nucleótidos que maduran mediante ajuste, pero que no codifican proteínas. Su número va en aumento, llegando a estimarse que

cubren unas 10 a 20 veces más de secuencia genómica que los ARNs codificantes de proteínas (Bernstein *et al.*, 2012).

En conjunto, los resultados del proyecto ENCODE son apasionantes y enriquecen enormemente nuestra visión del genoma humano, su regulación y funcionamiento, lo cual abrirá horizontes insospechados en la investigación genómica y tendrá fuertes implicaciones biomédicas en un futuro cercano (Bernstein *et al.*, 2012).

#### **1.4. Elementos móviles.**

Los elementos móviles, denominados a menudo como transposones o repeticiones, son secciones de ADN que pueden generar múltiples copias de ellos mismos en el genoma (Cordaux y Batzer, 2009; Kazazian, 2004). Cuando esto ocurre en la línea germinal, la copia adicional resultante puede ser transmitida a la progenie y esta mutación eventualmente puede fijarse en la población. La mayoría de los elementos móviles en el genoma humano son elementos dispersos largos (LINEs, Long Interspersed Elements), elementos dispersos cortos (SINEs, Long Interspersed Elements), o repeticiones terminales largas (LTR, Long Terminal Repeats). Todos estos elementos utilizan un proceso conocido como retrotransposición para propagarse. Un retrotransposón funcional debe tener un promotor interno que recluta a una polimerasa para transcribir una copia de ARN de sí mismo. Esta copia de ARN entonces debe reclutar una proteína que tiene actividad de endonucleasa que corta el ADN y de transcriptasa inversa que transcribe la copia de ARN de vuelta a otro lugar del genoma, lo que resulta en una copia genómica adicional del elemento móvil. Los LTRs y LINEs son autónomos, lo que significa que sus copias de ARN codifican la proteína con las actividades endonucleasa / transcriptasa inversa necesarias para la reinserción. Los SINEs no son autónomos porque no codifican una proteína con actividades endonucleasa/transcriptasa inversa y por lo tanto para su reinserción en el genoma dependen del reclutamiento de una proteína con dichas actividades que haya sido traducida de una copia de ARN de un LINE.

El ADN repetido disperso está formado por secuencias que se repiten miles de veces en el genoma humano, pero no en tándem sino de manera dispersa. Este tipo de

repeticiones constituyen un 45% de todo el genoma humano, y se clasifican en función del tamaño de la unidad repetida. Los SINE (Short Interspersed Nuclear Elements, elementos nucleares dispersos cortos) suponen un 13% del genoma humano. El principal SINE es la familia de elementos Alu, que es específica de primates y constituye un 10% de nuestro genoma. Un elemento Alu está formado por una secuencia de 250 - 280 nucleótidos, con unas 1.500.000 copias por genoma y una repetición cada 4 kb como promedio. Es un elemento relativamente rico en guaninas+citosinas (56% de contenido en CG, mientras que el contenido promedio del genoma humano es del 41%).

Los LINE (Long Interspersed Nuclear Elements, o elementos nucleares dispersos largos) constituyen un 20% del genoma humano. Son secuencias con un tamaño de varias kilobases, agrupados en distintas familias. El principal LINE es el llamado LINE-1 ó L1, formado por una secuencia de unas 6 kb repetida unas 800,000 veces en el genoma (aunque muchos de estos elementos están truncados y carecen de su región 5'), llegando a constituir alrededor de un 15% del genoma. Estos elementos, a diferencia de los SINE, no son ricos en guaninas+citosinas (tienen un 42% de citosinas+guaninas, que es cercano al contenido promedio del genoma humano). Un elemento L1 codifica dos proteínas: una proteína de unión a RNA en el marco de lectura ORF1 y una proteína con actividad endonucleasa y retrotranscriptasa en el marco de lectura ORF2. Está flanqueado por unas pequeñas repeticiones directas (en la misma orientación) y termina en una cola poli-A. Los elementos LINE son retrotransposones que pueden copiarse a sí mismos a través de un intermediario de ARN y transponerse a otras localizaciones genómicas. Según el modelo más aceptado, el elemento se transcribe por la ARN polimerasa II, a partir de un promotor interno, sus productos proteicos se unen a la cola poli-A de su propio ARN mensajero y el complejo se inserta en el ADN genómico por la acción combinada de la endonucleasa (que corta dentro de regiones ricas en AT que llevan la secuencia TTTTNA) y de la retrotranscriptasa. Las proteínas codificadas por los LINE son utilizadas también para la retrotransposición de elementos SINE y de pseudogenes procesados, por lo que pueden jugar un importante papel como elemento modificador del genoma. De hecho, se ha visto que la secuencia propia de los L1 tiene la propiedad de inhibir la transcripción, de ahí que los niveles de ARNm y proteínas codificadas por los L1 en las células sean muy bajos. Lo más interesante es que

también pueden modificar la transcripción de los genes en cuyos intrones hay abundancia de estos elementos: un 80% de los genes humanos tienen L1 en sus intrones, y la densidad en L1 correlaciona negativamente con los niveles de expresión de estos genes. Por tanto, su papel tanto en la evolución de genomas, como en la regulación génica, le confieren una gran importancia.

Los HERV (retrovirus endógenos humanos), representan copias de los retrovirus humanos que se han ido integrando en el genoma humano en el curso de la evolución y con frecuencia son el origen de proto-oncogenes celulares. Habitualmente representan copias truncadas del genoma de estos virus, y constituyen alrededor de un 8% del genoma (hay unas 450,000 copias). Por el hecho de que habitualmente conservan algunas de las repeticiones terminales largas de dichos genomas, son también denominadas LTRs (Long Terminal Repeats). Nuestro genoma también contiene unas 300,000 copias de elementos repetidos originados por transposones de ADN, lo que supone un 3% del total del genoma. Estos elementos contienen el gen de la transposasa (habitualmente truncado), flanqueado por repeticiones invertidas. De entre las distintas familias que existen cabe destacar el tipo MER1 ó MER2 y los elementos mariner (Hsmar2), responsables de algunas reordenaciones cromosómicas importantes en patología humana.

Cuando Barbara McClintock descubrió hace medio siglo los elementos móviles en el maíz, en realidad los denominó "elementos de control", debido a la forma en que influían en la expresión de los genes vecinos (McClintock, 1950; McClintock, 1956). Los investigadores contemporáneos de McClintock nunca aceptaron su idea acerca de que los transposones juegan un importante papel en la regulación génica (McClintock, 1973) y la desestimaron erróneamente en su totalidad cuando se publicó el operón lac y vieron que su regulación no estaba relacionada a eventos de transposición (Jacob y Monod, 1961).

En la década de 1990 algunos individuos revisaron y comenzaron a defender la idea de que las inserciones de elementos móviles estaban alterando los patrones de expresión de genes cercanos, de una manera que a veces era beneficioso para el huésped y resultaba en inserciones que ocurrían bajo selección purificadora (Britten, 1996-A, 1997; Brosius,

1999). Como se trataba todavía de la era pre-genómica, aumentaron los estudios de casos donde inserciones individuales fueron funcionales para la especie huésped. A este proceso donde un rasgo evolutivo es objeto de selección natural para proveer adaptación a determinadas condiciones y una vez que ya está consolidado, comienza a ser utilizado y perfeccionado en pos de una nueva finalidad (en ocasiones no relacionada con su "propósito" original), se le denomina exaptación (Gould y Vrba, 1982; Lowe *et al.*, 2007).

#### **1.4.1 Papel de los elementos repetitivos en la evolución del *locus* GH y en la expresión de sus genes.**

La genómica comparativa ha puesto de manifiesto que los principales linajes de vertebrados contienen cuantitativa y cualitativamente diferentes poblaciones de elementos retrotransponibles y transposones de ADN, con diferencias importantes también frecuentemente observadas entre especies del mismo linaje. Esto se debe fundamentalmente a: (i) la evolución diferencial de familias ancestrales de elementos transponibles, con escenarios de evolución que van desde la extinción completa a la invasión masiva; (ii) la introducción linaje-específica de elementos transponibles por infección y transferencia horizontal, por ejemplo por retrovirus endógenos; y (iii) la aparición linaje-específica de nuevos elementos de transposición, como se ha observado particularmente para retroelementos no codificantes SINEs. Durante la evolución de vertebrados, los elementos transponibles han contribuido repetidamente aportando secuencias reguladoras y codificantes para el huésped, guiando a la aparición de nuevos reguladores génicos y funciones linaje-específicas. En todos los linajes de vertebrados, hay evidencia de reordenamientos genómicos mediados por elementos transponibles, tales como inserciones, deleciones, inversiones y duplicaciones potencialmente asociadas con o después de eventos de especiación. En conjunto, estas observaciones indican que los elementos transponibles son los principales causantes de la diversidad genómica y biológica de los vertebrados, con posibles papeles importantes en la especiación y las principales transiciones evolutivas (Bohne *et al.*, 2008).

Los reordenamientos cromosómicos se producen en todo el genoma, sin embargo, la frecuencia observada de reordenamiento es mayor en ciertos intervalos. Estos intervalos

son típicamente regiones ricas en *repeticiones de bajo número de copias* (LCR) o de elementos repetitivos, tales como LTR, LINE, Alu, MIR y elementos de DNA MER2. Varios puntos de rupturas recurrentes en deleciones y duplicaciones cromosómica se han localizado para LCR y por lo general estos puntos son regiones grandes y altamente homólogas. Los mecanismos responsables de estos reordenamientos, a saber, recombinación homóloga no alélica entre las copias LCR, han sido bien establecidos. Sin embargo, pocos estudios han examinado los mecanismos responsables de los rearrreglos no recurrentes de puntos de ruptura en regiones no homólogas. Análisis previos han revelado deleciones originadas a partir de eventos de recombinación Alu/Alu y otras que resultan de un mecanismo de unión de extremos no homólogos. Los puntos de corte se han encontrado en LCRs y dentro de elementos repetitivos, incluyendo Alu y MER5B. Estos hallazgos sugieren que la arquitectura genómica de orden superior, tal como LCR, y las secuencias repetitivas pequeñas, como los elementos Alu, puede mediar deleciones cromosómicas a través de mecanismos homólogos y no homólogos (Shaw y Lupski, 2005).

En las proximidades de la región 3' flanqueante de los genes hGH-N y hGH-V se encuentra una repetición Alu Sx. La deleción de dicha repetición en construcciones reporteras que contienen secuencias 3' flanqueantes del gen hGH-N, aumentó la actividad reportera en las células hipofisarias GC transfectadas, lo que sugiere que esta región contiene un elemento represor. Esta región del gen hGH-N contiene la repetición Alu Sx entre los nucleótidos 2228 y 2501 (donde el codón de inicio se ubica en la posición +559 pb). Un mapeo refinado de este elemento silenciador reveló que el mismo comprende cuatro entidades discretas que incluyen un núcleo de dominio represor (CRD), un dominio antisilenciador (ASE) que contiene elementos que median la actividad silenciadora independiente de orientación y dos dominios que flanquean la CRD/ASE que modula la actividad silenciadora de una manera dependiente de CRD. El dominio modulador río arriba también es necesario para la función del silenciador independiente de la orientación (Burbano *et al.*, 2010)

Un elemento represor importante se encuentra entre los nucleótidos 2,306 y 2,465. Sin embargo, debido a que una serie de deleciones entre los nucleótidos 2,306 y 2,572

exhiben un aumento gradual de la actividad, se ha sugerido que múltiples elementos represores pueden estar distribuidos a través de esta secuencia de 266 pb. Esta región incluye la mayor parte del elemento Alu, indicando que un represor está localizado en dicho elemento.

Las repeticiones Alu del *locus* GH puede haber desempeñado un papel no sólo en la generación del *locus*, mediante la participación en la duplicación de genes, sino también, justo río abajo del gen hGH-N se han desarrollado elementos que actúan en cis que controlan la regulación transcripcional de los genes duplicados. Esto plantea la cuestión de si otras repeticiones Alu en el *locus* pueden afectar la transcripción de los genes vecinos. En este punto, sólo podemos especular basándose en las similitudes de las secuencias Alu que flanquean los genes GHs y CSHs. La repetición Alu ubicada en la región 3' de los genes hGH-N y hGH-V pertenece a la subfamilia Sx, mientras que las secuencias Alu que flanquean los genes CSHs son de la clase Sz en las regiones 5' y de la subfamilia Sq en las regiones 3'. De hecho, la región 3' flanqueante de los genes hGH-N y hGH-V son idénticas en alrededor del 92%. Por ello, es muy probable que el elemento Alu dentro de la región 3' flanqueante del gen hGH-V también posea un elemento silenciador (Trujillo *et al.*, 2006)

Los transposones MER20 tienen las firmas epigenética de potenciadores, aisladores y represores. Estos elementos se unen directamente a factores de transcripción esenciales para el embarazo y coordinadamente regulan la expresión génica en respuesta a la progesterona y cAMP. El elemento transponible MER20 ha contribuido al origen de una nueva red de regulación génica involucrada en el embarazo de los mamíferos placentarios; en particular mediante el reclutamiento de la vía de señalización del cAMP en células del estroma endometrial. Las novedades que definen a los mamíferos Euterianos (placentarios) incluyen el desarrollo interno prolongado, el reconocimiento materno de la gestación, una placenta invasiva y un endometrio uterino ricamente vascularizado que puede acomodar implantación del cigoto. Un paso esencial en el establecimiento de la gestación en muchos mamíferos placentarios, es la diferenciación (decidualization) de células del estroma endometrial (CEE) en respuesta a la hormona progesterona, el segundo mensajero cAMP y, en algunas especies, las señales fetales. La diferenciación de las CEE implica

reprogramación extensa de muchas funciones celulares, incluyendo el silenciamiento simultáneo de las vías de proliferación celular y la activación de vías de señalización de AMPc y progesterona. Así, la evolución de la gestación fue probablemente dependiente en la evolución de las CEE y la señalización celular mediada por hormonas y cAMP (Konopka *et al.*, 2009)

Aunque numerosos genes de respuesta a progesterona/cAMP se expresan en CEE humanas, uno de los más dramáticamente inducido es el de la prolactina (PRL). En particular, en este gen el potenciador sensible a progesterona/cAMP en las CEE se deriva de un transposón de ADN de la familia hAT-Charlie (MER20) encontrado sólo en la placenta de mamíferos, lo que sugiere que los MER20s han jugado un papel en la reconexión del paisaje de regulación génica de las CEE.

Explorando si otros genes de respuesta a la progesterona/cAMP se asociaban con MER20s, se realizaron búsquedas para MER20s río arriba, río abajo y dentro de las regiones codificantes e intrones de genes regulados diferencialmente en humanos. El 42% de MER20s se encontró dentro de 200 kb de los sitios de iniciación y terminación de la transcripción de los 6,504 genes regulados diferencialmente; mientras que sólo el 8% de MER20s se encontraron en la misma ventana alrededor de genes que no son regulados diferencialmente en decidualización. Estudios anteriores han demostrado que los elementos transponibles contienen sitios de unión a factores de transcripción que pueden ser donados para regular la expresión de genes cercanos, lo que sugiere que MER20s pueden haber sido reclutados por genes para actuar como elementos reguladores en la expresión endometrial (Lynch *et al.*, 2011)

Dos transposones de ADN insertados en el gen de la prolactina humana actúan en conjunto como un promotor alternativo adicional para dirigir la expresión de la PRL pituitaria. Se ha hipotetizado que este evento ha resultado en la expresión divergente (hipofisiaria versus extra hipofisiaria) de la prolactina en los primates, y en las acciones diferenciales de la prolactina hipofisiaria versus extra hipofisiaria en la lactancia versus la gestación, respectivamente. Sin embargo, la expresión extra hipofisiaria de PRL en

linfocitos y células deciduales humanas reveló que en los tejidos extra hipofisarios, el ARNm de la PRL era más largo que su homólogo hipofisario. La razón de esta discrepancia es que en células deciduales y linfocitos humanos, la transcripción de la PRL es dirigida por un promotor alternativo, que está situado 5.8 kb río arriba del sitio de inicio de transcripción hipofisaria, resultando en la transcripción de un exón adicional (Gerlo *et al.*, 2006)

### **1.5. Estado del *locus* GH de los primates al comenzar este trabajo.**

La familia de GH en los primates se originó a partir de una rama evolutiva inicialmente unigénica que continuó sin grandes cambios hasta la aparición de los platirrinos (Ceboidea). La evolución de la GH en los primates es única por su dramática evolución acelerada. En un principio se privilegiaron eventos de expansión del *locus* a causa de duplicaciones que se propiciaron por la presencia de secuencias repetitivas tipo Alu que rodeaban al *locus* ancestral (Chen *et al.*, 1989).

Después de la separación de los tarseros y antropoides, pero antes de la divergencia de los platirrinos y los catarrinos, múltiples eventos de duplicaciones génicas condujeron a un aumento del número de genes GH que evidencian la notable expansión que ha sufrido este *locus* en primates (Revol de Mendoza *et al.*, 2004; González Álvarez *et al.*, 2006; Wallis *et al.*, 2006). Además del gen del tipo GH, cuya expresión siguió restringida a la hipófisis, aparecieron genes tanto para CSHs como para GH-V (Revol de Mendoza *et al.*, 2004; Chen *et al.*, 1989; Golos *et al.*, 1993; Ye *et al.*, 2005).

La multiplicación génica del *locus* GH y la divergencia de sus miembros fueron acompañadas por cambios que resultaron en diferencias en cuanto a la expresión en los tejidos de los genes miembros del *locus*, con al menos dos especificidades: hipófisis y placenta.

Hasta el momento no está claro cómo algunos genes GH son expresados en placenta en platirrinos. Sin embargo, desde Cercopithicoidea, los miembros de esta familia de genes acumularon sustituciones aminoacídicas que predicen la adquisición de nuevas

características distintivas de las hormonas placentarias del *locus* humano. El origen de las diferencias en la especificidad tisular de la expresión génica mostrada por los diferentes miembros del *locus*, continúa siendo un enigma. Algunas otras interrogantes salen a la luz en los escenarios en los que ha tenido lugar la evolución del *locus* GH en los primates: ¿Cómo se llevó a cabo la fase de expansión en el *locus* GH?, ¿Fue igual o diferente la ruta seguida por los distintos subórdenes de primates?, ¿Qué cambios en su codificación han experimentado sus genes?

Antes de comenzar el presente trabajo, se conocía en detalle la secuencia y la organización del *locus* GH en el humano (Chen *et al.*, 1989). Se había reportado además, la anatomía de los *loci* GH de marmoset (*Callithrix jacchus*) y del mono capuchino (*Cebus albifrons*) (Wallis y Wallis; 2006), así como del mono rhesus (*Macaca mulata*) y el mono verde (*Cercopithecus aethiops*) (Revol De Mendoza *et al.*, 2004; González-Álvarez *et al.*, 2006). Para el chimpancé se había sugerido un *locus* GH pentagénico con una composición y organización similar a la del humano (Pérez-Maya, 2008) que discrepaba con el hallazgo de seis genes (las secuencias completas de dos genes GHs, tres CSHs y la secuencia parcial de otro CSH) (Revol *et al.*, 2004) y con el ensamblaje liberado por el consorcio de secuenciación del Genoma del Chimpancé, el cual muestra un ensamblaje de dicha región que sugiere un *locus* con siete secuencias relacionadas a genes GH/CSHs. Para el caso particular del gorila, previamente se sugirió que éste tenía una composición y organización similar a la del humano (Aguilera de Lira, 2009), sin embargo no se logró tener una buena cobertura, ni evidencias suficientes que apoyaran dicho arreglo de genes, por ello se decidió re-evaluar este *locus*.

### **1.6. Justificación del trabajo.**

La GH desempeña funciones claves en los diferentes estadios de la vida, siendo la más reconocida el regular el crecimiento lineal postnatal. Sin embargo, se le reconocen papeles en procesos tales como la síntesis proteica en el tejido muscular, la lipólisis en el adiposo, la reparación tisular y la regeneración ósea. Más recientemente, se le ha asociado a neurogénesis, neuroprotección, aprendizaje y memoria. La GH aumenta la eficiencia del sistema inmunitario y se considera que la reducción en los niveles circulantes de esta

hormona durante la vejez es responsable, en parte, del deterioro corporal. Por otra parte, los niveles de las hormonas relacionadas a la GH, como la GH-V y las CSHs, se han asociado con el buen funcionamiento placentario. Si bien se tienen evidencias de que estas hormonas actúan en la regulación del metabolismo materno durante el embarazo y sobre los tejidos fetales, asegurando el suministro de energía al feto y la sobrevivencia del mismo, todavía hay mucho que descubrirles de sus mecanismos de acción y potencial terapéutico.

Con el fin de comprender la diversidad de los primates, así como nuestro propio pasado, es menester investigar los mecanismos que crean y alteran las redes de regulación génica. Mientras que investigadores previos han planteado la hipótesis de que los elementos móviles han sido, y siguen siendo, una fuerza importante en la creación de nuevas regiones reguladoras en el linaje humano, esta hipótesis no ha sido explorada sistemáticamente a nivel genómico. Desde la divergencia del humano y un ancestro común de éste con el orangután, gorila y chimpancé, algunos genes han alterado sus patrones de expresión. Los genes miembros del *locus* GH parecen ser parte de estos genes, cuyo patrón de expresión ha sufrido importantes cambios. En el presente trabajo de tesis, se utilizaron secuencias genómicas que comprenden el *locus* GH y regiones flanqueantes que recientemente han sido obtenidas en nuestro laboratorio, para varias especies de primates, con la finalidad de comprender el aporte de los elementos móviles que se han “exaptado” como secuencias de regulación génica putativas en el linaje humano durante la evolución de los primates.

La motivación detrás de nuestros esfuerzos en la secuenciación y en la genómica comparativa del *locus* GH en primates, se enfoca en avanzar en la comprensión de dicho *locus* en el humano y dilucidar la vía que dictó su evolución en primates desde una composición unigénica hasta un arreglo multigénico de expresión tejido específica. El profundizar en la investigación de genómica comparativa, aunado al aprovechamiento de tecnología de expresión génica y biotecnología que se han desarrollado en nuestro laboratorio, nos permitirá diseccionar y con ello aspirar a develar los elementos regulatorios claves y las fuerzas que actuando sobre ellos moldearon la evolución de los genes en particular y del *locus* en general, tanto en la regulación, como en la relación de la

estructura-función de las hormonas. Esto como un prerrequisito para explotar este conocimiento y materiales en beneficio de la salud.

### **1.7. HIPÓTESIS.**

Proponemos que los elementos repetitivos han jugado un papel sumamente importante en los eventos que han tenido lugar durante la evolución de los primates, dirigiendo el movimiento de bloques de secuencias, que han permitido la adquisición de elementos reguladores y ganancia de su funcionalidad. Proponemos que el *locus* GH del estrepsirrino incluido en este estudio, contiene secuencias que pueden identificarse como elementos reguladores incipientes relacionados al inhibidor hipofisiario y al potenciador placentario descritos en los MVM y GM. Además proponemos que fue a partir de la divergencia de los tarsios y los MNM que se adquirieron los cambios necesarios para su función y con ello el origen de la especificidad tisular observada en los primates superiores.

## CAPÍTULO II: OBJETIVOS

### 2.1. OBJETIVO GENERAL.

Comparar los *loci* GH y sus regiones flanqueantes en especies representativas de primates de los subórdenes *strepsirrinos* y *haplorrinos*, tomando como referencia al humano (*Homo sapiens*), para identificar y describir los eventos evolutivos que acompañaron la fase de expansión del locus GH en primates y establecieron las bases para su regulación diferencial espacial y temporal.

### 2.2. OBJETIVOS ESPECÍFICOS.

- 2.2.1. Ensamblar los *loci* GH de las especies de primates de interés, a partir de las secuencias disponibles en los proyectos genómicos y las generadas en nuestro laboratorio.
- 2.2.2. Determinar: A) la composición y organización genómica de los *loci* GH ensamblados y B) los elementos reguladores claves (anotación estructural).
- 2.2.3. Establecer las similitudes y diferencias en la composición y organización genómicas de los *loci* ensamblados (genómica comparativa).
- 2.2.4. Investigar los posibles mecanismos evolutivos que moldearon la evolución del *locus* GH en primates (genómica evolutiva).

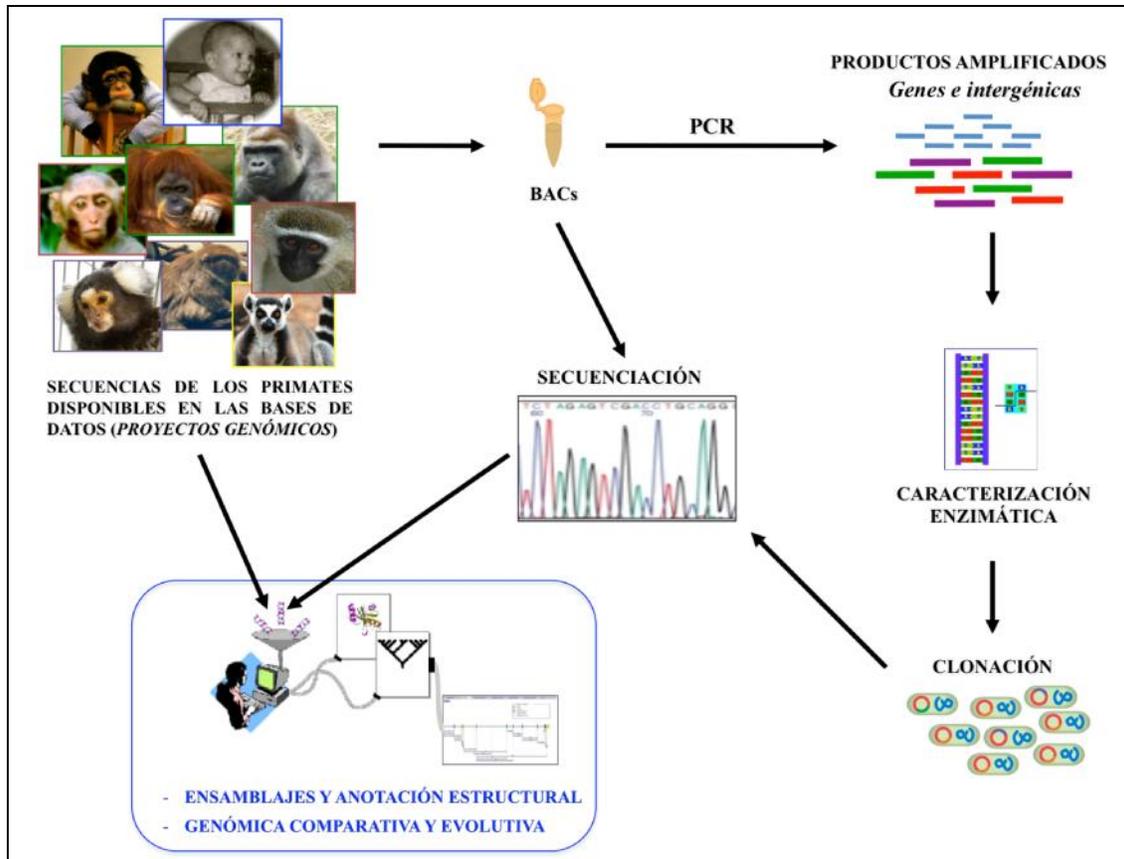
### CAPÍTULO III: ESTRATEGIA GENERAL

La estrategia general (figura 8) seguida para desarrollar el presente trabajo se abordó a través de dos enfoques: 1) el análisis de cromosomas artificiales bacterianos (BACs) que contienen los *loci* GH completos de varias especies de primates mediante amplificación de regiones específicas y por secuenciación directa y 2) el examen bioinformático detallado de los datos disponibles de los proyectos de secuenciación de sus genomas en los casos donde fue posible.

El trabajo comprendió varias etapas. Primero se recabó y ordenó la información sobre los *loci* GH de especies representativas de primates disponible en las bases de datos de secuencias. Luego se extrajo ADN a partir de cepas portadoras de BACs que contenían la región del *locus* GH de varias especies de primates, a partir del cual se llevó a cabo una secuenciación directa con el apoyo de Genome Quebec y empleando iniciadores consensos diseñados específicamente para este trabajo, se amplificaron los genes y las regiones intergénicas en nuestro laboratorio.

En seguida los productos amplificados se caracterizaron con enzimas de restricción, clonaron y secuenciaron. Con las secuencias obtenidas y las recabadas de las bases de datos, se procedió al ensamblaje y anotación estructural de la región génica en estudio. Los conflictos de los ensamblajes cuando más de un arreglo génico pudiera ser posible fueron resueltos mediante el diseño e implementación de estrategias basadas en amplificación y/o caracterización enzimática con endonucleasas de restricción.

Se concluyó con la genómica comparativa y evolutiva de estos *loci* tomando como referencia su contraparte en el humano, teniendo en cuenta sus genes, las proteínas que éstos codifican, los elementos involucrados en la regulación transcripcional y el patrón de elementos repetitivos presente en las regiones intergénicas. Finalmente se plantearon hipótesis para tratar de explicar los principales hallazgos.



**Figura 8. Estrategia general.** Comprende las etapas de extracción y purificación del ADN de los BACs, amplificación de los genes y regiones intergénicas, caracterización enzimática, clonación, secuenciación y análisis bioinformático de las secuencias.

## CAPÍTULO IV: MATERIAL Y MÉTODOS

### 4.1. Material.

#### 4.1.1. Origen de los reactivos.

Los reactivos utilizados en este trabajo para la preparación de las soluciones, medios de cultivo y geles, provinieron de las compañías Sigma Chemical Company (St, Louis, MO, EUA), Merck (Monterrey, México), Aldrich Chemical Company (Milwaukee, WI, EUA), United States Biochemical (Cleveland, Ohio, EUA) y DIFCO Laboratories (Detroit, MI, EUA).

El cloranfenicol presente en el medio de cultivo utilizado para crecer la cepa portadora de un Cromosoma Artificial Bacteriano (BAC, de sus siglas en inglés) que contenía el *locus* GH de chimpancé, provino de SPECTRUM (Gardena, CA, EUA) y la extracción de ADN BAC se realizó con el Kit BACMAX™ DNA Purification procedente de EPICENTRE (Madison, WI, EUA).

Los reactivos utilizados para la Reacción en Cadena de la Polimerasa (PCR, de sus siglas en inglés) se adquirieron de las siguientes casas comerciales: la enzima DNA polimerasa *Taq* y el MgCl<sub>2</sub> de Promega Corporation (Madison, WI, EUA), los dNTPs de New England Biolabs, Inc. (Beverly, MA, EUA) y los iniciadores de Invitrogen (La Jolla CA, EUA) y Sigma Chemical Co. Las amplificaciones por PCR largo se realizaron con el Kit MasterAmp™ Extra-Long PCR de EPICENTRE. La purificación de los productos amplificados de interés, separados por electroforesis en geles de agarosa, fue realizada con el Kit Wizard® SV Gel and PCR Clean-Up System de Promega.

El vector de clonación pTOPO-XL-3.5 fue de Invitrogen y la kanamicina utilizada en los medios selectivos provino de SIGMA. Las enzimas de restricción que se utilizaron en las caracterizaciones enzimáticas proceden de New England Biolabs (Beverly, MA, EUA) y la enzima RNase One™, empleada en el tratamiento de las preparaciones de ADN plasmídico, se obtuvo de Promega corporation. La secuenciación se llevó a cabo con los estuches comerciales de LiCor (LiCor, Lincoln, NE) y ABI Prisma (Perkin Elmer, Cetus, Norwalk, CT).

#### 4.1.2. Material Biológico.

En este trabajo se utilizó ADN procedente de cepas bacterianas portadoras de BACs, cuyos insertos incluyen al *locus* GH de varias especies de primates, obtenida por tamizaje con una sonda radioactiva del gen hGH de bancos genómicos en BACs del laboratorio del Dr. Pieter de Jong del Children Oakland Hospital Research Institute, en Oakland California, EUA. En la adquisición de este material biológico se contó con la asistencia técnica de Irám Rodríguez.

#### 4.2. Equipo.

Se utilizaron los siguientes equipos: un horno de microondas GoldStar modelo MA-875M, microcentrífuga Eppendorf modelo 5415 (Brinkmann Instruments, Inc. NY, EUA), centrífuga Labnet modelo Spectrafuge 24D (Edison, NJ, EUA) y ultracentrífuga Beckman modelo J2-M1(Beckman Instruments, Inc. CA, EUA), así como un agitador de tubos Barnstead Termolyne modelo Maxi Mix (Bubuque, IOWA, EUA), una placa de agitación magnética con control de temperatura marca Corning modelo PC320 (N.Y. EUA), una incubadora Shell Lab modelo 1535 (Sheldon Manufacturing, EUA), un termomixer Eppendorf (Westbury, NY, EUA) y un baño de agua giratorio modelo G76 (New Brunswick Scientific Co. inc. Edison, N.J. EUA).

Dependiendo de sus requerimientos los reactivos se almacenaron en un congelador marca TorRey de -20°C (Mty, Mex) o en un ultracongelador de -70°C de So Low Enviromental (Cincinnati, Ohio, EUA). La balanza analítica fue de la marca OHAUS modelo AP110S y la granataria de la marca Sartorius modelo 1206MP (Göttingen, Alemania). La electroforesis en geles de agarosa se realizó en cámaras Thermo EasyCast™ B1A y B2 con una fuente de poder BRL 500 (GIBCO-BRL). Para la documentación de la imagen de los geles se utilizó un fotodocumentador UVP Mini Darkroom de Bioimaging Systems (Upland, CA, EUA). El seguimiento del crecimiento bacteriano y la cuantificación del ADN se realizaron con un espectrofotómetro de la marca Eppendorf modelo BioPhotometer. Las amplificaciones por PCR se realizaron en un termociclador Eppendorf Mastercycler gradient (Amburgo, Alemania). El pulso eléctrico para la transformación de las bacterias se hizo en un electroporador modelo Gene Pulser II y celdas de 0.2 mm de la marca BIO-RAD (Hercules,

CA, EUA). La secuenciación se realizó en un secuenciador automático ABI Prisma Perkin Elmer. El procesamiento de los datos y los análisis bioinformáticos se realizaron en computadoras Macintosh eMac modelo A1002 y laptop Toshiba Satellite T2050.

#### 4.2.1. Programas computacionales.

Los programas de análisis de Biología Molecular utilizados fueron: Amplify versión 3.0 (University of Wisconsin, WI, EUA), OLIGO versión 6.40 (© 1989-2000 Wojciech Rychlik, Molecular Biology Insights, Inc., Cascade, CO, EUA), Mauve 2.1.0 (Darling *et al.*, 2004), GATAligner (Nix y Eisen, 2005), SinicView (Chun-Chieh *et al.*, 2006), CONSED versión 14.0 (University of Washington, Seattle, EUA) (Gordon *et al.*, 1998), DNAMAN versión 5.2.2 Demo (©1994-2001 Lynnon Biosoft Copyright) y MEGA 4.0 (Kumar *et al.*, 2004). El paquete de programas DNASTar Lasergene versión 7.0.0 (Copyright ©1989-2006 DNASTAR Inc., Madison, WI, EUA).

Para la construcción de los vectores y el análisis de patrones de restricción, fue empleado el programa Vector NTI Suite 9 (© 1994-2000 Informax, The University of North Carolina Shared BioInformatics Resource). Para el análisis densitométrico de los productos de PCR se empleó el programa computacional GEL PRO 3.2 (Media Cybernetics, Silver Spring, MD, USA). Para la cuantificación del ADN y la determinación de los tamaños se utilizó el software LabWorks versión 4.5.00.0 (©1996-2003, Media Cybernetics, Inc.).

Los programas bioinformáticos utilizados vía Internet fueron el BLAST disponible en <http://www.ncbi.nlm.nih.gov/BLAST/>, NEBcutter 2.0 de New England Biolabs a disposición en <http://www.neb.com>, ClustalW accesible en <http://www.ebi.ac.uk/Tools/clustalw/>, el programa de predicción de genes GENSCAN disponible en <http://genes.mit.edu/GENSCAN.html> (Stanford University, USA), el conjunto de bases de datos y programas para análisis comparativo de secuencias genómicas VISTA disponibles en <http://genome.lbl.gov/vista/index.shtml> y ECR Browser Genome Alignment a disposición en <http://ecrbrowser.dcode.org/genomeAlignment.php>. El programa Repeatmasker Open-3.0 (Smit, AFA., Hubley, R. y Green, P., 1996-2004) para identificación de elementos repetitivos fue

utilizado en <http://www.repeatmasker.org/>, mientras que los programas PipMaker y MultiPipMaker que permiten graficar el alineamiento entre dos o más secuencias genómicas similares fueron empleados en <http://pipmaker.bx.psu.edu/pipmaker/>.

### **4.3. Métodos.**

#### **4.3.1. Caracterización de los *loci* GH de las especies de primates de interés.**

El refinamiento de la caracterización del locus GH del chimpancé, la re-examinación del locus GH del gorila y el análisis por vez primera de los loci GH de otras especies de primates de interés, se abordaron a través de dos enfoques: 1) el análisis de cromosomas artificiales bacterianos (BACs) que contienen los loci GH completos de estas especies de primates mediante amplificación de regiones específicas y por secuenciación directa, y 2) el examen bioinformático detallado de los datos disponibles de los proyectos de secuenciación de sus genomas en los casos donde sea posible.

#### **4.3.2. Amplificación de los *loci* GH.**

Los BACs (CHORI-251-61D11, CHORI-255-4K22, CHORI253-28b13, LBNL-2-223J6, CHORI259-34A9 y CHORI250-110K24 + CHORI250-195F15) que contienen los loci GH del chimpancé (cGH), gorila (gGH), orangután (oGH), lémur (lGH), tití común (tcGH) y mono rhesus (rGH), respectivamente; fueron obtenidos por el tamizaje de una biblioteca de BACs procedente del Centro de Recursos BACPAC del Hospital Infantil del Oakland Research Institute, (Oakland, California), utilizándose para identificarles una sonda derivada del gen GH-N humano (hGH-N) marcada radiactivamente. El ADN extracromosómico se extrajo a partir de una población bacteriana portadora de estos BACs con el kit BACMAX DNA Purification, siguiendo las instrucciones del fabricante (EPICENTRE, Madison, WI, USA).

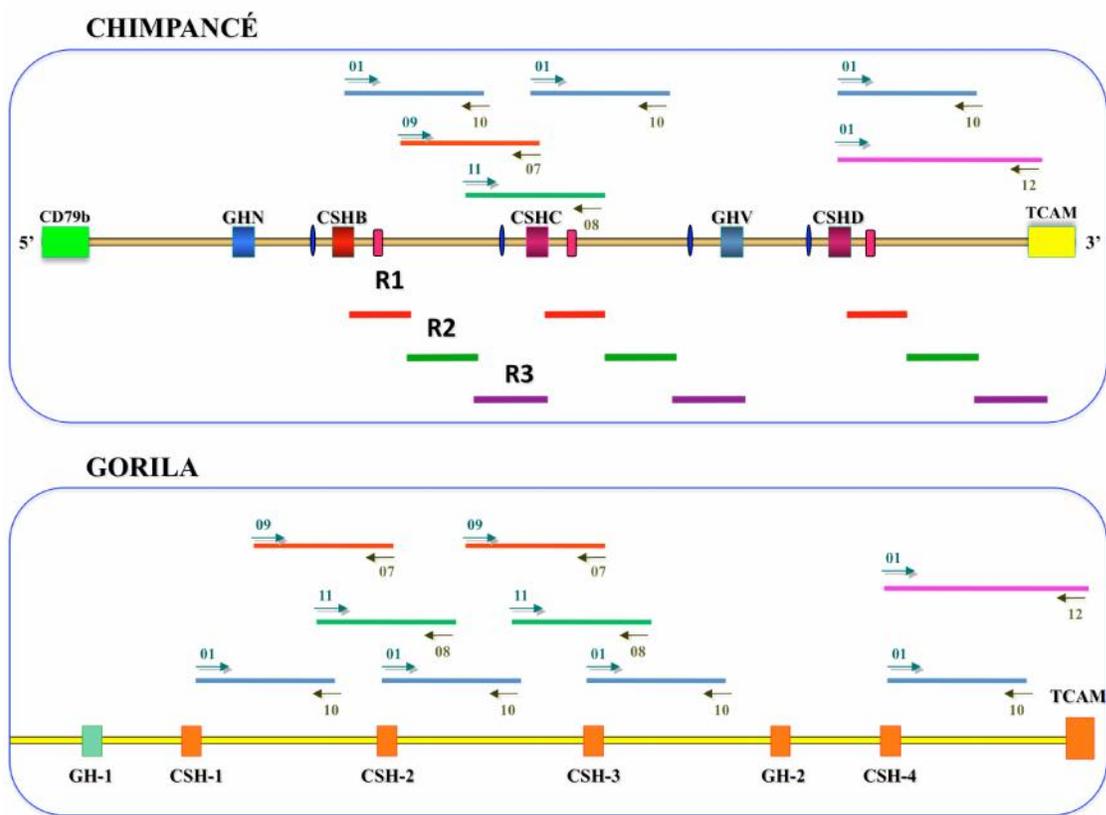
Los genes y las regiones intergénicas (RIG) de los *loci* de primates fueron amplificadas a partir de sus respectivos BACs utilizando la reacción en cadena de la polimerasa (PCR) y primers que figuran en el cuadro 1. En la figura 9 se representan los amplicones y las combinaciones de oligonucleótidos para amplificar los *loci* GH de

chimpancé y humano, pero en general estos oligonucleótidos también se usaron para el resto de los primates.

**Cuadro 1. Iniciadores utilizados en este estudio.**

CLAVE	NOMBRE	SECUENCIA	ORIENTACIÓN	BLANCO
00F	GHPRIMK	ATTAAGGGTACCGTGGACAGCTCACCTAG	Sentido	Genes
00R	GHPRIMEX	CCTCGAGATATCCTAGAAGCCACAGCT	Antisentido	Genes
01	GH/CS-Con	CATCTCCCTGCTGCTCAT	Sentido	RIGs
02	IGC-Con	GGGGAATTGAACACAAGGA	Antisentido	RIGs
03	PlacenCon	GGAGCTCAGGGTTGTTTTTC	Sentido	RIGs
04	IGC-Con	TCTCCCTGGAGAATAATCCC	Sentido	RIGs
05	CHElemPCon	GTAGAGACGGGGTTTCACTATG	Antisentido	RIGs
06	CHElemPCon	AGCAGGGGAAAGTGGTGA	Sentido	RIGs
07	PlacenCon	GATGAGCAGCAGGGAGATG	Antisentido	RIGs
08	IGL-Con1	CTTTTTCTTGCTCCTTCCC	Antisentido	RIGs
09	IGL-Con1	GAGAGCAGAAGTACCCTTC	Sentido	RIGs
10	IGL-Con2	GGTAACCTTTGCTGCCCTG	Antisentido	RIGs
11	IGL-Con2	GCTTCTCCTCTTTGTCCT	Sentido	RIGs
12	TCAM-Con	CCATAGCCACTTTTGATGC	Antisentido	RIGs
13	FDLCHIM	GCTCATGCCAGTAATCTCAGTMC	Sentido	RIGs
14	RDLCHIM	GGTCTGAGCACCATTGAGTC	Antisentido	RIGs
15	FLIPA3	TGTCGACCCATTAACCTGTC	Sentido	RIGs
16	IGLEcoF	TTTGAGATACAGGATATCTACAGCCCTGATG	Sentido	RIGs
17	IGLKpnR	CCCCACAGTTTGGTACCCTAAGATTAGGACTAC	Antisentido	RIGs
18	ENHF	CTGAGATTCTGATATAATTAGCCTGGAATGTGG	Sentido	RIGs
19	ENHR	CACACCAAGTTGATGAGCTGAAATGTAGACTAAC	Antisentido	RIGs
20	ELEMPF	GTTGCCAACACCACAGCCAACCAC	Sentido	RIGs
21	ELEMPR	CTGTGGCCTAGGATGTGAATGCTGG	Antisentido	RIGs

Para la amplificación de los genes GH se mezcló en una reacción de PCR 10 pmol de cebadores (00F-00R) (cuadro 1), 10 ng de ADN BAC, 5 µl de tampón 10X (proveído con la enzima), 0.2 mM de cada dNTP, 1.5 mM MgCl<sub>2</sub>, una unidad de Taq Gold ADN polimerasa (EPICENTRE, Madison, WI, USA) y agua Milli-Q hasta ajustar a un volumen final de 50 µl por reacción. Después de un paso inicial de desnaturalización a 95 °C durante 5 min, la PCR se llevó a cabo durante 30 ciclos utilizando una temperatura de desnaturalización de 94 °C durante 30 segundos, una temperatura de 64 °C durante 90 segundos y una temperatura de extensión de 72 °C durante 2 min. La reacción se detuvo después de un tiempo de extensión final de 10 min a 72 °C.



**Figura 9. Iniciadores utilizados para amplificar los *loci* GH de chimpancé y gorila.** Los productos amplificados se representan con líneas horizontales abarcando regiones específicas de los *loci* GH que son esenciales para resolver los conflictos y facilitar el ensamblado de las lecturas originadas de la secuenciación total de los BACs. Con flechas horizontales ubicadas sobre los extremos de los productos amplificados representados se indican los iniciadores utilizados.

Las RIG de los loci GH fueron amplificados utilizando el MasterAmp Extra-long PCR kit (Epicenter), siguiendo las instrucciones del fabricante. La mezcla de reacción de 10 pmol de cada cebador (cuadro 1, figura 9), 40 ng de ADN BAC, 25  $\mu$ l de Pre-mezcla 4 (2X) proveída con la enzima y 2.5 U de ADN polimerasa Extra MasterAmp, se ajustó con agua Milli-Q hasta un volumen final de 50  $\mu$ l. Después de un primer paso de desnaturalización a 94 °C durante 5 minutos, se llevaron a cabo amplificaciones durante 12 ciclos, que comprenden incubación a 94 °C durante 30 segundos (desnaturalización), 55-57 °C durante 40 segundos (hibridación) y 72 °C durante 7-14 minutos (elongación), seguido por 23 ciclos similares, pero con incrementos de 15 segundos por ciclo en el tiempo de extensión. El programa concluyó con un paso de extensión final a 72 °C durante 20 min. La amplificación se llevó a cabo en un termociclador ABI Veriti (Applied Biosystems, Foster City, CA). Los productos de PCR fueron analizados por electroforesis en gel de agarosa al

1% seguido por tinción con bromuro de etidio y exposición a luz ultravioleta en un transiluminador.

#### **4.3.3. Tamizaje de las clonas recombinantes portadoras de los genes GH/CSH.**

Las células transformadas fueron tamizadas por PCR en colonia con los iniciadores apropiados para confirmar la presencia de los genes GH/CSH. Los productos amplificados fueron analizados mediante electroforesis en geles de agarosa al 1 %, utilizando como marcador de peso molecular al bacteriófago  $\lambda$  digerido con la enzima de restricción PstI. Posteriormente, los amplicones fueron caracterizados con enzimas de restricción (PstI, BstEII y AvrII) y los patrones de digestión fueron comparados con los esperados por predicción en base a sus secuencias. Las clonas que resultaron positivas fueron crecidas a 37 °C en tubos con LB suplementado con kanamicina como medio selectivo y en agitación constante durante 18 horas. Finalmente, se procedió a la extracción del ADN plasmídico de las clonas recombinantes identificadas.

#### **4.3.4. Extracción del ADN plasmídico.**

Se centrifugaron 3 ml del cultivo bacteriano durante 5 min a 5,000 rpm y se removió el medio por aspiración, manteniendo el tubo en hielo. El ADN plasmídico se extrajo por el método de la lisis alcalina (Sambrook *et al.*, 2001). En breve, este constó de los siguientes pasos: las células se resuspendieron en 100  $\mu$ l de solución I (glucosa a 50 mM, Tris-HCl a 25 mM y EDTA a 10 mM y pH 8) con el vortex. Se agregaron 200  $\mu$ l de solución II (preparada en fresco mezclando 0.2 ml de NaOH a 10 N y 1 ml de SDS al 10% en 8.8 ml de agua), mezclando inmediatamente por inversión e incubando en hielo por 5 min. Pasado este tiempo, se agregaron 150  $\mu$ l de solución III (acetato de amonio a 7.5 M), se mezcló de inmediato por inversión y se incubó por 5 min. en hielo.

El tubo se centrifugó a 14,000 rpm por 5 minutos y el sobrenadante se recuperó en un tubo nuevo donde se le hizo una extracción fenol/Sevag ( $\frac{1}{2}$  volumen de fenol y  $\frac{1}{2}$  de Sevag). Se centrifugó a 14,000 rpm por 2 min. y el sobrenadante se transfirió a un tubo nuevo donde el ADN plasmídico se precipitó agregando dos volúmenes de etanol al 100% frío y mezclando por inversión completamente. Después de centrifugar a 14,000 rpm por 5 min, la pastilla se lavó dos veces con 1 ml de etanol al 70% previamente enfriado y se dejó

secar a temperatura ambiente durante 30 minutos. Finalmente, se resuspendió en 50  $\mu$ l de TE (pH 8.0) conteniendo RNAsa a 20 $\mu$ g/ml.

#### **4.3.5. Selección de las clonas recombinantes portadoras de fragmentos de RIGs.**

Los plásmidos fueron analizados mediante electroforesis en geles de agarosa al 0.8 %, utilizando como marcador de peso molecular al bacteriófago  $\lambda$  digerido con la enzima de restricción *Pst*I. Los plásmidos recombinantes que mostraron un retraso en su patrón de corrimiento en correspondencia con el tamaño del inserto, fueron seleccionados como candidatos y posteriormente se sometieron a caracterizaciones con enzimas de restricción (*Pst*I, *Bst*EII y *Avr*II). Los amplicones que originaron las clonas fueron digeridos en idénticas condiciones y corridos en el mismo gel. Los patrones de corrimientos fueron comparados con los esperados por los análisis *in silico* por predicción en base a sus secuencias.

#### **4.3.6. Clonación molecular.**

Los productos amplificados de los *loci* GH (amplicones) fueron clonados en el vector comercial pTOPO-XL-3.5, incluido en el TOPO <sup>®</sup> XL PCR Clonación Kit (Invitrogen, La Jolla, CA, USA), de acuerdo con las instrucciones del fabricante y transformados en células electrocompetentes de *Escherichia coli* suministradas con el kit.

#### **4.3.7. Secuenciación.**

Los amplicones clonados fueron secuenciados utilizando BigDye <sup>®</sup> Terminator (Applied Biosystems, Foster City, CA), cebadores específicos o los M13 universales, un analizador de ADN ABI PRISM <sup>®</sup> 3100 Genetic Analyzer y el software (Applied Biosystems). Además, los ADN BAC de las especies de primates estudiadas en este trabajo, fueron secuenciados con ayuda del Servicio de Secuenciación de la Universidad McGill y Genome Quebec Innovation Centre, de Montreal, Canadá, empleando un secuenciador FLX 454 Roche, con una cobertura de alrededor de 25X. Los cromatogramas fueron importados al CodonCode Aligner Versión 3.7.1 (CodonCode Corporation). La secuencia del *locus* GH del chimpancé ensamblada se ha depositado ya en las bases de datos

GenBank/EMBL/DDBJ, con el número de acceso JN622009 y la del resto de los primates serán depositadas también después de otra revisión rigurosa.

#### **4.3.8. Re-evaluación del ensamblaje genómico.**

Las secuencias de los *loci* GH disponibles en el Proyecto del genoma del chimpancé (Mikkelsen *et al.*, 2005), del gorila (Scally *et al.*, 2012) y del orangután (Locke *et al.*, 2011) (fueron re-evaluadas usando un método utilizado anteriormente (Wallis, 2008). Los ensamblajes genómicos disponibles en [http://www.ensembl.org/Pan\\_troglodytes](http://www.ensembl.org/Pan_troglodytes) (ensamblaje de marzo de 2006, Pan\_troglodytes\_2.1), [http://www.ensembl.org/Gorilla\\_gorilla](http://www.ensembl.org/Gorilla_gorilla) (ensamblaje de 2008), [http://www.ensembl.org/Pongo\\_abelii](http://www.ensembl.org/Pongo_abelii) (ensamblaje de 2007) y las secuencias trazas identificadas en la base de datos WGS Traces de Pan troglodytes, Gorilla gorilla y Pongo abelii (<http://www.ncbi.nlm.nih.gov/Traces>), utilizando los métodos de búsqueda BLAST y BLAT (Altschul *et al.*, 1990; Kent 2002), se integraron con los datos de secuencias de genes e intergénicas derivados en este estudio y la secuenciación 454 del BAC, para dar un ensamblaje final. Las WGS Traces se originaron en el Centro de Secuenciación de Genomas de la Universidad de Washington y en el Instituto Broad. Las secuencias fueron ensambladas a partir de las trazas utilizando el paquete Staden (<https://sourceforge.net/projects/staden/>).

#### **4.3.9. Identificación de las secuencias generadas.**

Las secuencias de genes y regiones intergénicas generadas en este estudio se compararon en el BLAST (Altschul, *et al.*, 1990) con las secuencias disponibles en las bases de datos del GenBank, con el fin de determinar su identidad. Adicionalmente fueron analizadas con el programa vecscreen para verificar que no hubiese contaminación con secuencias del vector. Finalmente se compararon con los ensamblajes genómicos disponibles, con el propósito de determinar su probable ubicación y facilitar el ensamblaje final.

#### **4.3.10. Análisis de las proteínas.**

Las secuencias aminoácidas se dedujeron de las secuencias codificantes mediante

el programa Genscan (Burge y Karlin, 1997), (Burge, 1998) y (Burge y Karlin, 1998) y con las primeras se hizo el alineamiento en el programa MEGALIGN del paquete de programas DNASTar Lasergene versión 7.0.0 por el método ClustalW (Higgins y Sharp, 1988) para determinar la similitud entre ellas e identificar los aminoácidos clave en la unión al receptor (Liu *et al.*, 2001). El significado de las diferencias de aminoácidos se evaluó mediante modelaje molecular utilizando el programa RasMol y la estructura 3D de modelo GH humana: receptor (de Vos *et al.*, 1992).

#### **4.3.11. Análisis del promotor proximal.**

El análisis se centró en identificar los sitios de unión a factores transcripcionales descritos como responsables de la transcripción basal del gen hGH-N y se enfocó en la región que va del nucleótido -134 al -1. Dentro de estos destacan los sitios Pit-1 proximal y distal, la caja TATA y los elementos SP1 e InrE. Las secuencias de cada uno de los promotores proximales de los genes GH/CSH se alinearon con su contraparte en el humano utilizando el programa MEGALIGN del paquete de programas DNASTar Lasergene versión 7.0.0 por el método ClustalW (Higgins y Sharp, 1988).

#### **4.3.12. Análisis de los elementos reguladores del control transcripcional.**

El análisis consistió en identificar con el programa Bl2Seq (Tatusova y Madden, 1999) los dos elementos claves de la especificidad tisular: el inhibidor o elemento P que reprime la expresión de los genes placentarios en la hipófisis (ubicado 2 kpb río arriba de dichos genes) y el potenciador o “enhancer” que incrementa la expresión de los genes CSHs en la placenta (ubicado a 2.3 kpb pero río debajo de los genes CSHs). Las secuencias de cada uno de los elementos reguladores del control transcripcional se alinearon con su contraparte en el humano utilizando el programa MEGALIGN del paquete de programas DNASTar Lasergene versión 7.0.0 por el método ClustalW (Higgins y Sharp, 1988).

#### **4.3.13. Identificación de los elementos repetitivos.**

La identificación de los elementos repetitivos se realizó con el programa RepeatMasker (Smit *et al.*, 1996-2010), utilizando al *locus* GH humano como referencia.

## CAPÍTULO V: RESULTADOS

### 5.1. Disección experimental y resolución de conflictos de ensamblajes.

#### 5.1.1. Iniciando el análisis de los datos.

Estudios previos realizados en nuestro laboratorio sobre el *locus* GH del chimpancé sugirieron que éste contenía seis genes (Revol *et al.*, 2004), mientras que el proyecto genómico muestra siete secuencias relacionadas a GH/CSHs; en contraste con los cinco genes bien caracterizados del *locus* humano (Chen *et al.* 1989). En una aproximación previa (Pérez Maya, 2008), tras ensamblar el borrador del *locus* GH del chimpancé, se sugirió y mostró evidencias de que dicho *locus* estaba formado por cinco genes, pero sólo se obtuvieron las secuencias definitivas de cuatro de ellos. La secuencia del quinto gen y parte de las regiones intergénicas, habían quedado pendientes de obtenerse.

Con los datos de secuencias generados, las evidencias previas y teniendo en cuenta la posibilidad de que el *locus* pudiese estar formado por cinco genes, en vez de seis o siete, se propusieron varios abordajes para comprobar el modelo de ensamblaje de cinco genes. A nivel de las regiones intergénicas, dicho arreglo tendría dos intergénicas largas. Basado en el ensamblaje propuesto, se predijeron *in silico* mapas de restricción con el objetivo de verificar de forma definitiva a cuál arreglo corresponderían los resultados experimentales. Los patrones de cortes obtenidos, tanto para los productos amplificados como para las clonas derivadas, siempre fueron consistentes con las predicciones realizadas para el arreglo de cinco genes. En este proyecto se continuó con la estrategia iniciada en la Tesis de Maestría del autor del presente trabajo, pero abordando aquellas regiones aún no resueltas y utilizando nuevos juegos de oligonucleótidos para obtener y caracterizar en detalle regiones genómicas clave para el objetivo de este trabajo.

Con el fin de re-evaluar el *locus*, fueron amplificados los genes y regiones intergénicas a partir de un BAC que contiene el *locus* cGH entero, usando cebadores basados en las secuencias humanas correspondientes y las secuencias previamente obtenidas en la referida Tesis de Maestría del autor. El BAC también estuvo sujeto a una secuenciación completa utilizando un secuenciador FLX 454 de Roche (Servicio de Genome Quebec). El ensamblaje automático de novo de las secuencias obtenidas no dio un

ensamblaje fiable, pero el análisis detallado de estas lecturas ayudaron a resolver una serie de regiones con problemas y a confirmar muchas partes del ensamblaje del *locus*.

Para el caso particular del gorila, estudios previos también de nuestro laboratorio sugerían que el *locus* de este primate tenía cinco genes, en una estructura similar a la del humano (Chen *et al.*, 1989) y a la del chimpancé (Pérez-Maya, 2008). Sin embargo, el proyecto genómico sugería una estructura de seis genes. Ante esta situación, también se re-evaluó el *locus* GH del gorila, siguiendo la misma estrategia diseñada para el chimpancé (Pérez-Maya, 2008). De esta forma, se logró obtener una mayor cobertura en la secuencia y apoyándonos en las trazas del proyecto genómico, fue posible ensamblar la secuencia del *locus* GH del gorila y se reporta completamente curada en la presente tesis, junto con la secuencia correspondiente del chimpancé.

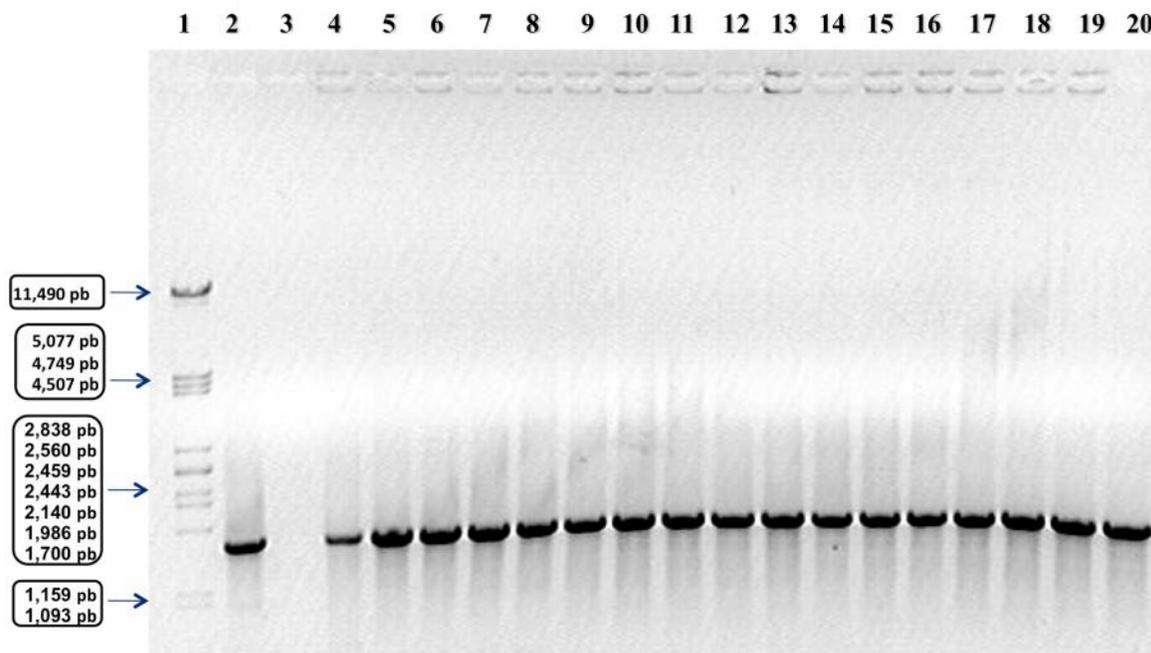
Para el orangután, el ensamblaje del *locus* GH a partir de las secuencias obtenidas arrojó dos posibles alternativas, un arreglo de cinco genes y otro de cuatro. Sin embargo, un abordaje experimental basado en la amplificación de las regiones intergénicas largas con flancos en genes CSHs resultó ser clave para resolver dicho conflicto. El enfoque utilizado permitió confirmar el arreglo génico presente en el *locus* GH del orangután y reportar su secuencia.

### **5.1.2. Aclarando dudas sobre los genes.**

El ADN extraído a partir de cepas de *E. coli* que contienen BACs portadores del *locus* GH de varios primates, resultó de muy buena calidad y utilidad. A partir de estos se amplificaron los genes GH/CSH de los primates, observándose una banda única de 1533 pb, misma que coincide con el peso molecular esperado para dichos productos y en algunos casos, como el títi común, se obtuvieron bandas de pesos moleculares diferentes correspondientes a pseudogenes.

Los productos amplificados se clonaron y las clonas positivas fueron tamizadas mediante PCR en colonia, con los iniciadores empleados para amplificar las unidades transcripcionales de los genes GH/CSH (tabla 1) y posteriormente verificados en un gel de

agarosa (figura 10). Se observó una banda de un solo tamaño de 1533 pb, la que delató a aquellas clonas que contenían vectores portando genes GH/CSH.

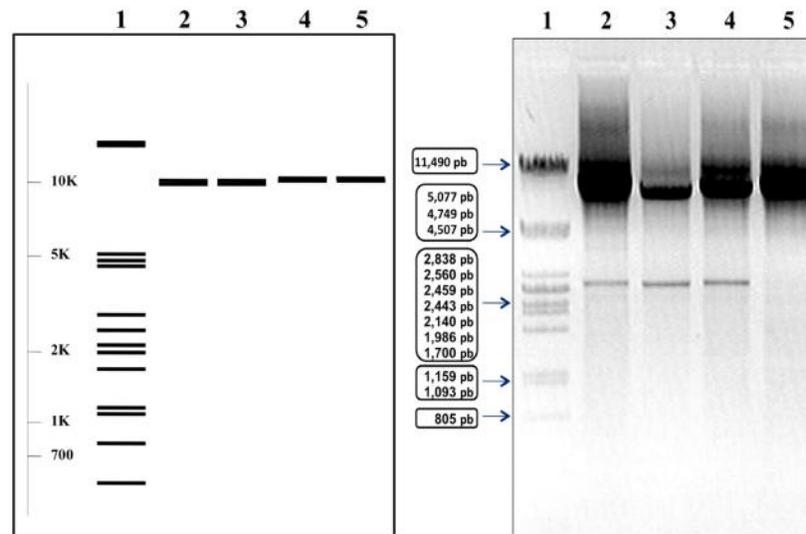


**Figura 10. Tamizaje mediante PCR en colonia de vectores recombinantes portadores de GHs.** Electroforesis en gel de agarosa al 1%, en el cual fueron identificadas las clonas que contenían los vectores portadores de las unidades transcripcionales de GH/CSHs. En el carril 1 se utilizó como marcador el bacteriofago  $\lambda$  digerido con la endonucleasa de restricción *Pst* I, en el 2 como control positivo el producto de 1533 pb amplificado con el par de iniciadores consenso 5'GHPRIMK y 3'GHPRIMEX para genes GH/CSH, en el 3 el control negativo para la PCR y en el resto de los carriles las PCRs de clonas.

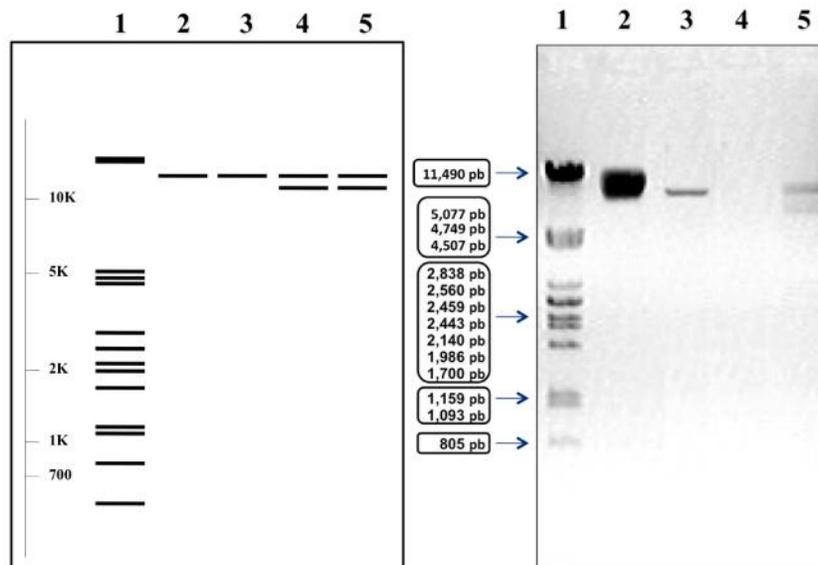
Posteriormente, los productos amplificados obtenidos fueron caracterizados mediante cortes con enzimas de restricción y los patrones de corte generados permitieron distinguir las clonas candidatas a ser secuenciadas.

### 5.1.3. Las regiones intergénicas como claves en el arreglo de los genes.

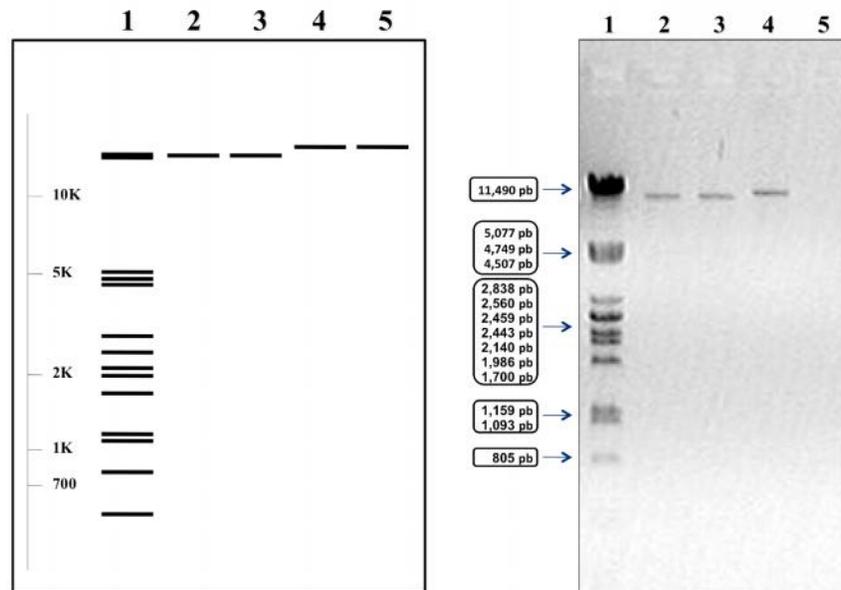
Buscando resolver la contradicción de los modelos de cinco, seis y siete genes, se recurrió a la amplificación de las regiones intergénicas a partir de secuencias conservadas en dichas regiones y en los genes. Por medio de la PCR larga, con múltiples combinaciones de iniciadores y utilizando como templado el ADN de los BACs portadores de los *loci* GH del chimpancé y del gorila, se obtuvieron productos amplificados que permitieran amplificar regiones específicas de los *loci*. Algunos ejemplos representativos son mostrados a continuación (figuras 11, 12, 13 y 14).



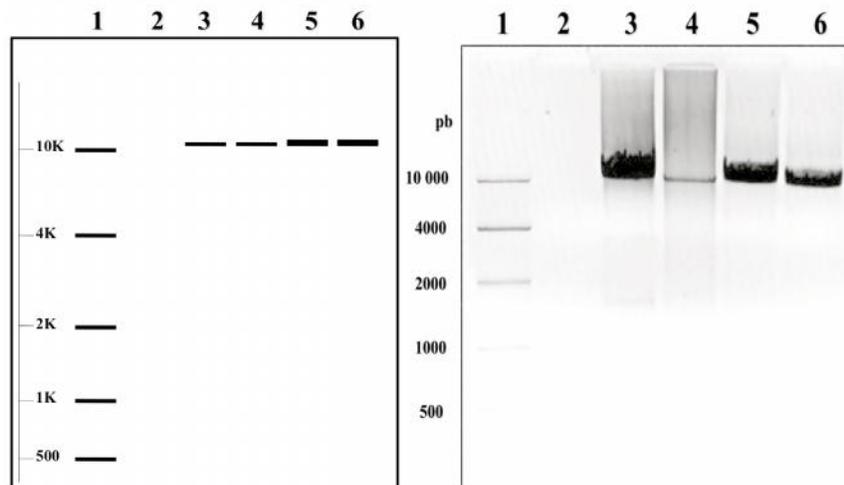
**Figura 11. Amplificación de las regiones intergénicas de los *loci* GH de chimpancé y gorila.** Productos de amplificación que abarcan los primeros dos tercios de las intergénicas largas y de la región posterior al último gen de los *loci* (obtenidos con los oligonucleótidos 01-10). Para los carriles 2 y 3 el amplicón lo conformó una mezcla de cuatro fragmentos cuyos pesos moleculares fueron: 9826, 9531, 9481 y 8830 pb; mientras que para los carriles 3 y 4, la mezcla esperada fue de tres fragmentos cuyos pesos son: 10199, 9428 y 4924 pb. Los productos de amplificación se resolvieron en geles de agarosa al 1%. Se utilizó el mismo marcador de peso molecular que el empleado en la figura anterior (como también lo será en lo sucesivo identificándose simplemente como  $\lambda$ ), en los carriles 2 y 3 los productos amplificados a partir del BAC de gorila con las premezclas 7 y 4, en los carriles 4 y 5 los productos amplificados a partir del BAC de chimpancé con las premezclas 7 y 4.



**Figura 12. Amplificación de las regiones intergénicas de los *loci* GH de chimpancé y gorila.** Productos amplificados que abarcan la región de la primera intergénica larga (10861 pb) para el chimpancé y la primera y segunda intergénica larga ( 10915 y 10579 pb) de gorila con los oligonucleótidos 09 – 07. Los productos de amplificación se resolvieron en geles de agarosa al 1%. En el carril 1 marcador de peso molecular  $\lambda$ , en los carriles 2 y 3 los productos amplificados a partir de dos diluciones del BAC del chimpancé con la premezcla 4 y en los carriles 4 y 5 los productos amplificados a partir del BAC del gorila con las premezclas 7 y 4.



**Figura 13. Amplificación de las regiones intergénicas de los *loci* GH de chimpancé y gorila.** Producto amplificado comprendido entre la segunda mitad del último gen del *locus* y secuencias más allá de éste (TCAM), obtenidos con los oligonucleótidos 01-12. En el carril 1 marcador de peso molecular  $\lambda$ , en los carriles 2 y 3 los productos amplificados (tamaño esperado de 12364 pb) a partir del BAC de gorila con las premezclas 7 y 4, en los carriles 4 y 5 los productos amplificados (tamaño esperado de 13645 pb) a partir del BAC de chimpancé con las premezclas 7 y 4.

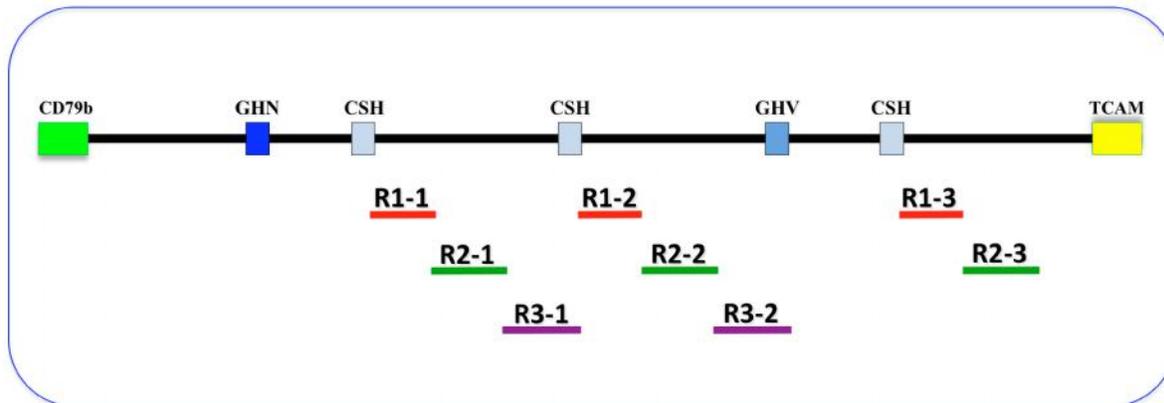


**Figura 14. Amplificación de las regiones intergénicas de los *loci* GH de chimpancé y gorila.** Productos amplificados identificados como FGTO10K, que abarcan regiones de dos intergénicas largas consecutivas y el gen incluido entre ellas, obtenidos con los oligonucleótidos 11-08. Los productos de amplificación se resolvieron en geles de agarosa al 1%. En el carril 1 como marcador de peso molecular una escalera de 500 a 10,000 pb, en el carril 2 el control negativo para la PCR, en los carriles 3 y 4 los productos amplificados (correspondiente a un único fragmento de 10272 pb) a partir de dos diluciones del BAC del chimpancé con la premezcla 4, en los carriles 5 y 6 los productos amplificados (conformados por la combinación de fragmentos de 10296 y 10277 pb) a partir del BAC del gorila con las premezclas 7 y 4.

Todos los productos amplificados obtenidos fueron clonados, las clonas portadoras de los mismos fueron caracterizadas y finalmente estas fueron secuenciadas.

#### 5.1.4. Un nuevo enfoque para verificar el ensamblaje del *locus*.

Un segundo enfoque abordado para verificar el modelo de cinco genes en el chimpancé y descartar el de seis (Revol de Mendoza *et al.*, 2004) y de siete genes (proyecto genómico del chimpancé), fue la amplificación de las regiones intergénicas largas.

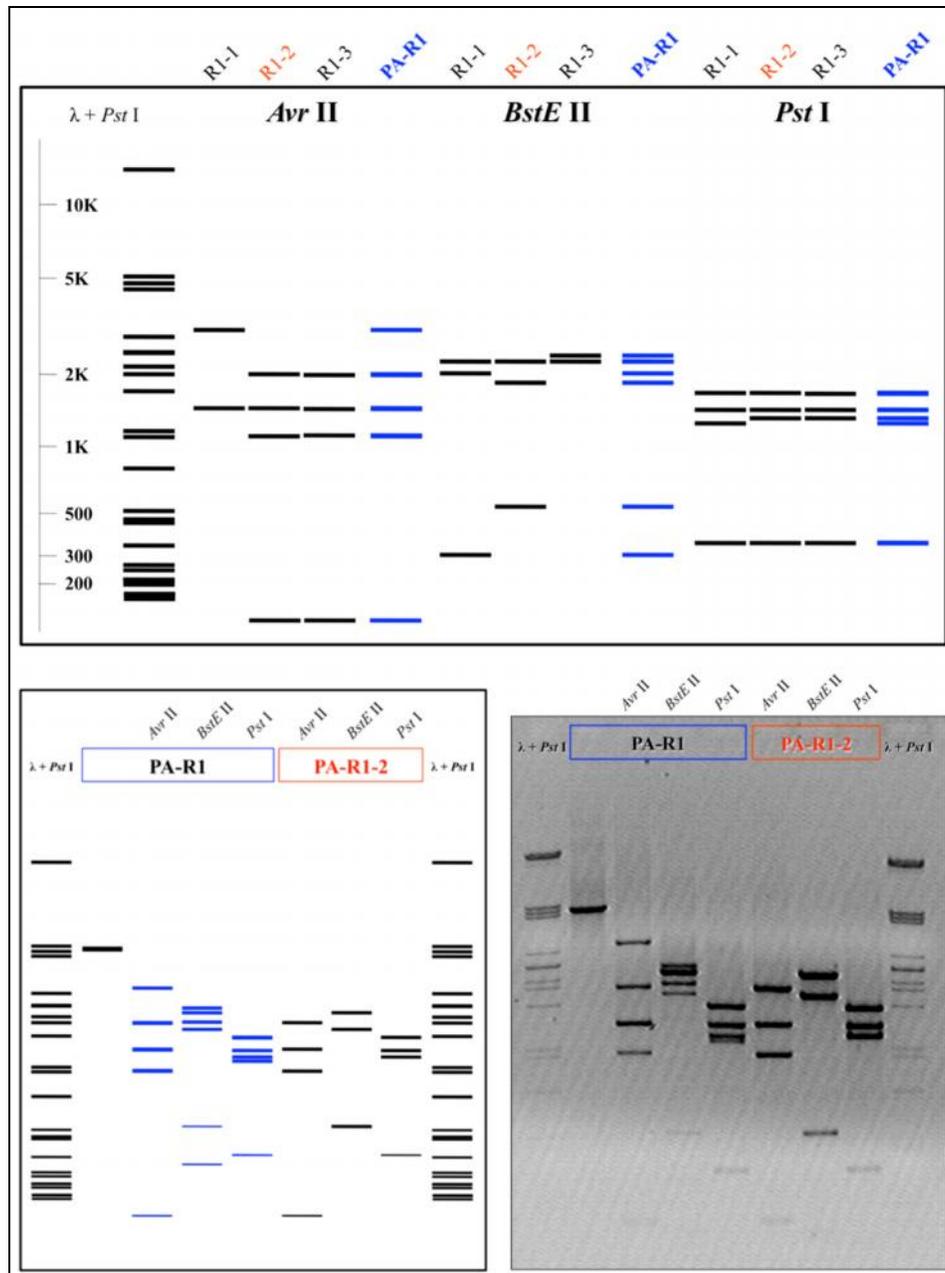


**Figura 15.** Representación de las regiones amplificadas con tres pares de iniciadores en el arreglo de genes propuestos para el *locus* GH del chimpancé. Se representan con líneas horizontales los probables productos amplificados a obtener. Los tres fragmentos R1 fueron amplificados con el par de iniciadores 03-08, tres fragmentos R2 con los iniciadores 09-10 y dos fragmentos R3 con los iniciadores 11-07.

Como se muestra en la figura 15, para un arreglo de cinco genes se esperaría la amplificación de dos regiones correspondientes a las intergénicas largas y la región posterior al último miembro del *locus*. Aproximadamente el primer tercio de estas regiones fueron amplificadas con los oligonucleótidos 03-08, generándose en la amplificación una mezcla de los productos R1-1, R1-2 y R1-3. La parte central de estas regiones fueron amplificadas con los oligonucleótidos 09-10, dando como resultado una mezcla de los fragmentos R2-1, R2-2 y R2-3. Mientras que con la pareja de oligonucleótidos 11-07 se amplificaron las partes finales de las dos regiones intergénicas largas, produciendo una mezcla de los amplicones R3-1 y R3-2. Todos los productos amplificados obtenidos se sometieron a caracterizaciones de forma independiente con las enzimas *AvrII*, *BstEII* y *PstI* y sus resultados experimentales fueron comparados con sus respectivas predicciones *in silico* derivadas del arreglo de cinco genes.

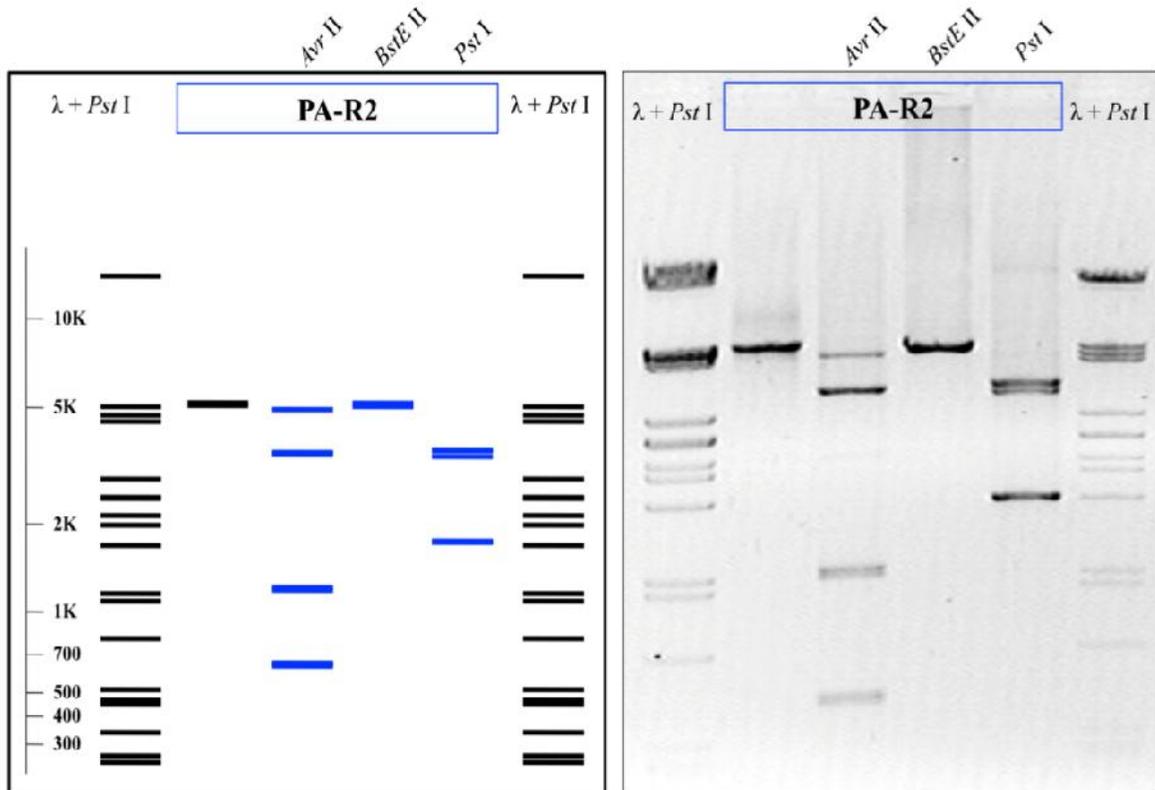
En las figuras 16, 17 y 18 se muestran las caracterizaciones con las enzimas de restricción *AvrII*, *BstEII* y *PstI*, que se obtuvieron al ser digeridas las mezclas de productos

amplificados generados (R1, R2 y R3). Los análisis con las enzimas de restricción para todas las clonas recombinantes concuerdan nuevamente con el arreglo de cinco genes.



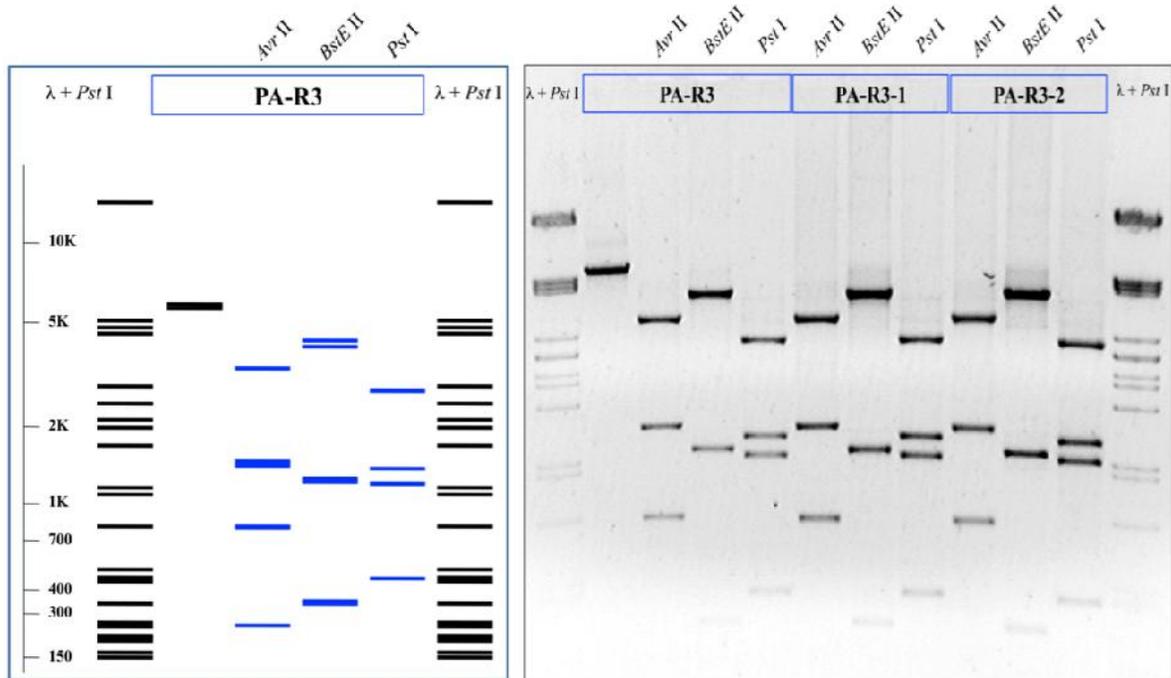
**Figura 16. Caracterización del producto amplificado de las intergénicas largas con flancos en genes.** En la parte superior se muestra cómo se realizó la predicción de los patrones de cortes para la mezcla de fragmentos. En la parte inferior izquierda se muestran los patrones de digestión predichos para la mezcla de fragmentos que se amplificarían (PA-R1) y para el producto amplificado de unas de las clonas generadas (PA-R1-2), según un arreglo de cinco genes y a la derecha se muestra un gel de agarosa al 1% en el cual se verificaron los productos de digestión. En los carriles 1 y 9 marcador de peso molecular; el carril 2 corresponde al producto amplificado PA-R1 sin digerir; los carriles 3, 4 y 5 corresponden al producto amplificado PA-R1 digerido con las enzimas de restricción *AvrII*, *BstEII* y *PstI*, respectivamente y en los carriles 6, 7 y 8, el amplicón PA-R1-2 digerido en el mismo orden con las mismas enzimas.

En la parte superior de la figura 16 se muestra cómo se realizó la predicción de los patrones de cortes primero para cada fragmento en lo individual y luego para la suma de estos. Esta figura constituye el ejemplo representativo del procedimiento seguido en aquellos casos en los que se amplificó más de un fragmento en una misma reacción de PCR.



**Figura 17. Caracterización del producto amplificado obtenido para R2.** A la izquierda se muestran los patrones de digestión predichos para R2 según el arreglo de cinco genes. A la derecha se muestra el gel de agarosa al 1% en el cual se verificaron los productos de digestión. En los carriles 1 y 6 marcador de peso molecular  $\lambda$ , carril 2 corresponde al producto amplificado sin digerir, los carriles 3, 4 y 5 corresponden al producto amplificado digerido con las enzimas de restricción *AvrII*, *BstEII* y *PstI*, respectivamente.

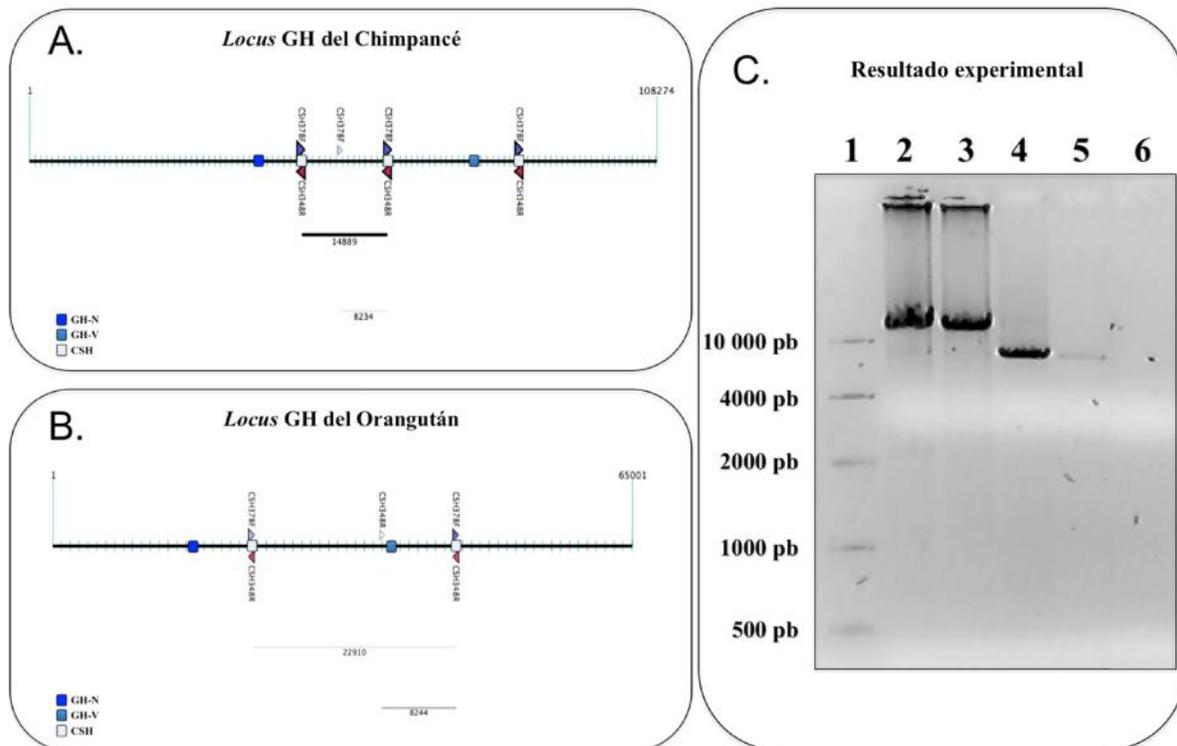
Como se puede observar en la foto del gel, el patrón de fragmento esperado coincidió con el observado con cada enzima de restricción. En el caso de que el *locus* se ajustara a un arreglo diferente de genes, este patrón se vería afectado con la presencia de varias bandas adicionales.



**Figura 18. Caracterización del producto amplificado obtenido para R3.** A la izquierda se muestran los patrones de digestión predichos para R3 según el arreglo de cinco genes. A la derecha se muestra el gel de agarosa al 1% en el cual se verificaron los productos de digestión. En los carriles 1 y 12 marcador de peso molecular  $\lambda$ , carril 2 corresponde al producto amplificado PA-R3 sin digerir, los carriles 3, 4 y 5 corresponden al producto amplificado PA-R3 digerido con las enzimas de restricción *AvrII*, *BstEII* y *PstI*, respectivamente, en los carriles 6, 7 y 8, el amplicón PA-R3-1 digerido en el mismo orden con las mismas enzimas y en los carriles 9, 10 y 11 el producto amplificado PA-R3-2 digerido de la misma forma.

### 5.1.5. Confirmación de los arreglos génicos para los *loci* GH del chimpancé y del orangután.

Otro enfoque importante para revelar los arreglos de genes presente en los *loci* de chimpancé y orangután, fue la amplificación de las intergénicas largas con flancos en genes CSHs y la posterior secuenciación de los productos amplificados. En la figura 19 (paneles A y B) se representan los productos amplificados esperados para el chimpancé y el orangután, prediciendo un arreglo pentagénico para el primero y uno tetragénico para el segundo. En el panel C se muestra el resultado experimental que confirma los arreglos predichos.



**Figura 19. Representación de las regiones amplificadas en los arreglos de genes propuestos para los *loci* GH del chimpancé (Panel A) y del orangután (Panel B).** Se representan con líneas horizontales los probables productos amplificados a obtener utilizando el par de iniciadores CSH378F-CSH348R y se indica el tamaño de los mismos en cada caso. **Panel C. Amplificación de regiones intergénicas de los *loci* GH del chimpancé y del orangután.** Producto amplificado distintivo de 14,889 pb que abarca la segunda mitad de un gen CSH, la IG larga posterior a éste y la primera mitad del siguiente gen CSH, obtenido con los oligonucleótidos CSH378F-CSH348R. Los productos de amplificación se resolvieron en geles de agarosa al 1%. En el carril 1 como marcador de peso molecular una escalera de 500 a 10,000 pb, en los carriles 2 y 3 los productos amplificados a partir del BAC del chimpancé con las premezclas 1 y 4, en los carriles 4 y 5 los productos amplificados a partir del BAC del orangután con las premezclas 1 y 4, en el carril 6 el control negativo para la PCR.

Adicionalmente, se caracterizaron otras regiones de los *loci* que corroboran dichos ensamblajes. En todos los casos, los resultados esperados (predicción *in silico* en base a secuencia) para los ensamblajes del *locus* del chimpancé basado en un arreglo pentagénico y del *locus* del orangután en uno tetragénico, coinciden con los resultados experimentales obtenidos.

## 5.2. Disección bioinformática.

### 5.2.1. Ensamblaje y anotación estructural de los *loci* GH de primates no homínidos.

Fueron ensambladas las secuencias de los BACs conteniendo los *loci* GH del lemur de cola anillada, el tití común y el mono rhesus, pero aún debe continuarse en mayor o

menor medida, con el proceso de curado de dichas secuencias. Además se inició su anotación estructural (figura 20), identificándose genes y elementos regulatorios. Adicionalmente se describieron las particularidades encontradas en sus *loci*.

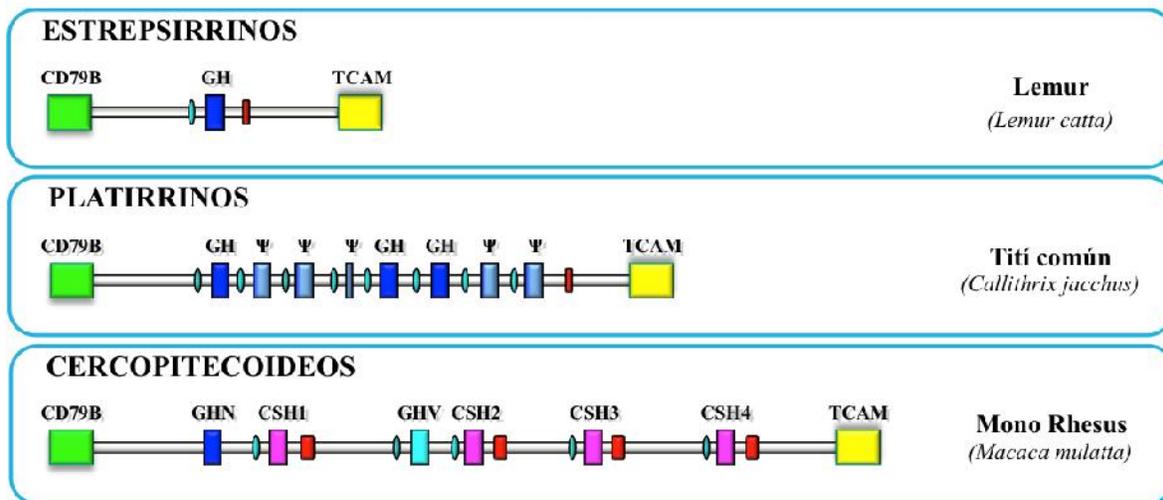


Figura 20. Anotación estructural de los *loci* GH de los primates no homínidos ensamblados. Se representan con rectángulos azules y rosados los genes GH/CSHs y pseudogenes que componen los *loci* GH. Los rectángulos verdes representan el gen CD79b no relacionado que se encuentra río arriba de los *loci* GH y los amarillos, el gen TCAM que se encuentra río abajo de los mismos. Con óvalos azules se representan las secuencias similares a los sitios del inhibidor hipofisiario descritos para el *locus* GH en el humano y con rectángulos rojos, las secuencias similares a los potenciadores placentarios descritos en el humano.

El ensamblaje del *locus* GH del lemur reveló que este es un *locus* unigénico. Se obtuvo para éste la secuencia continua que abarca desde el gen CD79b hasta el TCAM. Interesantemente su inspección nos permitió identificar secuencias similares al elemento P en un 83% y otras similares al potenciador placentario en un 78%. Para el tití común, la organización y composición génica del *locus* GH ya había sido reportada, sin embargo se extendió el ensamblaje hacia las regiones flanqueantes incluyendo a los genes CD79b y TCAM y se identificó una secuencia similar al potenciador placentario en un 88%.

En el caso del mono rhesus, se había reportado previamente la secuencia parcial del *locus* GH, sin embargo en el presente trabajo se reporta completamente dicho *locus*, que

incluye la segunda IG mediana y la tercera IG larga. Interesantemente, encontramos una inversión en dicha intergénica larga que involucra al elemento P del último gen. Además, se extendió el ensamblaje hasta incluir a los genes CD79b y TCAM.

### 5.2.1. Ensamblaje y anotación estructural de los *loci* GH de los homínidos. Genómica comparativa y evolutiva.

Los BACs conteniendo los *loci* GH de varias especies de homínidos fueron ensamblados, anotados y comparados en sus regiones codificantes, intrónicas, regulatorias y regiones flanqueantes. Se llevó a cabo un análisis detallado de los *loci* GH de los homínidos, el cual se describe a continuación. En el caso del chimpancé y el gorila, la secuencia se ha dejado totalmente curada, mientras que para el orangután dicha secuencia se encuentra muy próxima al curado definitivo, por ello no ha sido incluida en todos los análisis.

Con las secuencias obtenidas en este trabajo, las recuperadas del GenBank y las caracterizaciones enzimáticas, fue posible determinar la organización genómica de los *loci* GH de los homínidos analizados, tal y como se muestra en la figura 21. Adicionalmente, se estableció una nueva nomenclatura (cuadro 2) para los genes que lo conforman con el objetivo de uniformar ésta con respecto a la del humano y así facilitar su seguimiento y la mejor comprensión del manuscrito.

**Cuadro 2. Nomenclatura usada para los genes de los *loci* GH.**

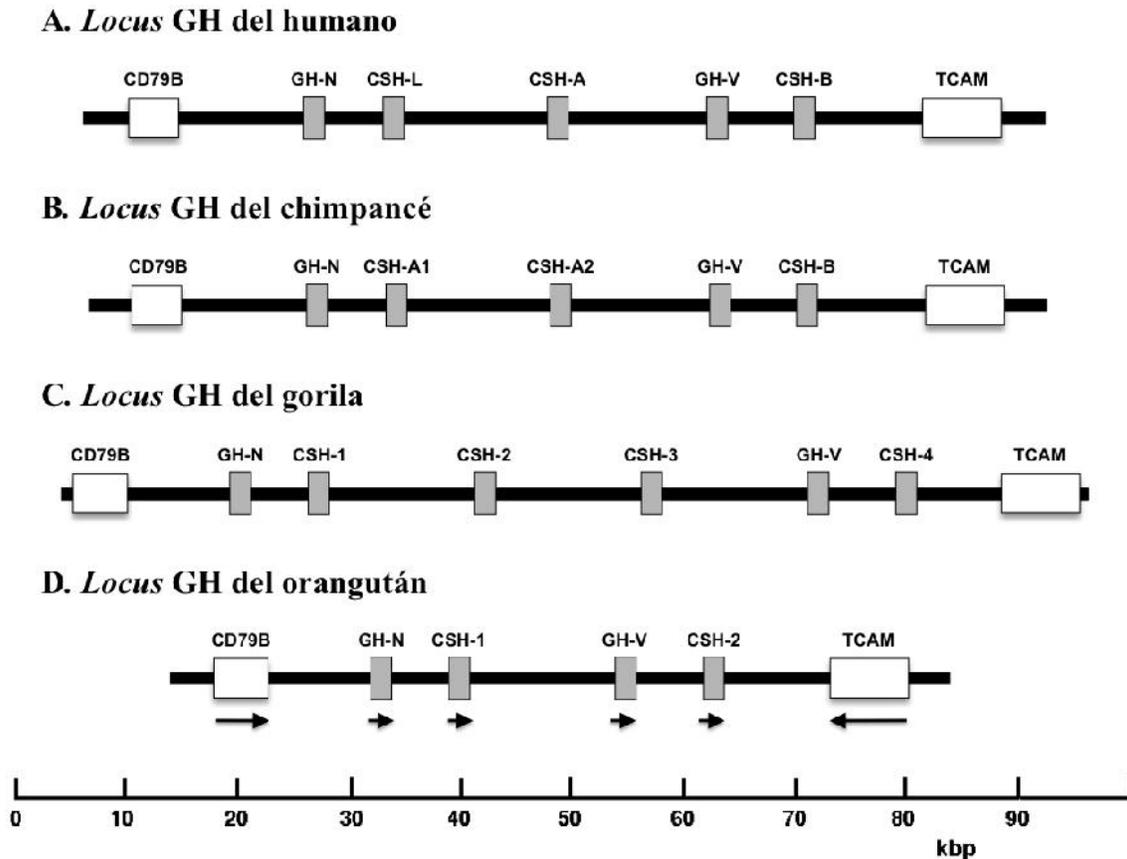
Posición en el <i>locus</i> H-C/G/O	Gen	Abreviaturas				Tejido de expresión
		Humano	Chimpancé	Gorila	Orangután	
1/1/1	Hormona de crecimiento “Normal”	<i>hGH-N</i>	<i>cGH-N</i>	<i>gGH-N</i>	<i>oGH-N</i>	hipófisis
2/2/2	Somatomotropina coriónica	<i>hCSH-L</i>	<i>cCSH-A1</i>	<i>gCSH-1</i>	<i>oCSH-1</i>	placenta
3/3	Somatomotropina coriónica	<i>hCSH-A</i>	<i>cCSH-A2</i>	<i>gCSH-2</i>		placenta
4	Somatomotropina coriónica			<i>gCSH-3</i>		placenta
4/5/3	Hormona del crecimiento Variante	<i>hGH-V</i>	<i>cGH-V</i>	<i>gGH-V</i>	<i>oGH-V</i>	placenta
5/6/4	Somatomotropina coriónica	<i>hCSH-B</i>	<i>cCSH-B</i>	<i>gCSH-4</i>	<i>oCSH-2</i>	placenta

La organización general del *locus* GH del chimpancé es similar a la de humano, con 5 genes, que codifican la GH hipofisiaria (cGH-N), la GH variante (cGH-V) y tres CSHs (cCSH-A1, cCSH-A2 y cCSH-B, equivalentes a hCSH-L, hCSH-A y hCSH-B, respectivamente; figura 21). La similitud general entre estos cinco genes es aproximadamente del 93%, reflejando su reciente generación.

El *locus* gGH, contiene seis genes GH/CSH, en contraste con los cinco genes del chimpancé y del bien caracterizado *locus* GH humano (Chen *et al.* 1989). Dada la similitud general entre los genomas del chimpancé y del gorila con el humano, las diferencias podrían ser potencialmente importantes y dignas de mayor investigación. Sin embargo, es evidente que la naturaleza repetitiva de la región del *locus* GH, que surge como consecuencia de varias rondas de duplicación génica (ver más abajo), conduce a problemas en la obtención de un ensamblaje exacto.

Los datos disponibles para los *loci* GH en los ensamblajes genómicos (<http://www.ensembl.org/gorilla/>), también fueron examinados en detalle, utilizando los datos disponibles en la base de datos WGS Trace (<http://www.ncbi.nlm.nih.gov/Traces>). Siete genes completos fueron bien soportados en el ensamblaje del *locus* GH del gorila. Su secuencia se completó sobre la base de los datos derivados de la secuenciación del BAC, en el cual se detecta un sólo gen para la sexta posición, debido a que el polimorfismo se ha perdido en la clonación del BAC. Una evaluación detallada permitió el ensamblaje de los datos de la secuenciación genómica para dar una secuencia completa para el *locus* GH del gorila. Los resultados de estos dos enfoques identificaron seis genes GH/CSHs en el gorila y cinco en el chimpancé, que se corresponden cercanamente con los cinco genes del *locus* GH humano.

En base a los hallazgos anteriores, se concluye que la organización general del *locus* GH del gorila (Fig. 21) es diferente a la del humano y chimpancé, con seis genes que codifican la GH pituitaria (gGH-N), GH variante (gGH-V) y cuatro CSHs (gCSH-1, gCSH-2, gCSH-3 y gCSH-4). La similitud global entre estos seis genes es de aproximadamente 94%, lo que refleja su reciente generación (similar a lo observado en el chimpancé).



**Figura 21. Organización genómica de los loci GH del humano, chimpancé, gorila y orangután.** A. El locus GH humano (basada en Chen *et al.* 1989). B. El locus GH del chimpancé basado en el presente trabajo. C. El locus GH del gorila basado en el presente trabajo. D. El locus GH del orangután basado en el presente trabajo. Los cuadros en gris indican las secuencias de los genes de GH/CSHs. Los cuadros blancos indican las secuencias de los genes CD79B y TCAM que flanquean el locus GH en los ensamblajes representados. El locus GH en el humano, el chimpancé, el gorila y el orangután se encuentra en el cromosoma 17 (17q24.2 en el humano) y se muestra aquí en la orientación inversa. Las flechas indican la dirección transcripcional. Adicionalmente fueron identificados en el BAC, pero no son mostrados, los genes SCN4A e ICAM2 río arriba de CD79b y SMARCD2, río abajo del pseudogen TCAM1.

En el caso del orangután se identificaron cuatro genes (figura 21) que codifican la GH hipofisiaria (oGH-N), GH variante (oGH-V) y dos CSHs (oCSH1 y oCSH2). La similitud global entre estos cuatro genes es del 91% aproximadamente (datos no mostrados). Uno de los genes CSHs tiene una alteración en un sitio 5' donador de splicing en el intrón 2 que genera una proteína incapaz de ser secretada quedando por tanto atrapada en el interior de la célula. El locus GH del orangután genera en total solo tres productos

proteicos funcionales. Los *loci* GH del gorila, chimpancé y humanos difieren en detalle en un número de aspectos, mismos que se discuten a continuación.

### 5.2.1.1. Análisis de las proteínas.

De las secuencias genómicas correspondientes a los genes que integran los *loci* GH de los primates analizados en este trabajo, con la ayuda del programa Genescan se predijeron las secuencias proteicas codificantes. En la figura 22 se muestra el alineamiento de las proteínas GH/CSH del gorila, del chimpancé y del humano, usando como referencia a la secuencia del gen de la GH hipofisaria del humano (hGH-N).

	10	20	30	40	50	60	70	80	90	100
<b>hGHN</b>	<b>FPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIPTPSNREETQOKSNLELLLRISLLLIQSWLEPVQFLRSVFANS</b>									
cGHN	.....									
gGHN	.....									
hGHV	.....R.Y...Y.....L.....VK.....L.....									
cGHV	.....R.Y...Y.....L.....VK.....L.....									
gGHV	..I.....R.D...Y.....L.....VK.....L.....									
hCSHL	VQ.V.....KE...Q...A...I.....T.....HDS...F...D...S...M.....H.....E.R...R...T.T.N									
hCSHA	VQ.V.....H...Q...A...I.....T...D.....HDS...F...D...M.....E.....R...M...N									
hCSHB	VQ.V.....H...Q...A...I.....T...D.....HDS...F...D...M.....E.....R...M...N									
cCSHA1	VQ.V.....H...Q...A...I.....D.....HDS...F...D...M.....E.....R...M...N									
cCSHA2	VQ.V.....H...Q...A...I.....D.....HDS...F...D...M.....E.....R...M...N									
cCSHB	VQ.V.....H...Q...A...I.....D.....HDS...F...D...M...H.....E.....R...M...N									
gCSH1	VQ.V.....KE.V.Q...A...I.....D.....HDS...F...D...M.....G.H.....E.....MR...I.T.N									
gCSH2	VQ.V.....H...Q...A...I.....D.....H.S...F...D...M.....E.....R...M...N									
gCSH3	VQ.V.....H...Q...A...I.....D.....H.S...F...D...M.....E.....R...M...N									
gCSH4	VQ.V.....H...Q...A...I.....D.....H.S...F...D...M.....E.....R...M...N									
	110	120	130	140	150	160	170	180	190	
<b>hGHN</b>	<b>LVYGASDSNVYDLLKDLLEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDALLKNYGLLYCFRKMDKRVETFLRIVQCRSVEGSCGF</b>									
cGHN	.....E.....									
gGHN	.....									
hGHV	.....RH.....W.....N.S.....K.....									
cGHV	.....RH.....W.....N.S.....K.....									
gGHV	.....RH.....W.....N.S...S...K.....									
hCSHL	..DT...DD.H.....M.....HL...TL.....H.....H.....M.....									
hCSHA	..DT...DD.H.....R.....L.....H.....H.....M.....									
hCSHB	..DT...DD.H.....R.....L.....H.....H.....M.....									
cCSHA1	..DT...DD.H.....M.....R.....L.....H.....H.....M.....									
cCSHA2	..DT...DD.H.....M.....R.....L.....H.....H.....M.....									
cCSHB	..DT...DD.H.....R.....R.....L.....H.....H.....M.....									
gCSH1	..DT...DD.H.....R.....L.....H.....H.....M...F.....									
gCSH2	..DT...DD.H.....R.....L.....H.....H.....M.....									
gCSH3	..DT...DD.H.....R.....L.....HE.....H.....M.....									
gCSH4	..DT...DD.H.....R.....L.....H.....H.....M.....									

**Figura 22. Alineamiento de las secuencias aminoacídicas de las proteínas maduras .** La hGH-N se utiliza como referencia; coincidencias con esta en las otras secuencias se muestran como “.” y deleciones como “-”. La secuencia dada para hCSH-L es una traducción conceptual, asumiendo un empalme equivalente al observado en los otros genes.

Resultó de interés investigar si las diferencias aminoacídicas encontradas ocurrían en los sitios clave de unión al receptor. Adicionalmente se le prestó una especial atención a

aquellos residuos que basado en datos experimentales se han reportado que juegan un papel determinante en la función, siendo estos en GHs: Ile4, Glu56, Arg64 e Ile179; mientras que en CSHs Val4, Asp56, Met64 y Met179 (Goffin *et al.*, 1996). También se tuvo en cuenta el residuo de fenilalanina de la posición 44 (Phe44), ya que se ha sugerido que puede tener una importancia estructural particular en el sitio I de unión de la hormona con su receptor y estar relacionado con la actividad lactogénica de la misma (Peterson y Brooks, 1997; Duda y Brooks, 2003).

El gen GH-N de chimpancé codifica una proteína que difiere de la correspondiente del gen humano en dos posiciones: una en el péptido señal (residuo 23 Pro en chimpancé, Thr en humano) y una en la proteína madura (residuo 132, Glu en chimpancé, Gly en humano) (figura 22). El gen GH-N del gorila codifica una proteína idéntica a la GH-N humana, excepto por una sustitución de Thr por Ala en el residuo -24 del péptido señal.

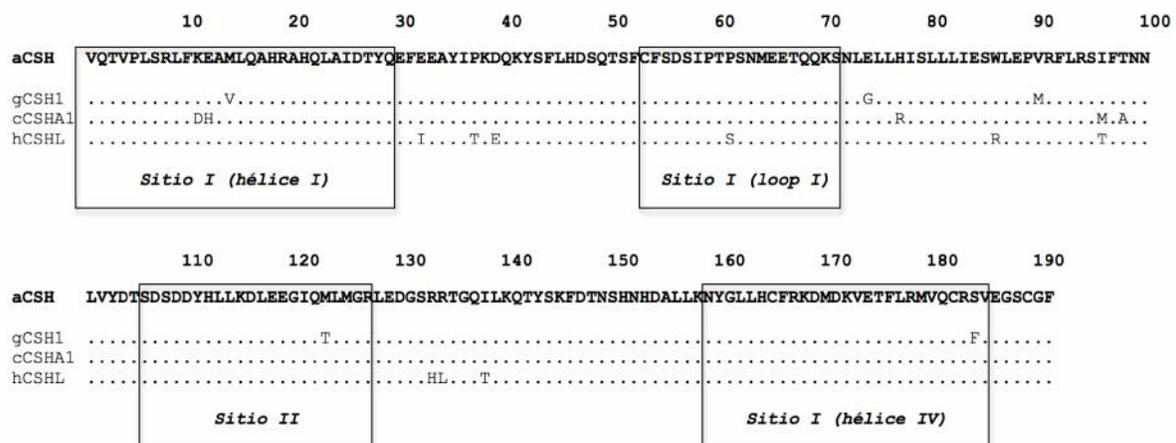
La proteína GH-V madura del chimpancé resultó ser idéntica a la del humano, sin embargo sus secuencias del péptido señal se diferencian por una sustitución de Pro por Ser en el residuo 8 (figura 23). El gen GH-V gorila codifica una proteína que difiere de los correspondientes genes humanos y de chimpancés en tres posiciones (Fig. 22), Ile-3, Asp-21 y Ser-146 en gorila, Thr, Tyr y Phe en chimpancés y humanos).

<b>hGHN-PS</b>	<b>MATGSRTSLLLA<b>F</b>GLLCLPWLQ<b>E</b>GSA</b>
cGHN-PS	MA <b>P</b> GSRTSLLLA <b>F</b> GLLCLPWLQ <b>R</b> GSA
gGHN-PS	MA <b>A</b> GSRTSLLLA <b>F</b> GLLCLPWLQ <b>E</b> GSA
hGHV-PS	MA <b>A</b> GSRTSLLLA <b>F</b> GLLCL <b>S</b> WLQ <b>E</b> GSA
cGHV-PS	MA <b>A</b> GSRTSLLLA <b>F</b> GLLCLPWLQ <b>E</b> GSA
gGHV-PS	MA <b>A</b> GSRTSLLLA <b>F</b> GLLCLPWLQ <b>E</b> GSA
hCSHL-PS	MA <b>A</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
hCSHA-PS	MA <b>P</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
hCSHB-PS	MA <b>A</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
cCSHA1-PS	MA <b>A</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
cCSHA2-PS	MA <b>P</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
cCSHB-PS	MA <b>P</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
gCSH1-PS	<b>F</b> APGSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
gCSH2-PS	MA <b>P</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
gCSH3-PS	MA <b>P</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
gCSH4-PS	MA <b>P</b> GSRTSLLLA <b>F</b> AL <b>L</b> CLPWLQ <b>E</b> AG <b>A</b>
	:*,*****.,****.,****.,.*

**Figura 23. Alineamiento de las secuencias aminoacídicas correspondientes al péptido señal.** El péptido señal de hGH-N se utiliza como referencia. La secuencia dada para hCSH-L es una traducción conceptual, asumiendo un empalme equivalente al observado en los otros genes.

El gen CSH1 de gorila codifica una proteína que es sustancialmente diferente de las otras CSHs. Las proteínas gCSH1 y gCSH2 difieren en doce posiciones (Figs. 22 y 23), una en el péptido señal (residuos -26, Phe en gCSH1, Met en gCSH2) y once en la proteína madura: Asp-11, His-12, Met-14, Glu-39, Asn-47, Glu-74, Arg-77, Val-90, Met-96, Ala-98 y Ser-184 en CSH1 son reemplazados por Lys, Glu, Val, Asp, Asp, Gly, His, Met, Ile, Thr y Phe, respectivamente.

El pseudogen putativo CSH-L humano codifica potencialmente una proteína que es sustancialmente diferente del resto de las CSHs humanas, debido a una sustitución en el sitio donador de “splicing” en el intrón 2. En los genes correspondientes de gorila y chimpancé (CSH1 y CSH-A1) el sitio donador del empalme en el intrón 2 está intacto. Se llevó a cabo la reconstrucción del potencial producto proteico (aCSH) del ancestro común de los genes que codifican para hCSH-L, cCSHA1 y gCSH1 con el propósito de entender mejor lo qué le sucedió a este potencial pseudogen durante la evolución de los homínidos (fig. 24).

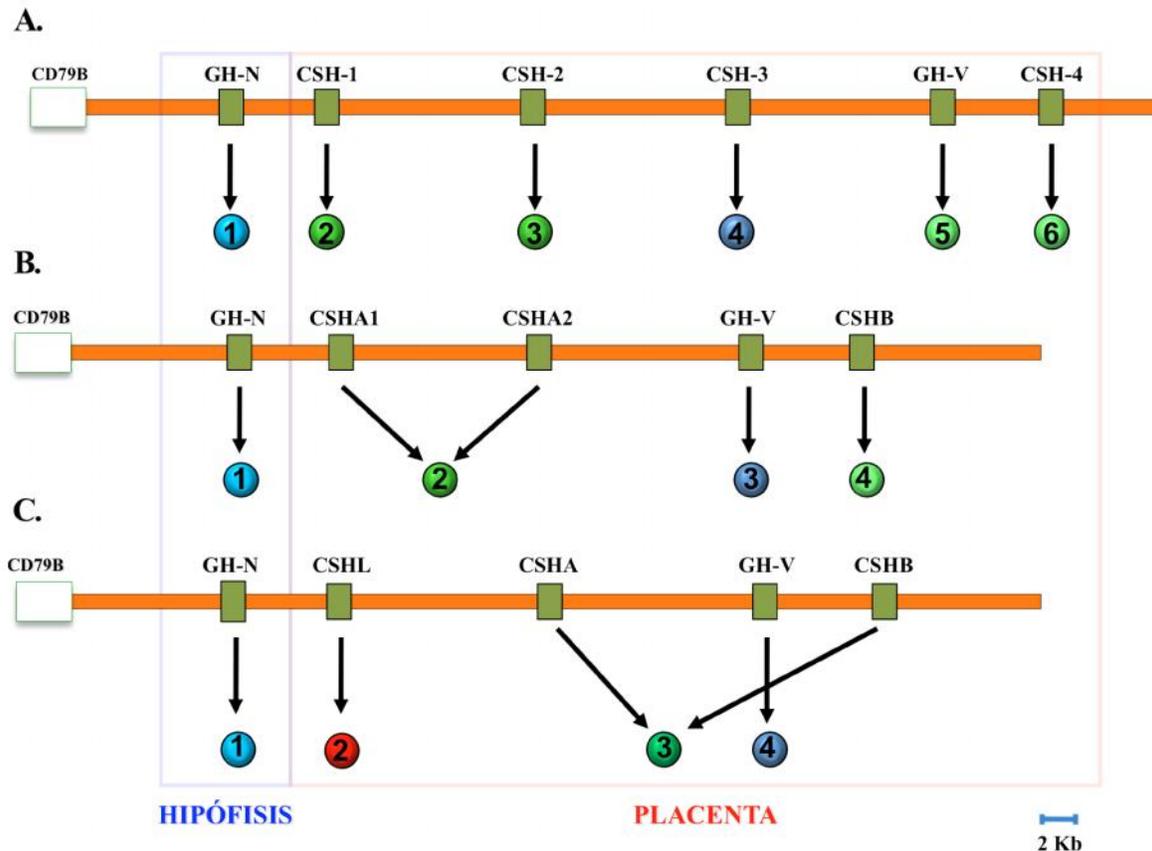


**Figura 24. Reconstrucción de la secuencia aminoacídica del ancestro común de los genes gCSH1, cCSHA1 y hCSH-L.** La secuencia reconstruida del ancestro común (aCSH) de hCSH-L, cCSHA1 y gCSH1 se utiliza como referencia; coincidencias con esta en las otras secuencias se muestran como “.” y deleciones como “-”. La secuencia dada para hCSH-L es una traducción conceptual, asumiendo un corte y empalme de exones equivalente al observado en los otros genes.

Las proteínas maduras gCSH2 y gCSH3 difieren en un residuo (el residuo 154 es Asp en CSH2 y Glu en CSH3). El gen CSH4 en gorila codifica una proteína madura que también difiere de gCSH2 en un sólo residuo (Arg-91 en CSH2 se sustituye por Gln).

Las proteínas maduras hCSH-A/B y gCSH2 difieren en los residuos, 34, 39, 47 y 164 (Thr, Asp, Asp y Tyr en humano, Ala, Glu, Asn y His en gorilla, respectivamente). Mientras que las proteínas cCSHA1/A2 y gCSH2 difieren en los residuos 39, 47 y 123 (Asp, Asp y Met en chimpancé, Glu, Asn y Thr en gorilla, respectivamente). En tanto que cCSHB y gCSH2 difieren en los residuos 39, 47, 69 y 126 (Asp, Asp, His y Arg en chimpancé, Glu, Asn, Gln y Gly, respectivamente).

En la figura 25 se representan las proteínas deducidas de los genes de los *loci* GH del humano, chimpancé y gorila. Como se puede observar en el humano, hay un potencial pseudogen representado con un círculo rojo sin número que correspondería a una proteína incapaz de secretarse y los dos genes CSH codifican la misma proteína madura (Barrera-Saldaña *et al.*, 1983). En contraste, en el chimpancé todos los genes se predicen funcionales y de las proteínas codificadas por los genes CSH dos codifican la misma proteína madura y comparten con la tercera una similitud superior del 98%. En el gorila, igualmente todos los genes se predicen funcionales y codifican proteínas diferentes. Tres de sus genes CSHs (gCSH2, gCSH3 y gCSH4) codifican proteínas maduras que tienen una similitud entre ellas superior al 98% y de alrededor del 93 y 94% con respecto a la codificada por gCSH1. Sin embargo en el orangután solo hay dos CSHs, de las cuales una es incapaz de secretarse, similar a lo que sucede con la CSH-L de humano, por lo que sólo tres de los cuatro genes que presentan dicho *locus* se predicen funcionales (datos no mostrados).



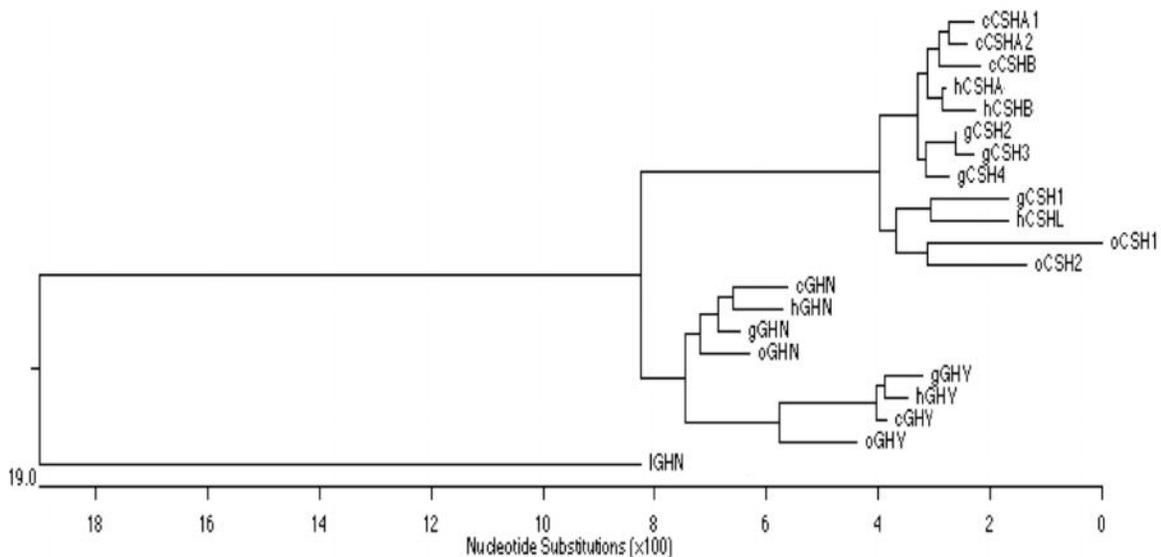
**Figura 25. Representación de los productos proteicos deducidos de los loci GH del gorila (A), del chimpancé (B) y del humano (C).** Las cajas coloreadas indican los genes que conforman los loci GH, mientras que las blancas al gen CD79b que flanquea la región 5' de cada locus y las líneas sólidas representan a las RIGs. Los círculos representan las proteínas maduras deducidas. **A.** Locus GH del gorila formado por seis genes que producen seis productos proteicos diferentes. **B.** Locus GH pentagénico del chimpancé que genera cuatro proteínas diferentes, debido a que dos de los genes codifican la misma proteína madura. **C.** Locus GH pentagénico del humano que produce cuatro proteínas maduras, dos genes codifican para la misma proteína madura y otro es incapaz de rendir una proteína secretada pues queda atrapada en el interior de la célula debido a una alteración en el gen a nivel del sitio 5' donador de "splicing" en el intrón 2, generándose en total solo tres productos proteicos funcionales en el locus.

Un análisis comparativo entre las proteínas ortólogas de humano y chimpancé muestra que todas las secuencias aminoacídicas del chimpancé exhiben una similitud arriba del 98% respecto a su ortólogo en el humano, excepto por CSH-A1, cuyo valor resultó del 94%. Interesantemente, esta proteína mostró mayor similitud con hCSH-A y hCSH-B (98%). La secuencia aminoacídica de la GH-V de humano y chimpancé son idénticas. La GH-N de humano y chimpancé comparten un 99.5% de similitud en sus secuencias proteicas (un solo cambio aminoacídico). En el humano los genes CSH-A y CSH-B codifican la misma proteína madura y en el chimpancé son los genes CSH-A1 y CSH-A2 los que codifican la misma proteína madura. Por otra parte, el gorila comparte con el

humano idéntica proteína madura (GH-N) y difiere en su GH-V, de éste y del chimpancé, en tres posiciones aminoacídicas. El orangután comparte la misma GH-N (proteína madura) con el humano y con el gorila y su GH-V madura difiere con la del chimpancé y el humano en un 6% y con la del gorila en un 7%.

### 5.2.1.2. Análisis filogenético de los genes GH/CSH de humano y chimpancé.

En la figura 26 se muestra el árbol filogenético de los genes GH/CSH de humano, chimpancé, gorila y orangután, en cuya construcción se incluyó como grupo fuera la secuencia del lémur. Sólo se observa clara ortología entre los genes GH-N y GH-V de los cuatro homínidos, mientras que para las CSHs se observa un agrupamiento especie-específico y no se aprecia una clara ortología entre ellas. Adicionalmente se utilizaron otros métodos como máxima parsimonia y "maximum likelihood" para construir el árbol, pero siempre se obtuvo la misma topología (González *et al.*, 2006).



**Figura 26. Árbol filogenético de los genes GH/CSH de humano, chimpancé, gorila y orangután.** El árbol fue construido con el programa MEGA 3.0 por el método de NJ, utilizando las secuencias codificantes de los genes del chimpancé, gorila, orangután y lémur generadas en este estudio y como referencia las del humano.

### 5.2.1.3. Un vistazo a los reguladores de la transcripción.

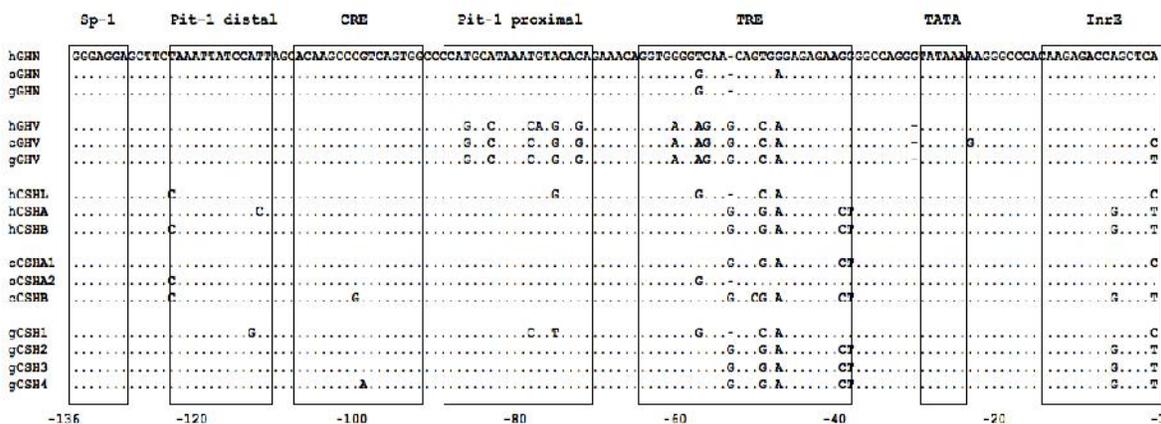
Muchos elementos reguladores se han identificado en el *locus* hGH, incluyendo los asociados con las regiones promotoras inmediatamente río arriba de cada gen, un elemento de respuesta a glucocorticoides en el intrón 1 del gen GH-N, un potenciador de la expresión placentaria río abajo de los genes CSHs, " elementos P " río arriba de cada gen, (excepto de

la GH-N) y una región de control del *locus* lejos río arriba de la GH-N (Ho *et al.* 2004). El último de éstos cae fuera del *locus* GH considerado aquí, por lo que el resto de los elementos se analizan en detalle a continuación.

#### 5.2.1.4. El promotor proximal.

En la figura 27 se muestra el alineamiento de las secuencias correspondientes al promotor proximal de los genes GH/CSH de humano, chimpancé y gorila, indicándose los sitios de unión a factores transcripcionales descritos para el primero.

Estos sitios incluyen la caja TATA, los elementos Pit-1 proximal y distal (Theill y Karin, 1993; Krawczak *et al.*, 1999), un elemento de respuesta a AMPc (CRE) (Eberhard *et al.*, 1996), un elemento de respuesta a hormona tiroidea (TRE) (Glass *et al.*, 1987), un sitio de unión del factor SP-1 (Lemaigre *et al.*, 1989) y un sitio de unión del iniciador (InrE).



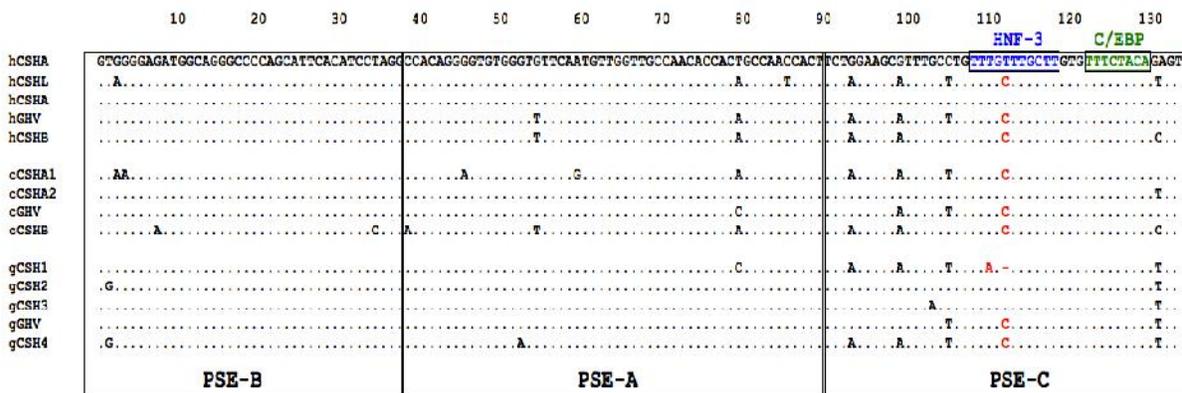
**Figura 27. Alineamiento de los promotores proximales.** El gen de hGH-N se utiliza como referencia; coincidencias a éste en otras secuencias se muestran como "." y deleciones como "-". Los elementos reguladores que han sido identificados en el gen hGH-N se muestran en los rectángulos.

Los promotores proximales exhiben entre 97 y 99% de similitud entre sí y todos poseen los elementos mínimos indispensables para la transcripción basal. Sin embargo, se requieren ensayos de transfección frente algún gen reportero para evaluar el significado biológico de los cambios puntuales observados. En el caso de los elementos Pit-1, se puede observar que tanto el distal, como el proximal se encuentran altamente conservados respecto a la secuencia consenso, excepto para las GH-V de humano y chimpancé, cuyos promotores exhiben algunas variaciones, coincidiendo en ambos casos.

En el caso particular del lemur, su promotor proximal exhibe una similitud entre el 70 y 77% con respecto al de los homínidos analizados (datos no mostrados)

**5.2.1.5. El inhibidor hipofisiario o elemento P.**

Este elemento que inhibe la expresión de los genes placentarios en la hipófisis se encontró en todas las regiones intergénicas del chimpancé. El elemento P se encontró en cada una de las regiones intergénicas del chimpancé. En el genoma humano el elemento P está situado en las posiciones -2,439; -2,386; -2,287 y -2,690 de los genes CSH-L, CSH-A, GH-V y CSH-B, respectivamente. En el chimpancé está en las posiciones -2,258; -2,249; -2,222 y -2,515 de los genes CSH-A1, CSH-A2, GH-V y CSH-B, respectivamente. En la figura 28 se muestra el alineamiento de las 90 pb que comprenden a esta secuencia reguladora para humano y chimpancé. Todas las secuencias entre sí tienen una similitud arriba del 95.6%.



**Figura 28. Alineamiento de los inhibidores hipofisiarios.** Las secuencias de los cuatro inhibidores hipofisiarios predichos para el locus del chimpancé y los cinco para el locus del gorila, se alinearon con sus homólogos en el humano. El gen de hCSH-A se utiliza como referencia; coincidencias con esta en otras secuencias se muestran como “.” y deleciones como “-”. Los rectángulos muestran los tres dominios de unión a proteínas (PSE-A, PSE-B y PSE-C) del elemento P (Norquay *et al.*, 2006). El sitio de unión predicho para HNF-3, C/EBP, la familia de RFX y la mitad del sitio para la familia NF-1 están encerrados en un rectángulo.

Interesantemente, en el lemur parece haber una secuencia similar a un elemento P con una similitud entre el 76 y el 80% (Datos no mostrados) con respecto a los elementos P de los homínidos alineados en la figura 28. Dicha secuencia no está presente en la rata, pero sí en el perro en el cual se presenta con una similitud entre el 39 y el 51% con respecto al

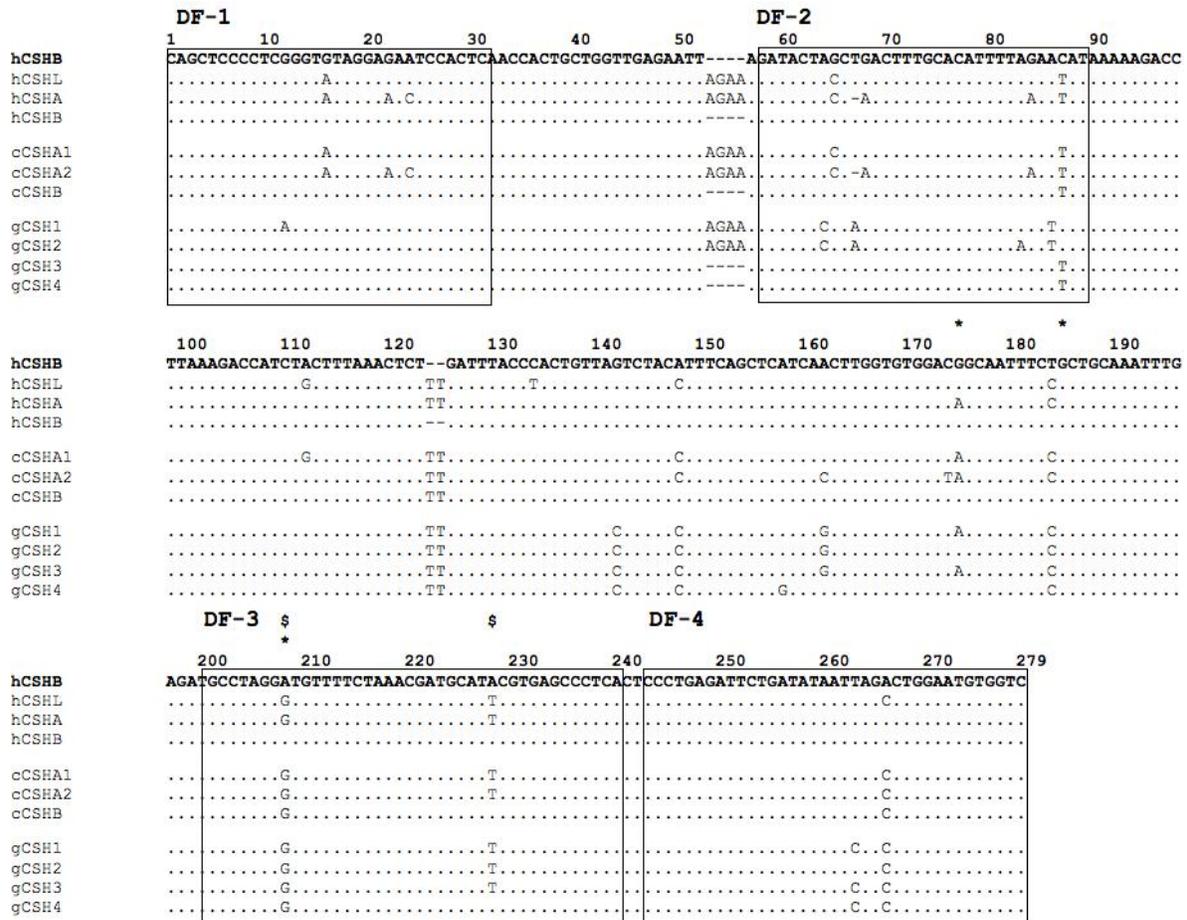
de los homínidos alineados. Quizás a este nivel el mecanismo de regulación es diferente al descrito en humanos.

#### **5.2.1.6. El potenciador placentario o enhancer.**

Este elemento que potencia la expresión de los genes CSH en la placenta, se encontró aproximadamente a 2.3 kpb después de estos genes en el chimpancé. En la figura 29 se muestra el alineamiento de las 280 pb que comprenden a esta secuencia reguladora en el chimpancé y en el humano, indicándose cuatro regiones identificadas en el humano por “footprinting” (Jacquemin *et al.*, 1994), de las cuales DF3 y DF4 parecen ser las más importantes.

Todas las secuencias entre sí exhibieron una similitud arriba del 97.9%. En el humano los potenciadores de los genes hCSH-L y hCSH-A presentan una inserción de cuatro pares de bases río arriba del dominio DF2, que se ha demostrado que reducen la funcionalidad de éstos (Jacquemin *et al.*, 1994). Esta misma inserción se encontró en los potenciadores de los genes cCSH-A1 y cCSH-A2 del chimpancé. En el dominio DF3 se indican dos nucleótidos presentes en el gen hCSH-B, que son reemplazados por G y T en los genes hCSH-L y hCSH-A en el humano. Estos cambios en el dominio DF3 de dichos potenciadores son responsables de su actividad reducida (Jacquemin *et al.*, 1996). Los potenciadores de los genes cCSH-A1 y cCSH-A2 del chimpancé también son portadores de estas dos mutaciones, sin embargo, el gen cCSH-B solo presenta una de estas.

Unas secuencias similares a este elemento fueron halladas en lemur y en el tití común, exhibiendo similitudes de 76-80 y 77-88%, respectivamente. Enfatizar las diferencias en los dominios DF3 y DF4 y sugerir que son necesarios ensayos de funcionalidad. Describir también si es posible la situación del orangután en este punto (datos no mostrados).

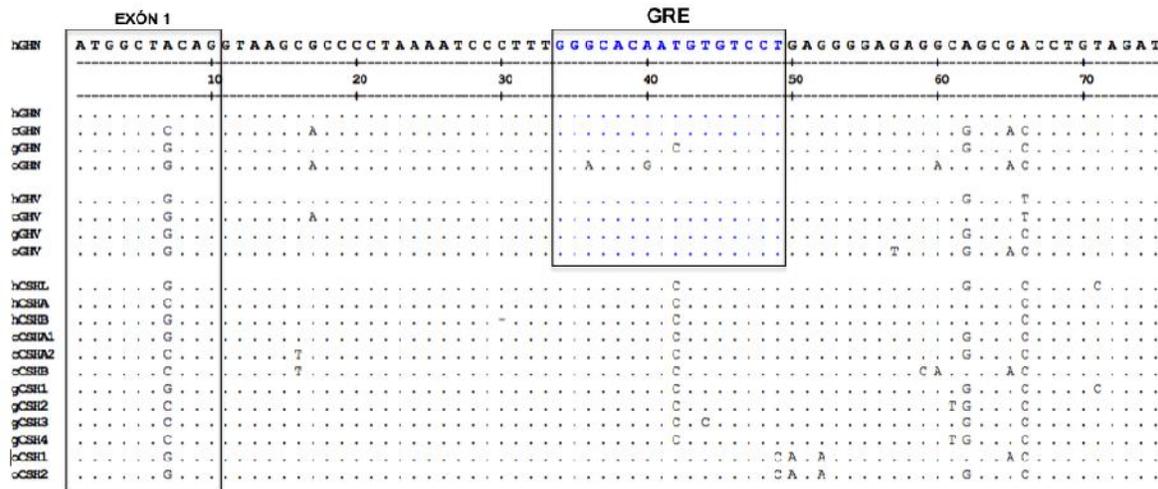


**Figura 29. Alineamiento de los potenciadores placentarios.** Las secuencias de los tres potenciadores placentarios predichos para el *locus* del chimpancé y los cuatro para el del gorila se alinearon con sus homólogos en el humano. El gen de hCSH-B se utiliza como referencia; coincidencias con esta en las otras secuencias se muestran como “.” y deleciones como “-”. Los rectángulos muestran cuatro dominios DF del potenciador placentario. Los nucleótidos que cuando están mutados causan una reducción en la actividad del potenciador se muestran como \* (Lytras *et al.*, 1996) y \$ (Jacquemin *et al.*, 1996).

### 5.2.1.7. Elementos de respuesta a glucocorticoides.

Se ha reportado la presencia de un elemento de respuesta a glucocorticoides (GRE) en el primer intrón del gen de la GH-N humana (Slater *et al.*, 1985). En la figura 30 se muestra parte de un alineamiento de las secuencias nucleotídicas de los genes GH/CSH de humano, chimpancé y gorila que abarca el exón 1 y parte del primer intrón. El GRE fue identificado en el alineamiento tomando como referencia el reporte previamente mencionado de su secuencia para el gen de la hGH-N. En este trabajo se determinó que dicho elemento se encuentra conservado en los genes GH-N y GH-V de chimpancé, notándose que también lo está en el gen GH-V humano y en el de la GH-V de gorila y

orangután. Sin embargo, sustituciones nucleotídicas presentes en el gen GH-N de gorila y orangután sugiere que probablemente dicho elemento en estas especies no es funcional.

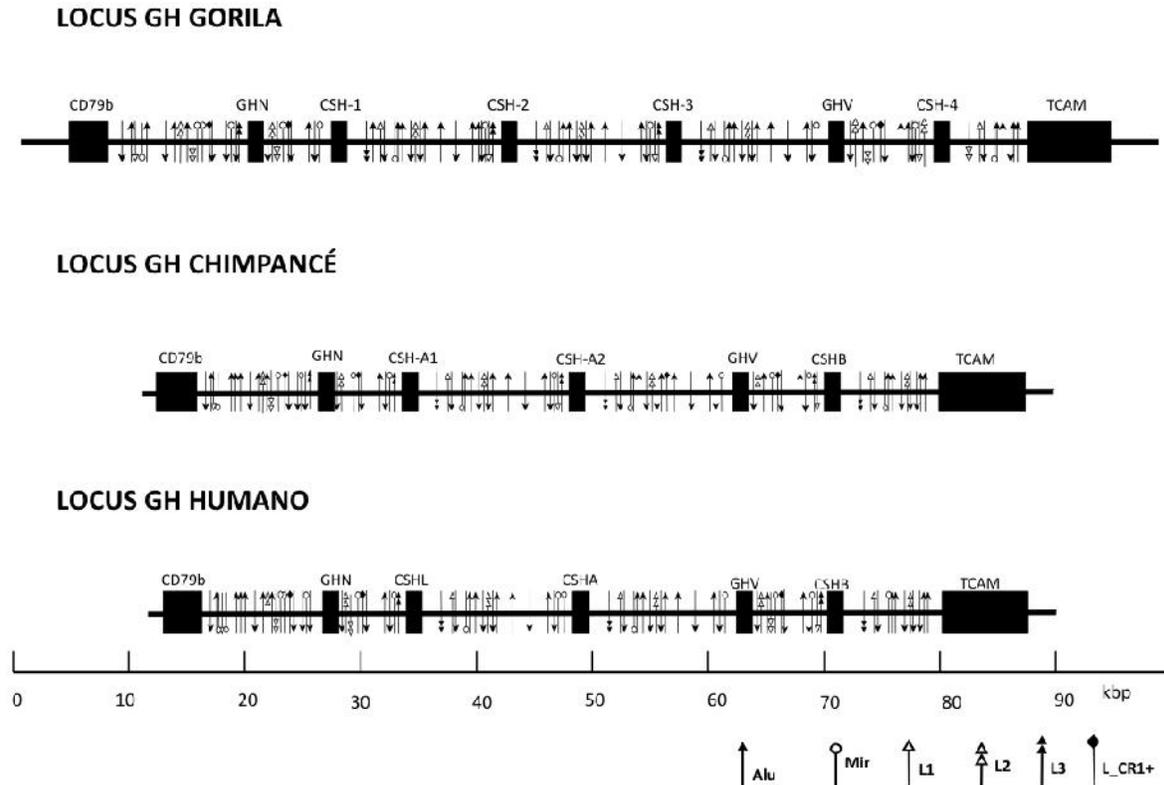


**Figura 30. Alineamiento del exón 1 y el comienzo del intrón 1 en los genes GH/CSH.** Las secuencias del exón 1 y el comienzo del primer intrón de los genes GH/CSH del chimpancé, el gorila y el orangután, se alinearon con sus homólogos en el humano. El gen de hGH-N se utiliza como referencia; coincidencias con ésta en otras secuencias se muestran como “.” y deleciones como “-”. En la parte superior de la figura se presentan los diez y seis nucleótidos de la secuencia reportada previamente para el elemento de respuesta a glucocorticoides (GRE) (Slater *et al.*, 1985) y se indica su identificación dentro del alineamiento. El primer rectángulo muestra el exón 1 y el segundo rectángulo identifica los GRE.

### 5.2.1.8. Comparaciones de las regiones intergénicas.

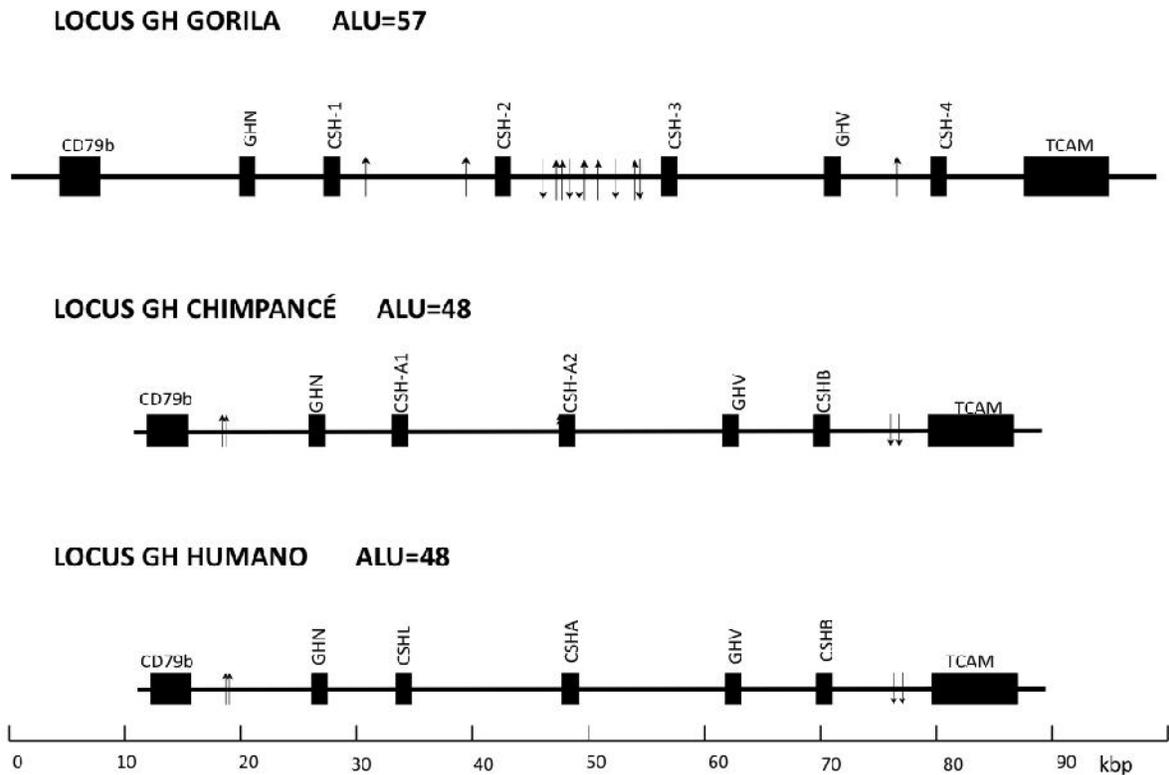
#### 5.2.1.8.1. Elementos repetitivos y eventos indels.

Todas las regiones intergénicas del chimpancé exhiben una similitud superior al 96.9% respecto a su ortólogo en el humano. En la figura 31 se muestra un análisis comparativo de la distribución de elementos repetitivos entre los *loci* GH del humano, del chimpancé y del gorila; no se han incluido en esta representación a los elementos hAT Charlies porque a estos se les dará una atención particular más adelante. Utilizando la herramienta bioinformática RepeatMasker se confirmó la presencia en el *locus* GH del chimpancé de 48 secuencias Alu equivalentes a las reportadas en el humano por Chen y colaboradores en 1989, junto con un número adicional de elementos relacionados a SINE y LINE.



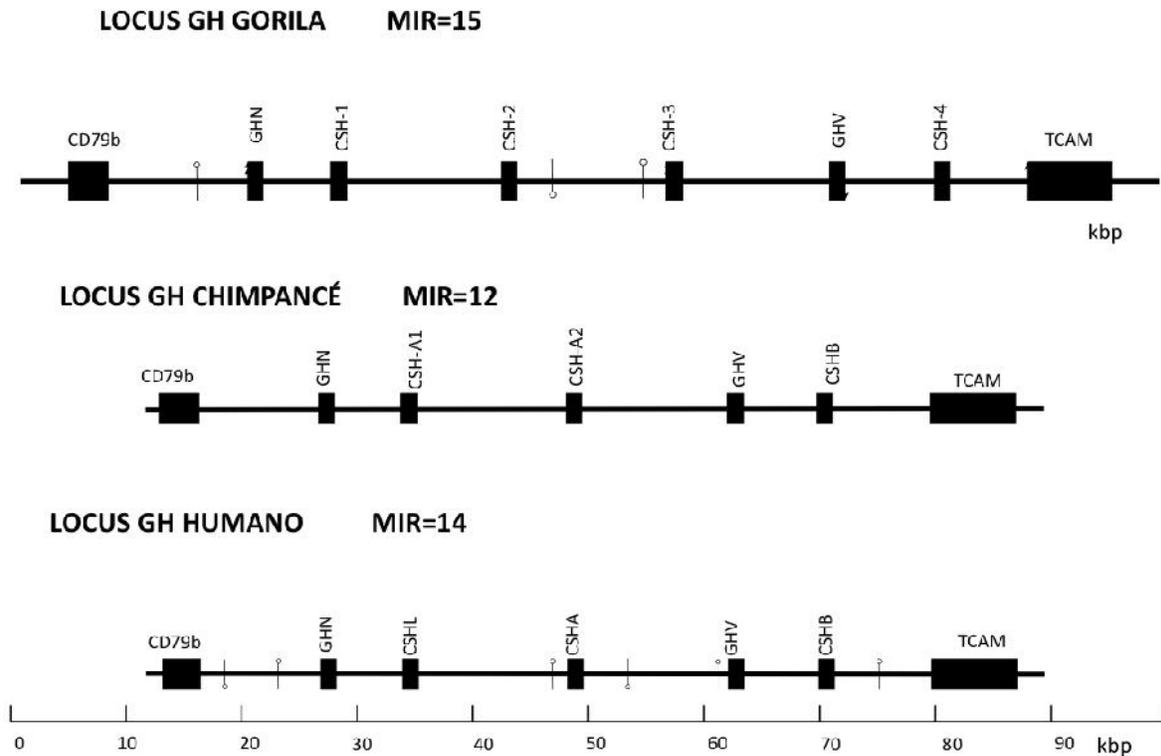
**Figura 31. Análisis comparativo de la distribución de elementos repetitivos.** Las cajas en negro indican los genes que forman los *loci* GH, mientras que las líneas sólidas representan a las RIGs. En la parte inferior se muestra una leyenda para la identificación de los elementos repetitivos.

Para una mejor comprensión, estos elementos han sido divididos para su análisis. Inicialmente se comenzó por analizar de forma comparativa los elementos Alu (figura 32) que es una familia que ha jugado un importante papel en la evolución de los primates. Para el caso del chimpancé, el número de estos elementos (48) coincide con el humano. Para el gorila y teniendo en cuenta que contiene una región intergénica adicional, posee un número mayor de repeticiones Alu (57). En la región 5' del *locus* tanto el chimpancé como el humano comparten dos elementos Alu con orientación sentido, que no están presentes en el gorila, dos elementos Alu deletados en la primera IGL, un elemento Alu deletado en la segunda IGM y dos inserciones en la región posterior.



**Figura 32. Identificación de las diferencias encontradas en la distribución de elementos Alus.** Las cajas en negro indican los genes que forman los *loci* GH, mientras que las líneas sólidas representan a las RIGs.

Adicionalmente fueron analizadas las repeticiones MIRs (fig.33). El *locus* GH del chimpancé cuenta con 12 repeticiones MIRs, a diferencia del humano en el cual hay 14 de estas repeticiones. En el gorila, a pesar de que tiene una intergénica adicional, cuenta con 15 repeticiones de este tipo. Con respecto al *locus* GH del gorila, en la región 5' flanqueante, el *locus* humano tiene una inserción de un elemento MIR y el *locus* del chimpancé tiene deletada una secuencia MIR. Adicionalmente en la primera IGL del *locus* humano hay un inserción de un elemento MIR y otra en la región 3' flanqueante con respecto al *locus* GH del gorila.



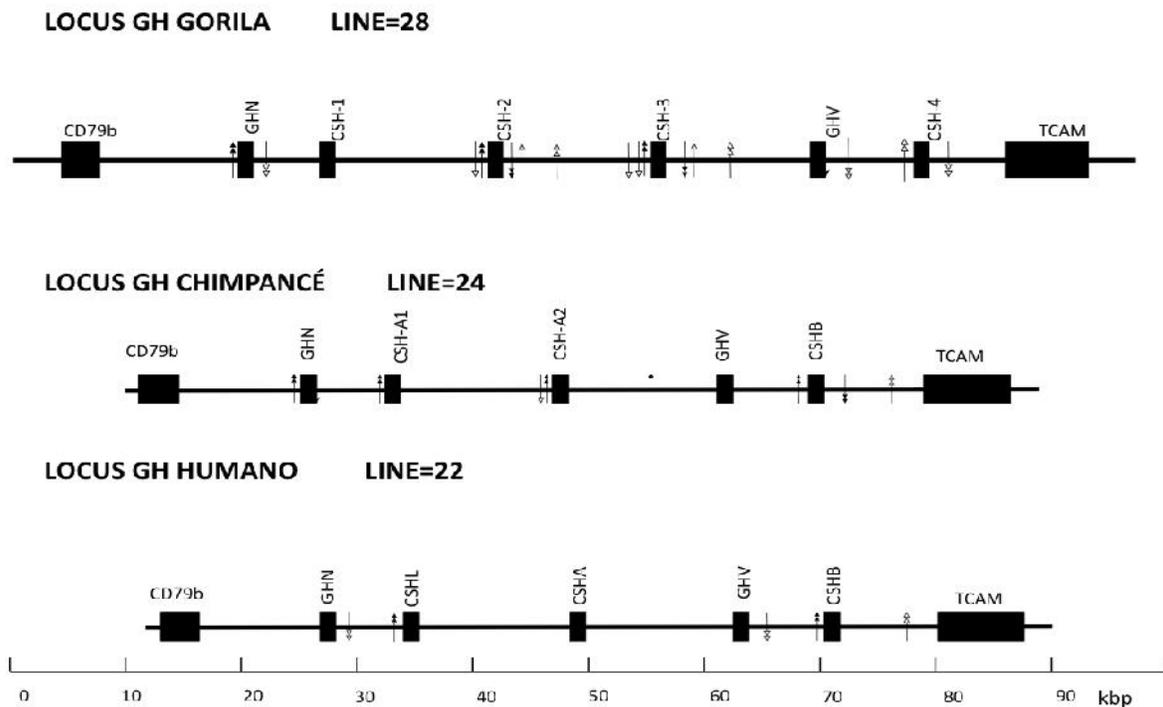
**Figura 33. Identificación de las diferencias encontradas en la distribución de los elementos MIRs.** Las cajas en negro indican los genes que forman los *loci* GH, mientras que las líneas sólidas representan a las RIGs.

Otra familia importante de elementos repetitivos que fue analizada fue la de los elementos LINEs (fig. 34). El número de estos elementos en el humano es 22 (cinco LINE1, nueve LINE2, cinco LINE3 y tres CR1), en el chimpancé es 24 (seis LINE1, siete LINE2, siete LINE3 y cuatro CR1) y en el gorila es 28 (ocho LINE1, once LINE2, seis LINE3 y tres CR1).

Con respecto a la distribución de elementos LINE, se halló una delección de un elemento LINE3 en la región 5'flanqueante en el humano, una inserción de un elemento LINE3 en la primera IGM del humano y una inserción (LINE3) y una delección (LINE2) en la primera IGM del chimpancé. Dos delecciones (LINE1 y LINE3) en la primera IGL del humano. Una inserción de un L-CR1 en el *locus* GH del chimpancé y una delección de tres elementos repetitivos (LINE, LINE2 y LINE3) en humano y chimpancé con respecto al gorila. En la región 3' flaqueante de humano y chimpancé hay una delección de un

elemento LINE2 y la inserción de un segundo elemento LINE2 con respecto al gorila. Adicionalmente en el chimpancé se halla una inserción de un elemento LINE3.

El análisis con RepeatMasker, junto con una inspección visual, mostró que la principal diferencia en el contenido de elementos repetitivos entre los *loci* de humano y chimpancé radica en un elemento relacionado a LINE 1 río arriba de CSH-A2/CSH-A, presente en chimpancé pero no en humano.



**Figura 34. Identificación de las diferencias encontradas en la distribución de los elementos LINES.** Las cajas en negro indican los genes que forman los *loci* GH, mientras que las líneas sólidas representan a las RIGs.

También analizamos los eventos Indels que han tenido lugar en los *loci* GH de estas tres especies. Para dicho análisis, se incluyó las regiones anterior y posterior que flanquean a estos *loci*.

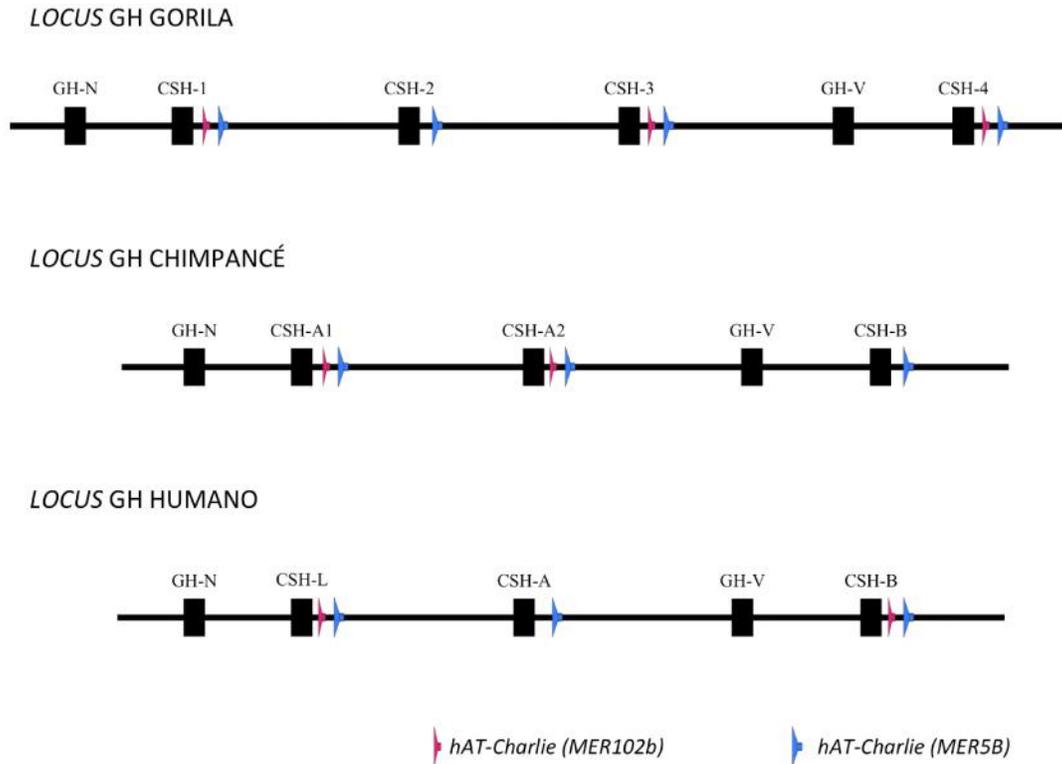
En el caso del gorila, su *locus* GH contiene un gen y una IGL adicional con respecto a los *loci* de humano y chimpancé. Tomando como referencia la primera IGL del gorila se identificaron los eventos Indels en las restantes IGL de gorila y en las de humano y

chimpancé. En todas las IGL se identificó un deleción de aproximadamente 324 pb. En la tercera IGL de gorila, en la segunda de chimpancé y en las dos IGLs de humano se identificó un deleción de 227 pb. La primera IGL de humano y chimpancé coinciden en una inserción de 61 pb. Por otra parte la segunda IGL de chimpancé se distingue del resto por un inserción de 11 pb y una deleción de 134 pb.

Otra diferencia importante entre humano y chimpancé lo constituyen cuatro eventos indels adicionales de más de 15 nt, incluyendo tres extensiones/reducciones de secuencias ricas en AG simples y una secuencia de 112 nt en el intrón 3. Esta última está presente en humano, pero no en chimpancé y no es identificado como un elemento repetitivo por RepeatMasker, a pesar de la presencia de un largo intervalo de poli-T.

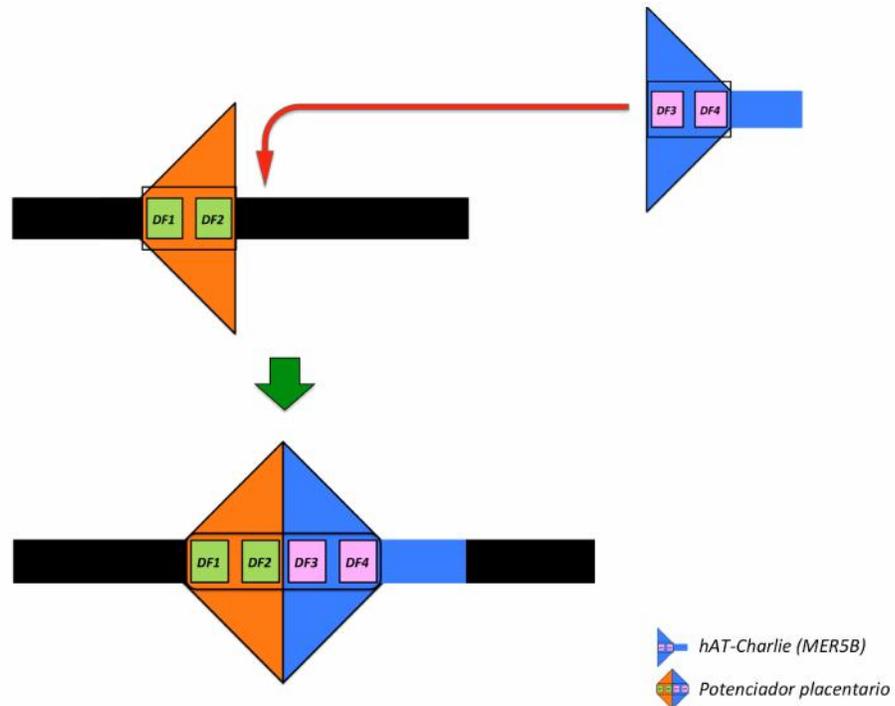
En la región 3' flanqueante de los *loci* GH de humano y chimpancé hay una inserción de 1325pb y 1338 pb, respectivamente, con respecto a la secuencia del *locus* GH del gorila. En la figura 13 se puede apreciar una de estas diferencias de pesos al comparar la migración de los fragmentos que contienen dicha región amplificados sobre los BACs de chimpancé y de gorila.

En último lugar hemos dejado el análisis de los elementos hAT Charlies (fig. 35), porque hemos hecho un hallazgo interesante en estos elementos. Se encuentran cinco elementos hAT-Charlie en el *locus* GH humano y de chimpancé aunque con diferente distribución, a diferencia de siete de estos elementos hallados en gorila. Solo dos tipos de elementos hAT-Charlie fueron identificados en los *loci* GH de los primates (MER102b y MER5B), de los cuales el MER5b parece estar relacionado con la adquisición de la funcionalidad del potenciador.



**Figura 35. Análisis comparativo de la distribución de elementos hAT-Charlie.** Las cajas en negro indican los genes que forman los *loci* GH, las líneas sólidas representan a las RIGs. En la parte inferior se muestra una leyenda para la identificación de los elementos repetitivos.

Hemos descubierto (fig. 36) que los enhancers de los genes CSHs se encuentran interrumpidos en su extremos 3' por elementos MER5b (hAT Charlie). Precisamente este elemento aporta los dominios más importantes descritos como responsables de la actividad potenciadora de estos elementos reguladores (citar).



**Figura 36. Contribución de los elementos hAT-Charlie al origen del potenciador placentario.**

Se llevó a cabo la comparación de la GH hipofisiaria de especies representativas de los diferentes grupos de primates como se muestra en la figura 37. El número de cambios aminoacídicos se corresponde con la distancia filogenética que los separa. Se analizaron los sitios de importancia en la unión hormona-receptor.



Figura 37. Comparación de las GHs hipofisiarias de especies representativas de los diferentes subórdenes de primates. Con flechas azules se indican los sitios claves de unión al receptor y con letras rojas se representan los aminoácidos que difieren con respecto a la GHN humana (GHN-Hs). En la esquina inferior derecha se muestra el número de cambios aminoacídicos de las hormonas hipofisiarias del chimpancé (Pt), mono rhesus (Mm), tití común (Cj) y lemur catta (Lc) con respecto a la del humano (Hs).

## CAPÍTULO VI: DISCUSIÓN

El análisis de las secuencias de un genoma ofrece una amplia visión de su estructura, pero muy limitada de su función y evolución. El principal obstáculo en la interpretación de dichas secuencias había sido la noción de que su mayor parte no parecía ser importante (se estima que solo el 2% del genoma humano codifica para proteínas), pero el Proyecto ENCODE está echando esto por tierra.

La genómica comparativa es el siguiente paso tras determinar la secuencia de un genoma o una región génica específica y constituye uno de los principales métodos para la anotación funcional de los genomas. Las secuencias funcionales son blancos sobre los cuales actúa la evolución, por ello mediante el alineamiento de secuencias es posible hallar huellas dejadas por presiones selectivas durante millones de años (Miller *et al.*, 2004). A través de la comparación y el análisis de las secuencias de genomas de especies cercanas a la nuestra, es posible comenzar a descifrar las claves de la reciente historia evolutiva de nuestro genoma (Alberts, 2006).

La gran mayoría de las sustituciones que se producen en las secuencias de ADN en una población o entre especies son selectivamente neutrales (Kimura, 1968; Khaitovich *et al.*, 2006). Las mutaciones en regiones funcionalmente importantes por lo general dan lugar a alelos deletéreos, haciendo que aquellos individuos que los hereden mueran sin dejar prole (Hardison 2003; Miller *et al.*, 2004). De esta forma la selección negativa o purificadora hace que las secuencias funcionales cambien más lentamente que las secuencias que no lo son (o que no realizan funciones significativas), asegurando la retención de dichas funciones. Otras secuencias en cambio quedan a merced de la selección positiva o Darwiniana, muy común después de los eventos de duplicación génica. Estos eventos son de suma importancia para la generación de nuevos genes, donde una de las dos copias del gen queda libre para mutar y especializarse con el fin de llevar a cabo una función diferente, representando un cambio ventajoso para el portador (Alberts 2006; Khaitovich *et al.*, 2006). La especialización se produce de forma gradual, permitiendo que series repetidas de duplicación y divergencia durante millones de años, conduzcan a que un

gen de origen a una familia de genes dentro de un único genoma (Alberts, 2006). Tal es el caso de la familia multigénica GH, la cual constituye un excelente modelo para estudiar estos y otros aspectos de la evolución de los genomas.

En la evolución humana solamente una pequeña fracción de todos los cambios en la secuencia del ADN entre los seres humanos y los chimpancés es relevante en las diferencias funcionales entre estas especies, haciendo la identificación de cambios no-neutrales una tarea desalentadora (Khaitovich *et al.*, 2006).

### **6.1. Obtención de los genes GH/CSH y sus regiones intergénicas: chimpancé.**

En el presente trabajo, se obtuvo para el chimpancé la secuencia completa de cinco genes relacionados a GH y se encontró que todos ellos están formados por cinco exones separados por cuatro intrones, en posiciones y tamaños similares a los de los genes del *locus* humano. Por más que se intentó, no fue posible amplificar el gen CSH-A reportado previamente (Revol *et al.*, 2004). Coincidentemente, en la información contribuida por el CSGCh tampoco se encontró secuencia alguna que pudiésemos relacionar con dicho gen.

Los seis genes previamente reportados para el chimpancé se obtuvieron a partir de ADN genómico (Revol *et al.*, 2004), lo cual puede que haya hecho más complicado el proceso de identificación del número de genes en el *locus*. Debido a la elevada similitud que muestran entre ellos los genes CSH de chimpancé, discriminar entre un alelo y un miembro no-alélico del *locus* puede convertirse en una tarea sumamente difícil. Los resultados obtenidos previamente contrastan en cuanto al número de genes con los previamente fueron reportadas cuatro intergénicas (Revol *et al.*, 2004), lo que apoya los resultados obtenidos en el presente estudio.

### **6.2. Organización de los loci GH: chimpancé y del gorila.**

La similitud global entre los cinco genes del *locus* GH del chimpancé es aproximadamente del 93 % y entre los seis genes del *locus* GH del gorila es de aproximadamente 94%, lo que refleja la reciente generación de estos a partir de múltiples rondas de duplicación génica (Chen *et al.*, 1989), y algunas secciones muestran una

similitud mucho mayor que esto, en parte debido a procesos de conversión génica. Los genes que flanquean el *locus* son similares en los homínidos estudiados.

Nuestros hallazgos nos llevaron a concluir que el *locus* GH del chimpancé (figura 19), con una organización casi idéntica a la del humano (Chen *et al.*, 1989) está formado por cinco genes, dos de ellos codifican GHs y los tres restantes CSHs. Utilizando la nomenclatura adoptada en la sección de resultados una vez completado el ensamblado (cuadro 3), en el *locus* el primer miembro corresponde a la GH-N, seguido por CSH-A1, CSH-A2, GH-V y CSH-B.

No encontramos evidencias que apoyen el reporte del gen CSH-A, por ninguno de los dos enfoques llevados a cabo y adicionalmente son propuestas correcciones para las demás secuencias de genes previamente reportadas. No puede descartarse la posibilidad de que el sexto gen sea una característica particular de la muestra de DNA genómico usada por Revol de Mendoza y colaboradores.

La organización y composición génica del gorila (figura 18) difiere de la de sus primos, el chimpancé y el humano, por presentar seis genes, el primero de ellos corresponde a la GH-N y los siguientes a gCSH-1, gCSH-2, gCSH-3, GH-V y gCSH-4. El orangután en cambio, está formado por cuatro genes, el primero también corresponde a la GH-N, seguido por CSH1, GH-V y CSH2 (fig. 34).

### 6.3. Análisis de las proteínas.

Como se ilustra en la figura 19, en el humano hay un potencial pseudogen (CSH-L) y los dos genes CSH, codifican la misma proteína (Barrera-Saldaña *et al.*, 1983), sin embargo, en el chimpancé todos los genes se predicen funcionales y de las proteínas codificadas por los genes CSH dos codifican la misma proteína madura y comparten con la tercera una similitud superior al 97.9%. Por otro lado se observa una marcada disminución en los valores de similitud para el pseudogen porque éste ha acumulado cambios que lo alejan de los demás (Gerstein y Zheng, 2006). El gen CSH-L humano codifica potencialmente una proteína que es sustancialmente diferente a las CSHs, debido a una

sustitución en el sitio donador de splicing en el intrón 2. El ARNm de este gen se expresa a un nivel bajo, pero no está claro si esto da lugar a niveles significativos de proteína funcional. En los genes correspondientes de gorila y chimpancé (CSH1 y CSH-A1) el sitio donador del empalme en el intrón 2 está intacto. El gen CSH1 de gorila codifica una proteína que es sustancialmente diferente de los otras CSHs.

Al compararse la secuencia de los productos proteicos potenciales de los genes hCSH-L, cCSHA1 y gCSH1, con la correspondiente a la de su ancestro común aCSH, para entender mejor lo que le sucedió a este gen durante la evolución de los homínidos (fig.20), observamos que tanto el gorila como el chimpancé difieren del ancestro común en cinco residuos aminoacídicos no compartidos entre ellos. Por otro lado, el producto proteico del humano difiere del ancestro en nueve residuos, no compartidos con chimpancé, ni con gorila. Algunos de los cambios identificados se encuentran en los sitios de unión del modelo humano GH:receptor (de Vos *et al.*, 1992).

Las proteínas maduras gCSH1 y gCSH2 difieren en once posiciones (Fig. 19), de las cuales las correspondientes a los residuos, His-12, Met-14 y Ser-184 en CSH1 son reemplazados por Lys, Glu, Val y Phe encontrándose dentro del sitio de unión 1 del modelo humano GH:receptor (de Vos *et al.*, 1992).

En el humano, los genes CSH-A y CSH-B codifican la misma proteína madura. En chimpancé los genes CSH-A1 y CSH-A2 codifican la misma proteína madura (nuevamente con el péptido señal difiriendo en el residuo 24 que en CSH-A1 es Ala y en CSH-A2 es Pro) y la proteína codificada por CSH-B difiere de CSH-A1/A2 en tres residuos, todos en la proteína madura: Gln-69, Met-123 y Gly-126 en CSH-A1/A2 son sustituidos en CSH-B por His, Thr y Arg, respectivamente. En el gorila los cuatro genes CSH codifican proteínas diferentes. Así, el gorila tiene cuatro CSHs diferentes, en contraste con las dos de chimpancé y solo una en el humano, y estas pueden ser diferentes en la actividad biológica. El *locus* GH de orangután tiene dos genes CSHs, uno de ellos contiene una mutación en el segundo intrón que no permite el correcto procesamiento de sus transcritos, sugiriendo que la mayoría de éstos serían incapaces de expresar una proteína secretable.

En primates la rápida evolución de la GH se produjo de forma simultánea a la de su receptor (GHR), lo cual es sustentado por evidencias que establecen una correspondencia entre los residuos aminoacídicos conservados o sustituidos en la hormona con respecto a su receptor. Un claro ejemplo es que en MNM con respecto a no-primates y prosimios, se ha observado una sustitución del residuo His por el residuo Asp en la posición 171 de la GH, la cual parece haber dictado en MVM un cambio de Leu por Arg en la posición 43 del GHR. Estos cambios que tuvieron lugar en aminoácidos que intervienen en la unión hormona-receptor han sido conservados por la evolución en órdenes superiores (Souza *et al.*, 1995; Liu *et al.*, 2001) y se traducen en la afinación de esta interacción molecular. A nivel fisiológico, la evolución de esta interacción se evidencia por la capacidad de la hGH de unirse al GHR de especies filogenéticamente inferiores y no a la inversa (Liu *et al.*, 2001).

El significado potencial de los cambios aminoacídicos de todas las proteínas respecto a la GH-N del humano, puede ser más revelador si se circunscribe a las posiciones o residuos aminoacídicos involucrados en la unión al receptor. El análisis comparativo entre las proteínas ortólogas de humano y chimpancé, reveló que entre ellos comparten la misma proteína para la GH-V y que las GH-N de ambas especies exhiben muy alta similitud en sus secuencias aminoacídicas (99.5%), diferenciadas sólo por una sustitución aminoacídica. Esta última sustitución está a distancia ( $> 5 \text{ \AA}$ ) de los dos sitios de unión al receptor y por lo tanto es poco probable que afecte directamente a la interacción hormona-receptor. Estos resultados sugieren que hay una presión selectiva para conservar las GHs (Miller *et al.*, 2004). De las tres posiciones aminoacídicas en que difiere la proteína GH-V madura de gorila de las de chimpancé y humano (Fig. 19), la sustitución de Asp-21 en gorila, por Tyr en chimpancés y humanos se encuentra dentro del sitio 1 de unión del modelo humano GH: receptor (de Vos *et al.*, 1992).

Como se puede deducir de las comparaciones entre estas, mientras que las GH-N y GH-V del chimpancé muestran uno o ningún cambio respecto a sus contrapartes en el humano (figura 19), los genes CSH de este primate muestran un mayor número de cambios, incluido uno en la posición 164 de las referidas regiones de importancia para la afinidad por

el receptor (Liu *et al.*, 2001). Este cambio presente en los tres genes CSH de chimpancé es una sustitución de un residuo tirosina por histidina que también está presente en todos los genes CSH del mono Rhesus (datos no mostrados) (González *et al.*, 2006). Es posible que este residuo aminoacídico no impida la unión al receptor, aunque quizás sí afecte la afinidad con la que se da (por su contribución a la conformación espacial del sitio de unión al receptor) y en consecuencia las potencias relativas de las actividades somatogénicas y lactogénicas (Duda y Brooks, 2003).

El residuo Thr en la posición 123 es también visto en la GH humana y se localiza dentro del segundo sitio de unión en el modelo GH:receptor en humanos (de Vos *et al.* 1992). Entonces esto podría afectar la interacción con el receptor. De esta manera, el chimpancé potencialmente tiene dos diferentes CSHs en contraste con el humano y estas pueden diferir en su actividad biológica.

Las diferencias entre las proteínas CSH de gorila, chimpancé y CSH humano se indican en la figura 19. Las proteínas maduras hCSH-A/B y cCSH-A1/A2 difieren en los residuos 34, 123, y 164 (para las respectivas posiciones en el humano se localizan Thr, Thr y Tyr, mientras en el chimpancé estas posiciones son ocupadas por Ala, Met e His, respectivamente). El residuo 164 está dentro del primer sitio de unión del modelo humano GH: receptor (de Vos *et al.*, 1992) y el residuo 123 está dentro del segundo sitio de unión, es por ello que estas sustituciones pueden afectar la interacción con el receptor.

Las proteínas maduras hCSH-A / B y gCSH2, difieren en cuatro residuos, de los cuales el residuo 164 (Tyr en humano, His en gorilla, respectivamente) está dentro del sitio de unión 1 del modelo humano GH:receptor (de Vos *et al.* 1992) (véase más arriba), por lo que esta sustitución puede afectar a la interacción con el receptor.

Las proteínas cCSHA1/A2 y gCSH2 difieren en tres residuos, de los cuales el residuo 123 (Met en chimpancé, Thr en gorilla, respectivamente) se encuentra dentro del segundo sitio de unión del modelo humano GH:receptor (de Vos *et al.*, 1992) (véase más arriba), por lo que estas sustitución puede afectar a la interacción con el receptor.

cCSHB y gCSH2 difieren en cuatro residuos, de los cuales el residuo 69 (His en chimpancé, Gln en gorila, respectivamente) está dentro del primer sitio de unión y el residuo 126 (Arg en chimpancé, Gly en gorila, respectivamente) está dentro del segundo sitio de unión del modelo humano GH:receptor (de Vos *et al.*, 1992).

#### **6.4. Análisis de los elementos reguladores.**

Entender la regulación de los genes ha sido y sigue siendo uno de los desafíos principales para la comunidad biomédica. La regulación génica es mediada por una variedad de sitios de unión a factores de la transcripción. Un primer paso en el avance hacia la comprensión de la regulación de los miembros del *locus* GH es la identificación de los elementos reguladores presentes en éste. El segundo paso es la identificación de huellas en el ADN que revelen acontecimientos evolutivos a los que éste fue sometido. La llamada huella digital filogenética se basa en la observación de los elementos reguladores que están bajo presión selectiva, que hace que ellos evolucionen a una tasa más lenta que la secuencia no funcional circundante. Así, comparando regiones reguladoras ortólogas de especies relacionadas, la huella digital filogenética predice las subsecuencias altamente conservadas que podrían funcionar como elementos reguladores potenciales (Prakash y Tompa, 2005).

Muchos elementos reguladores han sido identificados en el *locus* hGH (figuras 4 y 5), incluyendo aquellos asociados con las regiones promotoras inmediatamente río arriba de cada gen. Destacan un elemento de respuesta a glucocorticoides en el intrón 1 del gen GH-N, un potenciador de la expresión placentaria río abajo de hCSH-B, "elementos P" río arriba de cada gen (excepto en GH-N) y una región de control del *locus* más lejos río arriba de la GH-N (Ho *et al.*, 2004). El último de ellos queda fuera del *locus* GH considerado aquí, los otros elementos son considerados en detalle más adelante.

#### **6.5. Análisis del promotor proximal.**

Un número de elementos de respuesta han sido identificados en la región promotora de hGH-N, y se han comparado con las regiones correspondientes de otros genes en los *loci* hGH, cGH y gGH en el alineamiento de la figura 24. La caja TATA se conserva en todos

los genes. El elemento SP-1 (Lemaigre *et al.*, 1989) está completamente conservado, posiblemente reflejando su papel tanto en la expresión hipofisiaria, como en la placentaria.

Los elementos Pit-1 proximal y distal (Krawczack *et al.*, 1999; Theill and Karin 1993) están generalmente bien conservados, a pesar de las observaciones de que Pit-1 juega un importante papel en controlar la expresión de GH-N en la hipófisis, pero no de los genes GH/CSH en placenta (Jiang *et al.*, 1995). Además del efecto sobre la unión a Sp1 (interferencia de la unión de GHF-I/Pit-1 en la secuencia del sitio distal adyacente), la actividad del promotor parece ser independiente de la unión Pit-I/GHF-I en sitio Pit-1 distal. Esto sugiere que la expresión específica de la hipófisis está dirigida por la unión de GHF-I/Pit-1 al sitio Pit-1 proximal (Tansey y Catanzaro, 1991). Para los genes gCSH2, gCSH3 y gCSH4 del gorila, y el gen CSH-A1 del chimpancé, a diferencia de cualquiera de los genes CSH humanos, las secuencias de sus elementos Pit-1 proximal y distal son idénticas a las del gen GH-N de estas tres especies.

Una notable excepción de esta conservación de los elementos Pit-1 se ve en el elemento proximal de los genes GH-V, que muestran cinco sustituciones nucleotídicas, compartidas en gran parte entre gorilas, chimpancés y humanos. Este elemento ha sido objeto de una evolución adaptativa desde el evento de duplicación que dio lugar a la GH-N y GH-V, reflejando su expresión diferencial en la hipófisis y la placenta. Las cinco sustituciones nucleotídicas antes mencionadas afectan la fuerza de los promotores, siendo responsable de que el gen hGHV tenga niveles de expresión más bajos "in vitro" que los genes hCSHs. Esto eleva la posibilidad de que un mecanismo similar puede existir en la placenta (Nickel, BE. *et al.*, 1990).

El elemento de respuesta a hormona tiroidea (TRE; Glass *et al.* 1987) se ha reportado que funciona en la placenta, pero no en la hipófisis (Leidig *et al.*, 1992.) y en consecuencia este elemento muestra una considerable variación de secuencia (fig. 24). Este elemento tiene idéntica variación en el humano y en el gorila, lo cual sugiere que los genes expresados en placenta en gorila muestran el mismo patrón de respuesta a las hormonas tiroideas que el observado en el humano. Para el humano, en comparación con el

chimpancé, esta variación difiere de lo esperado en términos de estricta ortología. De esta manera, el elemento que precede a la CSH-A2 de chimpancé se asemeja más al que precede a la GH-N de chimpancé (o humano), que al que precede a las otras secuencias CSHs. Interesantemente, en esta región la secuencia de los genes CSH-A1 y CSH-B del chimpancé son más similares a CSH-A y CSH-B del humano que al resto de los genes. Las posiciones -49 y -47 identificadas como cruciales para la función de estos elementos (Leidig *et al.*, 1992) se hayan ocupadas por los nucleótidos G y A, respectivamente, como ocurre en los genes CSH humanos y no en la GH-N. Para el caso del gen CSH-B del chimpancé, en esta región su secuencia es similar a la del gen GH-N humano, por lo cual se predice una falta de funcionalidad para dicho elemento.

Por otro lado, este elemento en el caso de los genes GH-V del humano y del chimpancé, no es funcional. Sin embargo, los cambios que presenta se encuentran retenidos en ambas especies y también en el orangután (datos no mostrados), sugiriendo que esta secuencia puede jugar un papel funcional en la regulación de la expresión de dicho gen en placenta (Prakash y Tompa, 2005). Los eventos de conversión génica probablemente han influenciado la variación en esta región, pero una consecuencia puede ser que los genes placentarios expresados en chimpancé muestren un patrón de respuesta a la hormona tiroidea diferente a la observada en el humano.

El sitio de unión del iniciador (InrE), que se requiere para la actividad eficiente del promotor y la máxima actividad del potenciador (Jiang *et al.*, 1995), está en general bien conservada. El elemento de respuesta a AMP cíclico (CRE, Eberhardt *et al.* 1996) está altamente conservado, en contraste con la situación en mamíferos no primates (Wallis *et al.*, 2001).

#### **6.6. El potenciador placentario.**

Se ha reportado que el potenciador placentario se encuentra ~ 2,2 kb río abajo de CSH-B en humano, y aumenta notablemente la expresión de dicho gen en células de placenta, pero no en otros tejidos (Rogers *et al.*, 1986; Walker *et al.*, 1990.). Secuencias muy similares se encuentran río abajo de hCSH-L y hCSH-A, sin embargo estos tienen

muy poca actividad potenciadora. Un alineamiento de la región que incluye el potenciador aparece en la figura 26, mostrando las cuatro regiones identificadas por Jacquemin *et al.* (1994) por footprinting, de las cuales DF3 y DF4 parece ser las más importantes.

La inserción de la secuencia AGAA río arriba del dominio DF-2 (encontrado en hCSH-A y hCSH-L) se ha demostrado que disminuye la funcionalidad del potenciador (Jacquemin *et al.*, 1994), curiosamente esta secuencia no está presente ni en hCSH-B, cCSH-B, gCSH3, ni gCSH4. Esto sugiere que estos genes no han reducido la actividad de sus potenciadores. Los potenciadores de los genes CSHs restantes muestran las mismas inserciones 4-bp río arriba del dominio DF-2 postuladas para actuar como un inhibidor (Jacquemin *et al.*, 1994).

La baja actividad potenciadora de las secuencias hCSH-L y hCSH-A ha sido explicada en dos estudios independientes por unas pocas sustituciones (Lytras *et al.*, 1996; Jacquemin *et al.* 1996), que se indican en la figura 26. Una sustitución detectada en ambos estudios (posición 208 en la figura 26, en el dominio DF-3: A en hCSH-B, G en hCSH-A y hCSH L) se encuentra en todos los genes CSHs de gorila y chimpancé, lo que sugiere que el potenciador de aquí es mucho menos activo que en el gen CSH-B del humano.

Otra sustitución que le confiere la baja actividad potenciadora de los genes hCSHA y hCSHL se indica en la figura 26 (posición 228, en el dominio de DF-3: A en hCSHB, T en hCSHL y hCSHA), encontrándose en todos los genes CSHs de chimpancé y gorila, excepto en cCSHB y gCSH4, sugiriendo que esos potenciadores podrían ser ligeramente más activos que los potenciadores de los genes CSHL y CSHA del humano, CSHA1 y CSHA2 del chimpancé y CSH1, CSH2 y CSH3 del gorila.

Aunque en este momento, la funcionalidad de los potenciadores del gorila y el chimpancé no han sido probada, nuestros hallazgos sugieren que existen mecanismos de control similares para la transcripción diferencial de los miembros del *locus* GH y la conservación de la especificidad tisular.

### 6.7. El elemento P.

Los elementos P se encuentran río arriba de la GH-V y de los tres genes de CSH en el humano y se cree que inhiben la expresión de estos genes en la hipófisis (Norquay *et al.* 2006) y/o activan su expresión en la placenta (Elefant *et al.*, 2000; Ho *et al.*, 2004). Secuencias equivalentes son encontradas en chimpancé y gorila, y bien pueden desempeñar la misma función. En el genoma humano el elemento P está situado en las posiciones -2,439; -2,386; -2,287 y -2,690 de los genes CSH-L, CSH-A, GH-V y CSH-B, respectivamente. En el chimpancé está en las posiciones -2,258; -2,249; -2,222 y -2,515 de los genes CSH-A1, CSH-A2, GH-V y CSH-B, respectivamente. El elemento P se identificó en todos los RI de los gorilas (en la posición -2282, -2533, -2516, -2186 y -2556 delante de CSH-1, CSH-2, CSH-3, V-GH-CSH y 4 genes, respectivamente).

En la figura. 25 se muestra un alineamiento de las secuencias correspondientes al elemento P (Norquay *et al.* 2003, 2006). Los dominios PSE-A y PSE-B están altamente conservados. El dominio PSE-C muestra un número de sustituciones en gran parte compartidas entre humanos, chimpancés y gorilas, pero los sitios putativos de unión para HNF-3 y C/EBP ubicados en este dominio se encuentran altamente conservados. En el alineamiento de las secuencias correspondientes al elemento P, el mayor número de diferencias se observaron en el dominio PSE-C.

Cabe señalar, sin embargo, que aunque no hay elemento-P río arriba del gen GH-N en humano, chimpancé y gorila, tales secuencias se encuentran en esta posición en el tití y en el perro, haciendo que esta simple interpretación en términos de supresión pituitaria sea menos sostenible (Wallis y Wallis 2006). En estas especies se observan un gran número de cambios en las secuencias del elemento P en comparación con los grandes simios, esto sugiere que: a) no comparten la misma función, b) tengan diferencias en las proteínas de unión a estos sitios o c) su localización tejido específico dicta diferencias en la función entre los grandes simios y los monos del Nuevo Mundo y los perros.

### 6.8. Elemento de respuesta a glucocorticoides.

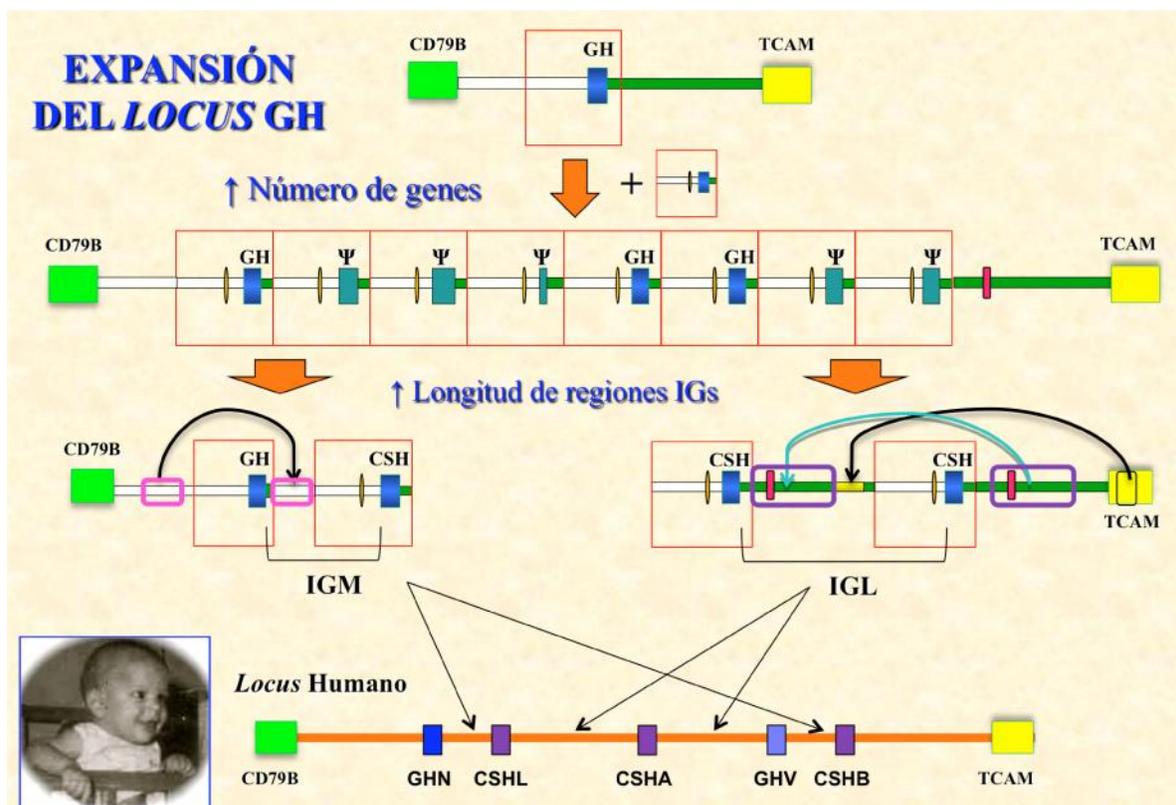
Se ha descrito ampliamente la capacidad reguladora transcripcional que tienen los glucocorticoides a través de los elementos de respuesta a los mismos (GRE). El GRE ha sido reportado en el primer intrón de la hGH-N y de la hGH-V (Slater *et al.* 1985). En la figura 27 se muestra un alineamiento de las secuencias nucleotídicas de los genes GH/CSH de humano, chimpancé y humano. El GRE fue identificado en este alineamiento tomando como referencia el reporte previo de su secuencia para la hGH-N. Estos elementos se encuentran completamente conservados en los correspondientes genes (cGH-N y cGH-V) del chimpancé, de igual forma se conserva en el gen GH-V de gorila, pero está ausente en la GH-N de esta última especie.

### 6.9. Elementos repetitivos y eventos indels.

Los elementos Alu han contribuido claramente a la evolución del genoma humano y en los primates han mostrado una importante actividad en los eventos evolutivos (Kehrer-Sawatzki y Cooper, 2007).

Chen *et al.* 1989 identificaron 48 elementos repetitivos Alu (SINE) en el *locus* GH humano, algunos de estos truncados. Como se muestra en la figura 28, el uso de la herramienta bioinformática RepeatMasker reveló diferencias en el contenido de elementos repetitivos entre los *loci* GH de humano, chimpancé y gorila y confirmó que 48 secuencias Alu equivalentes están también presentes en el genoma del chimpancé y 57 en el *locus* GH del gorila, junto con un número adicional de elementos SINE y LINE relacionados. Este análisis, junto con una inspección visual, mostró que la única diferencia clara en el contenido de elementos repetitivos entre los *loci* de humano y chimpancé, es un elemento relacionado a LINE 1 (224 nt correspondiente al extremo 3' de LINE 1) río arriba de los genes CSH-A2/CSH-A, presente en chimpancé pero no en humano. Una muy similar secuencia es encontrada río arriba de los genes CSH-B/CSH-B en ambos chimpancé y humano, sugiriendo que la diferencia se originó por delección de dicho elemento en el humano.

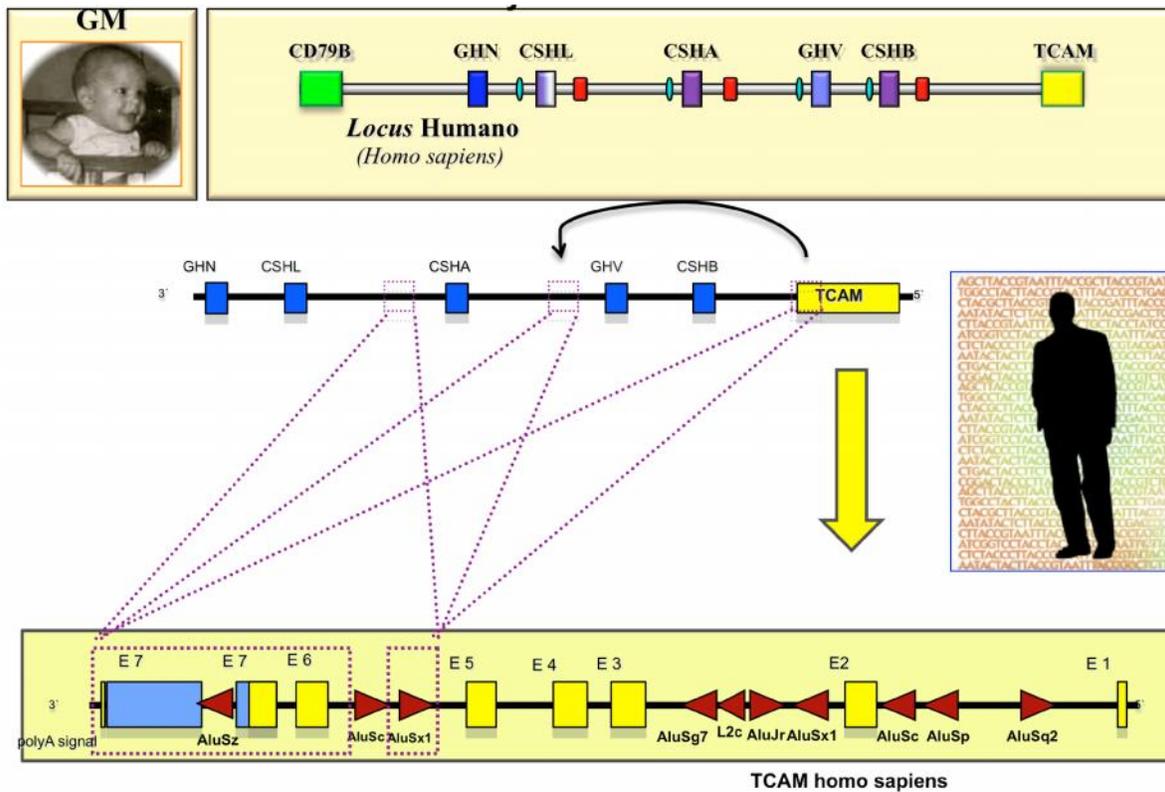
Este elemento LINE 1 río arriba de los genes CSH-A2/CSH-A, representa el evento de inserción/delección (indel) de mayor tamaño observado en una comparación de los *loci* de humano y chimpancé. Cuatro eventos indels adicionales de más de 15 nt fueron identificados, incluyendo tres extensiones/reducciones de secuencias ricas en AG simples y una secuencia de 112 nt en el intrón 3. Esta última está presente en humano, pero no en chimpancé y no es identificada como un elemento repetitivo por RepeatMasker, a pesar de la presencia de un a largo intervalo de poli-T. Los *loci* GH de humano y de chimpancé en sus regiones 3' tiene un inserto de ~ 1277 nucleótidos con respecto a la secuencia del gorila.



**Figura. 38. Evolución del locus GH.** La expansión del locus GH tuvo lugar por un incremento en el número de genes y un aumento en el tamaño de las regiones intergénicas.

Los elementos Alu no solo han jugado un importante papel en la generación del locus GH (Chen *et al.*, 1989), sino también, al menos en el caso del elemento Alu Sx, situado justamente detrás de los genes hGH-N y hGH-V, en la regulación transcripcional de estos genes. En el humano ha sido demostrado (Trujillo *et al.*, 2006) que este elemento Alu Sx contiene un silenciador compuesto por cuatro dominios de interacción que incluyen un dominio represor (CRD), un dominio antisilenciador (ASE) que contiene elementos que

median la actividad silenciadora independiente de la orientación y dos dominios flanqueantes que modulan la actividad silenciadora de manera dependiente del CRD. El silenciador es dependiente del promotor GH y su función es específica para las células que expresen hGH-N o hGH-V (Trujillo *et al.*, 2006). Resulta interesante que el silenciador contenido en la secuencia Alu Sx que ha sido reportado para el humano, también esté presente detrás de los genes GH-N y GH-V del chimpancé, sugiriendo una conservación de este mecanismo de regulación transcripcional entre ambas especies; aunque es muy posible que esa secuencia reguladora sea una adquisición evolutiva relativamente reciente (Trujillo *et al.*, 2006). El gorila y el chimpancé, también presenta la secuencia reportada en el humano, correspondiente a un silenciador que está integrado en una secuencia Alu Sx detrás de los genes GHN y GHV.



**Figura. 39.** Detalle de la evolución del locus GH. Locus GH y gen TCAM en el humano.

El gen GHV de gorila en su primer intrón tiene una repetición simple de  $(GA)_{38}$ .

### 6.10. Visión general (Reconstruyendo la historia evolutiva del *locus* GH humano).

Con este trabajo se ha completado en algunos casos las secuencias faltantes del *locus* GH y genes flanqueantes y en otros se ha agregado completamente dichas regiones. Se completaron el tití común, el mono rhesus y el chimpancé y se reportan los ensamblajes de dichas regiones para el lemur, el tití de vientre rojo, el mono vervet, el orangután y el gorila.

La historia comienza en los prosimios, en los cuales la GH es codificada por un sólo gen, cuya secuencia es muy similar a la ancestral encontrada en los mamíferos. Ya en los antropoides, los *loci* GH están integrados por múltiples genes no alélicos relacionados [23-25]. En general, las GHs en este grupo de primates están bastante conservadas.

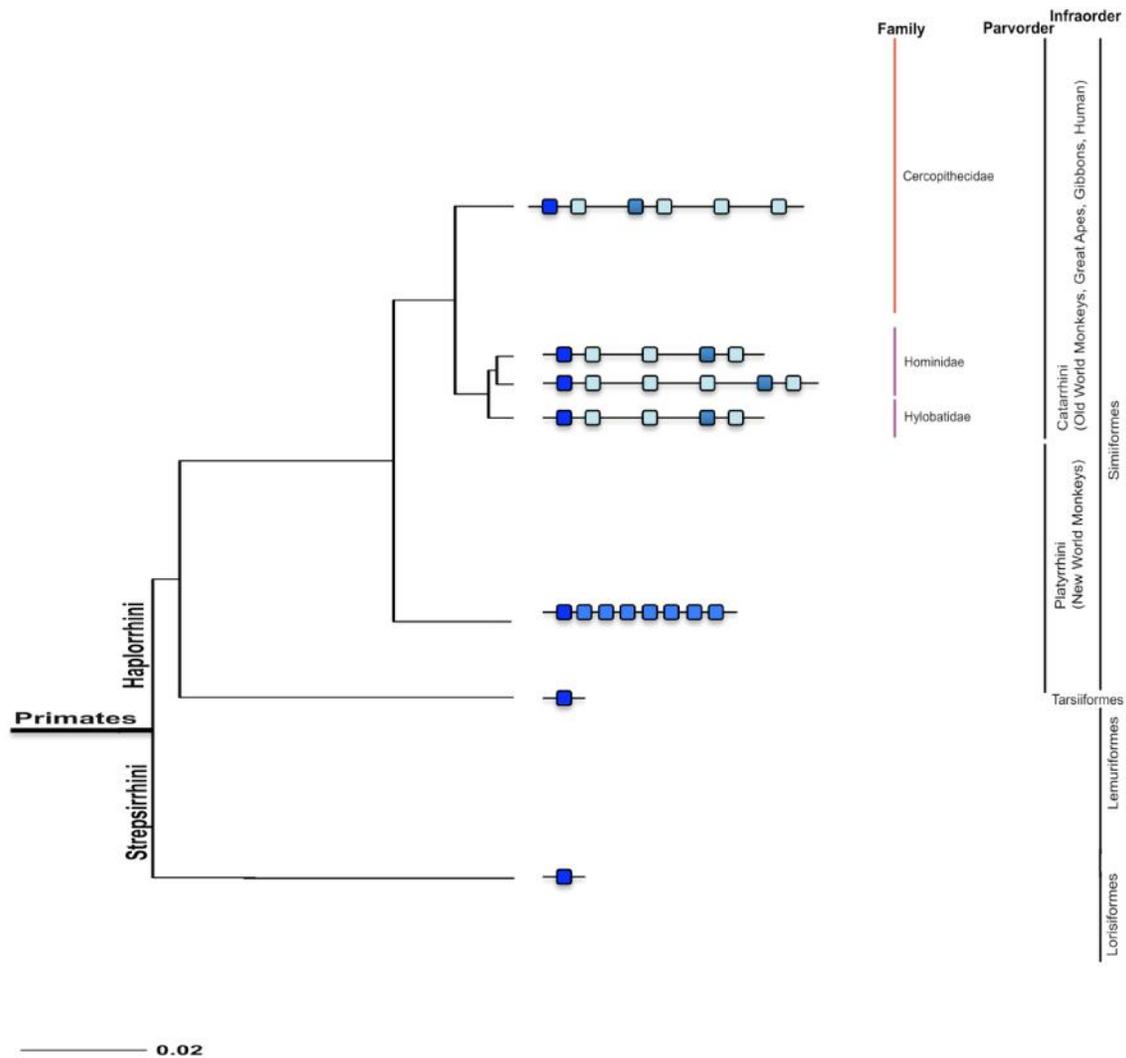
La inspección de los mismos revela que el gen situado en el extremo 5' de estos *loci* corresponde al de la GH hipofisiaria, referido comúnmente como GH-N o GH-1 (de expresión en hipófisis), y que los *loci* GH variaron en los diferentes grupos de primates experimentaron diferentes rutas evolutivas (no todos los genes de los *loci* resultan ortólogos). La aportación de la secuencia de varios estrepsirinos, proporcionará profundidad a las comparaciones genómicas y permitirá un entendimiento más completo de la evolución de los genomas en primates (Horvath, *et al.*, 2007).

La evolución del *locus* GH en primates (últimos 60 millones de años) se distingue por eventos dramáticos de multiplicación y divergencia génicas. A la vez, por el surgimiento de una diferenciación en la especificidad de la expresión de sus miembros en al menos dos tejidos: hipófisis y placenta. En cuanto al surgimiento de la expresión diferencial de los genes miembros del *locus*, la expresión en la placenta de los genes relacionados a GH se ha documentado en antropoides (Chen *et al.*, 1989; Golos *et al.*, 1993; Papper *et al.*, 2009; Rodríguez-Sánchez *et al.*, 2010). Sin embargo, hasta el momento no han aparecido reportes de genes GH expresados en placenta en prosimios. A la fecha no está del todo clara la funcionalidad del elemento P en los MNM; por lo que resulta imperativo continuar analizando a detalle el origen y función del elemento P y del potenciador en los primates,

para consecuentemente determinar cuándo surgió la regulación de la especificidad tisular de los genes relacionados a GH.

Finalmente, se sabe desde hace décadas que la GH en el humano afecta numerosos órganos y sistemas, actuando sobre diversos procesos metabólicos y homeostáticos. Pero más recientemente se ha acumulado evidencia de su participación en el desarrollo cerebral, la neurogénesis, la neuroprotección, el aprendizaje y la memoria. Por su lado, los niveles maternos de CSH se han asociado con el crecimiento del feto y la producción de IGF-I durante el embarazo (Sonntag, 2005). Dado el impacto de la GH en la salud, aun desde la embriogénesis y desarrollo fetal, no dudamos en que como fruto de este proyecto en el mediano plazo se desprendan muy posiblemente aplicaciones en la clínica y la biotecnología.

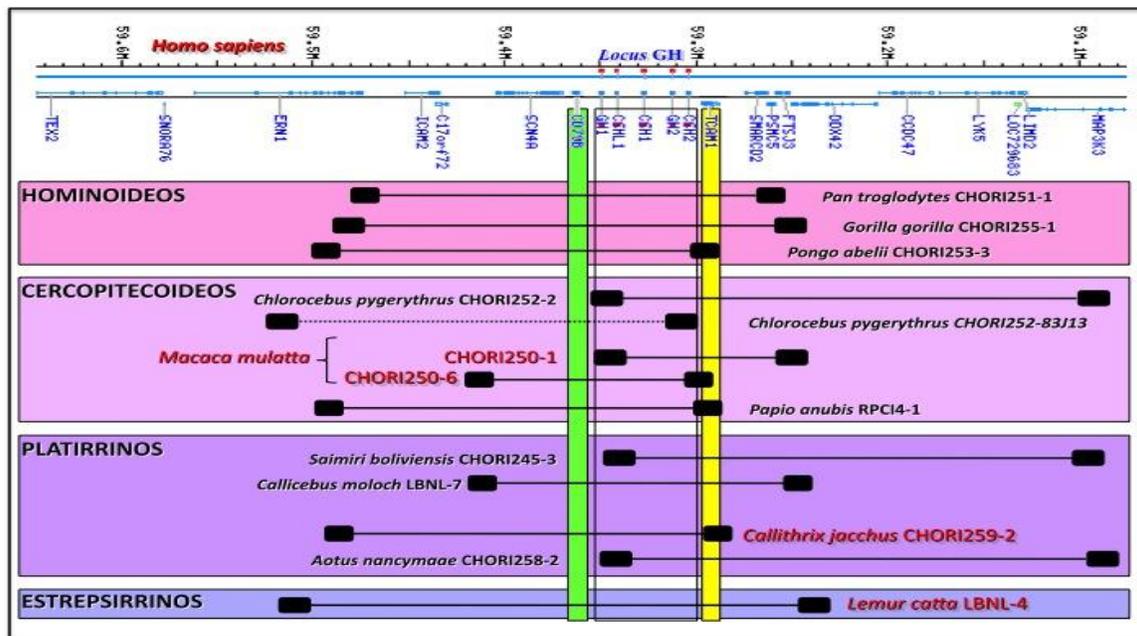
Se ha observado en el caso MVM y GM que secuencias similares a la región final del gen TCAM están presentes en las intergénicas largas del *locus* GH. Dichas secuencias probablemente fueron incorporadas durante los eventos de duplicación que han tenido lugar en este *locus*. El estudio de los *loci* GH en primates sugiere que los elementos repetitivos han desempeñado un importante papel en dichos eventos. El origen del potenciador que está presente en el *locus* GH es otra de las interrogantes en esta historia evolutiva. Nuestro análisis bioinformático ha permitido identificar una región de similitud (después del último gen del *locus* GH del marmoset) al potenciador identificado en MVM y GM, cuya función en estos se ha demostrado que es la de potenciar la expresión de los genes CSHs en placenta. De igual forma identificamos una región similar en el lemur (un prosimio), cuyo *locus* GH se caracteriza por presentar un único gen, frente al cual identificamos además una región similar al elemento P presente en MNM, MVM y GM y al cual en el caso de MVM y GM se le adjudica la capacidad de reprimir la expresión de los genes placentarios en hipófisis.



**Figura 40. Representación de las secuencias genómicas de los *loci* GHs.**

En nuestros esfuerzos por entender la evolución del *locus* GH, ahora colaboramos con el Dr. Michael Wallis, experto evolucionista de la GH, para confirmar la composición y organización génica de este *locus* en especies representativas de los diferentes grupos de primates, en un proceso minucioso de anotación de los *loci*, con el fin de descifrar la ruta evolutiva en primates y los mecanismos a través de los cuales se adquirieron los nuevos genes, estos divergieron para dar origen a genes tipo CSH y GH-V y se adquirió la especificidad tisular de expresión en primates. Nuestro material de partida clave, lo

constituyen las secuencias obtenidas de una colección de BACs portadores de los *loci* GH de once especies de primates (figura 40). Nuestros hallazgos iniciales confirman un sólo gen en lémures y más de una media docena (incluidos algunos pseudogenes) en platirrinos. A la vez revelan seis genes GH/CSHs en cercopitecoideos y entre cuatro y cinco genes GH/CSHs en hominoideos. La inspección de los mismos, muestra que inician con el gen GH-N y que su composición varía en los diferentes grupos de primates al experimentar diferentes rutas evolutivas (no todos los genes de los *loci* resultan ortólogos).



**Figura 41. Representación de las secuencias genómicas contenidas en nuestra colección de BACs.** Mapeo de los BACs de once especies de primates que forman parte de nuestra colección donada amablemente por el Dr Pieter de Jong del Children's Hospital Oakland Research Institute, de Oakland, EUA.

Las secuencias de múltiples genes relacionados a GH han sido descritas y analizadas en primates, pero hasta la fecha, la secuencia del *locus* GH sólo se ha resuelto por completo en nuestros laboratorios para el caso del humano, el tití y el chimpancé, pero pronto lo será para otras especies claves en este enigmático rompecabezas de la evolución de esta fascinante región del genoma humano.

### **6.11. El valor de la genómica comparativa.**

Recientes análisis de genómica comparativa indican que las marcadas diferencias entre los humanos modernos y los chimpancés, se deben más a cambios en la regulación génica que a las diferencias en el contenido génico. Casi semanalmente se publican noticias de genes y elementos reguladores identificados que parecen experimentar una fuerte selección positiva a lo largo del linaje humano. Los *loci* involucrados han sido de particular interés (Bradley, 2008). Investigaciones tales como el proyecto del genoma del hombre de Neanderthal prometen que los análisis moleculares continuaran proporcionando nuevas visiones acerca de nuestra evolución (Green *et al.*, 2006).

La secuencia del genoma de un individuo necesita solamente ser determinada una vez, sin embargo los niveles de expresión cambian en las diferentes etapas de desarrollo, en los diversos tejidos, y diversos ambientes (Khaitovich *et al.*, 2006). El presente trabajo señala cambios sutiles en los genes o incluso la ausencia de estos entre ambos *loci*. Aunque resultará importante evaluarlos, también será menester complementarlo con estudios de expresión de las mismas, ya sea en cultivo celular o de ser posible, y mejor aún, en tejidos como la placenta.

### **6.12. Perspectivas.**

Extender estudios similares a otras especies de primates de diferentes subórdenes, seguro arrojará luz sobre la evolución de esta familia génica, así como también del origen de su especificidad tisular y temporal.

Las clonas generadas en este trabajo con los plásmidos portadores de los genes GH/CSH de chimpancé con los sitios de restricción (KpnI en el extremo 5' y EcoRV en el 3') incorporados para permitir su posterior subclonación unidireccional en el vector de expresión pcDNA 3.1 hygro (+), pueden ser utilizadas en futuras investigaciones para la determinación del potencial codificante y funcional de dichos genes.

## CAPÍTULO VII: CONCLUSIONES

- 1.- Se ensamblaron los *loci* GH y regiones flanqueantes de especies representativas de primates.
- 2.- Se realizó la anotación estructural de los *loci* ensamblados y se analizaron sus particularidades.
- 3.- Todos los genes de los *loci* GH de los grandes simios se predicen funcionales.
- 4.- Se identificaron en el lemur, secuencias similares al inhibidor hipofisiario y al potenciador placentario.
- 5.- Se identificó en el tití común, una secuencia similar al potenciador placentario.
- 6.- Se identificó un evento de inversión en la IGL3 del *locus* GH del mono rhesus.
- 7.- El gorila produce cuatro CSHs, con propiedades biológicas potencialmente diferentes, mientras que en el chimpancé se producen dos y en el humano solo una.
- 8.- El promotor y otros elementos clave de la regulación génica, están conservados en los *loci* GH de los grandes simios.
- 9.- El gen hCSH-L ha cambiado sustancialmente desde la divergencia gorila / chimpancé / humano, hasta convertirse potencialmente en un pseudogen.
- 10.- La diferencia más importante observada entre los tres homínidos analizados, no lo constituye precisamente el número de genes que conforman al *locus* GH, sino la capacidad de codificación que estos tienen.

## CAPÍTULO VIII: BIBLIOGRAFÍA

- Abascal, F., 2003. Análisis de genomas. Métodos para la predicción y anotación de la función de las proteínas. Tesis doctoral, Universidad Autónoma de Madrid.
- Alsat, E., J. Guibourdenche, *et al.*, 1997. Human placental growth hormone. *Am J Obstet Gynecol* 177(6): 1526-34.
- Altschul, S. F., W. Gish, *et al.*, 1990. Basic local alignment search tool. *J Mol Biol* 215(3): 403-10.
- Barrera Saldaña HA, Seeburg PH, Saunders GF. Two structurally different genes produce the same secreted human placental lactogen hormone. *J Biol Chem.* 1983 Mar 25:258(6): 3787-93.
- Barrera-Saldana, H.A., D.L. Robberson, and G.F. Saunders, Transcriptional products of the human placental lactogen gene. *J Biol Chem*, 1982. 257(20): p. 12399-404.
- Barrera-Saldana, H.A., Growth hormone and placental lactogen: biology, medicine and biotechnology. *Gene*, 1998. 211(1): p. 11-8.
- Barrera-Saldaña, H.A., Seeburg, P.H. y Saunders, G.F., 1983. Two structurally different genes produce the same secreted human placental lactogen hormone. *J Biol Chem* 258 (6), 3787-3793.
- Barrera-Saldaña, HA; Harper ME y GF Saunders. 1982. Chromosome localization of the human placental lactogene-growth hormone gene cluster to 17q22-24. *Am. J. Human . Genet.* 34, 227-234.
- Baudet, B. Martin, Z. Hassanali, E. Parker, E. J. Sanders, and S. Harvey. Expression, Translation, and Localization of a Novel, Small Growth Hormone Variant. *Endocrinology* 148(1): 103–115.
- Baumann GP. Growth hormone isoforms. *Growth Horm IGF Res.* 2009 Aug 19(4): 333-40. Epub 2009 May 24.
- Bernstein, B.E., *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature*, 2012. 489(7414): p. 57-74.
- Bhagavan, N.V. and C.-E. Ha, Chapter 29 - Endocrine Metabolism II: Hypothalamus and Pituitary, in *Essentials of Medical Biochemistry* 2011, Academic Press: San Diego. p. 397-407.
- Birney, E., *et al.*, 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146): 799-816.

- Bohne, A., *et al.*, Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res*, 2008. 16(1): p. 203-15.
- Burbano, H.A., *et al.*, Targeted investigation of the Neandertal genome by array-based sequence capture. *Science*, 2010. 328(5979): p. 723-5.
- Burge, C. B. y Karlin, S., 1998. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* 8, 346-354.
- Burge, C. B., 1998. Modeling dependencies in pre-mRNA splicing signals. In Salzberg, S., Searls, D. and Kasif, S., eds. *Computational Methods in Molecular Biology*, Elsevier Science, Amsterdam, pp. 127-163.
- Burge, C. y Karlin, S., 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78-94.
- Cajiao, I., A. Zhang, *et al.*, 2004. Bystander gene activation by a locus control region. *EMBO J* 23(19): 3854-63.
- Carroll, S. B., 2003. Genetics and the making of Homo sapiens. *Nature* 422(6934): 849-57.
- Carroll, S. B., 2005. Evolution at two levels: on genes and form. *PLoS Biol* 3(7): e245.
- Caufriez, A., *et al.*, Placental growth hormone as a potential regulator of maternal IGF-I during human pregnancy. *Am J Physiol*, 1990. 258(6 Pt 1): p. E1014-9.
- Caufriez, A., F. Frankenne, *et al.*, 1990. Placental growth hormone as a potential regulator of maternal IGF-I during human pregnancy. *Am J Physiol* 258(6 Pt 1): E1014-9.
- Cesar L. Boguszewski, Per-Arne Svensson, Thomas Jansson, Ross Clark, Lena M. S. Carlsson and Björn Carlsson. Cloning of Two Novel Growth Hormone Transcripts Expressed in Human Placenta. *J. Clin. Endocrinol. Metab.* 1998 83: 2878-2885.
- Chellakooty, M., *et al.*, A longitudinal study of intrauterine growth and the placental growth hormone (GH)-insulin-like growth factor I axis in maternal circulation: association between placental GH and fetal growth. *J Clin Endocrinol Metab*, 2004. 89(1): p. 384-91.
- Chellakooty, M., Vangsgaard, K., Larsen, T., Scheike, T., Falck-Larsen, J., Legarth, J., Andersson, A.M., Main, K.M., Skakkebaek, N.E., Juul, A., 2004. A longitudinal study of intrauterine growth and the placental growth hormone (GH)-insulin-like growth factor I axis in maternal circulation: association between placental GH and fetal growth. *J Clin Endocrinol Metab* 89, 384-391.
- Chen, E.Y., *et al.*, 1989. The human growth hormone locus: Nucleotide sequence, biology, and evolution. *Genomics*, 4(4): p. 479-497.

- Choudhuri, S., 2003. The Path from Nuclein to Human Genome: A Brief History of DNA with a Note on Human Genome Sequencing and Its Impact on Future Research in Biology. *Bulletin of Science Technology Society* 23, 360-367.
- Consortio del Proyecto ENCODE, 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306(5696): 636-40.
- Cooke N.E. Ray J. Emery, JG, Liebhaber A., 1988. Two distinct species of human growth hormone-variant mRNA in the human placental the expression of novel growth hormone proteins . *J. Biol. Chem* 263:9001-9006.
- Cooke, N.E., et al., Human growth hormone gene and the highly homologous growth hormone variant gene display different splicing patterns. *J Clin Invest*, 1988. 82(1): p. 270-5.
- Corine M. Lecomte, André Renard and Joseph A. Martial. A new natural hGH variant—17.5 kd—produced by alternative splicing. An additional consensus sequence, which might play a role in branchpoint selection. *Nucl. Acids Res.* (1987) 15 (16): 6331-6348.
- Courtois, S. J., D. A. Lafontaine y G.G. Rousseau, 1992. Characterization of an alternative promoter in the human growth hormone gene. *J Biol Chem* 267(27): 19736-43.
- Davidson, Eric H., Jonathan P. Rast, Paola Oliveri, Andrew Ransick, Cristina Calestani, Chiou-Hwa Yuh, Takuya Minokawa, Gabriele Amore, Veronica Hinman, César Arenas-Mena, Ochan Otim, C. Titus Brown, Carolina B. Livi, Pei Yun Lee, Roger Revilla, Alistair G. Rust, Zheng jun Pan, Maria J. Schilstra, Peter J. C. Clarke, Maria I. Arnone, Lee Rowen, R. Andrew Cameron, David R. McClay, Leroy Hood, y Hamid Bolouri, 2002. A Genomic Regulatory Network for Development. *Science*, 295: 1669.
- DeNoto F. M., Moore D. D., and Goodman H.M., (1981) Human Growth Hormone DNA sequence and mRNA structure: Possible alternative splicing. *Nucleic Acid Res.* 9, 3719- 3728.
- Diamanti-Kandarakis, E., et al., Hormones in sports: growth hormone abuse. *Hormones (Athens)*, 2004. 3(1): p. 37-45.
- Duda, K. M. y C. L. Brooks, 2003. Differential effects of zinc on functionally distinct human growth hormone mutations. *Protein Eng* 16(7): 531-4.
- Elefant, F., Y. Su, *et al.*, 2000. Patterns of histone acetylation suggest dual pathways for gene activation by a bifunctional locus control region. *EMBO J* 19(24): 6814-22.
- Enard, Wolfgang; Svante Pääbo, Khaitovich, Philipp, Bjoern Muetzel, Xinwei She, Michael Lachmann, Ines Hellmann, Janko Dietzsch, Stephan Steigele, Hong-Hai Do, Gunter Weiss, Florian Heissig, Thomas Arendt, Kay Nieselt-Struwe, Evan E. Eichle, 2004. Regional Patterns of Gene Expression in Human and Chimpanzee Brains. *Genome Res.*, 14: 1462 - 1473.

- Evolución de los primates. Modificado de Great Ape Trust. Consulta Mayo 28 de 2008.  
<http://www.greatapetrust.org/primates/index.php>.
- Facts About Genome Sequencing. Octubre 2006. Consulta 22 de Mayo de 2008.  
[http://www.ornl.gov/sci/techresources/Human\\_Genome/faq/seqfacts.shtml#complete](http://www.ornl.gov/sci/techresources/Human_Genome/faq/seqfacts.shtml#complete).
- Farreras-Rozman, 2000. Medicina Interna. Capitulo 16. Endocrinología. Decimocuarta Edición. Ediciones Harcourt. ISBN Obra Completa: 84-8174-357-7. Publicación Harcourt: [www.harcourt.es](http://www.harcourt.es).
- Felber, J. P., N. Zaragoza, *et al.*, 1972. The double effect of human chorionic somatomammotropin (HCS) and pregnancy on lipogenesis and on lipolysis in the isolated rat epididymal fat pad and fat pad cells. *Horm Metab Res* 4(4): 293-6.
- Fleischmann RD, MD Adams, O White, RA Clayton, EF Kirkness, AR Kerlavage, CJ Bult, JF Tomb, BA Dougherty, JM Merrick. , 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496 – 512.
- Frankenne F, F. Rentier-Delrue, ML, Scippo, J. Martical, G. Hennen 1987. Expression of the growth hormone variant gene in human placenta. *J. Clin. Endocrin. Metab.* 64:635-637.
- Frankenne, F., *et al.*, Discovery of a placental variant of human growth hormone: biochemistry, physiology and implication in the secretion of hypophyseal forms. *Reprod Nutr Dev*, 1987. 27(2B): p. 523-4.
- Frohman, L.A., Growth Hormone, in *Encyclopedia of Neuroscience*, R.S. Editor-in-Chief: Larry, Editor 2009, Academic Press: Oxford. p. 993-998.
- G Untergasser. M Hermann<sup>1</sup>, H Rumpold and P Berger. Complex alternative splicing of the GH-V gene in the human Testis. *European Journal of Endocrinology* (1998) 139 424–427.
- Gent, J., *et al.*, Ligand-independent growth hormone receptor dimerization occurs in the endoplasmic reticulum and is required for ubiquitin system-dependent endocytosis. *Proc Natl Acad Sci U S A*, 2002. 99(15): p. 9858-63.
- George DL, JA Phillips III, U. Francke, PH Seeburg 1981. The genes from growth hormone and chorionic somatomammotropin are on the long arm of the chromosome 17 in region q21-q22. *Human. Genet.* 57:138-141.
- Gerlo, S., *et al.*, Prolactin in man: a tale of two promoters. *Bioessays*, 2006. 28(10): p. 1051-5.
- Gerstein, M. y Zheng, D., 2006. The real life of pseudogenes. *Sci Am* 295(2): 48-55.

- Gibbs, R.A. and J. Rogers, Genomics: Gorilla gorilla gorilla. *Nature*, 2012. 483(7388): p. 164-5.
- Giustina A. and Veldhuis JD. Pathophysiology of the neuroregulation of growth hormone secretion in experimental animals and the human. *Endocr Rev.* 1998 Dec; 19(6): 717-797.
- Glazko, Galina V., Masatoshi Nei Estimation of Divergence Times for Major Lineages of Primate Species. *Mol. Biol. Evol.*, Mar 2003; 20: 424 - 434.
- Goffin, V., Shiverick, K.T., Kelly, P.A., Martial, J.A., 1996. Sequence-function relationships within the expanding family of prolactin, growth hormone, placental lactogen, and related proteins in mammals. *Endocr Rev* 17, 385-410.
- Golos ,TG; Durning, M., Fisher, J.M. y Fowler, P.D., 1993. Cloning of four growth hormone/chorionic somatomammotropin-related complementary deoxyribonucleic acids differentially expressed during pregnancy in the rhesus monkey placenta. *Endocrinology*, 133: 1744.
- Gonzalez Alvarez, R., et al., Growth hormone locus expands and diverges after the separation of New and Old World Monkeys. *Gene*, 2006. 380(1): p. 38-45.
- Goodman, M., L. I. Grossman, *et al.*, 2005. Moving primate genomics beyond the chimpanzee genome." *Trends Genet* 21(9): 511-7.
- Green, P.; Gordon, D., Abajian, C., 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8, 195-202.
- Green, R. E., J. Krause, *et al.*, 2006. Analysis of one million base pairs of Neanderthal DNA. *Nature* 444(7117): 330-6.
- Green, R.E., et al., A draft sequence of the Neandertal genome. *Science*, 2010. 328(5979): p. 710-22.
- Han, K., S. K. Sen, *et al.*, 2005. Genomic rearrangements by LINE-1 insertion-mediated deletion in the human and chimpanzee lineages. *Nucleic Acids Res* 33(13): 4040-52.
- Handwerger, S., Freemark, M., 2000. The roles of placental growth hormone and placental lactogen in the regulation of human fetal growth and development. *J Pediatr Endocrinol Metab* 13, 343-356.
- Hardison, R. C., 2003. Comparative genomics. *PLoS Biol* 1(2): E58.
- Harper, M.E., H.A. Barrera-Saldana, and G.F. Saunders, Chromosomal localization of the human placental lactogen-growth hormone gene cluster to 17q22-24. *Am J Hum Genet*, 1982. 34(2): p. 227-34.

- Hartog, M. , 1972. Human chorionic somatomammotrophin and its clinical significance. Clin Endocrinol (Oxf) 1(2): 208-18.
- Henikoff, Steven, Elizabeth A. Greene, Shmuel Pietrokovski, Peer Bork, Teresa K. Attwood, y Leroy Hood. , 1997. Gene Families: The Taxonomy of Protein Paralogs and Chimeras. Science 278: 609.
- Herrington, J. y C. Carter-Su. 2001 Signaling pathways activated by the growth hormone receptor. Trends Endocrinol Metab. 12(6): p. 252-7.
- Herrington, J., et al., The role of STAT proteins in growth hormone signaling. Oncogene, 2000. 19(21): p. 2585-97.
- Horgan, Douglas J., H. Ohno, Thomas P. Singer, y John E. Casida., 1968. Studies on the Respiratory Chain-linked Reduced Nicotinamide Adenine Dinucleotide Dehydrogenase. XV. INTERACTIONS OF PIERICIDIN WITH THE MITOCHONDRIAL RESPIRATORY CHAIN. J. Biol. Chem., 243: 5967 - 5976. [http://inmegen.gob.mx/index.php?option=com\\_content&task=view&id=88&Itemid=213&mo=1&yr=2007&lang=es](http://inmegen.gob.mx/index.php?option=com_content&task=view&id=88&Itemid=213&mo=1&yr=2007&lang=es).
- Hu L, Lytras A, Bock ME, Yuen CK, Dodd JG, Cattini PA 1999. Detection of placental growth hormone variant and chorionic somatomammotropin-L RNA expression in normal and diabetic pregnancy by reverse transcriptase-polymerase chain reaction. Mol Cell Endocrinol 157:131-142.
- Huang, X. y Madan, A., 1999. CAP3: A DNA sequence assembly program. Genome Res., 9, 868-877.
- Ingraham, H.A., Chen, R.P., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L., Rosenfeld, M.G., 1988. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell 55, 519-529.
- Initial sequence of the chimpanzee genome and comparison with the human genome. Nature, 2005. 437(7055): p. 69-87.
- International Consortium Completes Human Genome Project. Geoff Spencer, NHGRI. <http://www.genome.gov/11510908>.
- Jacquemin, P., E. Alsat, *et al.*, 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(10): 845-54.
- Jacquemin, P., Oury, C., Peers, B., Morin, A., Belayew, A., Martial, J.A., 1994. Characterization of a single strong tissue-specific enhancer downstream from the three human genes encoding placental lactogen. Mol Cell Biol 14, 93-103.

- Jiang, S.W., Shepard, A.R., Eberhardt, N.L., 1995. An initiator element is required for maximal human chorionic somatomammotropin gene promoter and enhancer function. *J Biol Chem* 270, 3683-3692.
- Jones, B. K., Monks, B.R. *et al.*, 1995. The human growth hormone gene is regulated by a multicomponent locus control region. *Mol Cell Biol* 15(12): 7010-21.
- Josimovich, J. B. y Atwood, B.L., 1964. Human Placental Lactogen (Hpl), a Trophoblastic Hormone Synergizing with Chorionic Gonadotropin and Potentiating the Anabolic Effects of Pituitary Growth Hormone. *Am J Obstet Gynecol* 88: 867-79.
- Karabulut, A.K., Layfield, R., Pratten, M.K., 2001. Growth promoting effects of human placental lactogen during early organogenesis: a link to insulin-like growth factors. *J Anat* 198, 651-662.
- Karin M., Theill, L., Castrillo, J.S., McCormick, A., Brady, H. 1990. Cell type specific expression of the growth hormone gene and its control by GHF-1. *Folia endocrinol.* 66. 1205-1220.
- Kehrer-Sawatzki, H. y D. N. Cooper, 2007. Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum Mutat* 28(2): 99-130.
- Khaitovich, P., Enard, W., Lachmann, M. y Pääbo, S., 2006. Evolution of primate gene expression. Review. *Nature Reviews Genetics* 7, 693-702.
- Kim, Y. J. y Felig, P., 1971. Plasma chorionic somatomammotropin levels during starvation in midpregnancy. *J Clin Endocrinol Metab* 32(6): 864-7.
- Kimura, A.P., S.A. Liebhaber y N.E. Cooke. 2004 Epigenetic modifications at the human growth hormone locus predict distinct roles for histone acetylation and methylation in placental gene activation. *Mol Endocrinol.* 18(4): p. 1018-32.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217(5129): 624-6.
- King, M.C. y Wilson, A.C., 1975. Evolution at two levels in humans and chimpanzees. *Science* 188(4184): 107-16.
- Konopka, G., et al., Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature*, 2009. 462(7270): p. 213-7.
- Koonin EV., 2001. Computational genomics. *Curr Biol.* 11:R155-158.
- Kumar, S., Filipski, A. *et al.*, 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proc Natl Acad Sci U S A* 102(52): 18842-7.
- Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: Integrated software for Molecular

- Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5, 150-163.
- Lander, E.S., Linton, L.M. *et al.*, 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822): 860-921.
- Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M., 2000. Structural evidence for evolution of the beta/alphabarrel scaffold by gene duplication and fusion. *Science*. 289:1546-1550.
- Leakey, M.G., Feibel, C.S. *et al.*, 1998. New specimens and confirmation of an early age for *Australopithecus anamensis*. *Nature* 393(6680): 62-6.
- Lesk, Arthur M., Michael Levitt, and Cyrus Chothia, 1986. Alignment of the amino acid sequences of distantly related proteins using variable gap penaltie. *Protein Eng*, 1: 77 - 78.
- Li, Y., *et al.*, 2005. Independent origin of the growth hormone gene family in New World monkeys and Old World monkeys/hominoids. *Endocrinology J*. Vol. 35, p. 399-409.
- Liu, J.C., Makova, K.D., Adkins, R.M., 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 Evol* 18, 945-953.
- Locke, D.P., *et al.*, 2011. Comparative and demographic analysis of orang-utan genomes. *Nature*, 469(7331): p. 529-33.
- Lowe, C.B., G. Bejerano, and D. Haussler. 2007 Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. *Proc Natl Acad Sci U S A*. 104(19): p. 8005-10.
- Lynch, V.J., *et al.*, Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet*, 2011. 43(11): p. 1154-9.
- MacLeod JN, Lee AK, Liebhaber SA, Cooke NE 1992. Developmental control and alternative splicing of the placentally expressed transcripts from the human growth hormone gene cluster. *J Biol Chem* 267:14219–14226
- MacLeod, J.N., Worsley, I. *et al.*, 1991. Human growth hormone-variant is a biologically active somatogen and lactogen. *Endocrinology* 128(3): 1298-302.
- Männik J., Vaas P., Rull K., Teesalu P., Rebane T., and 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(5): 2433–2442.
- Marques-Bonet, T., O.A. Ryder, and E.E. Eichler, Sequencing primate genomes: what have we learned? *Annu Rev Genomics Hum Genet*, 2009. 10: p. 355-86.

- Meyer, M., et al., A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 2012. 338(6104): p. 222-6.
- Mikkelsen, S.T. *et al.*, 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055): p. 69-87.
- Misra-Press A., Cooke N.E., Liebhaber S.A. 1994. Complex alternative splicing partially inactivates the human chorionic somatomammotropin-like (hCS-L) gene. *J Biol Chem* 269:23220–23229
- Nachtigal, M.W., Nickel, B.E., Cattini, P.A., 1993. Pituitary-specific repression of placental members of the human growth hormone gene family. A possible mechanism for locus regulation. *J Biol Chem* 268, 8473-8479.
- New findings from the ENCODE project challenge established views on human genome. 2007. [http://www.cbse.ucsc.edu/news/2007/06/13/encodeposter\\_nature0607sm.jpg](http://www.cbse.ucsc.edu/news/2007/06/13/encodeposter_nature0607sm.jpg). Consulta Junio 1, 2008.
- Nickel, B.E., E. Kardami y P.A. Cattini, 1990. Differential expression of human placental growth-hormone variant and chorionic somatomammotropin in culture. *Biochem J*, 267(3): p. 653-8.
- Nickel, B.E., Kardami, E. *et al.*, 1990. Differential expression of human placental growth-hormone variant and chorionic somatomammotropin in culture. *Biochem J* 267(3): 653-8.
- Norquay, L. D., X. Yang, *et al.*, 2006. Hepatocyte nuclear factor-3alpha binding at P sequences of the human growth hormone locus is associated with pituitary repressor function. *Mol Endocrinol* 20(3): 598-607.
- Norquay, L.D., *et al.*, 2001. A member of the nuclear factor-1 family is involved in the pituitary repression of the human placental growth hormone genes. *Biochem J*. 354(Pt 2): p. 387-95.
- Norquay, L.D., Yang, X. *et al.*, 2003. RFX1 and NF-1 associate with P sequences of the human growth hormone locus in pituitary chromatin. *Mol Endocrinol* 17(6): 1027-38.
- Ohta T., 1989. Role of gene duplication in evolution. *Genome*. 31:304-310.
- Olson, MV; Varki, A. The chimpanzee genome - A bittersweet celebration. *Journal: Editorial Material. SCIENCE* 305 (5681): 191-192.
- Perez-Maya, A.A., et al., The chimpanzee GH locus: composition, organization, and evolution. *Mamm Genome*, 2012. 23(5-6): p. 387-98.
- Peterson, F.C. y Brooks, C.L., 1997. Identification of a Lactogenic Motif in Human Growth Hormone. *Journal of Biological Chemistry* 272: 21444-21448.

- Prakash, A. y Tompa, M., 2005. Discovery of regulatory elements in vertebrates through comparative genomics. *Nat Biotechnol.* 23(10):1249-56.
- Proyecto del Genoma Humano. Instituto Nacional de Medicina Genómica. Mexico. [http://inmegen.gob.mx/index.php?option=com\\_content&task=view&id=88&Itemid=213&mo=1&yr=2007&lang=es](http://inmegen.gob.mx/index.php?option=com_content&task=view&id=88&Itemid=213&mo=1&yr=2007&lang=es). Consulta 25 de Mayo de 2008.
- Prufer, K., et al., The bonobo genome compared with the chimpanzee and human genomes. *Nature*, 2012. 486(7404): p. 527-31.
- Ray J, Jones BK, Liebhaber SA, Cooke NE. Glycosylated human growth hormone variant. *Endocrinology*. 1989 Jul; 125(1):566-8.
- Reich, D., et al., 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. 468(7327): p. 1053-60.
- Reich, D., et al., 2011 Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *American Journal of Human Genetics*. 89(4): p. 516-528.
- Reséndez-Pérez D, Ramírez-Solis R. Varela-Echavarría A., Martínez-Rodríguez HG, Barrera-Saldaña HA. 1990. Coding potential of transfected human placental genes. *Nucleic Acids Res*, 18: 4665-4670.
- Revol de Mendoza, A., et al., 2004. Expansion and divergence of the GH locus between spider monkey and chimpanzee. *Gene*, 336(2): p. 185-193.
- Roberts et al., 2001. A History of the Human Genome Project. *Science*: Vol. 291. no. 5507, p. 1195.
- Roberts, Leslie, R. John Davenport, Elizabeth Pennisi, y Eliot Marshall, 2001. A History of the Human Genome Project. *Science*. Vol. 291. no. 5507, p. 1195.
- Rogers BL, MG, Sobnosky, GF, Saunders. 1986. Transcriptional enhancer within the human placental lactogen and growth hormone multigene cluster. *Nucleic Acids Res*. 14: 7647-7659.
- Rouze P, Pavy N, Rombauts S., 1999. Genome annotation: which tools do we have for it? *Curr Opin Plant Biol*. 2:90-95.
- Ruiz Parra, A.I., Metabolismo en el embarazo. Entre lo clásico y lo actual. . *Rev Col Obstet Ginecol*, 1999. 50(3): p. 125-126.
- Rust, A.G., E. Mongin, and E. Birney, Genome annotation techniques: new approaches and challenges. *Drug Discov Today*, 2002. 7(11): p. S70-6.

- Rygaard, K., A. Revol, *et al.*, 1998. Absence of human placental lactogen and placental growth hormone (HGH-V) during pregnancy: PCR analysis of the deletion. *Hum Genet* 102(1): 87-92.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425.
- Sakatani, S., *et al.*, Structure, expression, and conserved physical linkage of mouse testicular cell adhesion molecule-1 (TCAM-1) gene. *Genome*, 2000. 43(6): p. 957-62.
- Sakatani, S., R. Takahashi, *et al.*, 2000. Structure, expression, and conserved physical linkage of mouse testicular cell adhesion molecule-1 (TCAM-1) gene. *Genome* 43(6): 957-62.
- Salzberg S.L., 2003. Genomics: Yeast rises again. *Nature*. 423:233-234.
- Sambrook Joseph., R.W.D., 2001. *Molecular Cloning A Laboratory Manual*. Cold Spring Harbor, New York.
- Scally, A., *et al.*, Insights into hominid evolution from the gorilla genome sequence. *Nature*, 2012. 483(7388): p. 169-75.
- Schanke, J. T., C. M. Conwell, *et al.*, 1997. Pit-1/growth hormone factor 1 splice variant expression in the rhesus monkey pituitary gland and the rhesus and human placenta. *J Clin Endocrinol Metab* 82(3): 800-7.
- Sen, S. K., K. Han, *et al.*, 2006. Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet* 79(1): 41-53.
- Shaw, C.J. and J.R. Lupski, Non-recurrent 17p11.2 deletions are generated by homologous and non-homologous mechanisms. *Hum Genet*, 2005. 116(1-2): p. 1-7.
- Shewchuk, B. M., Y. Ho, *et al.*, 2006. A Single Base Difference between Pit-1 Binding Sites at the hGH Promoter and Locus Control Region Specifies Distinct Pit-1 Conformations and Functions. 26: 6535-6546.
- Shewchuk, B.M., S.A. Liebhaber, and N.E. Cooke, Specification of unique Pit-1 activity in the hGH locus control region. *Proc Natl Acad Sci U S A*, 2002. 99(18): p. 11784-9.
- Smit, A.F.A.; Hubley, R. y Green, P. Datos no publicados. RepeatMasker en <http://repeatmasker.org>. Current Version: open-3.2.6. 1996-2004.
- Sonntag, W. E., M. Ramsey, *et al.*, 2005. Growth hormone and insulin-like growth factor-1 (IGF-1) and their influence on cognitive aging. *Ageing Res Rev* 4(2): 195-212.
- Spencer, Geoff. International Consortium Completes Human Genome Project, NHGRI. Consulta 23 de Mayo de 2008. <http://www.genome.gov/11510908>.

- Stein L., 2001. Genome annotation: from sequence to biology. *Nat Rev Genet.* 2:493-503.
- Steiper ME, Y.N., Primates, in *The Timetree of Life*, K.S. Hedges SB, Editor. 2009, Oxford University Press: New York. p. 482–486.
- Tatusova, T.A. y Madden, T.L., 1999. Blast 2 sequences - a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett.* 174:247-250.
- Trujillo, M. A., M. Sakagashira, *et al.*, 2006. The human growth hormone gene contains a silencer embedded within an *Alu* repeat in the 3'-flanking region. *Mol Endocrinol* 20(10): 2559-75.
- Varki, Ajit; Tasha K. Altheide., 2005. Comparing the human and chimpanzee genomes: Searching for needles in a haystack. *Genome Res.*, 15: 1746 - 1758.
- Verhaeghe, J., Pintiaux, A., Van Herck, E., Hennen, G., Foidart, J.M., Igout, A., 2002. Placental GH, IGF-I, IGF-binding protein-1, and leptin during a glucose challenge test in pregnant women: relation with maternal body weight, glucosa tolerance, and birth weight. *J Clin Endocrinol Metab* 87, 2875-2882.
- Walker W.H., Fitzpatrick S.L., Barrera-Saldaña H.A., Reséndez-Perez D., Saunders G.F., 1991. The human placental lactogen genes: Structure, function, evolution and transcriptional regulation. *Endocrine Reviews* 12, 316-328.
- Wallis, O. C. y M. Wallis, 2002. Characterisation of the GH gene cluster in a new-world monkey, the marmoset (*Callithrix jacchus*). 29: 89-97.
- Wallis, O. C. y W. Michael, 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*). *Journal of Molecular Evolution* V63(5): 591-601.
- Wallis, O.C., *et al.*, 2006. Polymorphism of the growth hormone gene of red deer (*Cervus elaphus*). *Gen Comp Endocrinol*, 146(2): p. 180-5.
- Wallis, O.C., Z. Maniou, y M. Wallis, 2005. Cloning and characterization of the gene encoding growth hormone in finback whale (*Balaenoptera physalus*). *Gen Comp Endocrinol*, 143(1): p. 92-7.
- Watts RL, Watts DC., 1968. Gene duplication and the evolution of enzymes. *Nature.* 217:1125-1130.
- Williams, C. y T. M. Coltart, 1978. Adipose tissue metabolism in pregnancy: the lipolytic effect of human placental lactogen. *Br J Obstet Gynaecol* 85(1): 43-6.
- Ye, Chun, e.a., 2005. Molecular evolution of growth hormone gene family in old world monkeys and hominoids. *Gene*, 350: p. 183-192.

Yoo, E. J., I. Cajiao, *et al.*, 2006. Tissue-specific chromatin modifications at a multigene locus generate asymmetric transcriptional interactions. *Mol Cell Biol* 26(15): 5569-79.

Zhang, G., et al., Cross-comparison of the genome sequences from human, chimpanzee, Neanderthal and a Denisovan hominin identifies novel potentially compensated mutations. *Hum Genomics*, 2011. 5(5): p. 453-84.

Zuckerandl, Emile, Linus Pauling., 1965. The Molecular Evolutionary Clock. *Journal of the History of Biology*. Springer Nerherlands. Vol 31.

## **CAPÍTULO IX: ANEXOS**

# *The chimpanzee GH locus: composition, organization, and evolution*

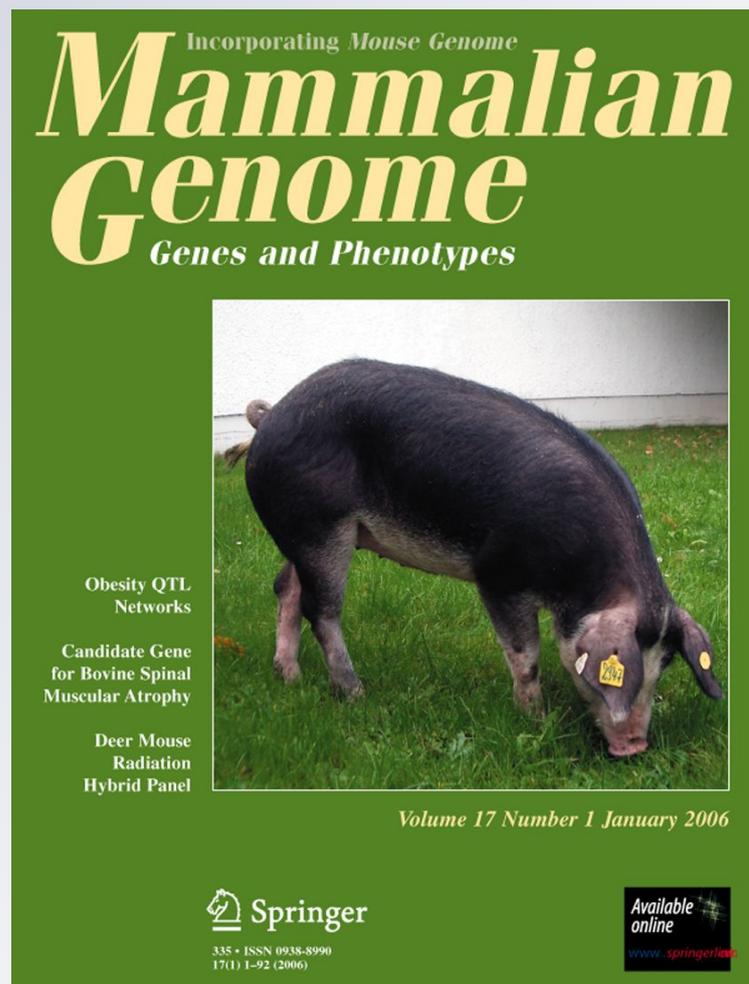
**Antonio A. Pérez-Maya, Irám  
P. Rodríguez-Sánchez, Pieter de Jong,  
Michael Wallis & Hugo A. Barrera-  
Saldaña**

**Mammalian Genome**

ISSN 0938-8990

Mamm Genome

DOI 10.1007/s00335-012-9392-4



# The chimpanzee *GH* locus: composition, organization, and evolution

Antonio A. Pérez-Maya · Irám P. Rodríguez-Sánchez ·  
Pieter de Jong · Michael Wallis ·  
Hugo A. Barrera-Saldaña

Received: 11 September 2011 / Accepted: 31 December 2011 / Published online: 31 January 2012  
© Springer Science+Business Media, LLC 2012

**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 five 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 vary. 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 fashion similar to that 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 remarkably polymorphic, unlike the corresponding region in human. The results shed new light on the complex evolution of the *GH* locus in higher primates.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00335-012-9392-4) contains supplementary material, which is available to authorized users.

A. A. Pérez-Maya · I. P. Rodríguez-Sánchez ·  
H. A. Barrera-Saldaña (✉)  
Departamento de Bioquímica y Medicina Molecular,  
Facultad de Medicina, Universidad Autónoma de Nuevo León,  
Av. Madero y Dr. Aguirre Pequeño, Col. Mitrás Centro,  
64460 Monterrey, NL, México  
e-mail: habarrera@gmail.com

A. A. Pérez-Maya  
e-mail: bioquimicomty@gmail.com

P. de Jong  
Children's Hospital Oakland Research Institute, Oakland,  
CA, USA

M. Wallis  
Biochemistry Department, School of Life Sciences, University  
of Sussex, Falmer, Brighton, Sussex BN1 9QG, UK

## 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 five genes in the same transcriptional orientation, with that encoding pituitary GH (*GH-N*) at the 5' end, followed by four 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 et al. 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 (Männik et al. 2010; Misra-Press et al. 1994). This gene is subject to an altered splicing pattern that suggests that the majority of the transcripts could encode a PL-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 six “genes,” including at least one pseudogene, while gibbon (*Hylobates leucogenys*) has seven *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 eight 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 six *GH*-like genes, while the draft chimpanzee genome ([http://www.ensembl.org/Pan\\_troglodytes](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 (Kehrer-Sawatzki and Cooper 2007; Mikkelsen et al. 2005), 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 five *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 analyzed 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 of 164 kbp) from the BACPAC Resource Center (Children’s Hospital Oakland Research Institute, Oakland, CA) 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 (Epicentre<sup>®</sup>, Madison, WI, USA), following the manufacturer’s instructions. The insert size of this BAC was about 215 kbp.

Genes and intergenic 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 ×10 buffer (provided with the enzyme), 0.2 mM of each dNTP, 1.5 mM MgCl<sub>2</sub>, one unit of Gold Taq DNA Polymerase (Epicentre), and Milli-Q water to a final volume of 50 µl per reaction. After an initial denaturation step at 95°C for 5 min, PCR was carried out for 30 cycles using a denaturation temperature of 94°C for 30 s, an annealing temperature of 64°C for 90 s, and an extension temperature of 72°C for 2 min. The reaction was stopped after a final extension time of 10 min at 72°C.

**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	GGAATGGTTGGAAGGCACTG	Antisense	IGRs

The IGRs of the *cGH* locus were amplified using the MasterAmp Extra-long PCR kit (Epicentre), following the manufacturer's instructions. The reaction mixture of 10 pmol of each primer, 40 ng of BAC DNA, 25  $\mu$ l of  $\times 2$  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  $\mu$ l. After an initial denaturation step at 94°C for 5 min, amplifications were carried out for 12 cycles, comprising incubation at 94°C for 30 s (denaturation), 55–57°C for 40 s (annealing), and 72°C for 7–14 min (elongation), followed by 23 similar cycles but with increments of 15 s per cycle in the extension time. Cycling concluded with a final extension step at 72°C for 20 min. Amplification was carried out on an ABI Veriti thermocycler (Applied Biosystems, Foster City, CA). PCR products were analyzed 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<sup>®</sup> XL PCR Cloning Kit (Invitrogen, Life Technologies, Carlsbad, CA, USA), according to the manufacturer's instructions, and transformed into electro-competent *Escherichia coli* cells supplied with the kit. Cloned amplicons were sequenced using BigDye<sup>®</sup> terminator (Applied Biosystems, Life Technologies, Foster City, CA) and universal M13 or specific primers and a DNA analyzer, ABI PRISM<sup>®</sup> 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 (454 Life Sciences, Roche, Branford, CT), with a depth of coverage of about  $\times 25$ . Chromatograms were imported into CodonCode Aligner ver. 3.7.1 (CodonCode Corporation, Dedham, MA).

#### 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 (Rodríguez-Sánchez et al. 2010; Wallis 2008). The genome assembly available at [http://www.ensembl.org/Pan\\_troglodytes](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 methods (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 center 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 (accession No. 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 modeling using Rasmol and the 3D structure of the human GH:receptor model of de Vos 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 nonsynonymous and synonymous sites in coding sequences (dN and dS) were analyzed using the codeml program in the paml package (Yang 2007).

## Results and discussion

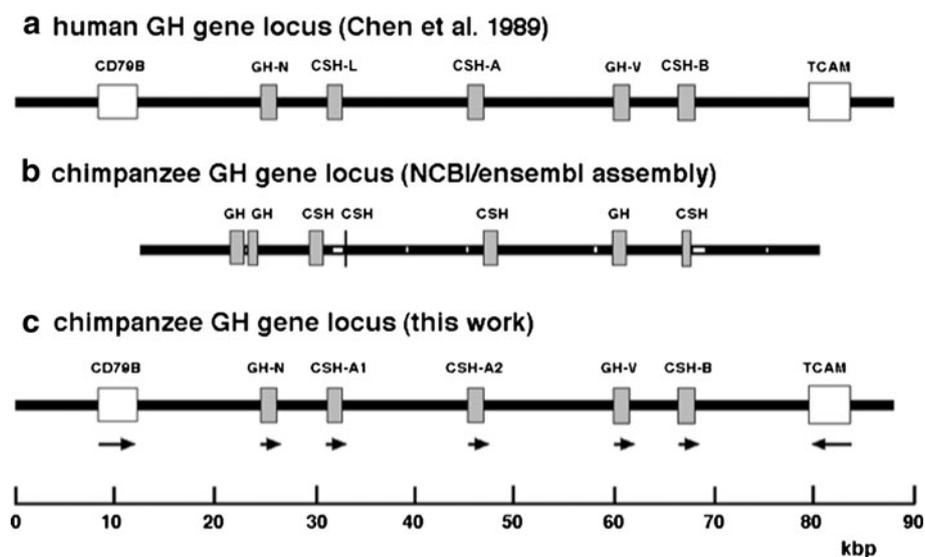
### Organization of the chimpanzee *GH* locus

Previous studies on the *cGH* locus suggested that it contains six (Revol de Mendoza et al. 2004) or seven (chimpanzee genome assembly 2.1; [http://www.ensembl.org/Pan\\_troglodytes/Info/Index](http://www.ensembl.org/Pan_troglodytes/Info/Index)) *GH*-like genes, in contrast with the five 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 five *GH*-related genes, corresponding closely to the five 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 sixth 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/](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 four 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 two other incomplete genes were not supported by this re-evaluation of the data. The first of these (Fig. 1b) is very similar to



**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](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 and 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)

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 and 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 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 five genes that encode 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 five 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; these are discussed below.

### Encoded proteins

The chimpanzee *GH-N* gene encodes a protein that differs 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.

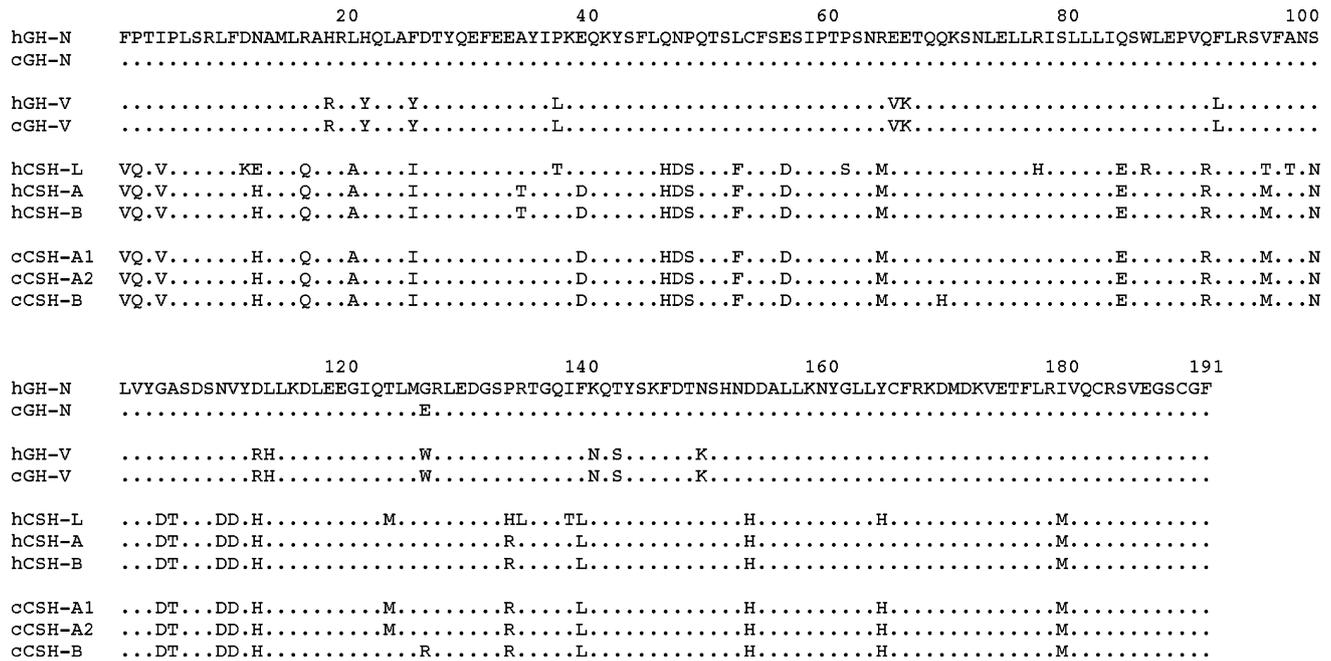
The human *CSH-L* gene potentially encodes a protein that is substantially different from placental lactogen because of a substitution in the splice donor site in intron 2. mRNA from this gene is expressed at low level, but it is not clear that this gives rise to significant levels of functional protein. 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 three residues, all in the mature protein: Gln-69, Met-123, and Gly-126 in CSH-A1/A2 are replaced by His, Thr, and 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, and Tyr in human, Ala, Met, and 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.

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

Position in locus	Gene	Abbreviations		Tissue of expression
		Human	Chimpanzee	
1	“Normal” growth hormone	<i>hGH-N</i>	<i>cGH-N</i>	Pituitary
2	Chorionic somatomammotropin	<i>hCSH-L</i>	<i>cCSH-A1</i>	Placenta
3	Chorionic somatomammotropin	<i>hCSH-A</i>	<i>cCSH-A2</i>	Placenta
4	Variant growth hormone	<i>hGH-V</i>	<i>cGH-V</i>	Placenta
5	Chorionic somatomammotropin	<i>hCSH-B</i>	<i>cCSH-B</i>	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 a 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



**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

**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*

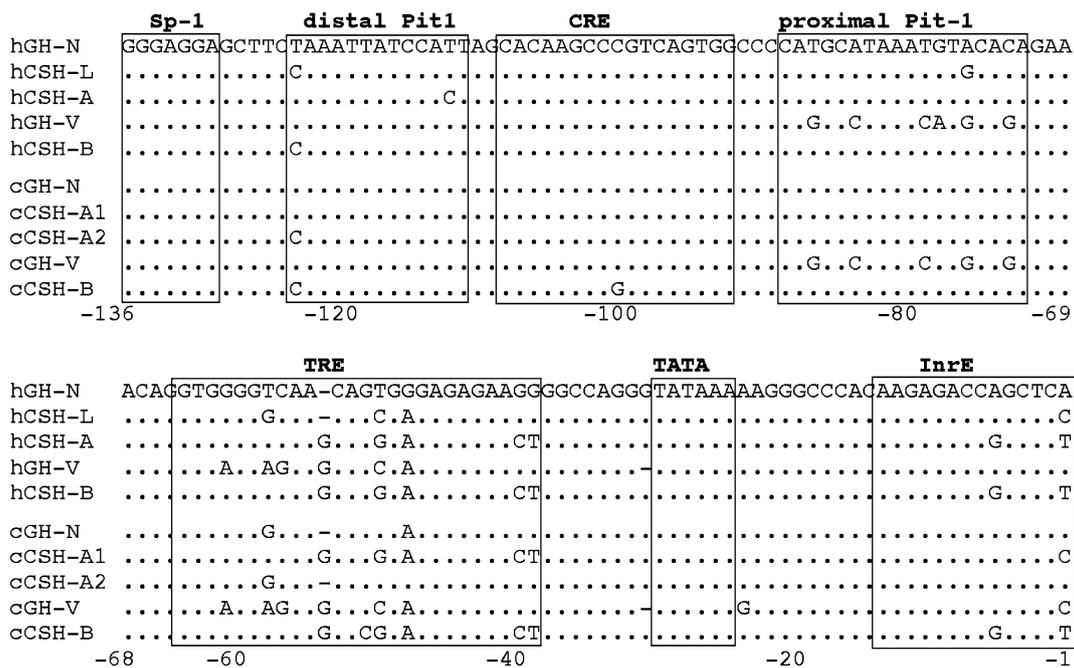
A number of response elements have been identified in the promoter region of *hGH-N* and are compared with the corresponding regions of other genes in the *hGH* and *cGH* loci in the alignment of Fig. 3. The TATA box is retained in all genes. The *Sp-1* element (Lemaigre et al. 1989) is completely conserved,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

The distal and proximal *Pit-1* elements (Krawczak et al. 1999; Theill and Karin 1993) 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 shows 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 vs. 50/811,  $p = 0.00072$ ; for chimpanzee, 5/18 vs. 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 to chimpanzee comparison, this variation differs from expectation in terms of strict orthology. Thus, the element preceding chimpanzee *CSH-A2* resembles that preceding chimpanzee (or human) *GH-N* rather than that preceding 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



**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

the enhancer (Jiang et al. 1995), is generally well conserved. The cyclic AMP response element (*CRE*, Eberhardt et al. 1996) is highly conserved, in contrast with the situation in nonprimate 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 four 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 (Jacquemin et al. 1996; Lytras et al. 1996), which are indicated in 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* and 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

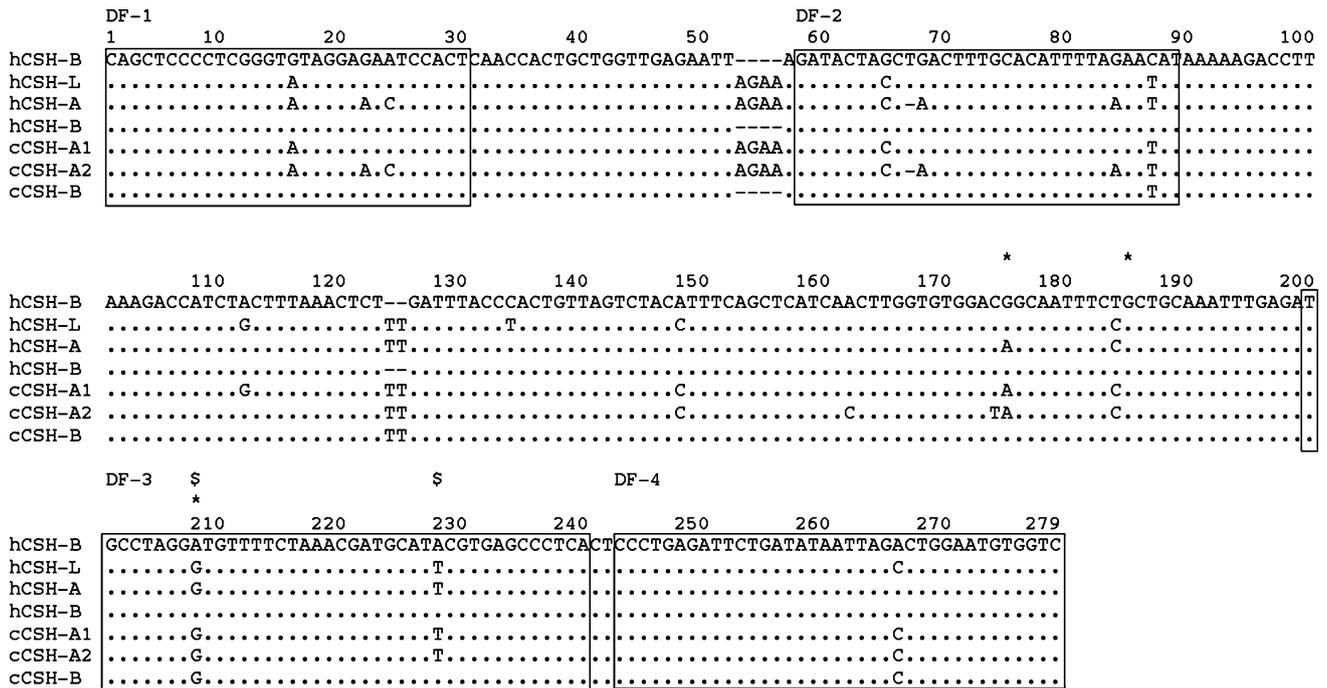
P-elements are found upstream of *GH-V* and the three *CSH* genes in human and are thought to inhibit expression of

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 the 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 nucleotides; 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 0.080–



**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)

0.176% of the chimpanzee genome (Mikkelsen et al. 2005). In the sequence of the *GH-N* gene, including 500 nt upstream and downstream of the coding sequence (~2,500 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 two 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*) suggest 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 analyzed using the GEN-ECONV program (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,

**Table 3** Gene conversion between *GH/CSH* genes

Sequences	nt Positions	Length	Location	<i>p</i> Value
Chimpanzee				
<i>CSH-A1:GH-N</i>	368–807	440	5' Upstream–promoter	<0.0001
<i>CSH-B:CSH-A1</i>	1198–1960	763	Intron 1–exon 4	0.0001
<i>CSH-A2:GH-N</i>	743–923	181	5' Upstream–promoter	0.0001
<i>CSH-A2:CSH-A1</i>	1240–1987	748	Exon 2–exon 4	0.0005
<i>CSH-A2:CSH-B</i>	1048–1965	918	Intron 1–exon 4	0.0120
<i>CSH-A2:CSH-A1</i>	2145–2632	487	Intron 4–3' downstream	0.0177
<i>CSH-B:CSH-A1</i>	2203–2665	463	Intron 4–3' downstream	0.0192
Human				
<i>CSH-A:GH-V</i>	1–172	172	5' Upstream	0.0004
<i>CSH-A:CSH-B</i>	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

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 they may have affected both expressed genes and regulatory sequences.

#### Repetitive elements and indels

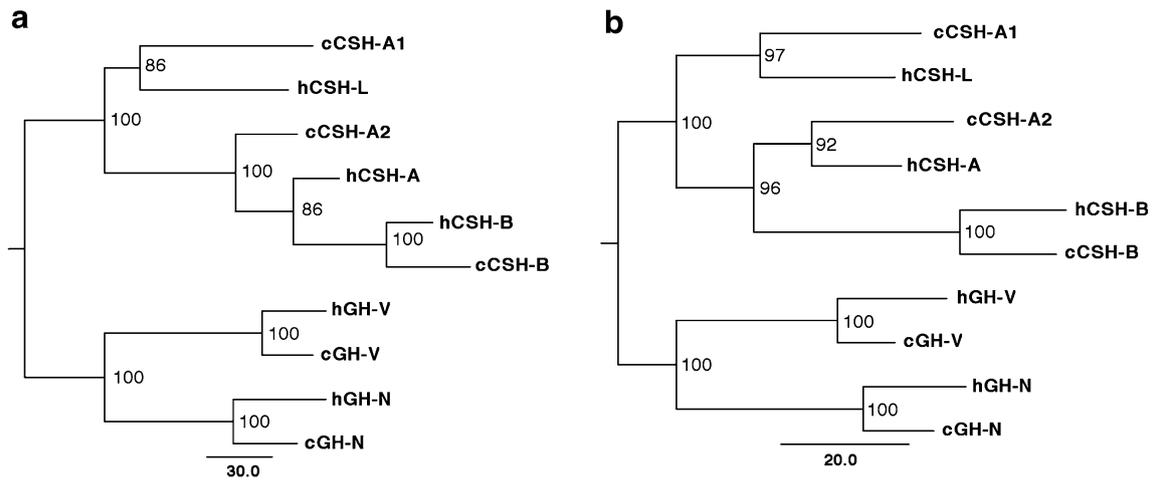
Chen et al. (1989) identified 48 *Alu* (SINE) repetitive elements in the human *GH* locus, some of them truncated. Use of RepeatMasker confirmed that 48 equivalent *Alu* sequences are also present in the chimpanzee genome, together with a number of additional SINE- and LINE-related elements. This analysis, 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. Four additional indels of more than 15 nt are seen, including three extensions/reductions of simple AG-rich sequences and a 112-nt sequence in intron 3. The latter 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 between chimpanzee and human *GH* loci implies that they shared the same evolutionary origin. In the chimpanzee locus, 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 five 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, 2,532 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 (1,803 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 1,803-nt region, where all the chimpanzee and human genes group according to the expected orthology. However, 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



**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

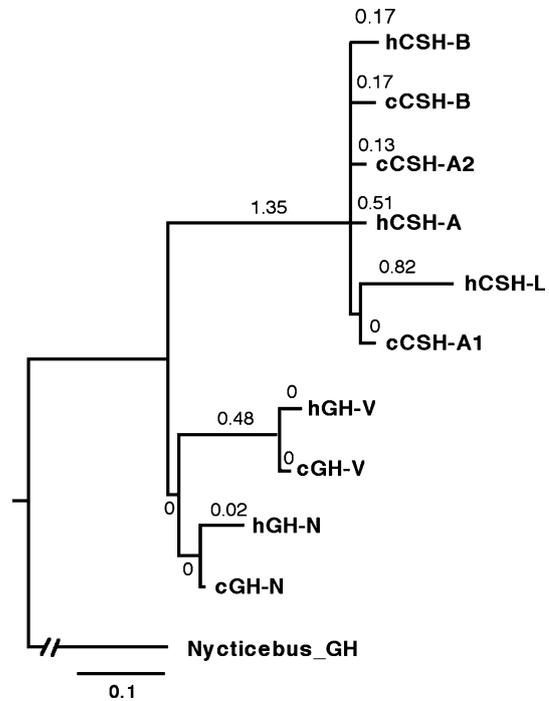
parsimony method in PAUP\*; neighbor-joining and maximum-likelihood methods gave similar results. Branch lengths are proportional to the number of substitutions, indicated by the *scale bars*. *Numbers* at nodes are percentages of at least 100 bootstrap replications supporting that clade

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).

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

*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 (Rodríguez-Sánchez et al. 2010; Wallis 1996). 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 in the lineage leading to *GH-V* after the duplication giving *GH-N* and *GH-V* (*dN/dS* = 0.48 for *GH-V* and 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,



**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

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 five 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 remarkably polymorphic, unlike the corresponding region in human.

**Acknowledgments** We thank Sergio Lozano-Rodriguez, MD, 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, Gelinias 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* 15:845–854
- 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, Frenzo JL, Muller F, Evain-Brion D (2002) Human placental growth hormone—a review. *Placenta* 23(Suppl A):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 XJ, Xiao R, Gong YY, Zhang YP (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 JC, Makova KD, Adkins RM, Gibson S, Li WH (2001) Episodic evolution of growth hormone in primates and emergence of the species specificity of human growth hormone receptor. *Mol Biol Evol* 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 et al (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-3 $\alpha$  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/geneconv>. Accessed 20 Jan 2012
- Slater EP, Rabenau O, Karin M, Baxter JD, Beato M (1985) Glucocorticoid receptor binding and activation of a heterologous promoter by dexamethasone by the first intron of the human growth hormone gene. *Mol Cell Biol* 5(11):2984–2992
- Smit AFA, Hubley R, Green P (1996–2010) RepeatMasker Open-3.0. Available at <http://www.repeatmasker.org>
- Swofford DL (1998) PAUP\*. Phylogenetic analysis using parsimony (\* and Other Methods), ver. 4. Sinauer, Sunderland
- Theill LE, Karin M (1993) Transcriptional control of GH expression and anterior pituitary development. *Endocr Rev* 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 Evol* 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