## Patrones de evolución molecular en genes del eje hormona de crecimiento / factor de crecimiento relacionado a insulina en caviomorfos y platirrinos neotropicales.

Lic. Natalia Rego natalia@pasteur.edu.uy

Orientador: Dr. Hugo Naya

Co-orientador: Dr. Enrique Lessa

Agosto, 2017

#### **Filiaciones:**

Proponente: Lic. Natalia Rego Unidad de Bioinformática Institut Pasteur de Montevideo. Orientador: Dr. Hugo Naya Unidad de Bioinformática Institut Pasteur de Montevideo. Co-orientador: Dr. Enrique Lessa Departamente de Ecología y Evolución Facultad de Ciencias, UdelaR.

### **Tribunal:**

Dr. Fernando Álvarez-Valín Sección Biomatemática Facultad de Ciencias, UdelaR. Dr. Héctor Romero Laboratorio de Organización y Evolución del Genoma Facultad de Ciencias, UdelaR. Dr. Arley Camargo Polo de Desarrollo Universitario Centro Universitario de Rivera, UdelaR.

### Resumen

Los roedores caviomorfos o histricomorfos sudamericanos (Rodentia, Hystricognathi) y los primates platirrinos (Primates, Platyrrhini) son clados de mamíferos que colonizaron Sudamérica durante su período de aislamiento de los restantes continentes, hace unos 45 a 30 millones de años atrás. Los caviomorfos comprenden alrededor de 240 especies actuales pertenecientes a cuatro linajes bien definidos: Cavioidea (e.g. Cavia, apereá), Chinchilloidea (e.g. Chinchilla), Erethizontoidea (e.g. Erethizon, coendú) y Octodontoidea (e.g. Echimys y Ctenomys, ratas espinosas y tucu-tucus). Existen unas 125 especies de platirrinos actuales clasificadas en tres familias: Atelidae (e.g. Ateles, monos araña), Cebidae (e.g. Cebus y Callithrix, capuchinos y marmositas) y Phiteciidae (e.g. Phitecia, sakis). Ambos grupos de mamíferos presentan un elevado número de anomalías en sus ejes endócrinos y vías metabólicas. Son resistentes a los glucocorticoides, son insulino-resistentes con insulinas de menor biopotencia y el glucagón es también divergente y con biopotencia disminuida. Además, los caviomorfos muestran cambios en la regulación del crecimiento: el crecimiento post-natal es independiente de la hormona de crecimiento y mantienen elevados niveles de IGF2 en el adulto. De particular interés resulta el rol que la selección positiva ha tenido en moldear estas alteraciones, seguramente asociadas a los peculiares patrones de desarrollo y morfológicos de estos mamíferos neotropicales.

Haciendo uso de los recursos genómicos y transcriptómicos disponibles en las bases de datos públicas, se obtuvieron grupos de ortólogos y alineamientos para 18 genes de interés: INS, IGF1, IGF2, INSR, IGF1R, INSRR, IGF2R, IGFBP1 a IGFBP7, GH, GHR, GHRH y GHRHR. Con el objetivo de identificar huellas de selección positiva a nivel molecular, los alineamientos se analizaron con modelos de sustitución de codones en un contexto filogenético (implementados en PAML y HyPhy). Las tasas de sustitución no-sinónima:sinónima obtenidas se compararon entre los diferentes genes y entre los diferentes clados de interés, así como con los resultados de la bibliografía disponible (principalmente Opazo *et al.*, 2005; Wallis, 2015).

Se estimó selección positiva en Platyrrhini (IGF2, IGF1R, INSRR, IGF2R, IGFBP1, IGFBP4, GHRH y GHRHR) y Caviomorpha (INS, INSR, IGFBP1, IGFBP2), siendo los resultados en general consistentes con la bibliografía disponible. En caviomorfos el receptor de insulina presentó evidencia de algunas sustituciones no-sinónimas moldeadas por selección episódica. Como ocurre con la insulina (este trabajo y Opazo *et al.*, 2005), estos cambios se presentaron en

forma linaje dependiente, espcíficamente en la rama ancestral de Octodontoidea, en concordancia con sus insulinas más divergentes.

Para los genes estudiados, el efecto de la selección positiva ha sido más extendido en Platyrrhini. Si bien aquí el muestreo taxonómico estuvo acotado a Cebidae, datos complementarios sugieren que ocurriría una clara concentración de cambios adaptativos en el ancestro del grupo, coincidiendo con un único pulso de diversificación propuesto para el infraorden. Por el contrario, los patrones de evolución molecular en Caviomorpha son linaje dependientes y muchos de los cambios no se restringen a lo ocurrido durante su diversificación en el Neotrópico, sino que aparecen durante su historia evolutiva compartida con los histricognatos africanos.

# Índice de Contenidos

Re	esume	en		i
Lis	sta de	e Figura	S	v
Lis	sta de	e Tablas		xi
1	Intro	oducción	1	1
	1.1	Anteced	lentes	1
	1.2	Objetivo	os generales y específicos	4
	1.3	Estructu	ra de la tesis	5
2	Met	odología	3	7
	2.1	Determi	nación de los genes y linajes a analizar	7
		2.1.1	Genes de interés	7
		2.1.2	Representación taxonómica	8
	2.2	Búsqueo	da de ortólogos y obtención de secuencias	10
	2.3	Alinean	niento	11
	2.4	Análisis	composicional	12
	2.5	Estimac	ión de tasas de sustitución no-sinónima: sinónima ( $\omega = dN/dS$ )	12
		2.5.1	CODEML: modelos de rama	13
		2.5.2	CODEML: modelos de rama y sitio	15
	2.6	HyPhy:	uso de modelos menos restrictivos	18
		2.6.1	BSREL, aBSREL y BUSTED	18
		2.6.2	МЕМЕ	20
		2.6.3	Relajación de la selección en Platyrrhini	21
	2.7	Filogeni	a mitocondrial	22
	2.8	Reconst	rucción de secuencias ancestrales	22
	2.9	Expresio	ón de INSR en <i>C. porcellus</i>	22
3	Resi	ultados		25
	3.1	Genes y	linajes	25
		3.1.1	Genes	25
		3.1.2	Linajes	27
	3.2	Grupos	de ortólogos y alineamientos	30

	3.3	Anális	is composicional	32
	3.4	Model	os de ramas y de ramas y sitios implementados en CODEML	34
		3.4.1	Modelo de ramas	34
		3.4.2	Modelo de ramas y sitios	35
	3.5	Model	os implementados en HyPhy	37
		3.5.1	BSREL, aBSREL y BUSTED	37
		3.5.2	MEME	38
3.5.3 Relajación de la selección en Platyrrhini				44
	3.6 Resultados específicos en Hystricognathi			46
	3.6.1 Insulina			
		3.6.2	IGF2	49
		3.6.3	Receptor de insulina	51
4	Disc	cusión		55
	4.1	Ortólo	gos y alineamientos	55
	4.2	INS e ]	INSR en Hystricognathi / Caviomorpha	57
	4.3	Patron	es en Platyrrhini	60
	4.4	Cavior	norpha versus Platyrrhini en el Nuevo Mundo	63
5	Con	clusion	es y perspectivas	67
	5.1	Conclu	isiones	67
	5.2	Perspe	ctivas	68
Re	Referencias 71			71
Α	APÉ	ÉNDICE	E: tablas suplementarias	95
в	APÉ	ÉNDICE	E: figuras suplementarias	127

# Lista de Figuras

2.1	Topología de referencia. El cladograma indica el muestreo taxonómico utilizado	
	en el estudio, así como la hipótesis filogenética empleada en los análisis de	
	modelos de codones.	9
2.2	Conformaciones de $\omega_1$ y $\omega_2$ utilizadas en los modelos de ramas de CODEML. Para	
	cada conformación, A a D, las ramas y clados blanco se indican en la topología de	
	referencia como $\omega_1$ y $\omega_2$ y están coloreadas en rojo. Las restantes ramas en negro	
	constituyen el <i>background</i> con $\omega_0$	15
2.3	Conformaciones de ramas blanco para los análisis de tipo ramas y sitios	
	implementados en CODEML (A) y BUSTED (A y B) y para RELAX (C). Las	
	ramas y clados en la categoría blanco se indican en rojo y el background en	
	negro (A y B). En el caso de RELAX (C), las ramas blanco se indican en rojo	
	(Platyrrhini), las ramas de referencia en azul (restantes euarcontos) y en negro se	
	indican los glires que se consideraron como ramas no clasificadas	17
3.1	El esquema indica los genes estudiados y sus relaciones funcionales. La GHRH	
	hipotalámica y la hormona de crecimiento hipofisiaria (GH o GH1) controlan la	
	producción endócrina de los factores de crecimiento IGF1 e IGF2 en el hígado.	
	La insulina es liberada de los islotes pancreáticos como respuesta a un incremento	
	en la glicemia. En las células diana, tanto INS como IGFs ejercen sus funciones a	
	través de los receptores diméricos INSR e IGF1R, así como a través de receptores	
	heterodiméricos INSR-IGF1R. Un tercer miembro de esta familia génica, el	
	INSRR, se considera un receptor huérfano y no tendría un rol en las vías de	
	señalización aquí estudiadas. El IGF2 también señaliza a través del INSR y del	
	IGF1R, siendo el IGF2R un receptor que no transduce señales y solo controla su	
	biodisponibilidad. En el plasma se encuentran otras proteínas de unión a IGF1 e	
	IGF2 que también intervienen en la regulación de sus niveles sanguíneos (IGFBP1	

a IGFBP7).

- 3.3 Análisis de correspondencia de todas las secuencias de codones. Los símbolos y colores identifican las secuencias pertenecientes al mismo gen.
  34

- 3.4 Sitios bajo selección positiva inferidos por MEME. Sobre el árbol mitocondrial de la Figura B-7 se dibujaron barras cuya altura es proporcional al número de sitios bajo selección positiva en todos los genes excepto IGF2R (rojo), al número de sitios bajo selección positiva en IGF2R (azul) y al número de genes con al menos un sitio detectado por MEME. Los valores graficados tienen rango de 0 a 27, correspondiendo el máximo al número de sitios en genes distintos de IGF2R en *Dipodomys*. El diseño resalta la contribución relativa de IGF2R *versus* los restantes genes. Los recuadros rojos destacan los resultados para tres ramas ancestrales de interés: la de Platyrrhini, la de Hystricognathi y la de Caviomorpha. 43

- 3.6 Alineamiento de las secuencias aminoacídicas del gen INS mostrando los sitios detectados bajo selección positiva en Hystricognathi. Se resaltan las secuencias correspondientes a Caviomorpha (rojo) y Phiomorpha (azul). En la estructura de la pre-pro-insulina, S, B, C y A indican los sitios correspondientes al péptido señal y a los péptidos B, C y A, respectivamente. Se observa la mayor conservación de las posiciones correspondientes a los péptidos B y A, que son los que aparecen en la hormona madura. Los sitios sobre fondo rojo son los que resultaron bajo selección positiva en dos instancias (B17, B18, B20 y A13 fueron detectados por Opazo et al., 2005 y CODEML en este trabajo; el sitio B29 fue detectado por MEME y CODEML en este trabajo); el sitio B22 sobre fondo naranja fue detectado solo por Opazo et al. (2005); los sitios con fondo gris fueron los detectados solo por MEME; los sitios con fondo verde fueron detectados solo por CODEML en este trabajo. Para los sitios mencionados, los asteriscos rojos indican que los sitios han sido referidos por De Meyts (2004) como participantes de la superficie formadora de hexámeros en las insulinas no histricognatas. Los asteriscos azules indican sitios que en humano tienen variantes alélicas en las que existe cambio de aminoácido y son conocidos sus efectos patogénicos o predichos por SIFT (Kumar et al., 2009) o PolyPhen (Adzhubei et al., 2013). Los datos de variantes (dbSNP; Sherry et al., 2001), SIFT y PolyPhen se accedieron a través de Ensembl
- 3.7 Alineamiento de secuencias aminoacídicas de pre-pro-IGF2. Los sitios están coloreados según el porcentaje de identidad. Se resaltan las secuencias de histricognatos. A partir de la reconstrucción de las secuencias de codones ancestrales realizadas, se confirman para Hystricognathi 11 sinapomorfías no sinónimas, correspondientes a los sitios 4, 60, 96, 143, 144, 156, 159, 160, 162, 170 y 174 (en naranja). El sitio 163 presenta una sinapomorfía para Rodentia (en verde). Los sitios 96 y 156 fueron indicados por MEME con eventos de selección episódica en la rama ancestral del clado (en rojo). Los asteriscos rojos indican sitios conocidos de O-glicosilación (residuos Ser71, Thr72, Thr75 y Thr139 según el pro-IGF2 humano; Duguay *et al.*, 1998).

48

50

B-1	Vía de señalización de insulina obtenida de KEGG (vía hsa04910; Kanehisa
	et al., 2015). Sobre el diseño original, genes proximales como INS e INSR se
	colorearon en rojo; algunos genes importantes que intervienen en la respuesta
	principalmente metabólica se indican en rosa y otros genes importantes en la
	respuesta de proliferación y crecimiento se colorearon en violeta. Cabe señalarse
	que el número total de genes es mucho mayor al indicado, ya que en general
	cada cuadrado representa múltiples genes. Por ejemplo, IRS representa cuatro
	parálogos diferentes (IRS1 a IRS4).

B-2 *Timetree*, cladograma calibrado según el tiempo geológico de los procesos de diversificación, para los *taxa* estudiados. La filogenia se obtuvo a partir de la base de datos TimeTree (http://www.timetree.org; Hedges y Kumar, 2009; Hedges *et al.*, 2015).

- B-4 Análisis de correspondencia del uso de aminoácidos de los genes y *taxa* estudiados. Los triángulos azules indican los aminoácidos mientras que los círculos rojos representan las especies. ANA: *A. nancymaae*; BOV: *B. taurus*; CAR: Carnivora; CJA: *C. jacchus*; CLA: *C. lanigera*; CPO: *C. porcellus*; CSO: *C. sociabilis*; DOR: *D. ordii*; FDA: *F. damarensis*; GGO: *G. gorilla*; HGL: *H. glaber*; HSA: *H. sapiens*; MMU: *M. mulatta*; MUS: *M. musculus*; NGA: *N. galili*; OCU: *O. cuniculus*; ODE: *O. degus*; OGA: *O. garnettii*; OPR: *O. princeps*; PER: Perissodactyla; SBO: *S. boliviensis*; TBE: *T. chinensis*; TSY: *T. syrichta.* . . . . . 131
- B-5 Análisis de correspondencia de todas las secuencias de codones. Los símbolos y colores identifican las secuencias pertenecientes al mismo *taxa*. Para mejor claridad se indican *taxa* de orden superior. EUARCHONTA-NP: euarcontos no Platyrrhini; GLIRES-NH: glires no Hystricognathi; OUTGROUP: Laurasiatheria. 132
- B-6 Alineamiento de IGF1R para una región que comprende los codones 291, 293, 294 y 295. En el recuadro a la derecha se indican los aminoácidos para esas posiciones. Si bien el modelo de ramas y sitios implementado en CODEML detectó selección positiva en cuatro codones muy cercanos, observando el alineamiento se descartan errores en alguna de las secuencias o en la hipóteiss de homología. En el caso del codon 291, catarrinos y platirrinos presentan codones sinónimos para serina, pero la sustitución AGC TCG involucra al menos tres pasos, siendo al menos uno de ellos un cambio no sinónimo.

- B-7 Topología de referencia con largos de rama obtenidos a partir de 12 genes mitocondriales de la hebra pesada. Un alineamiento construido a partir de los 12 genes codificantes de proteínas de la hebra mitocondrial pesada, con exclusión de la tercera posición de los codones, se utilizó para calcular en PhyML el largo de cada rama de la topología. Estos largos indican la tasa de cambio por rama, encontrándose la escala en unidades de sustituciones por nucleótido. Los números indican nodos que se utilizan para describir ramas internas en la Tabla 3.8.
- B-9 Visualización genómica del exón 11 y región adaycente del gen INSR en humano, tomada del UCSC Genome Browser (https://genome.ucsc.edu/). Debajo de la secuencia y modelo génico en humano (el gen tiene sentido de derecha a izquierda en la figura), se muestra el *track* de conservación y alineamiento para vertebrados, donde solo los *taxa* de interés han sido seleccionados. Para mayor claridad, en la parte inferior de la figura se incorporó un diagrama de esta región génica. Se puede observar un alto nivel de conservación general, así como la deleción en *C. porcellus* (flecha roja) de la región homóloga al inicio del exón 11 en los restantes mamíferos incluidos en el alineamiento. La secuencia identificada en UCSC Genome Browser como "brush-tailed rat" corresponde a *O. degus.*...
- B-10 Visualización en IGV (Robinson *et al.*, 2011) del scaffold\_42 de *C. porcellus* (cavPor3) para la región que contiene al INSR (se observan 11 de los 21 exones del gen). Se muestra el gráfico de cobertura y parte de los alineamientos obtenidos con Tophat (Kim *et al.*, 2013) y Rsubread (Liao *et al.*, 2013), de *reads* provenientes del secuenciamiento de dos muestras transcriptómicas, una de pulmón y otra de músculo esquelético. En la parte inferior se muestran los modelos de los transcriptos ensamblados por Cufflinks (Trapnell *et al.*, 2010) para ambas muestras. Las líneas verticales rojas están centradas alrededor de la región que es la homóloga a la región del cromosoma 19 de *H. sapiens* que contiene al exón 11 de 36 nucleótidos (scaffold\_42 14.479.982-14.480.021 en cavPor3). A pesar del alto nivel de transcripción e independientemente de la herramienta utilizada, no hay *reads* en ninguna de la muestras que mapeen a la región homóloga al exón 11 del INSR humano. En concordancia, el exón no aparece incluído en los modelos de los transcriptos ensamblados por Cufflinks. 137

# Lista de Tablas

2.1	Modelos de sustitución de codones empleados en este trabajo	14
3.1 3.2	Lista de los 18 genes analizados en el presente estudio	28
	Gblocks y BMGE.	30
3.3	Muestreo taxonómico asociado a cada gen.	31
3.4	Parámetros estimados con los modelos de ramas implementados en CODEML.	36
3.5	Valor de $\omega^+$ y sitios bajo selección según el modelo de ramas y sitios	
	implementado en CODEML. solo se muestran aquellos genes donde se infirió	
	selección positiva en al menos una de las conformaciones de linajes blanco probadas.	37
3.6	Resultados de BSREL para aquellos genes donde existieron ramas bajo selección	
	diversificadora para un $\alpha \leq 0, 1.$	39
3.7	Resúmen de resultados de los análisis de codones	40
3.8	Resultados de MEME por gen y por rama	42
A-1	Lista de ortólogos para INS (insulin)	95
A-2	Lista de ortólogos para IGF1 (insulin-like growth factor 1)	96
A-3	Lista de ortólogos para IGF2 (insulin-like growth factor 2)	96
A-4	Lista de ortólogos para INSR (insulin receptor)	97
A-5	Lista de ortólogos para IGF1R (insulin-like growth factor 1 receptor)	97
A-6	Lista de ortólogos para INSRR (insulin receptor-related receptor)	98
A-7	Lista de ortólogos para IGF2R (insulin-like growth factor 2 receptor)	99
A-8	Lista de ortólogos para IGFBP1 (insulin-like growth factor binding protein 1)	99
A-9	Lista de ortólogos para IGFBP2 (insulin-like growth factor binding protein 2)	100
A-10	Lista de ortólogos para IGFBP3 (insulin-like growth factor binding protein 3)	100
A-11	Lista de ortólogos para IGFBP4 (insulin-like growth factor binding protein 4)	101
A-12	Lista de ortólogos para IGFBP5 (insulin-like growth factor binding protein 5)	102
A-13	B Lista de ortólogos para IGFBP6 (insulin-like growth factor binding protein 6)	103
A-14	Lista de ortólogos para IGFBP7 (insulin-like growth factor binding protein 7)	103
A-15	b Lista de ortólogos para GH1 (growth hormone 1)	104
A-16	Lista de ortólogos para GHR (growth hormone receptor)	105
A-17	Lista de ortólogos para GHRH (growth hormone releasing hormone)	105

A-18 Lista de ortólogos para GHRHR (growth hormone releasing hormone receptor) .	106
A-19 Modelos de sustitución nucleotídica. Se presentan los modelos con mejor ajuste	
para cada gen según el valor de AICc. Los cálculos fueron realizados con	
Modelgeneartor v0.8.5 (Keane <i>et al.</i> , 2006)	106
A-20 Contenido de bases y G+C (%) para los diferentes <i>taxa</i> estudiados (suma de todas	
las secuencias por <i>taxa</i> )	107
A-21 Uso de aminoácidos para los diferentes taxa estudiados (suma de todas las	
secuencias por <i>taxa</i> )	108
A-22 Modelo de ramas y sitios (CODEML) para INS seleccionando el clado	
Caviomorpha como linaje blanco.	109
A-23 Modelo de ramas y sitios (CODEML) para IGF2 seleccionando la rama ancestral	
de Platyrrhini como linaje blanco.	109
A-24 Modelo de ramas y sitios (CODEML) para IGF2 seleccionando el clado	
Platyrrhini como linaje blanco	110
A-25 Modelo de ramas y sitios (CODEML) para INSR seleccionando la rama ancestral	
de Caviomorpha como linaje blanco	110
A-26 Modelo de ramas y sitios (CODEML) para IGF1R seleccionando la rama ancestral	
de Platyrrhini como linaje blanco.	111
A-27 Modelo de ramas y sitios (CODEML) para INSRR seleccionando el clado	
Platyrrhini como linaje blanco	111
A-28 Modelo de ramas y sitios (CODEML) para IGF2R seleccionando la rama ancestral	
de Platyrrhini como linaje blanco.	112
A-29 Modelo de ramas y sitios (CODEML) para IGFBP1 seleccionando la rama	
ancestral de Platyrrhini como linaje blanco.	112
A-30 Modelo de ramas y sitios (CODEML) para IGFBP1 seleccionando el clado	
Platyrrhini como linaje blanco.	113
A-31 Modelo de ramas y sitios (CODEML) para IGFBP1 seleccionando el clado	
Caviomorpha como linaje blanco.	113
A-32 Modelo de ramas y sitios (CODEML) para IGFBP2 seleccionando la rama	
ancestral de Platyrrhini como linaje blanco.	114
A-33 Modelo de ramas y sitios (CODEML) para IGFBP4 seleccionando la rama	
ancestral de Platyrrhini como linaje blanco.	114
A-34 Modelo de ramas y sitios (CODEML) para IGFBP6 seleccionando la rama	
ancestral de Caviomorpha como linaje blanco	115
A-35 Modelo de ramas y sitios (CODEML) para GHRH seleccionando la rama ancestral	
de Platyrrhini como linaje blanco.	115
A-36 Modelo de ramas y sitios (CODEML) para GHRH seleccionando el clado	
Platyrrhini como linaje blanco	116
A-37 Modelo de ramas y sitios (CODEML) para GHRHR seleccionando la rama	
ancestral de Platyrrhini como linaje blanco.	116

A-38 Modelo de ramas y sitios (CODEML) para GHRHR seleccionando el clado	
Platyrrhini como linaje blanco	117
A-39 Resultados de aBSREL para los genes INS, IGF1 e IGF2	118
A-40 Resultados de aBSREL para los genes INSR, IGF1R e INSRR	119
A-41 Resultados de aBSREL para los genes IGF2R e IGFBP1	120
A-42 Resultados de aBSREL para los genes IGFBP2, IGFBP3 e IGFBP4	121
A-43 Resultados de aBSREL para los genes IGFBP5, IGFBP6 e IGFBP7	122
A-44 Resultados de aBSREL para los genes GH1, GHR, GHRH y GHRHR	123
A-45 Resultados de BUSTED para los genes IGF2, INSR, IGF1R, INSRR e IGF2R.	124
A-46 Resultados de BUSTED para los genes IGFBP1, IGFBP2 e IGFBP6	125
A-47 Resultados de BUSTED para los genes GHRH y GHRHR	126
A-48 Prueba de relajación de la selección en Platyrrhini comparado con los restantes	
linajes de Euarchonta como referencia (RELAX)	126

### 1 Introducción

### 1.1 Antecedentes

Los roedores caviomorfos o histricomorfos sudamericanos (Rodentia; Hystricognathi) y los primates platirrinos (Primates; Platyrrhini) son clados de mamíferos que colonizaron Sudamérica durante su período de aislamiento de los restantes continentes, hace unos 45 a 30 millones de años atrás (en el Eoceno Medio - Oligoceno temprano). La hipótesis más aceptada y avalada por reconstrucciones paleogeográficas de la cuenca atlántica, hallazgos fósiles y estimaciones moleculares de tiempos de divergencia, es su llegada desde África en un evento biogeográfico de dispersión más o menos sincronizado, que involucraría el pasaje a través de un cordón de islas transatlántico así como rafting en islas flotantes impulsadas por paleocorrientes (Poux et al., 2006; de Oliveira et al., 2009; Loss-Oliveira et al., 2012). En el nuevo continente los caviomorfos sufrieron un proceso de diversificación inmediata y explosiva, siendo capaces de explotar los nuevos nichos que acompañaron los cambios climáticos del Oligoceno y Mioceno, mediante la aparición de nuevas características fenotípicas como hipsodoncia, diversos hábitos de vida o la expansión en el rango de los tamaños corporales (Vucetich et al., 2015). Los caviomorfos comprenden alrededor de 240 especies actuales (Wilson y Reeder, 2005) pertenecientes a cuatro linajes bien definidos: Cavioidea (e.g. Cavia e Hydrochoerus, apereás y carphincho), Chinchilloidea (e.g. Chinchilla, chinchillas), Erethizontoidea (e.g. Erethizon, puercoespines del Nuevo Mundo) y Octodontoidea (e.g. Echimys y Ctenomys, ratas espinosas y tucu-tucus). Por su parte, existen unas 125 especies de primates platirrinos actuales divididas en tres familias: Atelidae (e.g. Ateles, monos araña), Cebidae (e.g. Cebus y Callithrix, capuchinos y marmositas) y Pitheciidae (e.g. Pithecia, sakis). En Sudamérica y el Caribe experimentaron también una amplia radiación en nichos ocupados, morfología y tamaño, diversificación que habría ocurrido en el Oligoceno Tardío - Mioceno Temprano, representando los platirrinos actuales linajes longevos (Rosenberger, 1979). La mayoría de los estudios moleculares recientes confirman o al menos no refutan esta hipótesis (e.g. Opazo et al., 2006; Perez et al., 2013). En este sentido los platirrinos se distinguen de los caviomorfos, para los cuales las hipótesis evolutivas señalan la ocurrencia de tres pulsos de diversificación, apareciendo los octodontoideos y cavioideos modernos más recientemente, durante el Mioceno Medio (Pérez y Pol, 2012; Vucetich et al., 2015).

Debido a su importancia como modelo animal de laboratorio, una gran cantidad de información bioquímica se ha acumulado sobre el cobayo Cavia porcellus, haciéndose aparente que posee un elevado número de anomalías en sus ejes endócrinos y vías metabólicas. De destacar, el cobayo presenta relativa resistencia a glucocorticoides, manteniendo altos niveles de cortisol circulante así como poca afinidad al receptor de glucocorticoides. También presenta moléculas de insulina y glucagón divergentes, las cuales exhiben reducida biopotencia. C. porcellus también es notable porque no requiere hormona de crecimiento para su desarrollo postnatal. Para Keightley y Fuller (1996), las anomalías en esta especie así como en otros caviomorfos, son el resultado de las nuevas presiones selectivas a las que estuvo sometido el grupo durante el período de relativo aislamiento en el continente sudamericano. Más interesante aún, varias de estas anomalías son compartidas con Callithrix jacchus y otros platirrinos también diversificados en Sudamérica. De hecho, por ser los platirrinos similares a los humanos cortisol-resistente y tener también elevados niveles plasmáticos de cortisol (seis a quince veces mayor), han sido estudiados como modelo para la resistencia al cortisol, aunque tanto en estos como en C. porcellus, y a diferencia de humanos, esta condición representa el estado normal. Por otra parte, resulta interesante que los mecanismos subyacentes a los altos niveles plasmáticos difieren en ambos grupos de mamíferos neotropicales. Mientras que en C. porcellus los mecanismos son muy similares a los de otros roedores como Mus y la única excepción es la reducción en la afinidad del receptor, en platirrinos las diferencias son más numerosas y afectan también los niveles plasmáticos de otros esteroides incluyendo aldosterona, estradiol, progesterona y testosterona (Keightley y Fuller, 1996).

Retomando las anomalías en los ejes que regulan el metabolismo de los carbohidratos y el crecimiento, en *C. porcellus* estas son numerosas, hipotetizándose que una primera modificación haya desencadenado una casacada de alteraciones compensatorias. A modo de punteo (siguiendo a Keightley y Fuller, 1996), y como se había adelantado, *C. porcellus* es una especie insulino-resistente donde la hormona presenta múltiples sustituciones y una biopotencia menor al 5% de las restantes insulinas de mamíferos, mientras que muestra una mayor actividad mitogénica y promotora del crecimiento. Al respecto, Opazo y colaboradores (2005) evaluaron los patrones de evolución molecular adaptativa en el gen de insulina de 21 especies de caviomorfos pertenecientes a las superfamilias Cavioidea, Chinchilloidea y Octodontoidea, encontrando heterogeneidad en las tasas de sustitución no-sinónima:sinónima en concordancia con la heterogeneidad de las propiedades fisicoquímicas y de actividad biológica de las insulinas de los diferentes linajes. Una discusión actualizada sobre la evolución del gen de insulina en caviomorfos ha sido presentada por MacManes y colaboradores (2015). Con respecto a otras hormonas del eje gastro-entero-pancreático, el glucagón muestra modificaciones en su región C-terminal de unión al receptor y también un decrecimiento en bioactividad.

Referente al eje de la hormona de crecimiento y factores de crecimiento similares a insulina (IGF1 e IGF2), las características más sobresalientes son la independencia del crecimiento

y desarrollo postnatal de la hormona de crecimiento, la cual parece cumplir funciones principalmente en la regulación del metabolismo mientras que, por su parte, la acción del IGF1 se sabe independiente de la función hipofisiaria. Estos factores de crecimiento tendrían el papel principal en el crecimiento y desarrollo del individuo, y de hecho el IGF2 no desciende sus niveles postnatales sino que mantiene plenamente su importancia en el adulto. Muchas de estas alteraciones, que involucran un papel principal de la insulina y de ambos IGFs en el control del crecimiento, se repiten en alteraciones patológicas en humanos como cáncer (Belfiore, 2007). Una vez más, varias de las modificaciones del metabolismo de los carbohidratos y del crecimiento se reiteran en los primates del Nuevo Mundo como *Callithrix*, y de hecho en el pasado se plantearon algunas hipótesis que consideraban esos cambios como una respuesta adaptativa paralela a los niveles de zinc del suelo sudamericano con los que estos mamíferos se habrían encontrado en el proceso de colonización (Seino *et al.*, 1987).

De particular interés resulta el estudio del rol que la selección positiva ha tenido en estas alteraciones de los ejes endócrinos, seguramente involucrados en los peculiares patrones de desarrollo y morfológicos de caviomorfos y platirrinos. Una aproximación frecuente para analizar el cambio evolutivo a nivel molecular es el uso de modelos de sustitución de codones en un contexto filogenético (Yang, 1998; Yang y Bielawski, 2000). La identificación de selección positiva en evolución molecular ha sido acompañada del desarrollo de este tipo de modelos, donde se estiman las sustituciones no-sinónimas (dN) y sinónimas (dS) ocurridas. El cociente  $\omega$  dN a dS se considera una medida de la presión selectiva, de modo que  $\omega < 1$  (dN es menor que dS) indica que la selección purificadora está actuando para reducir la fijación de mutaciones deletéreas y  $\omega = 1$  (dN = dS) indica que los cambios son neutrales. Un valor de  $\omega > 1$ , es decir dN mayor que dS, indica la fijación de mutaciones ventajosas y se infiere selección positiva. Como se mencionó antes, esta aproximación fue utilizada por Opazo y colaboradores (2005) para el análisis de la evolución del gen de insulina en caviomorfos. En este trabajo los investigadores obtuvieron las secuencias codificantes del gen para 21 taxa mediante amplificación por PCR y secuenciación Sanger. En los años siguientes se produjo el desarrollo explosivo de las tecnologías de secuenciación masiva que permitieron que hoy existan disponibles en el dominio público numerosos recursos genómicos y transcriptómicos de especies de interés. El acceso a los mismos permite el análisis de patrones de evolución molecular ya no para un único gen, sino para una colección mayor de genes, incluyendo el repertorio entero de genes en un organismo de interés (análisis de selección positiva de tipo genome-wide scan). Existen varios ejemplos recientes de trabajos donde se analizan patrones de selección en múltiples genes que incluyen platirrinos o histricognatos (Worley et al., 2014; Harris et al., 2014; Davies et al., 2015; Wallis, 2015), varios de ellos impulsados por la aparición de un genoma de referencia para C. jacchus (el genoma de referencia de C. porcellus es anterior). De especial relevancia es el trabajo de Wallis (2015) donde analiza con foco en platirrinos los patrones coevolutivos de la insulina, los factores de crecimiento relacionados a insulina, y sus receptores. Las ventajas y perspectivas de la incorporación de datos y análisis a escala genómica son discutidas en Lessa y colaboradores (2014) y MacManes y colaboradores (2015). Sin embargo, cabe señalar que estos estudios tienen sus limitaciones. Además de los evidentes requerimientos bioinformáticos, en general abarcan un muestreo taxonómico poco extenso (*e.g.* comparado con el presentado por Opazo *et al.*, 2005) lo que puede conducir a inferencias erróneas y, en general carecen de una revisión exhaustiva de los grupos de ortólogos y alineamientos que son base de los análisis subsiguientes, lo cual puede conducir a falsos positivos o pérdida de potencia de las pruebas.

A la fecha (febrero 2016) existen disponibles en las bases de datos públicas genomas de múltiples mamíferos, incluyendo genomas de referencia para *C. porcellus* y *C. jacchus*, así como transcriptomas de los platirrinos *Saimiri boliviensis* y *Aotus nancymaae*, y de los caviomorfos *Octodon degus*, *Ctenomys sociabilis* y *Chinchilla lanigera*. Aún cuando no constituyen un muestreo representativo de todos los linajes de Platyrrhini y Caviomorpha, permiten una primera aproximación al estudio de los patrones de evolución molecular de los genes de los ejes endócrinos mencionados. Como hipótesis de trabajo se considera que varias de las anomalías fisiológicas en estos mamíferos neotropicales pueden asociarse a cambios en las fuerzas evolutivas que dejaron su huella a nivel molecular. Específicamente, se plantea que los cambios ocurridos en la insulina y factor de crecimiento podrían estar acompañados por cambios asociados en sus receptores y otros genes cercanos en la vía de señalización, y que estos cambios son factibles de ser detectados con los recursos genómicos/transcriptómicos y modelos de codones hoy disponibles.

### 1.2 Objetivos generales y específicos

### **Objetivos generales**

- Describir los patrones de evolución molecular de genes del eje hormona de crecimiento / factor de crecimiento relacionado a la insulina en dos grupos de mamíferos neotropicales (Caviomorpha y Platyrrhini).
- 2. Comparar los patrones obtenidos para caviomorfos y platirrinos y detectar procesos de evolución paralela.
- 3. Contextualizar resultados previos específicos para la insulina y la familia génica del receptor de insulina en el marco global de las vías de señalización que integran.
- 4. Contribuir al conocimiento de los procesos que acompañaron la diversificación de caviomorfos y platirrinos en la Región Neotropical.

### **Objetivos** específicos

- 1. Seleccionar la lista de genes y la lista de especies a estudiar en base a bibliografía consultada y a la disponibilidad de genomas y transcriptomas en las bases de datos públicas.
- 2. Obtener grupos de ortólogos en mamíferos para los genes y linajes de interés.
- 3. Para cada grupo de ortólogos, obtener un alineamiento afinado manualmente de las regiones codificantes.
- 4. Estimar mediante diferentes métodos que emplean modelos de codones en un contexto filogenético, las tasas de sustitución no-sinónima:sinónima como indicadoras del régimen selectivo al que han estado sometidos estos genes.
- 5. Comparar las tasas de sustitución no-sinónima:sinónima obtenidas entre los diferentes genes y entre las diferentes ramas y clados de interés.
- 6. Evaluar, para algunos genes, la correspondencia de los resultados aquí obtenidos con los de estudios previamente publicados (principalmente Opazo *et al.*, 2005; Wallis, 2015).

### 1.3 Estructura de la tesis

El presente trabajo tiene una estructura dividida en Introducción, Metodología, Resultados, Discusión y Conclusiones. Sin embargo, distintos aspectos de la metodología y de los resultados obtenidos se analizan y discuten a lo largo de los diferentes capítulos. Se trabajó con 18 genes y solo se discuten en profundidad algunos de ellos. El mayor esfuerzo se realizó en la discusión de los resultados para Caviomorpha, dado que existe bibliografía abundante en el caso de Platyrrhini. Esta tesis supone una primera aproximación al estudio de la hipótesis planteada, por lo que funciona como una herramienta que permite diseñar la propuesta de estudio definitiva (*i.e.* conjunto final de linajes y genes de interés, metodología refinada).

### 2 Metodología

Esta tesis utiliza la estrategia general de una búsqueda de patrones moleculares de evolución en secuencias codificantes en un contexto filogenético: obtención de secuencias ortólogas y una topología de referencia; alineamiento respetando el marco de lectura; uso de modelos de codones para la estimación de tasas de sustitución no-sinónima:sinónima; análisis de los resultados. Las secuencias se obtuvieron de repositorios públicos con genomas y transcriptomas disponibles. Si bien podría analizarse un alto número de genes en una aproximación más automática de tipo *genome-wide*, esto tiene como debilidad que errores como los presentes en las predicciones génicas, cambios en el marco de lectura o selección de diferentes isoformas, se acumulan en los alineamientos alterando los resultados de los análisis posteriores. Es así que en este estudio se optó por trabajar con un número reducido de genes (18), seleccionados en base a su asociación funcional, lo que permitió la edición manual de la construcción de cada grupo de ortólogos primero, y de los alineamientos después.

### 2.1 Determinación de los genes y linajes a analizar

### 2.1.1 Genes de interés

Una búsqueda preliminar en bases de datos como KEGG (Kanehisa *et al.*, 2015) y Reactome (Croft *et al.*, 2014) revela más de 250 genes vinculados a los ejes hipotalámico-hipofisiario-adrenal, hormona de crecimiento / factor de crecimiento relacionado a la insulina y gastro-entero-pancreático (a modo de referencia ver el esquema KEGG de la vía de señalización de la insulina en la Figura B-1). Sin embargo, en este estudio se propone realizar una edición manual de los grupos de ortólogos y alineamientos, por lo que se decidió acotar la lista a 18 genes de interés primario. Estos se seleccionaron en base a antecedentes bibliográficos, pertenencia a familias multigénicas y posición en las vías de señalización (se priorizaron hormonas y receptores, no así genes involucrados en la transducción de señales).

### 2.1.2 Representación taxonómica

El estudio de los patrones de evolución molecular por métodos filogenéticos requiere como primer paso definir los *taxa* a ser incluidos en el estudio y la correspondiente topología de referencia. Para ello se consideraron los *taxa* del superorden Euarchontoglires con genomas y/o transcriptomas disponibles en respositorios públicos como Ensembl (Cunningham *et al.*, 2014, http://www.ensembl.org/, último acceso 9 de marzo de 2016) y NCBI (Benson *et al.*, 2012, http://www.ncbi.nlm.nih.gov/, último acceso 9 de marzo de 2016). Para caviomorfos y platirrinos se trabajó con todos los *taxa* con secuencias disponibles (en su mayoría transcriptomas), mientras que para los restantes linajes se eligieron aquellos que presentaron un buen compromiso entre la calidad de las secuencias genómicas/transcriptómicas disponibles y su ubicación en la topología de referencia. La inclusión de al menos 17 ramas terminales, especialmente las que conducen a una reducción de las ramas largas del árbol, aumenta la potencia de los análisis de modelos de codones (Anisimova *et al.*, 2001; Kosakovsky Pond y Frost, 2005c; Gharib y Robinson-Rechavi, 2013, pero ver también Smith *et al.*, 2015). Como grupo externo se tomaron tres representantes del super orden Laurasiatheria: un cetartiodáctilo y dos representantes del clado Zooamata consistentes en un perisodáctilo y un carnívoro.

La Figura 2.1 muestra la filogenia de referencia utilizada, indicándose los géneros y especies priorizados así como grupos taxonómicos de orden superior. La topología es congruente con las hipótesis prevalentes sobre los patrones de diversificación de los placentados (*e.g.* Madsen *et al.*, 2001; Murphy *et al.*, 2001; Springer *et al.*, 2004; Meredith *et al.*, 2011; Song *et al.*, 2012). El principal conflicto se plantea a nivel de la raíz del orden Rodentia, donde ni los análisis morfológicos ni los moleculares han sido esclarecedores. Los datos moleculares apoyan la existencia de tres clados principales de roedores: *mouse-related, squirrel-related* y Ctenohystrica (cobayo y roedores cercanos), pero sus interrelaciones han permanecido controversiales (Huchon *et al.*, 2002, 2007; Montgelard *et al.*, 2008; Blanga-Kanfi *et al.*, 2009). En este trabajo se siguió a Churakov y colaboradores (2010), quienes a partir del estudio de cambios genómicos raros libres de homoplasia (*e.g.* retroposones e *indels*), encuentran evidencia para una asociación temprana entre el clado *mouse-related* y Ctenohystrica, siendo el clado *squirrel-related* su grupo hermano. La divergencia inicial entre el clado *squirrel-related* y el resto de los roedores fue primero propuesta por Waddell y colaboradores (2001) y luego apoyada en trabajos más recientes (Churakov *et al.*, 2010; Fabre *et al.*, 2012).

Las relaciones dentro de los clados de interés, Caviomorpha y Platyrrhini, siguen también las propuestas actuales de diversificación en ambos grupos de mamíferos neotropicales. En Caviomorpha existen cuatro superfamilias ampliamente reconocidas: Cavioidea, Erethizontoidea, Chinchilloidea y Octodontoidea. Mientras existe evidencia de que Octodontoidea y Chinchilloidea constituyen un grupo monofilético con exclusión de Cavioidea, Erethizontoidea ha alternado entre



Figura 2.1: Topología de referencia. El cladograma indica el muestreo taxonómico utilizado en el estudio, así como la hipótesis filogenética empleada en los análisis de modelos de codones.

una posición hermana a Octodontoidea + Chinchilloidea (y Cavioidea como linaje hermano de los restantes caviomorfos; Huchon y Douzery, 2001) y una propuesta más reciente donde constituye el grupo hermano de Cavioidea (Huchon *et al.*, 2007; Blanga-Kanfi *et al.*, 2009; Fabre *et al.*, 2012; Voloch *et al.*, 2013; Upham y Patterson, 2015). De todas formas, no existe a la fecha ningún representante de Erethizontoidea con abundantes secuencias en los repositorios como Ensembl o GenBank, por lo que la superfamilia no está incluida en este trabajo.

En Platyrrhini se reconocen tres familias monofiléticas (Cebidae, Atelidae y Pitheciidae), apareciendo Atelidae y Cebidae como grupos hermanos, aunque las relaciones entre familias son menos consistentes que en el caso de los caviomorfos (*e.g.* Opazo *et al.*, 2006; Wildman *et al.*, 2009; Springer *et al.*, 2012; Perez *et al.*, 2013; Aristide *et al.*, 2015; Jameson Kiesling *et al.*, 2015). En especial, no ha existido consenso para las afinidades filogenéticas de *Aotus* (Rosenberger y Tejedor, 2013). En la mayoría de los trabajos aparece como grupo hermano de Cebinae (Opazo *et al.*, 2006; Wildman *et al.*, 2009; Fabre *et al.*, 2009), pero también aparece como cercano a Callitrichinae (Perelman *et al.*, 2011; Springer *et al.*, 2012; Aristide *et al.*, 2015; Jameson Kiesling *et al.*, 2015), como un linaje hermano a Cebidae (Pérez y Pol, 2012) o incluso como un miembro de Pitheciinae (Rosenberger y Tejedor, 2013), siempre con bajo apoyo de nodo

en las filogenias moleculares. Con respecto al presente trabajo, los datos de secuencias genómicas y transcriptómicas disponibles a la fecha restringieron el muestreo taxonómico de platirrinos a representantes de la familia Cebidae: *Saimiri, Callithrix y Aotus*. Para la topología de referencia se consideró a *Aotus* como un linaje cercano a Cebinae (ver Figura 2.1; Wildman *et al.*, 2009, pero ver discusión en Jameson Kiesling *et al.*, 2015).

Finalmente, se generó en TimeTree (http://timetree.org/) la filogenia calibrada al tiempo de divergencia para los *taxa* elegidos (Figura B-2). TimeTree es una base de datos que guarda tanto las relaciones de parentezco entre todos los seres vivos como sus tiempos de divergencia, por lo que permite recuperar rápidamente, para un grupo de organismos, ambos componentes de su historia evolutiva en la forma de un cladograma escalado al tiempo geológico (Hedges y Kumar, 2009; Hedges *et al.*, 2015). Este *timetree* es presentado solo a modo de guía, dado que su topología es algo diferente a la seleccionada como referencia en este trabajo. En particular presenta a los histricognatos como el grupo hermano de los restantes roedores, lo cual podría afectar en parte las dataciones.

### 2.2 Búsqueda de ortólogos y obtención de secuencias

El identificador de cada gen en *Homo sapiens* según las bases de datos Gene (Maglott *et al.*, 2005), RefSeq (Pruitt *et al.*, 2014) y/o Ensembl (Cunningham *et al.*, 2014) se utilizó como punto de partida para extraer el grupo de ortólogos predefinido en Ensembl a través de BioMart (Smedley *et al.*, 2015). También se utilizaron y compararon los grupos de ortólogos disponibles en OMAbrowser (Altenhoff *et al.*, 2015, http://omabrowser.org/), un método (*Orthologous MAtrix*, Roth *et al.*, 2008; Altenhoff *et al.*, 2013) y base de datos especializada en la inferencia de ortólogos entre genomas completos. Para cada gen se asignó un transcripto de referencia en humano, ya sea el transcripto codificante más largo y/o con anotación en RefSeq y Havana (Harrow *et al.*, 2014). BLASTP (Altschul *et al.*, 1997) fue utilizado para combinar y refinar estos grupos de ortólogos con secuencias obtenidas de GenBank (Benson *et al.*, 2012). En aquellos casos en los que no se pudo recuperar un ortólogo para uno de los linajes, se intentó sustituir por su ortólogo en otra especie cercana. Por ejemplo, si para un gen no se encontró el ortólogo en *Otolemur garnettii*, este se sustituyó con el ortólogo en otro primate del suborden Strepsirrhini como *Microcebus murinus*, para el cual también existe genoma de referencia disponible.

En el caso de *Ctenomys sociabilis*, existen secuencias transcriptómicas en NCBI pero al momento (marzo 2016) no forman parte de las bases de datos nucleotídicas anotadas. Por este motivo, todos los transcriptos de *C. sociabilis* se seleccionaron en NCBI (Taxonomy=*Ctenomys sociabilis* **AND** Nucleotide) y se guardaron localmente para construir una base de datos de BLAST de tipo nucleotídico. De las 76.493 secuencias, 76.453 corresponden al BioProject PRJNA157895

Último acceso a estas bases de datos: 09/03/2016 (MacManes y Lacey, 2012). Las secuencias aminoacídicas de los genes de interés en humano fueron utilizadas como *query* para una búsqueda de similitud por TBLASTN de los transcriptos en *C. sociabilis*. Observando el *bit score, e-value* y el porcentaje de cobertura del *query*, se determinó para cada gen si su ortólogo estaba presente en la base de datos local, en cuyo caso se procedió a utilizar las coordenadas de los alineamientos para extraer la región con marco abierto de lectura en cada transcripto. Estos análisis se realizaron con la *suite* BLAST+ del NCBI C++ Toolkit (Camacho *et al.*, 2009) y permitieron recuperar las secuencias codificantes de 11 de los 18 genes de interés.

### 2.3 Alineamiento

Para cada grupo de ortólogos las homologías posicionales de las secuencias aminoacídicas se establecieron con MUSCLE v3.8.31 (Edgar, 2004). Con la ayuda del visualizador Jalview (Waterhouse et al., 2009) se editaron los alineamientos eliminándose aquellos taxa con muchos sitios ambiguos o altamente divergentes (indicadores de un transcripto incorrectamente ensamblado). En otros casos se detectaron ortólogos donde la alta divergencia quedaba restringida a un tramo de la secuencia y, cuando esto ocurrió en uno de los extremos, el mismo fue recortado (junto con la secuencia nucleotídica correspondiente). Mientras que estos patrones localizados pueden deberse a la inclusión de isoformas alternativas para alguno de los ortólogos, en general se deben a errores en los procesos de ensamblado y/o anotación. En instancias anteriores hemos detectado estos artefactos en transcriptos ensamblados (debido a un error de tipo indel) y en modelos génicos construidos a partir de secuencias genómicas, en este caso debido a dificultades en la anotación del primer exón 5' o a errores en la identificación exacta de los límites de alguno de los exones. Cuando el cambio en la secuencia aminoacídica debida al cambio en el marco de lectura está espacialmente acotado, se observa que en general no interfiere con la correcta inferencia de ortología y en el alineamiento se visualiza fácilmente como un ortólogo quimérico que puede ser editado y corregido, manteniéndose la información aportada por el resto de la secuencia.

Los alineamientos aminoacídicos revisados y editados se utilizaron a su vez para guiar los alineamientos de las secuencias nucleotídicas con PAL2NAL v14 (Suyama *et al.*, 2006), que permite mantener la correcta estructura en codones. Con el objetivo de hacer a los alineamientos nucleotídicos resultantes más apropiados para los análisis filogenéticos posteriores, se filtraron primero con Gblocks v0.91b (Castresana, 2000; Talavera y Castresana, 2007) para eliminar aquellos sitios con alto porcentaje de datos faltantes y luego con BMGE v1.12 (Criscuolo y Gribaldo, 2010), que detecta sitios con valores muy altos de divergencia. Ambas herramientas se utilizaron en modo de codones para mantener la integridad de las secuencias codificantes. En el caso de Gblocks, se eligieron opciones menos restrictivas, permitiendo la existencia de bloques

finales con algunos *gaps* presentes. Por otra parte, BMGE (*Block Mapping and Gathering with Entropy*; Criscuolo y Gribaldo, 2010) filtra sitios en un alineamiento múltiple basándose en un puntaje h, que es un valor de entropía pesado por una matriz de sustitución (matriz BLOSUM en el caso de alineamientos de codones), como forma de evaluar la variabilidad observada con respecto a la esperada biológicamente. En este estudio BMGE se utilizó con las opciones BLOSUM62 como matriz de sustitución y 0,5 como valor máximo permitido de h.

Para los alineamientos finales, los modelos de sustitución nucleotídica se obtuvieron con ModelGenerator v0.8.5 (Keane *et al.*, 2006) con seis categorías para la distribución *gamma* discreta y se consideró el valor de AICc (*corrected Akaike Information Criterion*) como criterio para definir el que mejor ajustaba a los datos en cada caso.

### 2.4 Análisis composicional

Las estimaciones de tasas de sustitución no-sinónima y sinónima dependen de factores composicionales como la frecuencia de las bases nucleotídicas, tasas de transición y transversión y uso de codones. Con el fin de revisar las características composicionales de los genes y *taxa* estudiados, el contenido en nucleótidos y el uso de codones y aminoácidos fueron calculados en el entorno R (R Core Team, 2015), con los paquetes **seqinR** v3.1 (Chrarif y Lobry, 2007) y **ca** v0.64 (Nenadic y Greenacre, 2007).

# 2.5 Estimación de tasas de sustitución no-sinónima:sinónima $(\omega = dN/dS)$

Para cada gen, se estudiaron las huellas moleculares del patrón de selección por medio de modelos de sustitución de codones en un contexto filogenético. Un grupo de métodos son parte del programa PAML v4.8 (Yang, 2007), y los restantes consisten en rutinas implementadas en HyPhy v2.2.6 (*Hypothesis testing using Phylogenies*; Kosakovsky Pond *et al.*, 2005). Todos los métodos se basan en la estimación del cociente  $\omega$ , permitiendo su variación en el tiempo (modelos de linajes, *branch models*) y en el tiempo y a lo largo de las secuencias codificantes (modelos de linajes y sitios, *branch-site models*). En algunos casos se asume una tasa de sustitución sinónima constante, mientras que en otros métodos se permite su variación (lo cual tiene mayor sentido biológico). Algunos de los análisis (como los implementados en PAML) requieren de la selección de ramas blanco (*foreground branches*), que son aquellas en las que se estima que pudo haber ocurrido un cambio en el régimen selectivo, mientras que otros no dependen de una selección *a priori* de ramas. Cuando hubo que seleccionar linajes blanco, la selección consistió en: la

rama ancestral de Platyrrhini, el clado Platyrrhini, la rama ancestral de Caviomorpha y el clado Caviomorpha (Figura 2.2 y Figura 2.3), que son las de interés en el presente trabajo. El análisis de varios genes y de varios modelos por gen (incluyendo diferentes conjuntos de ramas seleccionadas por gen), requiere indudablemente de una corrección final por múltiples pruebas de hipótesis (Anisimova y Yang, 2007). En general, una vez aplicada la corrección, no se recuperan valores de  $\omega$  mayor a 1 estadísticamente significativos (para los valores de  $\alpha$  comunmente usados), por lo que en los artículos se suele discutir el resultado "previo a la corrección". Por este motivo, en el presente trabajo no se efectuaron correcciones por pruebas de hipótesis múltiples (salvo las correcciones ya implementadas en algunas de las rutinas utilizadas), sino que los resultados se consideran principalmente en un contexto descriptivo que permita identificar resultados que ameriten un análisis más profundo (siguiendo la propuesta original de uso de *pvalues* de Ronald Fisher en 1925; Nuzzo, 2014). A continuación se detallan los modelos implementados y en la Tabla 2.1 se indican las principales características de cada uno.

### 2.5.1 CODEML: modelos de rama

Los primeros modelos de sustitución de codones (Goldman y Yang, 1994; Muse y Gaut, 1994) generaron el marco para el estudio de la evolución de secuencias por medio de tasas de sustitución sinónima (dS) y no-sinónima (dN). El modelo puede ser fácilmente extendido permitiendo la aplicación de una prueba de razón de verosimilitud (LRT, del inglés Likelihood Ratio Test) para la comparación de estas tasas (Muse y Gaut, 1994) o su cociente ( $\omega = dN/dS$ ; Yang, 1998) entre diferentes linajes. Así, Yang en su trabajo seminal de 1998 desarrolló y aplicó estos modelos de linajes en un juego de datos de lisozima de primates, corroborando los hallazgos obtenidos por Messier y Stewart (1997). Para ello estudió el comportamiento de las ramas ancestrales que conducían a homínidos y colobinos (ramas blanco) en comparación con las restantes ramas del árbol (background branches).

Para cada gen del presente trabajo, se usó el programa CODEML del paquete PAML v4.8 (Yang, 2007). Se corrió el modelo nulo con un único valor de  $\omega$  para todos los linajes, y cuatro modelos con tres valores de  $\omega$ :  $\omega_1$  para platirrinos (rama ancestral o clado),  $\omega_2$  para caviomorfos (rama ancestral o clado) y  $\omega_0$  para las ramas *background*. En la Figura 2.2 se muestran las cuatro conformaciones de  $\omega_1$  y  $\omega_2$  utilizadas. Para evaluar si un modelo alternativo de tres  $\omega$  ajusta a los datos mejor que el modelo nulo de un  $\omega$  se comparó el *LRT* con una distribución  $\chi^2$  con dos grados de libertad. En aquellos casos en los que el modelo alternativo resultó significativo para un  $\alpha = 0,05$  y algún valor de  $\omega$  resultó mayor que 1, la prueba de selección positiva se condujo entre el modelo alternativo de tres  $\omega$  sin restricciones y otro donde se fija  $\omega = 1$  para el linaje en cuestión (platirrinos o caviomorfos). En este caso el *LRT* resultante se compara con una distribución  $\chi^2$  con un grado de libertad. Todas las estimaciones por máxima verosimilitud de los modelos se

Tabla 2.1: I	Modelos de sustitució	ón de codones em	pleados en este	trabajo.				
		Uso de ramas	dS variable	Permite $\omega > 1$	Resultado	Resultado	Corrección	
	Tipo de modelo	blanco	entre sitios	en background	principal	secundario	múltiple hipótesis	Referencia principal
PAML: COI	DEML							
Branch	ramas	SI.	no	no	ε	no	no	Yang, 1998
Branch-site	ramas y sitios	S1.	no	no	8	<b>BEB</b> sitios	no	Yang y Nielsen, 2002
HyPhy (Dat	amonkey)							
BSREL	ramas y sitios	no	no	Si	ŵ	no	Si	Kosakovsky Pond et al., 2011
aBSREL	ramas y sitios	no	no	Si.	ε	no	Si	Smith <i>et al.</i> , 2015
BUSTED	ramas y sitios	Si.	no	Si.	ε	no	Si	Murrell et al., 2015
MEME	sitios	no	Si.	NA	ε	EBF sitios	si*	Murrell et al., 2012
RELAX	ramas	S1.	no	S1.	k	no	Si.	Wertheim et al., 2015
NOTA: Se indican q-values son calcu	algunas de sus principale nados pero solo se consid	s características como eran los <i>p-values</i> .	) el tipo de modelo	, el requerimiento de in	dicar ramas blan	co <i>a priori</i> o la imp	lementación de correcciór	ı por múltiple hipótesis. NA: no aplica; *:

valu	OTA
es se	Se
on c	indi
alcu	can
ladc	algı
s pe	ınas
ro s	de s
olos	d sn:
se cc	princ
onsic	ipal
lera	es c:
n los	aract
5 <i>p</i> -v	terís
alue	ticas
S.	cor
	no e
	i tip
	o de
	mo
	delc
	), el
	requ
	ıerin
	nien
	to d
	e inc
	licai
	ran
	nas t
	plane
	20 a
	pric
	ori o
	la i
	mple
	eme
	ntac
	ión (
	de co
	orre
	cció
	n po
	r mu
	íltip
	le h
	ipótu
	esis.
	NA
	: no
	apl
	ica;

realizaron partiendo de múltiples valores iniciales (valores de  $\omega = 0,5$ ;  $\omega = 1$ ;  $\omega = 2$ ) de modo de corroborar la convergencia a un único máximo global.



**Figura 2.2:** Conformaciones de  $\omega_1$  y  $\omega_2$  utilizadas en los modelos de ramas de CODEML. Para cada conformación, A a D, las ramas y clados blanco se indican en la topología de referencia como  $\omega_1$  y  $\omega_2$  y están coloreadas en rojo. Las restantes ramas en negro constituyen el *background* con  $\omega_0$ .

### 2.5.2 CODEML: modelos de rama y sitio

Otra extensión temprana de los modelos consisitió en la implementación de distribuciones estadísticas que permiten la variación de  $\omega$  entre sitios (*site-models*). Se asumen varias clases de sitios en la proteína con diferentes  $\omega$  y se prueba si existen sitios con  $\omega > 1$  mediante un *LRT* entre un modelo que no permite tales sitios y otro general que sí. En caso de detectarse la existencia de una categoría de sitios con  $\omega > 1$ , se utilizan métodos bayesianos para identificar los posibles codones bajo selección positiva (Nielsen y Yang, 1998; Yang *et al.*, 2000). Estos métodos son conservadores, porque requieren que la tasa de sustitución no-sinónima sea mayor que la sinónima, cosa que no ocurriría en casos de selección balanceadora, por ejemplo. Además, la detección de selección en algunos linajes o en algunos sitios requiere que el valor de  $\omega$  promediado entre todos los sitios o entre todos los linajes sea mayor a 1. Asumir constancia en las presiones selectivas

entre sitios resultó mucho más grave que asumir constancia entre los linajes y el uso de los modelos de sitios fue exitos en detectar selección positiva aún en casos donde el valor global de  $\omega$  era mucho menor que 1 (Yang y Bielawski, 2000; Anisimova et al., 2001). Se vislumbró entonces la potencia que podrían tener modelos que permitieran variar el cociente  $\omega$  tanto entre sitios como entre linajes y en el año 2002, Yang y Nielsen extienden los modelos previos permitiendo la variación simultánea del cociente  $\omega$  entre sitios y entre linajes. Los nuevos modelos (Yang y Nielsen, 2002; Zhang et al., 2005) permiten identificar selección positiva actuando en algunos linajes selecccionados a priori y operando en unos pocos sitios de la proteína. Cuatro clases de sitios son definidos en este modelo, que se caracterizan por diferentes valores de  $\omega$  estimados. La primera clase contiene a aquellos codones que están conservados en todos los linajes y tiene valores de  $\omega_0$  entre 0 y 1. La segunda clase contiene sitios que son neutrales y se fija el valor de  $\omega$  a 1. En las clases 2a y 2b restantes, los sitios en los linajes de referencia tienen  $\omega_0$  o  $\omega = 1$ , pero en los linajes blancos tienen  $\omega_2$  que podría ser mayor que 1. Se asume que cuando ocurre selección positiva en el linaje blanco, es igualmente probable que afecte a sitios que pertenecen a la clase 0 o 1 en los linajes de referencia. El modelo estima cuatro parámetros en la distribución de  $\omega$ :  $p_0$ ,  $p_1$ ,  $\omega_0$  y  $\omega_2$ . La prueba de selección positiva en los linajes seleccionados se construye comparando este modelo con un modelo nulo donde se fija  $\omega_2 = 1$ . En la práctica en el LRT se emplea una distribución  $\chi^2$  con un grado de libertad. Si la prueba de selección positiva rechaza la hipótesis nula, el modelo alternativo se usa para identificar los sitios bajo selección positiva en los linajes blanco. Con un procedimiento Bayes Empirical Bayes (BEB; Yang et al., 2005) se estima entonces para cada sitio la probabilidad posterior de pertenecer a cada clase, lo que permite identificar aquellos sitios aminoacídicos bajo selección positiva en las ramas blanco.

En este trabajo se usó el programa CODEML del paquete PAML v4.8 (Yang, 2007) para implementar estos modelos de linajes y sitios recién mencionados (opciones model=2 y NSites=2). Sobre cada alineamiento el modelo alternativo se corrió para cuatro selecciones de linajes blanco: rama ancestral de platirrinos, clado platirrinos, rama ancestral de caviomorfos y clado caviomorfos (Figura 2.3 A). Para el modelo nulo se fijó  $\omega_2 = 1$  y el *LRT* se comparó con una distribución  $\chi^2$  con un grado de libertad. En aquellos casos en los que se detectó selección positiva en el linaje blanco para un  $\alpha = 0,05$ , las probabilidades posteriores de pertenencia a la clase  $\omega_2$  estimada por BEB se usaron para identificar los posibles sitios bajo selección (*PP* > 0,95). Todas las estimaciones por máxima verosimilitud de los modelos alternativos se realizaron partiendo de múltiples valores iniciales (valores de  $\omega = 0,5$ ;  $\omega = 1$ ;  $\omega = 2$ ) de modo de corroborar la convergencia a un único máximo global.



Figura 2.3: Conformaciones de ramas blanco para los análisis de tipo ramas y sitios implementados en CODEML (A) y BUSTED (A y B) y para RELAX (C). Las ramas y clados en la categoría blanco se indican en rojo y el *background* en negro (A y B). En el caso de RELAX (C), las ramas blanco se indican en rojo (Platyrrhini), las ramas de referencia en azul (restantes euarcontos) y en negro se indican los glires que se consideraron como ramas no clasificadas.

### 2.6 HyPhy: uso de modelos menos restrictivos

El paquete HyPhy (Kosakovsky Pond et al., 2005) y su correspondiente interfaz web Datamonkey (Kosakovsky Pond y Frost, 2005b; Delport et al., 2010, http://www.datamonkey. org) implementan múltiples métodos aproximados o por verosimilitud para la detección de selección natural operando en sitios y/o linajes de un alineamiento. Algunos de los nuevos modelos de sitio como SLAC, FEL y REL (Kosakovsky Pond y Frost, 2005c) no asumen una tasa de sustitución sinónima dS constante entre sitios. En los modelos previos, como los implementados en PAML, la tasa de sustitución sinónima entre sitios se asume como constante, lo cual implica que la variación del cociente  $\omega$  entre sitios es resultado único de la variación correspondiente en las tasas no-sinónimas. Kosakovsky Pond y colaboradores incorporan en los modelos tasas de sustitución sinónima entre sitios variables y demuestran que la no consideración de esta variación puede conducir a errores en la estimación de  $\omega$  (Kosakovsky Pond y Muse, 2005; Kosakovsky Pond y Frost, 2005c). Otros modelos, como BSREL (Kosakovsky Pond et al., 2011), consituyen una mejora con respecto a limitaciones importantes de los modelos de sitios y ramas antes descriptos: no requieren de una selección de ramas blanco a priori y permiten selección positiva en los linajes de fondo (en contraste con las cuatro clases de sitios posibles en Yang y Nielsen, 2002). Si bien la elección *a priori* de ramas blanco puede ser de interés cuando existe una hipótesis biológica adecuada (e.g. una rama en la que un virus cambia de hospedero), la deshabilitación de la variación de los valores de  $\omega$  puede tener implicancias graves. Variación en la fuerza de la selección en los linajes blanco y/o de fondo, así como la existencia de selección positiva en el background, constituyen una violación de los supuestos del modelo y se demuestra que produce pérdida de potencia de las pruebas y aumento de la tasa de falsos positivos (Kosakovsky Pond et al., 2011). Algunos de los modelos implementados en HyPhy (a través de la interfaz web Datamonkey) fueron empleados en el presente trabajo y se detallan a continuación.

### 2.6.1 BSREL, aBSREL y BUSTED

Los métodos de efectos aleatorios fueron introducidos en el área por Nielsen y Yang (1998) y luego también empleados por Kosakovsky Pond y colaboradores (Kosakovsky Pond y Frost, 2005a,c). Consisten en ajustar una distribución de tasas de sustitución para todos los sitios y luego inferir la tasa a la que cada sitio individual evoluciona dada esta distribución. La inferencia sitio por sitio se basa en estimaciones de los parámetros de tasa por máxima verosimilitud a partir de los datos y se conoce como *empirical Bayes* (en contraste con una aproximación bayesiana estándard; Nielsen y Yang, 1998). En 2011 Kosakovsky Pond y colaboradores proponen BSREL (del término en inglés *Branch-Site Random Effects Likelihood*), un método que permite inferir a partir de un alineamiento y sin requerimiento de determinar ramas blanco, la presencia de selección positiva episódica para alguno de los linajes. A cada rama se adjudican tres valores de  $\omega$  y se permite que cada sitio adopte cualquiera de esos tres valores de  $\omega$ . Para un árbol con *B* ramas,

el modelo considera  $3^B$  configuraciones diferentes de  $\omega$  por sitio, a diferencia de las solo cuatro configuraciones posibles en el modelo de ramas y sitio original. Pese a la complejidad del modelo, el uso de algoritmos como la extensión del de poda de Felsenstein, permiten que su aplicación sea eficiente (Kosakovsky Pond *et al.*, 2011). Los tres valores de  $\omega$  corresponden a  $\omega_b^- \leq \omega_b^N \leq \omega_b^+$ representando selección purificadora fuerte, débil y selección positiva, respectivamente. Estos valores de  $\omega$  se presentan en proporciones asociadas  $q_b^- + q_b^N + q_b^+ = 1$ . Para el *LRT* se consideran *B* modelos nulos que restringen  $\omega_b^+ = 1$  sucesivamente en cada rama, dejando el resto de las ramas sin restricciones.

Como se indicaba, BSREL es un modelo más complejo pero debido a la estrategia algorítmica, no necesariamente más lento que el modelo de ramas y sitios con solo cuatro configuraciones de  $\omega$ permitidas. Sin embargo cabe preguntarse si toda la complejidad agregada es realmente necesaria. En este sentido, una rama muy corta tendrá escasos cambios y no será posible estimar con precisión diferentes categorías de tasas para diferentes sitios. En una rama más larga, la heterogeneidad de tasa entre sitios será evidente y se podrá inferir un mayor número de categorías de sitios. En 2015 se presenta aBSREL (adaptive BSREL; Smith et al., 2015) en el que los autores explotan este fenómeno y adaptan la complejidad del modelo al juego de datos en estudio, infiriendo el número óptimo de categorías de  $\omega$  para cada rama. La optimización hace uso del criterio de información Akaike corregido por tamaño de muestra (AICc). Así, se aplican modelos de diferente complejidad a las diferentes ramas de la filogenia, se mantiene el desempeño estadístico de los modelos anteriores (como BSREL), pero se resuelve en un orden de magnitud más rápido (se usan 20 a 60% menos parámetros). La prueba de selección positiva es mediante un LRT por rama con un modelo nulo en el que no se permiten valores de  $\omega_k > 1$ . El análisis por aBSREL aplicado a casi nueve mil alineamientos de genes de euteleostomos mostró que el 80% de las ramas pueden ser modeladas adecuadamente con un único cociente  $\omega$  y solo una relativamente pequeña cantidad de ramas claves requieren el modelado para mayor complejidad evolutiva, aunque no se encontraron ramas que requirieran ser modeladas con más de tres cocientes  $\omega$  (Smith *et al.*, 2015).

Si bien métodos como BSREL y aBSREL permiten evaluar la existencia de selección positiva en al menos un linaje del árbol sin división *a priori* de ramas blanco y *background*, en casos donde sí existe una hipótesis biológica adecuada, se puede llevar a cabo una prueba focalizada que permita identificar selección en las ramas de interés. De hecho el clásico modelo de ramas y sitios de Nielsen y Yang demostró que existe un incremento en la potencia de la prueba proveniente de no tener que compartir parámetros entre diferentes partes de la filogenia que de antemano se conoce como sujetas a diferentes presiones selectivas. Basados en BSREL, Kosakovsky Pond y colaboradores (Murrell *et al.*, 2015) presentan BUSTED (del término en inglés *Branch-site Unrestricted Statistical Test for Episodic Diversification*) que incorpora la especificación *a priori* de ramas blanco pero en el contexto de un modelo de selección flexible en el resto del árbol. Para cada partición blanco y *background* se estiman tres valores  $\omega_1 \le \omega_2 \le 1 \le \omega_3$  teniendo cada
categoría un parámetro de peso  $p_c$  asociado ( $\sum_{c=1}^{3} p_c = 1$ ). Así, la selección positiva no tiene que estar restringida a las ramas blanco y  $\omega$  puede variar tanto entre sitios como en el tiempo en ambas particiones. La selección positiva se infiere mediante un *LRT* que usa un modelo nulo donde en las ramas blanco  $\omega_3 = 1$ .

Los alineamientos de los genes de interés del presente trabajo fueron analizados según los modelos BSREL, aBSREL y BUSTED recién descriptos, haciendo uso de la implementación web de estos análisis (http://www.datamonkey.org). En los tres casos se incluyó la topología de referencia para ser usada por los análisis *online*. Para correr BUSTED, las ramas blanco se seleccionaron a través del *widget* interactivo que incorpora la interfaz web. Se consideraron las mismas conformaciones usadas en el modelo de ramas y sitio de CODEML, pero se incorporó una quinta configuración en la que se toma como rama blanco el linaje ancestral a Hystricognathi (Figura 2.3 B).

### 2.6.2 MEME

En forma complementaria a BSREL y extensiones derivadas, MEME se designa como un método para detectar selección a nivel de sitios individuales (Mixed Effects Model of Evolution; Kosakovsky Pond et al., 2011; Murrell et al., 2012). Está basado en la clase de métodos de efectos aleatorios de rama y sitio (Kosakovsky Pond et al., 2011) y permite capturar tanto las huellas de selección positiva sostenida como episódica. En MEME la distribución de  $\omega$  varía de sitio en sitio (el efecto fijo) y también de rama en rama para un sitio dado (el efecto aleatorio). MEME puede inferir selección diversificadora aún cuando esta afecta a solo unas pocas ramas en el sitio individual, situación en la que frecuentemente los métodos de sitio (e.g. Nielsen y Yang, 1998) reportan selección purificadora. Para cada sitio se infieren cuatro parámetros estimados a partir de todas las ramas del árbol:  $\beta^-, \beta^+, q^-$  y  $\alpha$ , cumpliéndose  $\omega = \beta/\alpha, \ \beta^- \le \alpha$  y  $\beta^+$ sin restricciones. Además,  $q^-$  es la probabilidad de  $\beta = \beta^-$ . La estimación de un  $\alpha$  para cada sitio permite incorporar la variación sitio a sitio de las tasas de sustitución sinónima (Kosakovsky Pond y Muse, 2005). Para la posterior prueba de selección, se ajusta un modelo nulo en el que se restringe  $\beta^+ < \alpha$ . Si el LRT indica que un sitio es sujeto de selección diversificadora, se adapta un método de Bayes empírico para explorar en qué ramas habría ocurrido la selección. Valores de EBF (Empirical Bayes Factor) positivos indican que los datos aumentan las probabilidades establecidas en la prior de observar selección en esa rama. Sin embargo, aún para valores grandes de EBF (e.g. EBF > 100) la evidencia en la que se basa esa estimación consiste en un único par rama-sitio observado, por lo que se recomienda considerar la inferencia de selección en las ramas como un dato exploratorio y MEME en sí se define como un método de sitios.

Para cada gen, MEME se corrió en el servidor Datamonkey, a partir del alineamiento, la topología de referencia y el modelo de sustitución nucleotídica que presentó mejor ajuste a los datos. Se cuantificó el número de sitios inferidos con episodios de selección positiva previo a las correcciones por hipótesis múltiples (es decir, se consideraron *p*-values < 0, 1 y no los *q*-values que en general fueron igual a 1 o apenas menores). Para estos sitios, se exploró qué linajes presentaban un valor de  $EBF \ge 20$ , observando si estos coincidían con el grupo externo (Laurasiatheria) o interno, y con las ramas y clados de interés o no (Figura 2.1).

### 2.6.3 Relajación de la selección en Platyrrhini

Con o sin evidencia significativa de selección positiva se encontró que los valores de  $\omega$  para las ramas de platirrinos en la filogenia eran en general mayores que el de los linajes restantes del árbol. Un incremento en el valor del cociente  $\omega$  puede responder a la presencia de algunos sitios bajo selección positiva o a una relajación de la selección purificadora (que conduce a un incremento del valor de  $\omega$  aunque este se mantiene < 1). Para cuantificar la contribución de ambas fuerzas en Platyrrhini se utilizó el modelo RELAX del paquete HyPhy (Wertheim et al., 2015). Si bien está basado en el marco metodológico de BSREL (Kosakovsky Pond et al., 2011), en lugar de focalizarse en estimaciones de dN y dS y evaluar la desviación de su cociente del valor esperado bajo neutralidad  $\omega = 1$ , RELAX calcula un parámetro de intensidad de selección  $k \ge 0$ . RELAX prueba si la selección se relaja o intensifica en un subgrupo de ramas de una topología predefinida. En el modelo nulo, la intensidad de selección es restringida a k = 1 en todas las ramas, mientras que en el modelo alternativo se permite que k difiera entre los grupos de ramas seleccionadas y de referencia. Para la prueba se realiza un LRT entre ambos modelos pero también se incluyen cálculos de AICc para evaluar el ajuste de ambos modelos. Cuando la introducción de k produce una mejora significativa en el ajuste del modelo, se concluye que la selección en las ramas de prueba se intensificó (k > 1) o relajó (k < 1) relativo a las ramas de referencia.

Los análisis de RELAX se llevaron a cabo en el servidor Datamonkey (http://test. datamonkey.org/relax) para los genes de interés, excepto GH1 que presenta expansiones génicas independientes en platirrinos y catarrinos (Wallis *et al.*, 2001; Li *et al.*, 2005; González Alvarez *et al.*, 2006; Petronella y Drouin, 2011). Los alineamientos y topologías de referencia fueron reducidos a los linajes de euarcontoglires (*i.e.* se eliminaron los linajes del grupo externo). Usando el *widget* del servidor web se seleccionó sobre la topología el grupo de ramas del clado Platyrrhini como grupo de prueba y el resto de las ramas de Primates más Scandentia como el grupo de referencia (Figura 2.3 C). Los glires se mantuvieron como ramas "no-clasificadas". Esta configuración de ramas permite poner a prueba si exite evidencia de relajación de la selección en Platyrrhini con respecto a los restantes euarcontos.

### 2.7 Filogenia mitocondrial

Como una aproximación a la cuantificación del cambio evolutivo ocurrido en cada rama del árbol de especies, se estimaron los largos de rama a partir de genes mitocondriales del conjunto de mamíferos definido como referencia. Con los mismos métodos antes descriptos, se obtuvieron alineamientos de codones para los 12 genes codificantes mitocondriales ubicados en la cadena pesada del genoma mitocondrial (con exclusión de *ND6* que se encuentra en la hebra liviana y estaría sometido a sesgos composicionales distintos). Con ayuda de los paquetes de R **seqinR** v3.1 (Chrarif y Lobry, 2007) y **apex** v1.0.1 (Jombart *et al.*, 2015), se removieron de cada alineamiento las bases correspondientes a la tercera posición de los codones (y así evitar posible homoplasia por saturación) y se concatenaron estos alineamientos reducidos en un alineamiento final de 24 secuencias con 7.218 sitios homólogos. ModelGenerator se utilizó con seis categorías de la distribución *gamma* discreta y según el criterio AICc se seleccionó GTR + I + G como modelo de sustitución nucleotídica. En PhyML v3.1 (Guindon *et al.*, 2010), con el alineamiento mitocondrial obtenido, las opciones correspondientes al modelo seleccionado y la topología de referencia como topología fija, se calcularon y optimizaron por máxima verosimilitud los largos de rama del árbol de especies de los *taxa* estudiados.

### 2.8 Reconstrucción de secuencias ancestrales

Se realizó una reconstrucción de secuencias ancestrales de IGF2, en el espacio de codones, usando los métodos de verosimilitud implementados en el módulo ASR (*Ancestral Sequence Reconstruction*) del servidor Datamonkey (Pupko *et al.*, 2000; Kosakovsky Pond y Frost, 2005c).

### 2.9 Expresión de INSR en C. porcellus

Davidsen y colaboradores (2014) publicaron datos de RNA-Seq de pulmón y músculo esquelético de *C. porcellus*, que utilizaron para el desarrollo y validación de un *microarray* de expresión para este animal modelo. La secuenciación de la fracción polyA del transcriptoma se realizó en una plataforma Illumina HiSeq 2000 y se generaron *reads* pareados de 101x2 bases, accesibles desde el repositorio SRA del NCBI (http://www.ncbi.nlm.nih.gov/sra) bajo los identificadores SRR1200908 y SRR1200910 (44.173.720 y 88.539.498 millones de *reads* para pulmón y músculo, respectivamente). Para esta tesis, los *reads* originales fueron mapeados en el genoma de *C. porcellus* y luego los transcriptos ensamblados para el scaffold\_42, que es el que contiene anotado el gen INSR en el cobayo. El filtrado de calidad de los *reads* se realizó con Trimmomatic-0.33 (Bolger *et al.*, 2014), usando los parámetros para *reads* 

pareados: LEADING:3, TRAILING:3, SLIDINGWINDOW:4:15 y MINLEN:60. Para el mapeo y ensamblado de los transcriptos se utilizó parte del protocolo Tuxedo (Trapnell *et al.*, 2012): Tophat-2.1.1 (Kim *et al.*, 2013) y Cufflinks-2.2.1(Trapnell *et al.*, 2010). El mapeo con Tophat sobre el genoma cavPor3 y transcriptoma de referencia (obtenidos de UCSC Genome Browser) incluyó las opciones -microexon-search y -b2-very-sensitive. En ambas muestras, el 89% de los *reads* presentaron mapeos genómicos concordantes en el par. Con SAMtools (Li *et al.*, 2009) se filtraron los archivos BAM para el scaffold\_42. Sobre estos BAM reducidos se corrió Cufflinks con las opciones -frag-bias-correct y -multi-read-correct. Los alineamientos junto a los modelos de transcriptos resultantes se visualizaron en IGV (*Integrative Genome Browser*; Robinson *et al.*, 2011). También se utilizó TransDecoder-3.0.0 (Haas *et al.*, 2013) para obtener el tramo codificante de los transcriptos y verificar las isoformas expresadas del INSR mediante BLASTP (Altschul *et al.*, 1997; Camacho *et al.*, 2009).

Debido a lo pequeño del exón 11 que diferencia ambas formas del INSR (36 nucleótidos) y a la presencia de posibles sitios dadores y aceptores de *splicing* no canónicos, el mapeo de los *reads* al genoma cavPor3 se repitió con la herramienta Rsubread (Liao *et al.*, 2013), un alineador que sigue una estrategia elegante y muy eficiente de múltiples semillas. Se utilizó la versión de Rsubread para el entorno R (R Core Team, 2015) y se corrió la rutina subjunc con las opciones unique=TRUE, indels=5, minFragLength=30 y reportAllJunctions=TRUE. Los alineamientos también fueron visualizados en IGV.

# 3 Resultados

### 3.1 Genes y linajes

### 3.1.1 Genes

Los genes a estudiar se presentan en la Tabla 3.1 y Figura 3.1. Se seleccionaron en base a antecedentes bibliográficos, pertenencia a familias multigénicas y posición en las vías de señalización.

Como continuación del trabajo de Opazo y colaboradores sobre evolución molecular de insulina en Caviomorpha (Opazo et al., 2005), se incluyó tanto este gen como sus dos parálogos IGF1 e IGF2 (LeRoith et al., 1993; Olinski et al., 2005). Estos factores de crecimiento relacionados a insulina tienen un rol fundamental en el crecimiento pre y post-natal y forman parte del eje de la hormona de crecimiento (GH1). IGF1 es producido en múltiples tejidos ejerciendo una acción parácrina, y en mamíferos tiene acción endócrina al ser producido y liberado por el hígado en respuesta a GH1. Sin embargo, en caviomorfos el crecimiento post-natal es independiente de GH1 (Mitchell et al., 1954; Adkins et al., 2000), por lo que este gen se incluyó entre los estudiados, así como su receptor (GHR) y otra dupla hormona-receptor que modula su nivel sanguíneo (GHRH y GHRHR; Carter-Su et al., 2016). La insulina y los factores de crecimiento ejercen sus funciones a través de los receptores INSR e IGF1R (Hernández-Sánchez et al., 2008). Estos son dos receptores tirosina quinasas diméricos que se originaron por duplicación génica temprano en la evolución de los vertebrados. Si bien INS e IGFs ejercen sus funciones principalmente a partir de la unión a los receptores INSR e IGF1R respectivamente, también son capaces de unirse en forma cruzada y con diferente afinidad al otro receptor, lo mismo que a receptores heterodiméricos INSR-IGF1R, añadiendo mayor complejidad al sistema. Existe un tercer receptor perteneciente a esta familia génica, el INSRR, cuya función principal se desconoce, considerándose un receptor huérfano (Hernández-Sánchez et al., 2008). Las IGFBPs comprenden una familia de siete proteínas sanguíneas de unión a IGF1 e IGF2 (IGFBP1 a IGFBP7), que potencian o atenúan la señalización de los IGFs según sea el contexto fisiológico (Oh et al., 1996; Daza et al., 2011). Finalmente, el receptor de IGF2 (IGF2R), es un receptor de manosa-6-fosfato que en amniotas habría incorporado un dominio adicional de unión a IGF2 interviniendo también en la regulación de su biodisponibilidad (Ghosh et al., 2003; ElShewy y Luttrell, 2009; McGaugh



*et al.*, 2015). IGF2R funciona como un receptor transmembrana que une IGF2 en la superficie celular y lo internaliza en vesículas de clatrina para su degradación, atenuando su señalización.

Figura 3.1: El esquema indica los genes estudiados y sus relaciones funcionales. La GHRH hipotalámica y la hormona de crecimiento hipofisiaria (GH o GH1) controlan la producción endócrina de los factores de crecimiento IGF1 e IGF2 en el hígado. La insulina es liberada de los islotes pancreáticos como respuesta a un incremento en la glicemia. En las células diana, tanto INS como IGFs ejercen sus funciones a través de los receptores diméricos INSR e IGF1R, así como a través de receptores heterodiméricos INSR-IGF1R. Un tercer miembro de esta familia génica, el INSRR, se considera un receptor huérfano y no tendría un rol en las vías de señalización aquí estudiadas. El IGF2 también señaliza a través del INSR y del IGF1R, siendo el IGF2R un receptor que no transduce señales y solo controla su biodisponibilidad. En el plasma se encuentran otras proteínas de unión a IGF1 e IGF2 que también intervienen en la regulación de sus niveles sanguíneos (IGFBP1 a IGFBP7).

Los 18 genes estudiados corresponden a proteínas transmembrana con dominios extracelulares de unión a insulina y factores de crecimiento, o péptidos y proteínas extracelulares con funciones señalizadoras o transportadoras en la circulación. Desde este punto de vista, es conocido que en eucariotas multicelulares y especialmente en mamíferos, las proteínas extracelulares presentan una

mayor tasa evolutiva que las intracelulares (Julenius y Pedersen, 2006; Liao *et al.*, 2010; Feyertag *et al.*, 2017; Sojo *et al.*, 2016). Esta observación sería independiente del nivel de expresión, del número de interacciones proteicas, de la importancia funcional del gen y existe la hipótesis de que corresponde simplemente a una mayor selección adaptativa frente a la variabilidad del medio extracelular que con respecto a la homeostasis intracelular (Sojo *et al.*, 2016). En el contexto de las proteínas aquí seleccionadas, el fenómeno sugiere una mayor probabilidad de detección de cambios en los patrones evolutivos y convierte este conjunto de genes en una interesante prueba de concepto para la aproximación metodológica empleada. De hecho, si se hubiesen incluido genes y proteínas que integran las vías de transducción de señales de INS e IGFs, también se encuentra descripto, tanto para *Drosophila* como para vertebrados, un gradiente en la fuerza de la selección purificadora a lo largo de la vía, que se incrementa desde los genes proximales a las hormonas a los más distales o *downstream* (Alvarez-Ponce *et al.*, 2009, 2011).

Todos los genes presentan ortología uno-a-uno en los taxa analizados, excepto INS que se encuentra duplicado en M. musculus y GH1 que presenta expansiones génicas en Catarrhini y Platyrrhini (Soares et al., 1985; Wallis et al., 2001). En el caso de INS, se mantuvieron ambos co-ortólogos en M. musculus salvo en los análisis de RELAX donde se removió INS1, el duplicado específico de Murinae originado por retroposición de INS2 (Beintema y Campagne, 1987; Shiao et al., 2008). Con respecto a la duplicación del locus de GH en antropoideos, las relaciones de co-ortología son bien complejas debido a expansiones independientes en platirrinos y catarrinos, aí como abundante conversión génica (Chen et al., 1989; Wallis et al., 2001; Li et al., 2005; González Alvarez et al., 2006; Petronella y Drouin, 2011). Los nuevos genes cumplen funciones de somatotropina y lactogénicas en la placenta, con un rol en la interfase materno-fetal (Chen et al., 1989; Papper et al., 2009). A pesar de esta dinámica evolutiva del loci, tanto en Catarrhini como en Platyrrhini es fácil identificar el co-ortólogo que mantiene la expresión hipofisaria y funciones propias de la GH1, y solo este homólogo fue incluido en los análisis. Sin embargo, en los análisis que involucraron una selección de ramas blanco (modelos de ramas y ramas y sitios en CODEML, modelos en BUSTED y RELAX) los platirrinos no fueron considerados como linaje blanco para el análisis de GH1 (se tuvo en consideración la abundante y reciente bibliografía al respecto).

### 3.1.2 Linajes

Los *taxa* incluidos en este trabajo así como sus relaciones filogenéticas se muestran en la Figura 2.1. Con respecto a los órdenes en Euarchonta, falta representación del Orden Dermoptera (lemures voladores) aunque en NCBI existe un borrador genómico con 32.223 predicciones de transcriptos codificantes para proteína de *Galeopterus variegatus*, por lo que este dermóptero podría incluirse en el futuro cercano. Con respecto a Platyrrhini, la topología de referencia solo incluye secuencias de *A. nancymaae*, *C. jacchus* y *S. boliviensis*. Como se discutió en el capítulo Metodología, la posición filogenética de *Aotus* varía en los diferentes trabajos revisados,

		(	,						
		HGNC				Ens	embl (GRCh38): Homo	sapiens	
Símbolo	₽	Nombre oficial	RefSeq	ID gen	coordenadas genomicas	# exones	ID transcripto	Largo pb (AA)	Otro nombre
INS	6081	insulin	NM_001185098	ENSG00000254647	11: 2159779-2161341 (-)	3	ENST00000397262	639 (110)	proinsulina
IGF1	5464	Insulin-like growth factor 1	NM_000618	ENSG0000017427	12: 102395867-102480645 (-)	S	ENST00000337514	7359 (153)	somatomedina C
IGF2	5466	Insulin-like growth factor 2	NM_001127598	ENSG00000167244	11: 2129112-2141238 (-)	8	ENST00000434045	1583 (236)	somatomedina A
INSR	6091	insulin receptor	NM_000208	ENSG00000171105	19: 7112255-7294034 (-)	22	ENST00000302850	4721 (1382)	S
IGF1R	5465	Insulin-like growth factor 1 receptor	NM_000875	ENSG00000140443	15: 98648971-98964530 (+)	21	ENST00000268035	11803 (1367)	S
INSRR	6093	insulin receptor-related receptor	NM_014215	ENSG0000027644	1: 156840063-156859018 (-)	21	ENST00000368195	5101 (1297)	receptor huérfano
IGF2R	5467	Insulin-like growth factor 2 receptor	NM_000876	ENSG00000197081	6: 159969099-160113507 (+)	48	ENST00000356956	14044 (2491)	receptor manosa-6P catión-independiente
IGFBP1	5946	Insulin-like growth factor binding protein 1	NM_000596	ENSG00000146678	7: 45888357-45893668 (+)	4	ENST00000275525	1653 (259)	proteína placental 12
IGFBP2	5471	Insulin-like growth factor binding protein 2	NM_000597	ENSG00000115457	2: 216632828-216664436 (+)	4	ENST00000233809	1439 (325)	ζ
IGFBP3	5472	Insulin-like growth factor binding protein 3	NM_001013398	ENSG00000146674	7: 45912245-45921874 (-)	S	ENST00000381083	1050 (297)	S
IGFBP4	5473	Insulin-like growth factor binding protein 4	NM_001552	ENSG00000141753	17: 40443461-40457731 (+)	4	ENST00000269593	2200 (258)	S
IGFBP5	5474	Insulin-like growth factor binding protein 5	NM_000599	ENSG00000115461	2: 216672105-216695525 (-)	4	ENST00000233813	6215 (272)	S
IGFBP6	5475	Insulin-like growth factor binding protein 6	NM_002178	ENSG00000167779	12: 53097436-53102345 (+)	4	ENST00000301464	1169 (240)	S
IGFBP7	5476	Insulin-like growth factor binding protein 7	NM_001553	ENSG00000163453	4: 57030773-57110385 (-)	4	ENST00000295666	1427 (282)	S
GH1	4261	growth hormone 1	NM_000515	ENSG00000259384	17: 63917200-63918838 (-)	S	ENST00000323322	803 (217)	hormona de crecimiento hipofisiaria; somatotropina
GHR	4263	growth hormone receptor	NM_000163	ENSG00000112964	5: 42423777-42721878 (+)	9	ENST00000230882	4560 (638)	S
GHRH	4265	growth hormone-releasing hormone	NM_021081	ENSG00000118702	20: 37251082-37261835 (-)	S	ENST00000237527	461 (108)	somatocrinina
GHRHR	4266	growth hormone-releasing hormone receptor	$NM_{000823}$	ENSG00000106128	7: 30938669-30993254 (+)	10	ENST00000326139	1602 (423)	ζ
				a affaial annón UCN	C. al idantificadan dal transa	into Do		nísti sos shtani	4 4. F

bla 3.1:
Lista
de los
18
genes
analizados
en el
presente estudio.

NOTA: Se indican para cada gen: el símbolo, el identificador y el nombre oficial según HGNC; el identificador del transcripto RefSeq; algunas características obtenidas de Ensembl v83 (http://www.ensembl.org); nombre alternativo, en caso de tenerlo. Los nombres oficiales se mantuvieron en inglés.

pero en general hay concordancia en que pertenece a la familia Cebidae. Esto implica que el presente trabajo carece de *taxa* representantes de las otras dos familias de platirrinos: Atelidae y Pitheciidae. Es así que todos los resultados que se presentan y discuten, si bien se consideran a nivel del parvaorden Platyrrhini estrictamente abarcan, en el mejor de los casos, inferencias únicamente a nivel de la familia Cebidae. Los análisis efectuados con CODEML y MEME, fueron también realizados para una topología en la que *Aotus* se presenta como linaje hermano del clado *Callithrix* + *Saimiri*. Estos resultados no se incluyen en este manuscrito, pero son básicamente iguales a los obtenidos para la topología ((*Aotus, Saimiri*), *Callithrix*) de la Figura 2.1.

Los roedores constituyen el grupo más diversificado de mamíferos vivientes, dando cuenta del 40% del total de las especies de la clase Mammalia. Los tres linajes principales están representados en la topología de referencia (Figura 2.1). Para el clado mouse-related existen recursos genómicos y transcriptómicos adicionales que podrían incluirse para aumentar la diversidad cubierta. Por ejemplo, las siguientes especies tienen cada una al menos 25.000 transcriptos codificantes de proteínas en NCBI: Cricetulus griseus, Mesocricetus auratus, Microtus ochrogaster, Peromyscus maniculatus y Jaculus jaculus. Todos son integrantes de la familia Cricetidae, salvo J. jaculus que pertenece a la familia Dipodidae. Es así que J. jaculus y alguno de los cricétidos podrían incluirse en una versión actualizada de este trabajo. Con respecto a Hystricognathi, se prestó atención a la existencia de secuencias de algún puercoespín del Viejo Mundo (familia Hystricidae) pero como mejor resultado se obtuvieron solo 157 secuencias nucleotídicas en GenBank para Hystrix, por lo que este linaje no está considerado en el muestreo taxonómico final. Con respecto a Caviomorpha, y según se mencionó en el capítulo anterior, no se contó con *taxa* pertenecientes a la superfamilia Erethizontoidea (e.g. Coendou y Erethizon), la cual tiene apenas 170 secuencias nucleotídicas depositadas en GenBank. Otro aspecto notable es la ausencia de representantes de Echimyidae (Octodontoidea). La familia está integrada por las ratas espinosas neotropicales, constituyendo la familia con mayor riqueza específica de caviomorfos y la que presenta también la mayor diversidad ecológica (Upham y Patterson, 2012; Fabre et al., 2013; Emmons et al., 2015; Upham y Patterson, 2015; Ojeda et al., 2015; Fabre et al., 2017). A pesar de esta riqueza, solo 1.550 secuencias se encontraron depositadas en GenBank, siendo Proechimys cuvieri la especie con más datos a la fecha de búsqueda (291 secuencias). Estas secuencias están dominadas por fragmentos de los genes CYTB e IRBP, ampliamente usados como marcadores moleculares. Finalmente, existe un gran número de roedores, especialmente histricognatos, con secuencias disponibles del receptor de la hormona de crecimiento, debido también a su uso como marcador molecular (e.g. Upham y Patterson, 2012). Estas secuencias son parciales y abarcan entre 200 a 300 codones de los más de 600 que tiene el receptor completo utilizado en este trabajo, por lo que no fueron incluidas en el análisis del gen GHR.

Último acceso a NCBI y Ensembl: 09/03/2016

### 3.2 Grupos de ortólogos y alineamientos

Los grupos de ortólogos se presentan como material suplementario en las tablas A-1 a A-18. El número de *taxa* muestreado por gen, así como los tamaños de los alineamientos luego de cada etapa de procesamiento (*i.e.* PAL2NAL – > Gblocks – > BMGE), se indican en la Tabla 3.2. La información se complementa con la Tabla 3.3 que muestra con colores qué tanto se desvió el muestreo taxonómico de cada gen con respecto al planteado en la Figura 2.1. *Gorilla gorilla* y *O. garnettii* fueron los *taxa* que más veces fueron reemplazados por secuencias provenientes de una especie con un género distinto (*Pan troglodytes* y *Pongo abelii* en el caso de *Gorilla* y *M. murinus* y *Propithecus coquereli* en el caso de *Otolemur*). Para *Macaca* y *Equus*, los reemplazos, si bien fueron frecuentes, implicaron especies pertenecientes al mismo género (*M. fascicularis* y *M. nemestrina*; *E. asinus* y *E. przewalskii*). Por otra parte, la mayor cantidad de datos ausentes fue presentada por *C. sociabilis* (siete de 18 secuencias faltantes) y *T. syrichta* (cinco de 18). Con respecto a los clados de interés, en Platyrrhini siempre hubo al menos dos representantes, con una ausencia de secuencia en *Saimiri* y cinco en *Aotus*. En Caviomorpha (y también Phiomorpha), la cobertura fue completa, salvo las faltantes en *C. sociabilis* recién mencionadas.

Gen	# Taxa	PAL2NAL (pb)	Gblocks (pb)	BMGE (pb)	Codones finales
INS	22	339	327	327	109
IGF1	23	459	459	459	153
IGF2	23	720	705	537	179
INSR	22	4167	4143	4032	1344
IGF1R	24	4143	4101	4101	1367
INSRR	23	3903	3897	3897	1299
IGF2R	24	8388	7458	7350	2450
IGFPB1	22	894	783	738	246
IGFBP2	23	1002	936	741	247
IGFBP3	21	1020	870	606	202
IGFBP4	23	774	774	699	233
IGFBP5	23	819	810	810	270
IGFBP6	24	741	708	705	235
IGFBP7	22	861	846	594	198
GH1	24	690	648	645	215
GHR	24	2235	1911	1890	630
GHRH	22	327	321	318	106
GHRHR	23	1305	1269	1257	419
TOTAL	NA	32787	30966	29706	9902

 Tabla 3.2: Métricas de los alineamientos nucleotídicos de los 18 genes. Se indica el número de *taxa* por gen, así como el largo después de aplicar sucesivamente PAL2NAL, Gblocks y BMGE.

NOTA: INS tiene 23 secuencias pero dos pertenecen a *M. musculus* que presenta dos parálogos, *ins1* e *ins2* para este gen; NA: no aplica.

Como se observa en la Tabla 3.2, IGFBP3 tuvo el menor número de *taxa* con solo 21; el máximo de 24 *taxa* lo presentaron IGF1R, IGF2R, IGFBP6, GH1 y GHR. El tamaño inicial de los

ipodomys ikomys eterocephalus etonnys inchilla avia ss mis									

. ÷ . A.L. Tabla 2 3. NOTA: Clave de colores con respecto a los *taxa* indicados en Figura 2.1. Azul: misma especie. Azul claro: igual género. Violeta: género diferente. Blanco: sin representante. Rojo indica casos especiales; en *M. musculus* existen dos secuencias correspondientes a *ins1* e *ins2*; en Catarrhini y Platyrrhini GH1 no mantiene una relación de ortología uno-a-uno con los homólogos en los restantes *taxa*. Para mayor información con respecto a los grupos de ortólogos y el muestreo taxonómico, revisar Tabla A-1 a Tabla A-18.

alineamientos (luego de ejecutado PAL2NAL) estuvo en el rango de 339 a 8.388 pares de bases (pb), siendo GHRH el gen más pequeño e IGF2R el de mayor tamaño. Luego de procesados con Gblocks y BMGE, el nuevo rango fue de 318 a 7.350 pb o de 106 a 2.450 codones, para un total de 9.902 codones finales a ser analizados en el marco de los modelos de sustitución de codones. Con respecto a BMGE, el análisis removió entre el 0 y el 30% de los nucléotidos, dado las opciones elegidas para correr el programa (BLOSUM62 como matriz de sustitución de aminoácidos). Para estos alineamientos se calculó el modelo de sustitución nucleotídica que mejor ajusta según AICc y los resultados se indican en la Tabla A-19.

### 3.3 Análisis composicional

La composición nucleotídica y el uso de aminoácidos se indican en las tablas suplementarias A-20 y A-21, respectivamente. Todos los taxa presentan GC2 con valor 15%, salvo Ctenomys que presenta 14%. En la primera posición de los codones, el contenido GC es 18 o 19%. Según lo esperado, los valores de GC3 son más variables con un rango de 54 a 61% y un promedio de 57%. El menor contenido GC3 se observa en Mus y Nannospalax, mientras que el mayor en Oryctolagus. Romiguier y colaboradores (2010) analizaron más de mil ortólogos en 33 mamíferos encontrando un rango del contenido GC3 de 46 a 53% para Boroeutheria, así como su mayor valor en Lagomorpha. Estos valores de dispersión del contenido GC3, así como el valor máximo en Lagomorpha, se repiten en el presente trabajo. Notoriamente, los 18 genes estudiados tienen un contenido GC3 en promedio 8% mayor que el estimado por Romiguier para los más de mil ortólogos y pueden considerarse como pertenecientes al grupo de los genes más ricos en GC. Esto podría explicar los bajos valores observados de GC3 en Mus y Nannospalax: mientras que se espera un incremento en el GC3 global en el linaje de los roedores mouse-related, se observa una erosión linaje específica en el subgrupo de genes de alto contenido GC3 (Romiguier et al., 2010). La Figura 3.2 y las figuras suplementarias B-3 y B-4 muestran los resultados de los análisis de correspondencia para la composición nucleotídica, de codones y aminoácidos, respectivamente. A nivel de codones se observa principalmente una correspondencia con la identidad de las bases en posiciones 1 a 3 y en definitiva con el contenido GC3. A nivel de aminoácidos esto también es evidente (e.g. observar MUS y NGA versus OCU y OPR separados según la primera dimensión), pero también se observan algunos taxa aislados, como CSO, TSY y ANA. Dado que estos taxa presentaron la mayor ausencia de datos, y además CSO presenta una asociación con un aminoácido poco frecuente como M (metionina), es posible que el resultado se encuentre sesgado por problemas de muestreo (aunque ver las coordenadas de los aminoácidos K e I que son ricos en AT).



Figura 3.2: Análisis de correspondencia de la composición nucleotídica de los genes y taxa estudiados. Los triángulos azules indican los nucleótidos en las posiciones 1 a 3 de los codones. Los círculos rojos representan las especies, algunas de las cuales están etiquetadas. ANA: A. nancymaae; BOV: B. taurus; CAR: Carnivora; CJA: C. jacchus; CPO: C. porcellus; CSO: C. sociabilis; GGO: G. gorilla; MUS: M. musculus; NGA: N. galili; OCU: O. cuniculus; OGA: O. garnettii; OPR: O. princeps; PER: Perissodactyla; TBE: T. chinensis.

Volviendo al análisis de codones, en la Figura 3.3 cada símbolo corresponde a una secuencia, las que se encuentran coloreadas según su identidad génica. En la Figura B-5 se observa el mismo gráfico pero las secuencias están coloreadas por *taxa*. De ellas se desprende que hay una mayor similitud en el uso de codones debido a la identidad del gen que con respecto al linaje muestreado. Esta homogeneidad por gen permite descartar sesgos importantes en la composición de codones que puedan afectar los cálculos posteriores de las tasas de sustitución no-sinónima:sinónima.



Figura 3.3: Análisis de correspondencia de todas las secuencias de codones. Los símbolos y colores identifican las secuencias pertenecientes al mismo gen.

# 3.4 Modelos de ramas y de ramas y sitios implementados en CODEML

### 3.4.1 Modelo de ramas

Los resultados del análisis de ramas implementado en CODEML se muestran en la Tabla 3.4. Se indican los resultados para el modelo nulo y el modelo alternativo con la conformación de ramas blanco con mayor incremento de verosimilitud con respecto al modelo nulo de un solo  $\omega$ . Las estimaciones presentaron convergencia sin importar el valor inicial de  $\omega$  (se muestran los resultados para  $\omega_{inicial} = 1$ , no así para 0,5 y 2). En tres casos (IGFBP1, IGFBP7 y GHRH) el

modelo alternativo infirió un valor de  $\omega > 1$  y los resultados para el modelo con  $\omega = 1$ , requerido para la prueba de selección positiva también se indican en la tabla. Como se puede observar, no hubo evidencia para rechazar el modelo nulo con un único valor de  $\omega$  en el caso de IGF1 y GH1. Para IGFBP1, IGFBP7 y GHRH el valor de  $\omega$  estimado para platirrinos fue mayor a 1 usando la rama ancestral de platirrinos como linaje blanco. Para IGFBP1 y GHRH el valor de  $\omega$  es  $\omega = 2$ y  $\omega = 1,7$ , respectivamente. Para IGFBP7 el valor estimado de  $\omega$  es muy alto (999), y se asocia a un problema de estimación debido a una tasa  $dS \approx 0$ . Efectivamente, al realizar la prueba de selección positiva, solo IGFBP1 mostró una estimación de  $\omega$  significativamente mayor que 1. En diez genes, el valor de  $\omega$  inferido para platirrinos ( $\omega_1$  en la Tabla 3.4), aunque menor que 1, resultó mayor que  $\omega_0$  y  $\omega_2$ , lo que condujo a los análisis con RELAX (ver más adelante). Finalmente, debido a que la detección a nivel de ramas requiere que en promedio, a lo largo de las secuencias, el valor de  $\omega$  sea mayor que 1, la prueba es poco potente para detectar selección positiva y el resultado obtenido concuerda en este sentido con lo esperado de antemano.

### 3.4.2 Modelo de ramas y sitios

La implementación de modelos de ramas y sitios resultó más potente en la identificación de linajes con huellas de selección positiva (Tabla 3.5). Para un  $\alpha = 0, 1$ , IGF2, IGF1R, INSRR, IGF2R, IGFBP1, IGFBP2, IGFBP4, GHRH y GHRHR mostraron evidencia de selección positiva en Platyrrhini (a nivel de la rama ancestral y/o del clado). Caviomorpha solo mostró  $\omega^+ > 1$  significativo en INS, INSR, IGFBP1 e IGFBP6. Sin embargo, en IGFBP6 ningún sitio presentó valores de BEB con PP > 0,95. Viendo el conjunto de datos para los genes estudiados, existe mayor evidencia para inferir procesos de selección positiva en la historia evolutiva de Platyrrhini que en Caviomorpha.

A modo de comentario, al observar los patrones de los posibles sitios bajo selección, se detectan casos de mucha cercanía en el alineamiento. Por ejemplo, considérense los sitios 26, 27 y 28 en INS o los sitios 291, 293, 294 y 295 en IGF1R. Si bien podría tratarse de dominios hipervariables bajo la fuerza selectiva (como sería el caso en IGF1R; ver Figura B-6), también es factible de que existan errores en las secuencias obtenidas de las bases de datos o en el alineamiento propuesto. Es recomendable al menos tener este punto en cuenta previo a extender en profundidad la discusión de estos resultados.

Las tablas A-22 a A-38 muestran datos más completos sobre los resultados de estos modelos de ramas y sitios en este subconjunto de genes que resultaron con evidencia de selección positiva.

Gen	Modelo	NP	Ts/Tv	TL	$\omega_0$	$\omega_1$	$\omega_2$	lnL	LRT	pvalue
INS	nulo alternativo C	45 47	4,0647 4,1438	7,0398 7,1302	0,1822 0,1478	0,1822 0,2837	0,1822 0,4459	-3118,8971 -3108,6684	20,4575	<0,0001
IGF1	nulo	45	3,0600	3,2399	0,0775	0,0775	0,0775	-2612,1745		
IGF2	nulo alternativo C	45 47	4,2537 4,2496	3,7639 3,7616	0,1823 0,1811	0,1823 0,7710	0,1823 0,1040	-3516,4150 -3511,6327	9,5647	0,0084
INSR	nulo alternativo C	43 45	3,0541 3,0639	5,0504 5,0769	0,0383 0,0332	0,0383 0,2706	0,0383 0,0457	-26697,6916 -26636,6139	122,1554	«0,0001
IGF1R	nulo alternativo C	47 49	3,3258 3,3360	4,0830 4,1000	0,0372 0,0328	0,0372 0,2939	0,0372 0,0315	-23953,4943 -23879,2225	148,5435	«0,0001
INSRR	nulo alternativo D	45 47	4,3954 4,4026	3,9136 3,9218	0,0755 0,0720	0,0755 0,1332	0,0755 0,1896	-24480,9216 -24470,3219	21,1993	<0,0001
IGF2R	nulo alternativo B	47 49	2,9506 2,9581	6,3559 6,3818	0,1357 0,1287	0,1357 0,1754	0,1357 0,1796	-63981,7675 -63963,3583	36,8185	«0,0001
IGFBP1	nulo alternativo C $\omega_0, \omega_1 = 1, \omega_2$	43 45 44	3,2502 3,2466 3,2351	8,3040 8,3195 8,3152	0,2279 0,2028 0,2031	0,2279 2,0450 1,0000	0,2279 0,1922 0,1921	-7596,1579 -7557,3512 -7560,2982	77,6133 0,0302	«0,0001 0,0152
IGFBP2	nulo alternativo C	45 47	3,3489 3,3169	4,2598 4,2288	0,0996 0,0978	0,0996 0,4269	0,0996 0,0443	-4890,2186 -4874,3524	31,7324	«0,0001
IGFBP3	nulo alternativo A	41 43	3,4981 3,5229	5,2803 5,3130	0,0842 0,0796	0,0842 0,7253	0,0842 0,0001	-4256,7709 -4243,9038	25,7342	«0,0001
IGFBP4	nulo alternativo A	45 47	5,6768 5,6808	3,2904 3,2904	0,0564 0,0481	0,0564 0,7226	0,0564 0,0001	-3806,6361 -3784,0806	45,1109	«0,0001
IGFBP5	nulo alternativo A	45 47	3,8911 3,8971	2,7272 2,7243	0,0372 0,0323	0,0372 0,3882	0,0372 0,0028	-3905,6066 -3891,4413	28,3306	«0,0001
IGFBP6	nulo Alternativo C	47 49	3,4372 3,4311	5,6604 5,6545	0,2766 0,2785	0,2766 0,9693	0,2766 0,1897	-6049,4156 -6044,2519	10,3274	0,0057
IGFBP7	nulo Alternativo A $\omega_0, \omega_1 = 1, \omega_2$	43 45 44	2,9351 2,9353 2,9335	2,7933 2,7882 2,7894	0,0886 0,0851 0,0853	0,0886 0,0001 1,0000	0,0886 999,0000 0,0001	-2893,9596 -2889,0919 -2889,5479	9,7354 0,9119	0,0077 0,3396
GH1	nulo	47	4,1743	5,7120	0,1330	NA	0,1330	-5192,4224		
GHR	nulo alternativo C	47 49	3,5840 3,5867	4,3834 4,3875	0,3407 0,3193	0,3407 0,6353	0,3407 0,4545	-14297,8530 -14291,9551	11,7959	0,0006
GHRH	nulo alternativo C $\omega_0, \omega_1 = 1, \omega_2$	43 45 44	4,6458 4,6461 4,6459	5,7530 5,7655 5,7699	0,2768 0,2616 0,2615	0,2768 1,7078 1,0000	0,2768 0,3049 0,3047	-2531,3036 -2528,6843 -2528,8348	5,2386 0,3010	0,0221 0,5833
GHRHR	nulo alternativo B	45 47	3,7518 3,7612	5,2971 5,3157	0,1753 0,1629	0,1753 0,2529	0,1753 0,2372	-10218,3101 -10213,4109	9,7984	0,0017

Tabla 3.4: Parámetros estimados con los modelos de ramas implementados en CODEML.

NOTA: NP: número de parámetros del modelo; Ts/Tv: cociente entre las tasas de transición y traversión; TL: largo del árbol en unidades de número de sustituciones nucleotídicas por codon; NA: no aplicable. Los modelos alternativos A, B, C y D hacen referencia a las conformaciones de ramas blanco y del *background* indicadas en la Figura 2.2. Para simplificar la tabla, en los casos de IGF1 y GH1 solo se indican los resultados del modelo nulo. Para GH1 no se consideraron los platirrinos, por lo que el modelo alternativo solo presentó un  $\omega$  asociado a la rama ancestral de Caviomorpha o a todo el clado, y otro valor de  $\omega$  para las ramas del *background*.

Tabla 3.5:	Valor de $\omega^+$ y sitios bajo selección según el modelo de ramas y sitios implementado en
	CODEML. solo se muestran aquellos genes donde se infirió selección positiva en al menos
	una de las conformaciones de linajes blanco probadas.

Gen	Linaje blanco	w+	pvalue	Sitios bajo seleccion (#)
INS	clado Caviomorpha	2,0176	0,0484	9, 22, 26, 27, 28, 34, 41, 42, 44, 45, 53, 70, 97, 101, 102 (15)
IGF2	ancestro Platyrrhini clado Platyrrhini	25,5621 5,3933	0,0013 0,0570	129, 136 (2) 129, 136 (2)
INSR	ancestro Caviomorpha	17,9400	0,0743	185 (1)
IGF1R	ancestro Platyrrhini	3,4362	0,0008	24, 280, 291, 293, 294, 295, 315, 320, 442, 558, 694 (11)
INSRR	clado Platyrrhini	2,7946	0,0906	710, 718, 1008 (3)
IGF2R	ancestro Platyrrhini	21,8997	0,0048	2023 (1)
IGFBP1	ancestro Platyrrhini clado Platyrrhini clado Caviomorpha	5,8759 2,8630 6,3711	<pre>«0,0001 0,0001 0,0880</pre>	40, 42, 69, 74, 85, 113, 122, 128, 138, 158, 177, 184, 201 (13) 40, 42, 70, 74, 98, 99, 113, 117, 118, 122, 125, 126, 128, 138, 150, 195, 201, 234, 242 (19) 7 (1)
IGFBP2	ancestro Platyrrhini	3,4911	0,0894	134, 139 (2)
IGFBP4	ancestro Platyrrhini	7,2375	0,0174	20, 164, 193 (3)
IGFBP6	ancestro Caviomorpha	64,4237	0,0759	NA (0)
GHRH	ancestro Platyrrhini clado Platyrrhini	64,0947 11,4519	0,0013 0,0657	105 (1) 105 (1)
GHRHR	ancestro Platyrrhini clado Platyrrhini	531,7692 4,0045	0,0085 0,0012	205 (1) 128, 205, 249, 256, 259, 260, 267 (7)

NOTA: Los linajes blancos siguen las conformaciones especificadas en la Figura 2.3. Se indican las estimaciones de  $\omega^+$  y el valor de *pvalue* resultante del *LRT* entre el modelo que deja libre  $\omega^+$  y uno nulo en el que  $\omega^+$  se fija a 1. Con fondo naranja se indican *pvalues* menores a 0,05. Finalmente se indica la identidad de los sitios inferidos con  $\omega > 1$  en el linaje blanco (con *BEB* > 0,95) y el total entre paréntesis. Los sitios se encuentran numerados según su posición en los alineamientos obtenidos en este trabajo. NA: no aplica.

### 3.5 Modelos implementados en HyPhy

### 3.5.1 BSREL, aBSREL y BUSTED

Los resultados obtenidos por BSREL y aBSREL fueron muy similares y se presentan en la Tabla 3.6 (BSREL) y en las tablas suplementarias A-39 a A-44 (aBSREL). Con respecto a los *taxa* de interés, los resultados son iguales salvo dos linajes en los genes IGF2 y GHRH que presentan selección positiva según BSREL y  $\omega > 1$  aunque no estadísticamente significativo según aBSREL. Esta similitud es concordante con lo propuesto por Kosakovsky Pond y colaboradores (2015), donde el método aBSREL se plantea directamente como una estrategia optimizada de BSREL, adaptada a cada juego de datos. En los genes estudiados, la mayor parte de las ramas se modelaron mediante una distribución con una única clase de sitios, y solo unas pocas ramas requirieron ser ajustadas a una distribución con dos valores de  $\omega$ .

En histricognatos estos métodos de efectos aleatorios han detectado evidencia de selección positiva solo en INSR (en la rama que conduce a Octodontoidea), IGFBP2 (en *Fukomys*) e IGFBP6 (en *Octodon*). Siete veces fue detectada selección diversificadora en platirrinos (IGF2, IGF1R, INSRR, IGF2R, IGFBP1, GHRH y GHRHR), en general en la rama ancestral del grupo (IGF2, IGF1R, IGF1R, IGFBP1 y GHRH). En IGF1, IGF1R, IGF2R, IGFBP2, IGFBP5 y GHRH se infirieron valores de  $\omega > 1$  en linajes distintos de Platyrrhini e Hystricognathi, resaltando la importancia de

utilizar modelos de sustitución de codones que no restringen la presencia de selección positiva a unas pocas ramas de la topología.

Los resultados de BUSTED para las combinaciones de genes y configuración de ramas blanco que resultaron bajo selección positiva se indican en las tablas A-45 a A-47. En Platyrrhini los genes IGF2, IGF1R, INSRR, IGF2R, IGFBP1, GHRH y GHRHR mostraron evidencia de selección episódica diversificadora en la rama ancestral y/o en todo el clado. En caviomorfos INSR e IGFBP6 mostraron ocurrencia de selección positiva a nivel de clado mientras que, IGFBP2 resultó bajo selección positiva cuando la rama ancestral de Hystricognathi se configuró como rama blanco. Con respecto a los linajes de interés, los resultados de BUSTED son totalmente concordantes con los obtenidos mediante BSREL/aBSREL.

Con respecto a la comparación entre estos resultados obtenidos con HyPhy y los obtenidos con el modelo de ramas y sitios implementado en CODEML la concordancia es alta (Tabla 3.7): siete genes (IGF2, IGF1R, INSRR, IGF2R, IGFBP1, GHRH y GHRHR) en Platyrrhini y dos en Caviomorpha (INSR e IGFBP6) presentaron huellas de selección con ambos grupos de análisis. En dos casos en platirrinos (IGFBP2 e IGFBP4) y dos en caviomorfos (INS e IGFBP1) se obtuvo evidencia de selección positiva con CODEML pero no así con los métodos implementados en HyPhy. Con respecto a IGFBP2, se detectó selección positiva en Fukomys y en la rama ancestral de Hystricognathi con los métodos implementados en HyPhy, pero no se puso a prueba esta rama al momento de correr el modelo de ramas y sitios en CODEML. Este podría ser un caso donde la existencia de selección positiva en ramas seleccionadas como background (como Fukomys y el ancestro de histricognatos) genera un sesgo en las estimaciones realizadas por CODEML, conduciendo a un resultado falso positivo para la rama ancestral de Platyrrhini. Algo similar podría estar sucediendo con IGFBP1, donde la ocurrencia de selección positiva en platirrinos generaría falsos positivos cuando en CODEML se analiza Caviomorpha como el linaje blanco. Con respecto a INS en Caviomorpha, los modelos implementados en HyPhy detectan valores de  $\omega$  mayores a 5, pero en ningún caso el resultado es estadísticamente significativo.

### 3.5.2 MEME

MEME es un método de sitios, aunque también permite explorar en qué ramas de la filogenia el episodio selectivo habría ocurrido. El número de sitios detectados en cada gen dependerá en parte del número de codones analizados por gen, así como de la cantidad de cambio acumulado en cada rama de la filogenia. Es por esto que los resultados de MEME son presentados en el contexto del número de codones analizado por gen y de los largos de rama de la filogenia. En la Figura B-7 se muestra el árbol mitocondrial obtenido a partir de los 12 genes codificantes de proteínas de la hebra pesada (tercera posición excluida); las ramas representan tasas de sustitución por nucleótido

	O. INCSULTATION OF DOINTLE PARA AYAN				aju suru		11/auvia para u	τ /  + /  +			
Cen	Kama bajo seleccion	(0) promedio	ω	$p_1$	$\omega_2$	$p_2$	8	b	TKI	pvatue	<i>pvalue</i> corregido
IGF1	Mus	0,3078	0,2294	0,9770	0,2295	0,0044	85,2073	0,0186	8,5180	0,0018	0,0756
IGF2	ancestro Platyrrhini	0,8889	0,3870	0,9535	0,3798	0,0150	35,4230	0,0315	8,6483	0,0016	0,0704
INSR	ancestro Octodon + Ctenomys	0,1152	0,0424	0,9796	1,0000	0,0006	7,7409	0,0198	10,2692	0,0007	0,0277
IGF1R	Oryctolagus ancestro Platyrrhini	0,0614 $0,3469$	0,0208 0,2494	0,9804 0,9772	0,1406 0,8357	0,0053 0,0091	10000,0000 32,1416	0,0143 0,0137	34,4679 18,3592	«0,0001 «0,0001	«0,0001 0,0004
INSRR	Callithrix	0,2386	0,1915	0,9961	0,1915	0,0005	450,7020	0,0034	13,2389	0,0001	0,0059
IGF2R	Pan Lemur Ictidomys <b>Saimiri</b> <b>Callithrix</b> ancestro Homo + Pan Oryctolagus	0,4124 0,1823 0,0935 0,2856 0,1056 0,2114 0,1612	0,2737 0,1034 0,0547 0,0547 0,2580 0,0971 0,1779 0,0748	0,9591 0,9018 0,9592 0,9935 0,9988 0,9968	0,3223 0,8179 0,7576 0,7575 0,2575 0,0972 0,1788 0,1788	0,0356 0,0364 0,0303 0,0007 0,0007 0,0008 0,0902	1729,5500 1726,4500 1188,0600 108,4980 10000,0000 54,8815 69,4943	0,0053 0,0118 0,0105 0,0030 0,0005 0,0005 0,0092	70,5766 26,2248 16,3187 12,1305 9,9631 9,4547 9,4547	*0,0001 *0,0001 *0,0001 0,0002 0,0008 0,0011	«0,0001 «0,0001 0,0012 0,0104 0,0327 0,0487
IGFBP1	ancestro Platyrrhini	1,9421	1,0000	0,9008	1,0000	0,0008	15,9258	0,0984	12,4903	0,0002	0,0084
IGFBP2	<b>Fukomys</b> Tarsius	0,3416 0,1518	0,0000 0,0099	0,9219 0,9189	0,0000 0,0102	0,0312 0,0020	10000,0000 $49,9455$	0,0469 0,0792	35,3818 27,6039	«0,0001 «0,0001	«0,0001 «0,0001
IGFBP5	Pongo	0,2106	0,0000	0,9863	0,0000	0,0096	10000,0000	0,0041	8,8878	0,0014	0,0617
IGFBP6	Octodon	0,5163	0,4349	0,9507	0,4358	0,0435	10000,0000	0,0058	12,6773	0,0002	0,0083
GHRH	<i>Mus</i> ancestro Platyrrhini	0,3548 2,4278	0,2889 1,0000	0,9143 0,9747	0,3196 1,0000	0,0000 0,0048	10000,0000 167,7000	0,0857 0,0205	10,7908 9,1976	0,0005 0,0012	0,0209 0,0485
GHRHR	Callithrix	0,4373	0,1002	0,9031	1,0000	0,0605	156,0510	0,0364	49,1801	«0,0001	«0,0001
NOTA: INS, I negrita.	IGFBP3, IGFBP4, IGFB <i>P</i> 7, GH1 y GHR n	o presentaron ramas	s bajo selecci	on diversific	adora para <i>o</i>	$\leq 0, 1.$ Los	linajes bajo selecc	ión que son c	le interés princ	sipal para este	trabajo se resaltaron en

39

BUSTEL	aBSREL	BSREL	CODEM	CODEM	Hystrico	RELAX	MEME	BUSTEI	aBSREL	BSREL	CODEM	CODEM	Platyrrh	Metodo
_			IL: BS	L: Branch	gnathi (Cavi						IL: BS	IL: Branch	<b>E</b> .	
_			C (15)		iomorpha	k=0,61								INS
_	_					k=0,71					_			IGF1
						k=1,35		A	А	A	A y C (2, 2)			IGF2
С	AOC	AOC	A (1)			k=0,21	A,C							INSR
_						k=0,35	A	A	A	A	A (11)			IGFIR
						k=0,84	C	C	CX	CX	C (3)			INSRR
						k=0,59	A,C	AyC	S y CX	S y CX	A (1)			IGF2R
			C (1)			k~0	A,C	AyC	А	A	A y C (13, 19)	А		IGFBP1
Η	Ъ	ч		_		k=0,23	A				A (2)			IGFBP2
			_	_		k=0,63	A							IGFBP3
	_	_	_			k=0.18	A				A (3)			IGFBP4
_	_	_	_			k=0,43	A			_				IGFBP5
C	0	0	A (0)	_		k=0,91			_			_		IGFBP6
_				_		k=0,85				_	_	_		IGFBP7
_	_					NA	A,C	NA			NA	NA		GH1
_						k=0,83	A,C							GHR
						k=0,66	А	AyC	А	A	A y C (1, 1)			GHRH
_						k=21,57	A,C	A y C	CX	CX	A y C (1,7)			GHRHR

# Tabla 3.7: Resúmen de resultados de los análisis de codones.

rama ancestral (A) o en el clado (C), según fuera el linaje blanco utilizado (Figura 2.2 y Figura 2.3). En BUSTED la rama ancestral de Hystricognathi se simbolizó como H. Para BS se muestra entre paréntesis el número de sitios bajo selección detectados por BEB. En el caso de BSREL y aBSREL que ponen a prueba todas las ramas de la topología, las ramas bajo selección incluyeron, además de la ancestral (A), a En RELAX se indica el valor k resultante de la comparación de Platyrrhini con respecto a los restantes euarcontos. En color se indican los resultados significativos para  $\alpha = 0.1$ . En MEME no se consideraron los restantes euarcontos. rama ancestral (A) y/o en el clado de interés (C). En histricognatos los resultados de MEME se indican como ancestro de Caviomorpha (A), Caviomorpha (C), Phiomorpha (P) o ancestro de Hystricognathi (AH) Callithrix (CX), Saimiri (S), Fukomys (F), Octodon (O) y a la rama ancestral de Octodon + Ctenomys (AOC). En Platyrrhini se indica si existió al menos un sitio bajo selección positiva sugerido por MEME en la NOTA: Los resultados se separaron por clados de interés, aún cuando varios de ellos se generaron al unísono. Para los análisis en CODEML (Branch y BS) y BUSTED, se indicó si se detectó selección en la valores corregidos por múltiple prueba de hipótesis y para los análisis de CODEML estos no se calcularon.

MEME

C,P,AH

-0

C,P,AH

A,C,P,AH C,P,AH A,C,P,AH A,C,P,AH

C,P,AH

Ç,P

C,P

C,P,AH

C

C,P

С,АН,Р

A,C

ya que no se calcularon largos absolutos. Vale aclarar que esta filogenia mitocondrial es pertinente en la interpretación de todos los resultados obtenidos, no solo en la inferencia con MEME.

En la Tabla 3.8 se indica el número de sitios inferidos bajo selección positiva para cada gen. El valor se presenta como el total de sitios bajo selección episódica y también como el porcentaje de sitios con respecto al número de codones analizado por gen. De los 9.902 codones analizados, 640 presentaron evidencia de selección positiva (6.5%). Algo más de la tercera parte de estos sitios (252) pertenecen al gen IGF2R, el gen más largo de los estudiados, que aporta alrededor del 25% de los codones analizados. En general los genes presentaron pocos sitios bajo selección positiva, aunque cuatro mostraron valores cercanos al 10% (INS, IGF2R, IGFBP6, GHRH) y dos entre 13 y 15% (GHR e IGF2, respectivamente). La Tabla 3.8 también muestra las posibles ramas en las que la selección natural habría actuado sobre estos sitios. En las ramas internas se utilizó el nombre del clado y un asterisco para indicar la rama ancestral del clado, mientras que para otras ramas se indicaron los nodos limitantes, siguiendo la numeración de la Figura B-7. Esta información se complementa con la Figura 3.4, donde en cada rama se grafica el número de sitios bajo selección episódica en todos los genes excepto IGF2R, el número de sitios bajo selección episódica en IGF2R y el número de genes distintos que presentaron al menos un sitio bajo selección. En líneas generales se observa una correspondencia entre el largo de las ramas (que es proporcional a la tasa de sustitución por nucleótido) y la cantidad de sitios detectados por MEME. Sin embargo se pueden detectar patrones particulares. Por ejemplo, para un largo de rama similar, Fukomys presenta el triple de sitios que su linaje hermano Heterocephalus. En los grupos de interés, Platyrrhini, Caviomorpha y también Hystricognathi, los episodios selectivos se distribuyen tanto en las ramas internas como en la ancestral del clado. La rama que conduce a Hystricognathi y la que conduce a Caviomorpha aparentan una acumulación de sitios bajo selección positiva para el largo que presentan si se las compara, por ejemplo, con lo que ocurre en la rama que conduce al clado mouse-related o en la rama ancestral de Simiiformes.

Si bien el método de ramas y sitios implementado en CODEML y el de sitios MEME son diferentes, la existencia de concordancia en algunos sitios señalados bajo selección por ambos métodos podría tomarse como un mayor apoyo a que esos sitios presenten efectivamente huellas de selección positiva. Para IGF1R los dos sitios bajo selección según MEME, que ocurrirían en la rama ancestral de Platyrrhini, estuvieron entre los 11 sitios señalados por BEB en CODEML con  $PP \ge 0,95$  (codones 315 y 442 del alineamiento). Para INSRR, tanto CODEML como MEME identificaron tres sitios en platirrinos, siendo dos de ellos coincidentes (sitios 718 y 1008). En IGF2R, de los nueve sitios identificados por MEME con un EBF mayor a 20 en la rama ancestral de platirrinos, el sitio 2.023 coincidió con el único sitio identificado bajo selección por CODEML. De los cinco sitios de IGFBP1 detectados por MEME bajo selección en alguna rama de platirrinos, cuatro están comprendidos en los casi 20 sitios detectados por BEB (sitios 40, 74, 117 y 201). Para IGFBP4, el sitio 164 es señalado bajo selección positiva por ambos métodos (de uno y tres sitios

Tabla 3.8: I	Resulta	ngos q	e MEN	ME por	gen y I	or rama	ŗ												
Rama	Total	INS	IGF1	IGF2	INSR	IGF1R	INSRR	IGF2R	IGBP1	IGFBP2	IGFBP3	IGFBP4	IGFBP5	IGFBP6	IGFBP7	GH1	GHR	GHRH	GHRHR
Homo	6	0	0	0	0	0	-	2	0	0	0	0	0	0	0	-	0	0	2
Gorilla	11	0	0	1	0	0	1	8	0	1	0	0	0	0	0	0	0	0	0
Macaca	010						- 0	, _	- 0					0		- 0	– د		
Aotus Saimiri	° 19		0 0	0 0	20	0 0		11 0	0 -	0 0	0 0	0 0		0 0	0 0	⊣ در	- 1		
Callithrix	13	0	0	0	0	0	ω	S	0	0	0	0	0	0	0	0	-	0	4
Tarsius	17	0	0	0	1	0	0	4	1	4	0	1	0	0	0	2	2	0	2
Microcebus	39	0	0	ω	0	2	ω	18	1		0	0	0	ω	0	-	S.	. 1	1
Tupaia	28 วา	) O	0	0		<u>×</u> 12	<u> </u>	12	<b>~</b> –	- 0	0	00	> —	, o	0	0 0	<u> </u>	0	> —
Ocnotolante	2N 17	) r		<b>ں</b> د	) F	4 6		- ר		л I				) r			4 C		
Oryctotagus Ictidomys	20 24	0 0	0 0	0 U	4 0	2 ~	20	9 ~		0 0	0 0	0 0		0 ٢	0 0	2	2	0 -	
Mus	30	-	<u> </u>	2	2	1	2	12	0	-	0	0	1	0	<u> </u>	0	2	2	2
Nannospalax	16	0	0	1	з	0	0	5	0	0	0	0	0	-	1	0	з	1	1
Dipodomys	35	0	0	4	2	S	1	10	0	0	0	1	0	2	1	1	7	2	1
Fukomys	25	0	0	1	S	1	0	9	1	S	0	0	0	1	0	0	2	0	0
Heterocephalus	8	-	0	0		·	0	2	0	0	0	-	·		0	0	0	0	0
Octodon	11	0 -		- 0	<u> </u>	- 0	0 2	אש			0 0			0 1	0 1		) с	0 0	0 1
Chinchilla	20	່ນເ	0 0		0 -	0 -	- 0	<i>د</i> م	0 0	0 0	_ <	0 0	_ 0	0 0	_ 0	20	1 00	- 0	0 0
Cavia	23	0	0	2	0	2	2	S	1	1	0	0	0	1	1	1	7	0	0
Bos	24	0	0	S	1	1	з	10	0	0	0	0	0	0	0	0	ω	1	0
Equus	19	<b>—</b>	0	• 0	. 2	) <u>–</u>	2	; œ	, –	0	0	0	0		0	0	, –	• 0	. 2
canis	2 2 2			o -				ر 13	0 14					0 -			סים		
*Primates	ωı	0	0	0	0		0	21	0	0	0	0	0	0	0	0	0	0	0
2829	2	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
2930	0 13				- 0		-	∠ u	- 0					o –		0 10	<b>ა</b> თ		
*Hominidae	14	0 0	0 0	0 0		0 0		0 I	0 -	0 0	0 0	0 0	0 0	0 0	0 0	0 0	1 00	0 0	0 0
*Platyrrhini	28	0	0	0	2	2	0	9	4	<u> </u>	<u> </u>		<u> </u>	1	0	-	ω	1	<u>,                                    </u>
*Glires	8	0	0	0	2	1	0	ы	0	0	-	0	0	0	0	0	0	0	1
*Lagomorpha	12	0	0	1	2	1	1	S	0	2	0	0	0	0	0	0	0	0	0
*Rodentia	4 -	0	00	0 0	0 1	0 1	0 0	- 0	0 0	0 0	0 0	0 0	0 0	0 1	0 0	0	0	0 0	0 1
05 36 00.17	- در			0 0				<b>)</b> -											
3940	12	0 0	0 0	0 0	- 0		- 0	ωι	0 0	0 0	0 0	0 0	0 0	ω	- 0	0 0	20	0 0	0 0
*Hystricognathi	22	1	0	2	0	1	2	7	2	0	0	0	0	2	0	0	S	0	0
*Phiomorpha	9	-	-	0	1	1	1	1	0	0	0	2	0	0	0	1	-	0	0
*Caviomorpha	. 9	) C			2		4 •	2		0	0 0	0 0			0 0		0 0	0 0	° –
4344	= -				n C			ى د								- 0			
*Laurasiatheria	= :	0 0	0 0	0 0	12 (	0 0	0 0	ος 1	0 0	0 0	0 0		0 0	0 0	0 0	0,	0 0	0 0	0 0
4647	4	0	0	0	1	0	0	ω	0	0	0	0	0	0	0	0	0	0	0
Total de sitios	640	Ξ	2	28	50	40	39	252	17	17	з	7	6	25	7	19	85	Ξ	24
# codones	9902	109	153	179	1344	1367	1299	2450	246	247	202	233	270	235	198	215	630	106	419
% sitios	6,5	10,1	1,3	15,6	3,7	2,9	3,0	10,3	6,9	6,9	1,5	3,0	2,2	10,6	3,5	8,8	13,5	10,4	5,7
		-			-		-		-	•	-				-	1			-
NOTA: Para cada rama, indicándose	gen se las ram	indica e 1as segú	l total c in la Fig	le sitios <i>a</i> gura B-7.	unalizado Las ram	s (número as internas	s de codon s se identif	es) y el to ican usand	tal de sitic	s inferidos l nodos que la	bajo selecci as delimitan	ón positiva j o con el no	por MEME, mbre del cla	así como si do que defi	1 porcentaje 1en (acompa	. El resu uñado po	ıltado tar r un aste	nbién se d risco). Así	esglosa por *Primates,
*Catarrhini, *Hor 3033, 2635, 35	ninidae, 36, 35.	*Platyr .37, 40.	rhini, *0 .41, 41	Glires, *L 42, 414	.agomorj 3, 4445	pha, *Rod y 2546 c	entia, *Hy le la Figur:	stricognath a B-7, resp	ui, *Phiom ectivamen	orpha, *Cav te.	iomorpha, *	Octodontoi	dea y *Laur	asiatheria, se	correspond	en con la	as ramas	2728, 30.	.31, 3132,
3033, 2630, 30.	30, 33	.3 /, 40.	.41, 41	42, 414	3,4445	y 2546 c	le la Figuri	а в-7, resp	ectivamen	te.									



43

detectados por MEME y CODEML, respectivamente). En GHRH, el único sitio detectado usando cada modelo es coincidente (sitio 105) mientras que, en GHRHR, cuatro sitios (posiciones 205, 249, 259 y 267 del alineamiento) coinciden entre los ocho detectados por MEME y los siete por el modelo *branch-site* de CODEML. Con respecto a Caviomorpha, la cantidad de sitios inferidos bajo selección fue menor y las únicas coincidencias se dan en el sitio 53 de INS (de un total de cuatro sitios detectados por MEME y 15 por CODEML) y en la posición 185 de INSR (entre cuatro sitios inferidos por MEME en alguna rama de caviomorfos y uno por el modelo de ramas y sitios implementado en CODEML).

### 3.5.3 Relajación de la selección en Platyrrhini

RELAX permite estimar una constante k que indica el cambio en la intensidad de la selección de un grupo de ramas de prueba con respecto a otro grupo de ramas de referencia del árbol. Se concluye que la selección en las ramas de prueba se intensificó si k > 1 o relajó si k < 1, relativo a las ramas de referencia. En la Tabla 3.7 y en la suplementaria A-48 se indican los valores inferidos de k resultantes de la comparación de Platyrrhini con respecto a los restantes euarcontos (Figura 2.3 C). Para un valor de significancia de  $\alpha = 0, 1$ , los genes INS, INSR, IGF1R, INSRR, IGF2R, IGFBP1, IGFBP2, IGFBP3, IGFBP4, IGFBP5 y GHRHR mostraron cambios en la intensidad de la selección, con valores de k siempre < 1, excepto GHRHR que presentó un valor de k mayor a 20. Estos resultados son concordantes con lo observado con los primeros resultados obtenidos a partir del modelo de ramas en CODEML (comparar los valores de  $\omega_1$  con los de  $\omega_0$ y  $\omega_2$  en la Tabla 3.4). RELAX calcula también los valores del cociente  $\omega$  y proporción de stios bajo selección purificadora fuerte ( $\omega_1$ ), bajo selección purificadora débil ( $\omega_2$ ) y bajo selección positiva ( $\omega_3$ ). Los cambios en los valores de  $\omega$  entre las ramas del clado Platyrrhini y los restantes euarcontos se muestran en la Figura 3.5 y la Tabla A-48 para estos genes con valores de  $k \neq 1$ en forma significativa. Como se desprende de la figura, en todos los casos excepto GHRHR, los valores de  $\omega_1$  y  $\omega_2$  son mayores en platirrinos, mientras que el valor de  $\omega_3$  disminuye con respecto a las ramas de referencia. Aún cuando existan algunos sitios bajo selección positiva, el mayor cambio estaría asociado a una relajación de la selección purificadora. El cambio más llamativo es en el gen IGFBP1, donde los tres valores de  $\omega$  en Platyrrhini convergen a 1, según las estimaciones del modelo. En el caso de GHRHR, la selección purificadora se intensifica, pero ocurre un mayor aumento de la fuerza de la selección positiva, conduciendo a un valor final del parámetro k elevado (k = 21, 57).





45

### 3.6 Resultados específicos en Hystricognathi

### 3.6.1 Insulina

La evidencia indica que existe una regulación diferente en el crecimiento pre y post-natal en histricognatos (e.g. Mitchell et al., 1954; Wriston, 1981; Levinovitz et al., 1992; Keightley y Fuller, 1996). Con respecto a la insulina, la de histricognatos es incapaz de asociarse en hexámeros de almacenamiento, tiene menor actividad biológica (la respuesta glicémica máxima es como en otros mamíferos, aunque para dosis mayores de insulina), mayor capacidad promotora del crecimiento y se podría unir a receptores adicionales como el de PDGF (platelet-derived growth factor; Bajaj et al., 1986; Horuk et al., 1979; King y Kahn, 1981; King et al., 1983; Opazo et al., 2004; Zimmerman et al., 1974). Opazo y colaboradores (2005) obtuvieron las secuencias completas de los péptidos de insulina para 21 especies de caviomorfos, y encontraron evidencia de selección positiva en los sitios B17, B18, B20, B22 y A13. Todos ellos forman parte del segundo sitio de unión de la insulina a su receptor y están también involucrados en la formación de hexámeros en las insulinas de los otros mamíferos, por lo que su cambio contribuiría en la incapacidad de formar hexámeros así como en las diferentes propiedades biológicas de estas insulinas de histricognatos (De Meyts, 1994, 2004; Opazo et al., 2005). Los autores reportan valores heterogéneos del cociente  $\omega$  entre los diferentes linajes de caviomorfos, en concordancia con los diferentes niveles de cambio en las propiedades biológicas de las insulinas reportados (e.g. menor cambio en Chinchilloidea y mayor cambio en Octodontoidea).

Excepto el sitio B22, los restantes cuatro sitios encontrados bajo selección positiva por Opazo y colaboradores (B17, B18, B20 y A13), fueron también indicados bajo selección positiva por los análisis de ramas y sitios con CODEML realizados en este trabajo (Tabla A-22). Otros sitios indicados bajo selección positiva por el análisis de ramas y stios fueron: S9, S22, B2, B3, B4, B10, B21, B29, C14, A9 y A12. MEME reportó los sitios B29 (también en CODEML), B30 y C12 bajo selección positiva en algún linaje de Hystricognathi (Tabla 3.8). El alineamiento construido para este gen de insulina se presenta en la Figura 3.6, donde también se indica la estructura de la pre-proinsulina y los sitios identificados bajo selección positiva según MEME y CODEML (este trabajo) y Opazo et al. (2005). De los sitios detectados en este trabajo y no en Opazo et al. (2005), el sitio B10 es de especial interés. En las insulinas no histricognatas, este sitio coordina zinc e interviene en la formación de los hexámeros de almacenamiento (Bajaj et al., 1986; De Meyts, 2004). En humanos el cambio de histidina a ácido aspártico es una variante conocida (rs121918101 en dbSNP 146; Sherry et al., 2001), con efectos deletéreos predichos según SIFT (SIFT = 0,01; Kumar et al., 2009) y PolyPhen (PolyPhen = 0,934; Adzhubei et al., 2013), y evidencia de patogenicidad con hiper-proinsulinemia y superactividad (Chan et al., 1987; Schwartz et al., 1987). Es de esperar entonces que el cambio de histidina a asparagina en Octodontoidea y Cavioidea tenga consecuencias en las propiedades biológicas de estas insulinas.

Los sitios C12 y C14 pertenecen al péptido C, un segmento de conexión que tiene un papel clave en el correcto plegamiento de la proinsulina y apareamiento de las cadenas A y B en el retículo endoplasmático, previo a su escisión para la generación de la insulina madura. Entre otras restricciones estructurales, el péptido C debe ser flexible durante el plegamiento y una región central rica en glicinas lo permite (Steiner, 2004). En el plegamiento final, los péptidos C se encuentran como bucles sin estructura en el exterior de la molécula, de modo de no interferir con las interacciones entre las cadenas A y B, ni con el proceso de dimerización o hexamerización estabilizada por zinc. Los sitios C12 y C14 señalados por los análisis de selección forman parte de esta región rica en glicinas y, considerando además el alineamiento de la Figura 3.6, se puede observar la disminución del contenido de este residuo en histricognatos. En humanos se conocen variantes deletéreas para estos sitios, lo que refuerza la idea de que las sustituciones producirían un cambio en las propiedades de plegado de la prohormona, contribuyendo también en las diferentes características de las insulinas de histricognatos, incluida su incapacidad de auto-asociación en hexámeros. Finalmente, luego de la escisión, el péptido C es almacenado y liberado junto a la insulina activa habiéndose acumulado en los últimos años evidencia de posibles funciones fisiológicas para el mismo (ver bibliografía citada en Steiner, 2004; Wang et al., 2012). En el caso de confirmarse actividades endócrinas propias del péptido C, los sitios indicados bajo selección positiva deberían discutirse en el nuevo contexto (e.g. sitios de unión a un receptor aún no identificado).

Volviendo al alineamiento de la Figura 3.6, se observa cómo los diferentes linajes en Hystricognathi presentan diferentes sustituciones aminoacídicas en INS, en concordancia con lo previamente establecido (Bajaj *et al.*, 1986; Opazo *et al.*, 2005; Zimmerman *et al.*, 1974). Se confirma que no existe ninguna sustitución aminoacídica que sea característica de Hystricognathi o Caviomorpha (Horuk *et al.*, 1980). Por otra parte, en Octodontoidea la fenilalanina B24 está ausente mientras que ganaron los residuos valina y prolina en el extremo C-terminal del péptido A (Nishi y Steiner, 1990; Opazo *et al.*, 2005). Si bien el número de sustituciones en Octodontoidea no se aparta mucho de lo que ocurre en Cavioidea, estas inserciones y deleciones la convierten en la insulina más divergente del grupo, en concordancia con sus propiedades biológicas más extremas (Horuk *et al.*, 1979; King y Kahn, 1981; King *et al.*, 1983; Bajaj *et al.*, 1986). Cabe recordar que los cambios de tipo *indel* no son considerados en los modelos implementados por CODEML o MEME, por lo que el nivel de divergencia está siendo subestimado cuando se calcula  $\omega$  (Opazo *et al.*, 2005).

tolemur topaia chotona chotona sis1_mus sis2_mus elerocephalu machilla elerocephalu eleroceph	omo iorilla aimiri almiri allithrix
<ul> <li>MAXYMELLE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS - OVV REP- AGSLOP LALEVPROKED VECCT BIC GV VLEUVCI.</li> <li>MALWIR LE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS - OVV REP- AGSLOP LALEVPROKED VECCT BIC GV VLEUVCI.</li> <li>MALWIR LE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG - AGSLOP LALEVAQUEED VOOLIG I COVULEUVCI.</li> <li>MALWIR LE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG - AGSLOP LALEVAQUEED VOOLIG I COVULEUVCI.</li> <li>MALWIR LE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULEUVCI.</li> <li>MALWIR LE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULEUVCI.</li> <li>MALWIR LE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULEUVCI.</li> <li>MALWIR LE LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULE VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIGSSI LVAALV COERGEFYTP KIRKVEDS VOOLIG I COVULE ALEVAND KIRKI VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIGSSI VAALV COERGEFYTF KIRKVEDS VOOLIG I COVULE ALEVAND KIRKI VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIGSSI VAALV COERGEFYTT KIRKVEDS VOOLIG I COVULE ALEVAND KIRKI VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIG I COSTI I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIGSSI VAALV COERGEFYTT KIRKVEDS VOOLIG I COVULE ALEVAND KIRKI VOOLIG I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIG I COSTI I COVULEUVCI.</li> <li>MALWIR LI LILLIAUNGEEP PAAR VOOLIG I COSTI VAALV COERGEFYTT KIRKVEDS VOOLIG I COUCIC I COVUCITII COVULEUVCI.</li> <li>MALWIR LI LIAUNGEEP PAAR VOOLIG I C</li></ul>	*10 20 * * * 30 * 40* * 50 * 60 * 50 80 90 10* 110 MALWMRLLPLLALLALWGPDPAAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPG-AGSLQPLALEGSLQKRGIVEQCCTSICSLVQLENYCN - MALWMRLLPLLALLALWGPDPAAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPG-AGSLQPLALEGSLQKRGIVEQCCTSICSLVQLENYCN - MALWMHLLPLLALLALWGPEPAPAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQAELGGGSI - TGSLPP - LEGLLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLLPLLALLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLPP - LEGLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLLPLLALLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLPP - LEGLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLLPLLALLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLPP - LEGLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLLPLLALLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLLPLLALLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLLPLLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLLPLLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLPLLALLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN - MALWMHLPLLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN MALWMHLPLLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQAELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN MALWMHLPLLALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQVELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN MALWMHLPLLALALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQVELGGGSI - TGSLP - ALEGPLQKRGVVDQCCTSICSLVQLQNYCN MALWMHLPLLALALWGPEPAPAFVNQHLCGPHLVEALYLVCGERGFFYAPKTRREAEDPQVGQVELGGGSI - TGSLP ALEGPLQKRGVVDQCCTSICSLVQLQNYCN



### 3.6.2 IGF2

IGF2 es otra hormona peptídica vinculada a insulina, altamente conservada en mamíferos (Rotwein, 1999), con función reguladora del crecimiento y actividad metabólica similar a la de insulina y mitogénica. Su función la ejerce a través de la unión al receptor IGF1R o receptor de insulina (isoforma A del INSR, ver más adelante), mientras que su unión al IGF2R e IGFBPs afecta y regula la disponibilidad de IGF2 en la circulación. En roedores murinos IGF2 tiene actividad principal durante el desarrollo prenatal mientras que, en histricognatos, la hormona mantiene su actividad en el adulto, como ocurre en humanos (Levinovitz *et al.*, 1992). El hígado es tradicionalmente el principal productor de IGFs, por lo que es interesante notar que los niveles adultos de IGF2 son mayores en el músculo esquelético de cobayos, lo que indicaría algún rol particular (Levinovitz *et al.*, 1992).

Para todos los sitios, los tres métodos implementados en ASR coincidieron en cada codón ancestral reconstruído.

El alineamiento en aminoácidos de IGF2 se presenta en la Figura 3.7. A partir de esta visualización y de las secuencias de codones ancestrales reconstruidas (Kosakovsky Pond y Frost, 2005c; Pupko et al., 2000), se determinó que la rama ancestral de Hystricognathi acumuló 29 sinapomorfías en IGF2, siendo 18 sinónimas y 11 no sinónimas. Otro sitio (posición 163 del alineamiento) presenta una sinapomorfía para Rodentia. Por otra parte, si bien los resultados de los análisis de BSREL y aBSREL no fueron informativos, MEME detectó 11 sitios bajo selección positiva en IGF2 (Tabla 3.8). Para dos de ellos (sitios 96 y 156 del alineamiento en Figura 3.7), el evento de selección se sugiere en la rama que conduce a Hystricognathi, mientras que en otros cuatro sitios la selección habría ocurrido en al menos un linaje de histricognatos. La mayoría de las sinapomorfías y ambos sitios detectados por MEME forman parte del péptido E de la prohormona y no de la hormona madura, que incluye solo los péptidos B, C, A y D. Dos de las sustituciones, incluyendo una de las señaladas por MEME, afectan sitios sujetos a O-glicosilación (sitios 96 y 163 en este alineamiento, que se corresponden con los residuos Thr72 y Thr139 de la prohormona humana de referencia; Duguay et al., 1998). En Hystricognathi estas treoninas son reemplazadas por serinas, también factibles de ser O-glicosiladas. La acumulación de cambios en el péptido E y posibles alteraciones en su glicosilación podrían interferir con el normal procesamiento y escisión, produciendo diferentes isoformas de IGF2 de mayor largo. La existencia de estas isoformas más largas fueron primero descriptas en humanos por Daughaday y colaboradores (Daughaday et al., 1988, 1993), asociadas con la presencia de tumores y fuerte hipoglicemia. La evidencia apunta a que estas isoformas que mantienen parte del péptido E tienen poca capacidad de unión al IGF2R y de formación de complejos terciarios con IGFBPs, por lo que escaparían de las vías normales de secuestro y se mantendrían disponibles para sostener la señalización activa y potente asociada a múltiples formas de cáncer (Greenall et al., 2013).

En forma interesante, este patrón de acumulación de sustituciones en IGF2 en la rama ancestral de Hystricognathi es un fenómeno bien distinto de lo anteriormente planteado para insulina, donde





los diferentes linajes de histricognatos tienen diferentes cambios en la secuencia aminoacídica de la hormona. En el péptido glucagón sucede algo similar a lo que ocurre en insulina. Glucagón es un péptido altamente conservado en muchos mamíferos, pero en el alineamiento de la Figura B-8 se pueden observar múltiples cambios en diferentes linajes de Caviomorpha (Nishi y Steiner, 1990). Estos cambios han sido asociados a una menor afinidad a su receptor y a efectos negativos en su potencia biológica (Huang *et al.*, 1986). Mientras que en insulina y glucagón los cambios en Caviomorpha e Hystricognathi aparecen como respuestas independientes a una presión selectiva común, los cambios en IGF2 habrían ocurrido principalmente en el ancestro de Hystricognathi.

Otra característica del IGF2, junto al IGF2R, es la regulación de su expresión por *imprinting* parental, siendo expresado el alelo paterno en IGF2 y el materno en IGF2R (Haig *et al.*, 1991). Este *imprinting* recíproco habría aparecido con los mamíferos terios y estaría asociado a la evolución estructural del receptor y a un aumento en su afinidad por el IGF2 (Williams *et al.*, 2012). El *imprinting* no ocurre por igual en todos los placentados: mientras que IGF2 tiene *imprinting* tanto en *Mus* como en humano, el de IGF2R se presenta siempre en *Mus* pero constituye un caracter polimórfico en humanos (Xu *et al.*, 1993; Killian *et al.*, 2001). No se encontraron datos sobre *imprinting* del par IGF2 e IGF2R en histricognatos.

IGF2R es una glicoproteína multifuncional (Morgan *et al.*, 1987; Roth, 1988) con un gran dominio extracelular integrado por 15 repetidos homólogos, siendo el repetido 11 el de unión al IGF2 (Schmidt *et al.*, 1995). La formación de un homodímero así como la cooperación entre el repetido 11 y el 13 estabilizan dicha unión (Linnell *et al.*, 2001; Williams *et al.*, 2007). Los análisis realizados con MEME señalaron para IGF2R múltiples sitios bajo selección positiva, 43 de los cuales tienen cierta evidencia de estar asociados a cambios en alguna de las ramas de Hystricognathi (Tabla 3.8 y Figura 3.4). En una revisión muy rápida de la salida de MEME se encontró que solo dos de estos sitios pueden asociarse a los repetidos 11 y 13 que intervienen en la unión a IGF2. Usando la secuencia humana como referencia (NP\_00867.2), el sitio 1.687 del repetido 11 cambia de asparagina a serina en la rama que conduce a Octodontoidea, mientras que el sitio 2.030 del repetido 13 cambia en Histricognathi, siendo el patrón de cambio más complejo (resultados no mostrados). Una evaluación más profunda es requerida para determinar si alguno de los sitios detectados por MEME tiene la posibilidad de modificar la actividad secuestradora de IGF2R y por lo tanto la biodisponibilidad de IGF2 en estos roedores.

### 3.6.3 Receptor de insulina

El receptor de insulina es una glicoproteína con actividad tirosina quinasa que media la transducción de señales de la insulina y del IGF2, y también del IGF1. Comparte una estructura proteica conservada con los otros dos miembros de la familia génica, el IGF1R y el INSRR

(Hernández-Sánchez *et al.*, 2008), y juntos forman parte de la superfamilia de los receptores tirosina quinasa (Hubbard y Till, 2000). Estos tres receptores forman en las membranas celulares estructuras diméricas  $\alpha 2/\beta 2$  que pueden presentarse tanto como homodímeros, así como heterodímeros formados por mónomeros  $\alpha/\beta$  diferentes (*e.g.* INSR/IGF1R). La unión del ligando al dominio extracelular desencadena un cambio conformacional que permite la autofosforilación de residuos de tirosina del dominio citoplasmático del receptor, iniciándose la cascada de señales intracelulares que conducen a las diversas respuestas biológicas (metabólicas, de proliferación celular, diferenciación y crecimiento), según sea el tipo celular y el estado de desarrollo y funcional.

Las cadenas  $\alpha$  y  $\beta$  son traducidas de un único transcripto de 22 exones en INSR e INSR y 21 exones en IGF1R. Tanto en INSR como en INSRR el exón extra es el 11, aunque presenta un origen independiente. El exón 11 en INSRR aparece ya en anfibios mientras que, el exon 11 en INSR se encuentra exclusivamente en mamíferos (Hernández-Sánchez et al., 2008). En INSRR este micro-exón se expresa en forma constitutiva, pero en INSR sufre splicing alternativo, resultando dos isoformas (INSRA e INSRB) que difieren en la presencia o ausencia de 12 aminoácidos en el extremo C-terminal de la unidad  $\alpha$  (la forma INSRB es la que incluye el exón 11; Seino y Bell, 1989). Las isoformas presentan diferencias en la afinidad del ligando, en la actividad quinasa, en la internalización del receptor y en la distribución tisular (McClain, 1991; Vogt et al., 1991; Kellerer et al., 1992; Frasca et al., 1999; Leibiger et al., 2001; Uhles et al., 2003). Por ejemplo, ambas isoformas unen insulina en forma similar, pero la isoforma B tiene mucha menor afinidad para el IGF2, por lo que es un receptor más específico para insulina (McClain, 1991; Frasca et al., 1999). INSRB se expresa predominantemente en tejidos blanco de insulina e involucrados en la homeostasis de glucosa (e.g. en ratón el tejido adiposo y el músculo expresan ambas isoformas, en el hgado solo la INSRB y en el cerebro se detecta solo la INSRA; Hernández-Sánchez et al., 2008). Esta nueva isoforma del INSR en mamíferos aparece como un receptor especializado en roles metabólicos, al mismo tiempo que la diversificación funcional contribuye en una regulación más precisa de la señalización por IGF2 (recordar que en mamíferos también aparecen otras dos novedades con respecto a IGF2 como lo son el imprinting materno del gen y el nuevo receptor de unión IGF2R).

McGaugh *et al.* (2015) extienden la función IGF2R a amniotas

A pesar de las anomalías que presentan sus insulinas, los roedores histricognatos son capaces de regular su glicemia como los restantes mamíferos (Opazo *et al.*, 2004; Vera *et al.*, 2008), lo que ha conducido a hipotetizar sobre el desarrollo de cambios compensatorios en el receptor de insulina de este grupo de roedores. Sin embargo, son escasos los trabajos sobre estos receptores en histricognatos y los hallazgos poco concluyentes. Experimentos de afinidad de unión de insulina a membranas mostraron que la insulina bovina tiene una similar afinidad por los INSR de histricognatos que por los receptores de insulina de otros mamíferos (Horuk *et al.*, 1979; Zimmerman *et al.*, 1974), aunque ensayos con análogos de insulina mostraron diferentes

propiedades entre los INSR de histricognatos y otros mamíferos, y a su vez entre los INSR de chinchillas y cobayos (Horuk *et al.*, 1979). Ensayos con anticuerpos para la subunidad  $\alpha$  reportan la precipitación del INSR de *C. porcellus* y estimulación de lipogénesis en adipocitos, lo cual no ocurre en *Chinchilla* (Tong *et al.*, 1994). Ensayos inmunológicos para la subunidad  $\beta$  no mostraron diferencias, en concordancia con la alta constancia evolutiva en el mecanismo de la transducción de señales (Muggeo *et al.*, 1979). Los cambios compensatorios descriptos con mayor claridad consisten en el incremento, para un tejido dado, del número de receptores INSR en cobayo con respecto a otros mamíferos (Muggeo *et al.*, 1979). Esto se enmarca en una relación inversa, a nivel de vertebrados, entre la actividad biológica de la insulina y el número de receptores presentes (Dufty *et al.*, 2002; Muggeo *et al.*, 1979).

De los taxa estudiados en este trabajo, F. damarensis, H. glaber y C. lanigera tienen predicciones génicas para las isoformas corta y larga en NCBI. C. sociabilis no cuenta con borrador genómico mientras que C. porcellus presentaría una deleción en la región homóloga al inicio del intrón 11 del gen, lo cual interferiría la señal del sitio dador para el proceso de *splicing* (Figura B-9). En el caso de O. degus, la secuencia intrónica está conservada sobre el posible sitio dador del intrón 11 (Figura B-9), pero la isoforma larga no ha sido predicha en los procesos automáticos de anotación génica. Además de los sitios dadores y aceptores de splicing, otros elementos silenciadores o potenciadores de la inclusión del exón fueron reportados en humano y ratón en los intrones 10 y 11 (Kosaki et al., 1998; Sen et al., 2009), pero los mismos no fueron estudiados en el presente trabajo. Para evaluar la expresión del INSRB, partiendo de NP 000199.2 (INSRB en humano), se realizaron búsquedas por BLASTP y TBLASTN sobre los repositorios nr, nt, ESTs y TSA del NCBI (secuencias proteicas, nucleotídicas, Expressed Sequence Tags y Transcriptome Shotgun Assembly, respectivamente). Solo se encontró evidencia de expresión de la isoforma B en C. lanigera, en músculo esquelético, riñón, mucosa nasofaríngea y mucosa tubotimpanasal (estos tejidos también expresan la isoforma A). Con respecto a los transcriptos ensamblados de novo de C. sociabilis (muestra de hipocampo; MacManes y Lacey, 2012), se encontraron *contigs* correspondientes a ambas isoformas. En el caso de C. porcellus, luego de que la búsqueda en los repositorios recién mencionados fuera negativa, se bajaron secuencias transcriptómicas de pulmón y músculo esquelético del SRA del NCBI (Davidsen et al., 2014), se mapearon los reads y se ensamblaron los transcriptos. Como se puede observar en una visualización en IGV de los reads alineados y de los modelos transcriptómicos (Figura B-10), a pesar del alto nivel de expresión del gen (FPKM = 535 y FPKM = 926 en pulmón y músculo esquelético, respectivamente), solo se constata la expresión de la isoforma A que no incluye el exón alternativo, incluso en músculo esquelético, un tejido que expresa INSRB en otros mamíferos. La evidencia actual indicaría que la expresión de la isoforma metabólica del INSR en Hystricognathi es linaje dependiente y/o que en algunos de los linajes su expresión está muy acotada con respecto al estado de desarrollo y fisiológico del animal.

Con respecto al uso de modelos de sustitución de codones en las secuencias codificantes recuperadas, el sitio 185 del alineamiento fue indicado bajo selección positiva por los modelos de ramas y sitios implementados por CODEML, cuando se seleccionó la rama que conduce a Caviomorpha como la rama blanco (Tabla A-25). Para este sitio los caviomorfos sustituyen isoleucina con alanina, otro aminoácido no polar de menor tamaño. Este sitio corresponde a I222 del INSR humano, se encuentra en el dominio extracelular de tipo furina rico en cisteínas (De Meyts, 2004, 2008) y en principio no se corresponde con ninguno de los residuos descriptos como críticos para el funcionamiento del receptor. Los análisis con MEME indicaron 22 sitios bajo selección episódica para INSR, y en 11 de los casos, las ramas sugeridas bajo selección incluyen un histricognato (Tabla 3.8). El sitio 185 del alineamiento también fue detectado por MEME, así como otros seis sitios donde el cambio se produce en la rama que conduce a Octodontoidea (sitios 382, 467, 528, 537, 579 y 745 del alineamiento, correspondientes a los residuos T419, Y504, L565, H574, Q616 y T683 de la secuencia NP\_000199 humana). Si se considera que Octodontoidea presenta la insulina más divergente del grupo y las propiedades biológicas más extremas (Horuk et al., 1979; King y Kahn, 1981; King et al., 1983; Bajaj et al., 1986), este hallazgo podría señalar la existencia de cambios compensatorios en el receptor. Estudios en profundidad que integren análisis con modelos estructurales y/o de cambios coevolutivos deberán implentarse en el futuro.

## 4 Discusión

En este capítulo se extiende la discusión en otros aspectos relevantes del trabajo, uno metodológico y otros específicos a los resultados. En particular se extiende la discusión sobre insulina y receptor de insulina en Hystricognathi / Caviomorpha que fue el objetivo inicial de este trabajo. También se incluye una breve discusión sobre los resultados obtenidos para Platyrrhini con respecto a la bibliografía disponible, así como un comentario final sobre los patrones de evolución en Caviomorpha y Platyrrhini.

### 4.1 Ortólogos y alineamientos

En esta tesis se analizaron patrones moleculares de selección en 18 genes para 24 *taxa* determinados *a priori*. Como se describe en el capítulo de Metodología, el trabajo manual en la construcción y edición de los alineamientos permitió: trabajar con secuencias que no estaban directamente accesibles en las bases de datos, reemplazar datos faltantes con secuencias provenientes de otra especie cercana/equivalente, eliminar secuencias muy divergentes, reemplazar isoformas y corregir algunas secuencias "quiméricas" (resultantes de un incorrecto ensamblado y/o anotación). Una vez obtenidos los alineamientos nucleotídicos, los mismos se filtraron combinando el uso de Gblocks (Castresana, 2000; Talavera y Castresana, 2007) y BMGE (Criscuolo y Gribaldo, 2010), manteniendo la estructura de codones.

Si bien aquí se analizaron solo 18 genes, el análisis de una vía de señalización como la IIS/TOR (*insulin/insulin-like signaling and target of rapamycin*) o INS/FOXO (*insulin/forkhead box O*), comprende el estudio de 50 a 70 genes (Alvarez-Ponce *et al.*, 2011; Wang *et al.*, 2012; McGaugh *et al.*, 2015). El análisis conjunto de las vías de señalización y transducción conjunta para INS, IGFs y GH involucra alrededor de 250 genes (*e.g.* Sano *et al.*, 2016) mientras que, el análisis de estas vías en conjunto con otros genes asociados a longevidad, inmunidad innata u otros contextos, implica el estudio de más de 300 genes (*e.g.* GenAge, http://genomics.senescence.info/genes/; de Magalhães *et al.*, 2009). Por otra parte, el incremento en la disponibilidad de datos transcriptómicos y genómicos conlleva al manejo de un número mayor de *taxa* por grupo de ortólogos. En definitiva, para la continuación del presente
trabajo, se vuelve un requisito la elección de un sistema más automatizado para la construcción de los grupos de ortólogos y alineamientos.

Por otra parte, diferentes autores han discutido la importancia del alineamiento en los análisis posteriores de selección (Wong et al., 2008; Mallick et al., 2009; Schneider et al., 2009; Fletcher y Yang, 2010; Markova-Raina y Petrov, 2011; Kosiol y Anisimova, 2012; Jordan y Goldman, 2012; Privman et al., 2012; Bielawski et al., 2016). PRANK en el modo empírico de codones (Löytynoja y Goldman, 2005; Kosiol et al., 2007; Löytynoja, 2014) ha resultado el que genera menos falsos positivos; otras herramientas tienden a "sobrealinear" colocando codones no homólogos en la misma columna (produciendo alineamientos más cortos), de modo que el mayor número de cambios artificialmente generado en esos sitios conduce a falsos positivos en los análisis de selección (Fletcher y Yang, 2010; Jordan y Goldman, 2012; Privman et al., 2012). El nuevo algoritmo en Clustal Omega también parece tener un buen desempeño y para secuencias más divergentes podría ser preferible frente a PRANK (Sievers et al., 2014). Además, es recomendable realizar un filtrado posterior de los alineamientos siendo ALISCORE, GUIDANCE, MaxAlign, PSAR, StatSigMA-w, SWAMP y trimAl, algunas de las herramientas disponibles (Gouveia-Oliveira et al., 2007; Capella-Gutierrez et al., 2009; Chen y Tompa, 2010; Kück et al., 2010; Penn et al., 2010; Kim y Ma, 2011; Harrison et al., 2014; Sela et al., 2015). De estas, se resalta GUIDANCE que evalúa qué tanto cambia el alineamiento frente a perturbaciones y en su modo residuo específico permite enmascarar solo los residuos poco confiables, en lugar de eliminar columnas enteras del alineamiento. También se destaca SWAMP, que recorre el alineamiento múltiple con ventanas deslizantes encontrando pequeñas regiones con un número de sustituciones no sinónimas inesperadamente alto, de modo que detecta la presencia de secuencias sospechosas de error (denominadas previamente como quiméricas en este trabajo; Harrison et al., 2014). La observación general es que para el alineamiento filtrado, la reducción de resultados falsos positivos (improved accuracy) compensa el posible aumento en falsos negativos (reduced power; Fletcher y Yang, 2010; Jordan y Goldman, 2012; Privman et al., 2012).

Sin lugar a dudas, los grupos de ortólogos y los alineamientos finales dependerán principalmente de los propios datos iniciales. Por ejemplo, durante la obtención de las secuencias para formar los grupos de ortólogos, aún sin considerar cuestiones de calidad de ensamblados, los datos genómicos con respecto a los transcriptómicos tienen la ventaja de permitir incluir información de sintenia en la determinación de las relaciones de homología. Además, para los datos transcriptómicos, la no expresión de un gen en un tejido puede conducir a incluir erróneamente un parálogo poco divergente que sí se exprese. Sin embargo, debido al menor costo y dificultad de ensamblado, es esperable que sean cada vez más las secuencias de origen transcriptómico en los datos a ser analizados. Por otra parte, la definición del grupo de ortólogos y el alineamiento final dependerán en gran medida de las características de cada gen: pertenencia o no a familias multigénicas, pérdida de parálogos en algunos linajes, velocidad de cambio evolutivo, estructura proteica, etc. A modo

de ejemplo, se sabe que las regiones desordenadas de las proteínas tienden a diverger más rápido y generan dificultades en la etapa de alineamiento, por lo que son esperables mayores errores en los alineamientos y sesgos en la etapa posterior de detección de selección en genes que codifiquen para proteínas con estas regiones desordenadas (Markova-Raina y Petrov, 2011; Jordan y Goldman, 2012).

Un nuevo pipeline de construcción de ortólogos y alineamientos podría incluir tres etapas diferentes. En una primera etapa, se propondría partir de un grupo amplio de secuencias por gen y usar herramientas como PALO (Villanueva-Cañas et al., 2013) o IsoSel (Philippon et al., 2017), que optimizan la selección de las isoformas más equivalentes entre sí (en lugar de seleccionar simplemente las secuencias más largas que suele conducir a la inclusión de exones especie-específicos o de baja frecuencia). En una segunda etapa, para este grupo de ortólogos filtrado, las secuencias aminoacídicas serían alineadas con MUSCLE seguido de una herramienta como SWAMP para identificar secuencias problemáticas remanentes (e.g. secuencias con regiones equivocadas por problemas de anotación). Las secuencias problemáticas deberían ser reemplazadas cada vez que sea posible o en su defecto, eliminadas. Para el grupo de ortólogos refinado se repetiría el duo MUSCLE - SWAMP hasta obtener un conjunto final de secuencias. En una tercera etapa, las secuencias nucleotídicas correspondientes serían alineadas con PRANK en modo codones (o Clustal Omega) y luego los alineamientos filtrados (filas con alto porcentaje de *indels* eliminadas, codones enmascarados) con herramientas como trimAl y GUIDANCE2. En paralelo, para las secuencias aminoacídicas, se podría estimar la presencia de regiones desordenadas e indentificar a priori alineamientos que requieran una mayor evaluación manual (e.g. DisEMBL; Linding et al., 2003). Recientemente se ha desarrollado PosiGene, un pipeline automatizado para la detección a escala genómica de genes con señales de selección positiva (Sahm et al., 2017), y podría evaluarse su utilidad en este estudio particular.

### 4.2 INS e INSR en Hystricognathi / Caviomorpha

otros autores proponen que los cambios se deben a relajación de la selección, luego de perderse la capacidad de formar hexámeros de almacenamiento (Beintema y Campagne, 1987).

 *al.*, 2008), la disminución de las funciones metabólicas en histricognatos representaría, desde este punto de vista, una reversión a una condición de tipo ancestral.

A pesar de la menor actividad biológica de la insulina con respecto a su función de reducir la glucosa en sangre, los histricognatos son capaces de regular la glicemia (al menos en condiciones fisiológicas normales para los animales; Opazo *et al.*, 2004; Vera *et al.*, 2008), para lo cual presentan diferentes estrategias compensatorias. Estas incluyen hiperinsulinemia (Zimmerman *et al.*, 1974; Kind *et al.*, 2003), aumento en la cantidad de receptores de insulina (Muggeo *et al.*, 1979) y cambios en la cantidad de transportadores de glucosa (Horuk *et al.*, 1983). La hiperinsulinemia puede ser consecuencia del incremento en la producción de insulina y/o de la disminución de su tasa de desaparición y degradación, fenómenos que también ocurren en pacientes humanos con insulinas mutantes que muestran afinidad alterada por el receptor (Haneda *et al.*, 1984).

Con respecto a los cambios a nivel de secuencia del INSR, estos se concentran en el dominio extracelular y son sutiles, ya que el receptor mantiene la capacidad de unir insulinas de otros mamíferos con alta afinidad (Zimmerman *et al.*, 1974), y los ensayos immunológicos no muestran grandes diferencias (Zimmerman *et al.*, 1974; Horuk *et al.*, 1979; Tong *et al.*, 1994). De todas formas, pequeños cambios en la secuencia aminoacídica podrían ser suficientes para desencadenar efectos mayores en la transducción de la señal. Pequeños cambios en la cinética de activación y/o en la trayectoria de internalización, pueden conducir a la activación alternativa de diferentes efectores de las vías de señalización (Hansen *et al.*, 1996; Jensen *et al.*, 2008; Jensen y De Meyts, 2009). De hecho se reconoce a la INSRA como "más mitogénica" que la isoforma INSRB, cuando solo se diferencian por la secuencia extra en la porción C-terminal de la subunidad  $\alpha$  extracelular (Belfiore *et al.*, 2009; Giudice *et al.*, 2011). Cabe considerar además, que el efecto último de mutaciones puntuales en las propiedades del receptor dependerán de la identidad y abundancia de las isoformas expresadas (Whittaker *et al.*, 2002).

Otros cambios compensatorios en Hystricognathi podrían incluir la modificación de los patrones de expresión de las isoformas A y B del receptor de insulina, pero como se describe en este trabajo, la evidencia al momento indicaría que el cambio es en sentido contrario (*i.e.* una disminución en la expresión de la isoforma larga con función metabólica). No existe evidencia de que las diferentes isoformas de INSR o el IGF1R sean capaces de desencadenar alguna acción exclusiva, por lo que las diferentes señales mitogénicas o metabólicas que desencadenan serían reflejo de una combinación entre la eficiencia relativa en la activación de ambos tipos de señales y los niveles de expresión de los respectivos receptores (Belfiore *et al.*, 2009; Boucher *et al.*, 2010; Siddle, 2012). Para el caso del INSR en humano, son múltiples las evidencias de una asociación entre hiperinsulinemia, sobre-expresión de INSR, sobre-expresión de la isoforma A y el desarrollo de cáncer, aterosclerosis y envejecimiento (Belfiore, 2007; Belfiore *et al.*, 2009), de modo que

se puede inferir que la alteración de la relación entre INSRA e INSRB por deficiencia en la producción y expresión de la isoforma B en Hystricognathi, tendría consecuencias fenotípicas.

Finalmente, se pueden producir cambios compensatorios en otros genes. Por ejemplo, se sabe que el glucagón en histricognatos presenta también menor actividad (Huang *et al.*, 1986). Además, están los reportes de la unión de insulina de *C. porcellus* a receptores de PDGF (King *et al.*, 1983), aunque esto también apuntaría en la dirección de incrementar más las funciones mitogénicas que las metabólicas de la hormona. De todas formas, no existen hallazgos más recientes sobre este fenómeno y en principio las funciones promotoras del crecimiento de la insulina estarían ocurriendo principalmente a través de su propio receptor.

A pesar de la insulina divergente y de que existen, como se mencionó recién, compensaciones varias que permiten una adecuada regulación del metabolismo de la glucosa, no hay un único patrón de cambio molecular en insulina o glucagón para los diferentes linajes de histricognatos analizados. Esto podría sugerir la existencia de una presión selectiva común, pero que el desencadenante sería ajeno a estas dos hormonas. En 1980, Horuk y colaboradores proponían:

"We speculate that hystricomorph rodents, both New and Old World, have low potency insulins because of a common evolutionary pressure. This pressure has resulted in a reduction of the role of insulin as a modifier of glucose metabolism, as well as a change in the insulin receptor. We speculate that another function of insulin has been consequently upgraded. One such function could be its mitogenic activity."

En este trabajo solo 18 genes de las vías de señalización en las que interviene la insulina fueron analizados, y un número aún menor estudiado con mayor detalle. Sin embargo, en el contexto de todas las anomalías endócrinas y fisiológicas reportadas para *C. porcellus* (Keightley y Fuller, 1996), los cambios a nivel de secuencia detectados en el IGF2 presentan un patrón consistente con una única ocurrencia en la historia evolutiva de los histricognatos. Al mismo tiempo, se detectaron algunos cambios adaptativos en el receptor de insulina y se infirieron cambios en los patrones de expresión de INSRA e INSRB. Como el IGF2 desencadena su señal a partir de la unión a receptores de INSR (por unión a dímeros que incluyan la isoforma INSRA; Belfiore, 2007; Belfiore *et al.*, 2009), los cambios en el receptor de insulina y en las actividades mitogénicas mencionadas por Horuk *et al.*, podrían constituir una respuesta a la presión vía IGF2. Esta hipótesis de que las insulinas histricognatas respondan a cambios iniciales en somatomedinas (IGF1 e IGF2), PDGF o sus receptores, fue previamente planteada por Chan y colaboradores (1984).

recordar que IGF2 ejerce su función por unión a dímeros IGF1R, INSR e INSR/IGF1R, no así a través del IGF2R.

Recientemente, Fang *et al.* (2014) estudiaron mediante genómica comparativa y transcriptómica, las adaptaciones a la vida subterránea y longevidad en los histricognatos africanos

*H. glaber* y *F. damarensis*. Entre otros puntos, los autores discuten la insulina altamente divergente en Hystricognathi y especialmente, el sitio 22 de la cadena B que cambia en estos roedores (ver Figura 3A en Fang *et al.*, 2014) y que ya fuera detectado bajo selección positiva por Opazo *et al.* (2005). Como también detectan expresión de IGF2 e IGFBP2 en hígado de individuos adultos, los autores hipotetizan sobre la evolución de una insulina diferente, acompañada por la secreción autócrina-parácrina de IGF2 en el hígado, que sustituiría en funciones a la insulina, mediante un manejo del nivel de glucosa similar al modo fetal. Si bien la discusión del vínculo INS-IGF2 en histricognatos coincide con lo planteado en los párrafos anteriores, la hipótesis de Fang y colaboradores parece apuntar a un cambio inicial más a nivel de insulina que de IGF2.

De todas formas, será necesario esperar a analizar todos los genes involucrados en estas vías de señalización para obtener la visión completa y un mejor indicio de la respuesta correcta. Por ejemplo, resta evaluar en profundidad el rol de la hormona de crecimiento en estos roedores. Se sabe que en Hystricognathi el crecimiento post-natal es independiente de IGF1, que las estructuras de GH y GHR son conservadas y que la hipófisis produce y libera GH en forma normal (Adkins *et al.*, 2000). Esto hace que se considere que el crecimiento corporal sea independiente de GH, pero no queda clara la importancia de esta hormona en el control metabólico de histricognatos, aún cuando es conocido su rol directo en la señalización de insulina (Clemmons, 2004; Xu y Messina, 2009). También será necesario esperar a tener una mejor cobertura taxonómica del grupo de forma de evitar inferencias erróneas, especialmente aquellas que se relacionen a la distribución filogenética de los sitios de interés (Davies *et al.*, 2015; Delsuc y Tilak, 2015; Thomas *et al.*, 2017).

#### 4.3 Patrones en Platyrrhini

En esta sección se discuten los hallazgos del presente trabajo en el contexto de la bibliografía relevada. Siete genes en Platyrrhini (IGF2, IGF1R, INSRR, IGF2R, IGFBP1, GHRH y GHRHR) mostraron consistentemente señales de selección positiva (Tabla 3.7). Para GHRH y GHRHR no se encontró información adicional en la bibliografía, pero los restantes genes han sido incluidos anteriormente en análisis de selección (Wallis, 2009, 2015). De estos, IGF1R ha presentado clara y reiterada evidencia de selección positiva en la rama ancestral de Platyrrhini (Worley *et al.*, 2014; Wallis, 2015). El otro gen con clara evidencia de selección positiva en la misma rama es IGFBP1 (Wallis, 2015). Si bien fueron analizados los sitios bajo selección, no fue evidente asociación alguna con aspectos funcionales de estas proteínas (Wallis, 2015), aunque para el caso de IGF1R Worley *et al.* (2014) proponen que la afinidad en la unión receptor-ligando estaría afectada. En la búsqueda de selección positiva a escala genómica, estos autores encuentran otros cuatro genes que habrían sufrido cambios adaptativos en la rama de *Callithrix* y que forman parte de los ejes GH/IGFs: IGF2, IGFBP2 e IGFBP7 y GHSR (*growth hormone secretagogue receptor*). Se plantea

que estos cambios estarían asociados a la miniaturización corporal característica de Callitrichinae (Worley *et al.*, 2014). En un trabajo complementario, Harris *et al.* (2014) secuencian los exones de estos cinco genes (IGF1R, IGF2, IGFBP2, IGFBP7 y GHSR) en primates adicionales, incluyendo otros callitrichinos, cébidos no callitrichinos y atélidos. A partir de los alineamientos resultantes, encontraron que ocho cambios no-sinónimos eran específicos de Callitrichinae, mientras que 131 eran específicos de Platyrrhini (Harris *et al.*, 2014), aumentando la probabilidad de que los episodios de selección positiva hayan afectado globalmente los patrones de crecimiento de todos los monos neotropicales y no exclusivamente los de los callitrichinos.

Basándose en la secuencia de insulina de *Aotus trivirgatus*, Seino y colaboradores (1987) la describen como la más divergente dentro de Primates y reportan su notoria menor afinidad de unión ligando-receptor. En otros trabajos posteriores, tanto INS como IGF1 son identificados como blancos de evolución adaptativa episódica en la rama que conduce a Platyrrhini (Wallis, 2009). Este resultado se contradice totalmente con los obtenidos en esta tesis, dado que ni INS ni IGF1 resultaron señalados bajo selección positiva por ninguno de los métodos aquí empleados (Tabla 3.7). Si bien existe cierta diferencia en los muestreos taxonómicos empleados (*e.g.* Wallis solo dispone de secuencias de *C. jacchus* para platirrinos), la diferencia principal radica en el incorrecto uso e interpretación de los modelos de selección en CODEML. Wallis basa sus discusiones en el hallazgo de valores de  $\omega$  significativamente mayores en la rama de *Callithrix* con respecto al del *background*, pero ambos valores son siempre menores a 1. Viendo el alineamiento de INS en la Figura 3.6 se puede encontrar algún cambio no-sinónimo exclusivo de Platyrrhini (*e.g.* sitio 92 correspondiente al residuo A2), pero no grandes divergencias en las secuencias en comparación con las restantes insulinas de primates. Claramente estos genes y la bibliografía disponible requieren una evaluación en mayor profundidad.

Con respecto a la hormona de crecimiento en Platyrhini, es extensa la bibliografía sobre la duplicación del locus de GH en antropoideos, con expansiones independientes en platirrinos y catarrinos, así como ocurrencia de abundante conversión génica (Chen *et al.*, 1989; Wallis *et al.*, 2001; Li *et al.*, 2005; González Alvarez *et al.*, 2006; Petronella y Drouin, 2011). La comparación del locus de GH entre *C. jacchus* y *Cebus albifrons* muestran expansión adicional en esta última especie, consistiendo el locus en unos 40 genes y pseudogenes. De este modo, incluso dentro de Platyrrhini (o más precisamente, dentro de Cebidae), el cluster de genes de GH es altamente variable (Wallis y Wallis, 2006). Si bien las relaciones de co-ortología son complejas, es identificable el gen de expresión hipofisiaria o GH1 (ver análisis en Premzl, 2015). Este gen es muy conservado en la mayoría de los mamíferos, incluyendo primates strepsirrinos, pero sufrió un proceso de evolución adaptativa en la rama que conduce a antropoideos y que precedió los eventos de duplicación en Platyrrhini y Catarrhini (Ohta, 1993; Wallis *et al.*, 2001). Durante este proceso, 35% de la secuencia aminoacídica resultó alterada.

Los análisis de RELAX mostraron cambios significativos en la intensidad de la selección de platirrinos con respecto a los restantes euarcontos. Los valores de k para INS, INSR, IGF1R, INSRR, IGF2R, IGFBP1, IGFBP2, IGFBP3, IGFBP4 e IGFBP5 resultaron menores a 1, indicando relajación de la selección purificadora (Tabla 3.7 y Figura 3.5). RELAX confirmó la observación inicial de que el valor estimado de  $\omega$  en los análisis de rama con CODEML fue mayor en platirrinos que en caviomorfos o en los restantes linajes (Tabla 3.4). Una posibilidad es que este patrón sea particular a estos genes. En este sentido, los genes están funcionalmente relacionados y además todos codifican para proteínas extracelulares o con dominios extracelulares, habiéndose ya discutido la correlación positiva entre extracelularidad y mayor tasa de evolución molecular (Julenius y Pedersen, 2006; Liao et al., 2010; Feyertag et al., 2017; Sojo et al., 2016). Sin embargo, las tasas de evolución dependen también de la interacción de factores como el coeficiente de selección, el tamaño poblacional efectivo y la tasa de mutación (Ohta, 1987; Lanfear et al., 2014). Por ejemplo, en linajes con tamaños poblacionales menores, las variantes pueden fijarse debido a procesos estocásticos, aún si son levemente deletéreas para los organismos. Este tipo de fuerzas demográficas dejarían una huella extendida en el genoma, por lo que el estudio de la relajación de la selección en muchos más genes (incluyendo elementos genómicos no codificadores de proteínas) serían reveladores en aportar una explicación. Anteriormente estudios a escala genómica en mamíferos han confirmado una correlación positiva entre la tasa evolutiva (medida como  $\omega$ ) y la masa corporal (usada como un *proxy* del tamaño poblacional efectivo; Figuet et al., 2016).

Wildman y colaboradores reportan diferentes tasas evolutivas en los diferentes linajes de Platyrrhini (Jameson Kiesling et al., 2015). En particular, la tasa de sustitución nucleotídica en Callitrichinae es significativamente mayor que en los restantes platirrinos (medida como sustituciones por sitio por año). Los autores asocian este incremento a su modo de vida "más rápido": estos pequeños primates tienen una edad de primera reproducción de 21,5 meses e intervalos entre nacimientos de siete meses, lo cual es mucho menor que en los restantes monos del Nuevo Mundo (en promedio 48,4 meses y 17,6 meses, respectivamente; Jameson Kiesling et al., 2015). Es decir, consideran que los menores tiempos generacionales pueden están asociados a la mayor tasa evolutiva de la subfamilia, vía un incremento en la tasa mutacional. Si bien son de interés, estos aportes de Jameson Kiesling et al. (2015) no son comparables con los resultados de esta tesis. Mientras ellos tienen una gran representación taxonómica de Callitrichinae, así como otros cébidos, atélidos y pitécidos, en este trabajo los platirrinos están representados por un único miembro de Callitrichinae (C. jacchus) y otros dos integrantes de la familia Cebidae. Por otra parte, los cambios en la tasa evolutiva mediados por un incremento de cambios mutacionales se verían reflejados como un aumento en la estimación de las tasas de sustitución sinónimas y por lo tanto una disminución en los valores de  $\omega$ , algo contrario a lo observado en los resultados de esta tesis.

#### 4.4 Caviomorpha versus Platyrrhini en el Nuevo Mundo

En 1975 Blundell y Wood proponían que el aislamiento conjunto que caviomorfos y platirrinos experimentaron en el Neotrópico sugería la existencia de un único factor de selección como la causa de las altas tasas de divergencia en las insulinas de estos mamíferos. Los autores elegían la escasez de zinc local como este factor común, idea que surge de la observación de la sustitución de la histidina B10, que en los islotes  $\beta$  pancreáticos aporta los ejes de unión al zinc requeridos para el almacenamiento en hexámeros. Posteriormente Chan et al. (1984) descartaban esta posibilidad. En primer lugar, el zinc presenta distribución geográfica ubicua y es componente esencial de muchas enzimas. En segundo lugar, la insulina de un histricomorfo africano (Hystrix), también es monomérica y de baja actividad metabólica, a pesar de mantener la histidina en la posición B10. En 1987, Seino y colaboradores estudian la insulina de Aotus trivirgatus, encontrando que también retiene la histidina B10 y que posee propiedades de unión a zinc normales. Insisten en descartar el zinc como factor selectivo, pero retoman la idea de la causa común para las insulinas divergentes. Como propuesta alternativa plantean la existencia de otro factor ambiental como la presencia de parásitos con requerimientos de insulina. Si un parásito presentase un receptor similar al receptor de insulina para incrementar la captación de nutrientes, la alteración de las insulinas de los caviomorfos y platirrinos hospederos resultaría en una estrategia evasiva frente al parásito.

en 1986 Rodríguez y Jungery descubren que *Plasmodium falciparum* produce un receptor de transferrina en el eritrocito

El rol de patógenos y parásitos en la evolución adaptativa de los organismos es bien conocido (*e.g.* Jiggins y Kim, 2007; Barreiro y Quintana-Murci, 2010) y resultan cada vez más importantes los vínculos entre inmunidad y metabolismo (*e.g.* McNelis y Olefsky, 2014; Lackey y Olefsky, 2015). Sin embargo, parece poco probable una hipótesis tan específica como la de Seino *et al.* (1987). En primer lugar, tanto de este trabajo como de la literatura revisada (especialmente la más reciente), no parece desprenderse que la insulina en Platyrrhini tenga huellas de selección positiva. Si bien es cierto que los muestreos taxonómicos no son idóneos y que no pueden descartarse cambios pequeños con efectos fenotípicos mayores, no queda claro que la insulina en los primates neotropicales sea divergente. En segundo lugar, el efecto de una nueva variante dependerá de la variación preexistente, por lo que aún frente a la misma presión selectiva, es muy poco probable que dos grupos de organismos respondan en forma convergente a nivel molecular (Storz, 2016). En este sentido, el análisis de patrones a nivel de vías metabólicas y otros conjuntos génicos (como ser genes vinculados a inmunidad), tenga seguramente mejores chances de éxito.

Al observarse los resultados resumidos en la Tabla 3.7, siete genes en Platyrrhini y tres en Caviomorpha muestran evidencia importante de un régimen de selección positiva. Dejando de lado las debilidades del muestreo taxonómico (especialmente para Platyrrhini que solo incluyó cébidos) y que fueron analizados únicamente 18 genes de las vías de insulina y hormona de crecimiento/factores de crecimiento vinculados a insulina, puede decirse que la selección adaptativa fue más penetrante en Platyrrhini que en Caviomorpha. Más interesante aún, los



cambios adaptativos no se restringen a Caviomorpha sino que se presentan también en los histricomorphos africanos.

Figura 4.1: Valores de peso corporal adulto, edad de maduración reproductiva en la hembra y máxima longevidad para glires. En A los datos se muestran para Ctenohystrica (Ct) y Glires no-Ctenohystrica (G). En B, la división es entre Caviomorpha (C), Ctenohystrica no-Caviomorpha (H) y glires no-Ctenohystrica (G). Debido a las relaciones alométricas, la edad de maduración y la máxima longevidad se muestran en función del logaritmo del peso corporal. Los datos se obtuvieron de la base de datos AnAge (http://genomics.senescence.info/species/; de Magalhães et al., 2007; de Magalhães y Costa, 2009). Se consideró solo una especie por género, la primera en presentar datos de alta calidad.

Los platirrinos se conocen solo en el Neotrópico y su historia evolutiva estaría asociada a un linaje inicial que se diversificó rápidamente acompañando el establecimiento y expansión de la selva amazónica (Jameson Kiesling *et al.*, 2015). No existe conocimiento sobre los proto-platirrinos, que habrían habitado África o Asia según las diferentes hipótesis. Por otra parte, resulta más difícil separar la historia evolutiva de los caviomorfos arribados y diversificados en Sudamérica de los restantes histricognatos (y ctenohistricos) africanos. Por ejemplo, en la Figura 4.1 se muestran distribuciones de peso corporal adulto (gramos), edad de maduración reproductiva de la hembra (años) y máxima longevidad (años) para múltiples especies de glires actuales. En las partes A y B de esta figura se analizan los mismos datos, pero en A las especies están clasificadas como "Hystricognathi" y "Glires no-Hystricognathi" mientras que en B los datos se dividen en "Caviomorpha", "Hystricognathi no-Caviomorpha" y "Glires no-Hystricognathi".

necesariamente mucho más grandes que los histricognatos africanos. Además, cuando se observan los valores de edad de maduración y longevidad máxima corregidos por el peso corporal, los de caviomorfos no se apartan mucho de los restantes glires, mientras que los histricognatos africanos sí lo hacen. La pregunta a plantearse es si los caviomorfos siguen siendo peculiares cuando se los observa en el contexto de todos los Hystricognathi (o Ctenohystrica).

Con respecto a los cambios moleculares en los genes de interés, antes se había señalado que la insulina en *Hystrix* tiene sus particularidades, y que en este y otros genes aquí analizados es poco frecuente encontrar cambios que sean sinapomorfías de Caviomorpha. Otros datos notorios son los provenientes de estudios en Phiomorpha. Estos histricognatos africanos comprenden especies subterráneas y recientemente han sido muy estudiados por presentar gran longevidad sin mostrar senecencia y no desarrollar cáncer (Kim y Ma, 2011; Fang *et al.*, 2014; Davies *et al.*, 2015). Los análisis de datos genómicos y transcriptómicos han revelado múltiples adaptaciones al modo de vida subterráneo y otros cambios asociados posiblemente a su mayor longevidad, incluidas sus insulinas divergentes (Fang *et al.*, 2014) y cambios en el GHRHR (Davies *et al.*, 2015). En suma, al menos parte de las peculiaridades moleculares de las vías de interés no están restringidas a los histricognatos neotropicales sino que tienen una historia evolutiva más larga.

Si bien Platyrrhini y Caviomorpha corresponden a dos clados de mamíferos neotropicales que arribaron al continente aislado y se diversificaron en forma más o menos paralela, sus historias evolutivas son diferentes. En Platyrrhini se habría sucedido un único pulso de diversificación inicial (Perez *et al.*, 2013; Jameson Kiesling *et al.*, 2015) mientras que, en Caviomorpha habrían ocurrido al menos tres pulsos (Pérez y Pol, 2012; Vucetich *et al.*, 2015), siguiendo un proceso de diversificación iniciado previamente en el continente africano. Los patrones moleculares de selección de los genes estudiados en este trabajo reflejan las diferentes historias evolutivas de estos mamíferos.

# 5 Conclusiones y perspectivas

### 5.1 Conclusiones

Esta tesis constituyó una primera aproximación al estudio de los patrones de evolución molecular en genes del eje hormona de crecimiento / factor de crecimiento relacionado a insulina en caviomorfos y platirrinos neotropicales. Se cumplió con los objetivos generales y específicos planteados.

- 1. Se seleccionaron 18 genes de interés, muchos de los cuales presentaron huellas de selección positiva en Platyrrhini (al menos siete genes) y Caviomorpha (al menos tres).
- La metodología empleada fue apropiada. Para aquellos casos en los que existían antecedentes, los resultados obtenidos fueron en general consistentes con la bibliografía disponible.
- 3. Con respecto a la construcción de grupos de ortólogos, alineamiento de secuencias codificantes y filtrado de los mismos, se plantean mejoras al método utilizado. Estas permitirán mayor automatismo y el análisis de un conjunto de genes mayor.
- 4. En cuanto al muestreo taxonómico utilizado, las secuencias disponibles permitieron realizar mejores inferencias sobre caviomorfos que sobre platirrinos. En el caso de los monos neotropicales, los *taxa* empleados se restringen a la familia Cebidae por lo tanto, estrictamente, los resultados e inferencias también se limitan a esta familia de platirrinos.
- 5. Se detectó selección positiva en insulina para caviomorfos, y parte de los sitios detectados bajo selección coinciden con los reportados por Opazo *et al.* (2005). No se dectaron huellas de selección adaptativa en las secuencias de insulina de platirrinos, constituyendo este un resultado discordante con parte de los antecedentes bibliográficos.
- 6. También se detectó selección adaptativa en el factor de crecimiento relacionado a insulina
  2, con varios cambios acumulados en la rama ancestral de Hystricognathi.
- 7. En cavimorfos el receptor de insulina presentó evidencia de algunas sustituciones no sinónimas moldeadas por selección episódica, así como posibles diferencias en la expresión

#### 5 Conclusiones y perspectivas

de las isoformas INSRA e INSRB del receptor. Como ocurre con la insulina, estos cambios se presentaron en forma linaje dependiente.

- 8. Se plantea para Caviomorpha/Hystricognathi la hipótesis de cambios iniciales en el factor de crecimiento relacionado a insulina 2, seguidos de cambios en la insulina y en el receptor de insulina. Como consecuencia de estos cambios, se produciría un detrimento en la capacidad de señalización metabólica del par insulina receptor de insulina.
- 9. Para los genes estudiados, el efecto de la selección positiva ha sido más penetrante en Platyrrhini. Si bien en general los resultados están acotados a lo que ocurre en Callitrichinae o Cebidae, datos complementarios sugieren que ocurriría una clara concentración de cambios adaptativos en el ancestro del grupo, coincidiendo con un único pulso de diversificación propuesto para el infraorden.
- 10. Por el contrario, los patrones de evolución molecular en Caviomorpha son linaje dependientes y muchos de los cambios no se restringen a lo ocurrido durante su diversificación en el Neotrópico, sino que aparecen durante su historia evolutiva compartida con los histricognatos africanos.
- Finalmente, se resalta que el estudio comparativo en profundidad de caviomorfos e histricognatos africanos aportaría en forma importante al conocimiento de estos mamíferos neotropicales.

#### 5.2 Perspectivas

En lo inmediato, se plantea mejorar la metodología para la construcción de los grupos de ortólogos y los alineamientos de secuencias codificantes. Para ello se podría utilizar de referencia estos mismos 18 genes ya conocidos e ir evaluando la mejora con respecto al método más manual antes utilizado.

Con la nueva metodología se podrá extender el estudio a vías de señalización completas e incluso conjuntos de genes de interés mayores. En particular se plantea el análisis del conjunto de genes resultante de la unión de las vías de señalización de insulina y hormona de crecimiento /factores de crecimiento relacionados a insulina, que incluya desde las hormonas y restantes componentes proximales a los factores de transcripción últimos en la transducción de las señales. En una etapa posterior se plantea estudiar otro conjunto de genes que involucre a aquellos asociados a inmunidad innata.

Según lo discutido en este trabajo, y en el contexto actual donde las secuencias genómicas y transcriptómicas de Platyrrhini están en general acotadas en su representación taxonómica, se plantea priorizar el estudio de caviomorfos e histricognatos como grupos de interés.

## Referencias

- Adkins, R. M., Vandeberg, J., y Li, W. H. (2000). Molecular evolution of growth hormone and receptor in the guinea-pig, a mammal unresponsive to growth hormone. *Gene*, 246(1-2):357–63.
- Adzhubei, I., Jordan, D. M., y Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*, Chapter 7:Unit7.20.
- Altenhoff, A. M., Gil, M., Gonnet, G. H., y Dessimoz, C. (2013). Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One*, 8(1):e53786.
- Altenhoff, A. M., Škunca, N., Glover, N., Train, C.-M., Sueki, A., Piližota, I., Gori, K., Tomiczek, B., Müller, S., Redestig, H., Gonnet, G. H., y Dessimoz, C. (2015). The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. *Nucleic Acids Research*, 43(Database issue):D240–9.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., y Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17):3389–402.
- Alvarez-Ponce, D., Aguadé, M., y Rozas, J. (2009). Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 Drosophila genomes. *Genome Research*, 19(2):234–42.
- Alvarez-Ponce, D., Aguade, M., y Rozas, J. (2011). Comparative Genomics of the Vertebrate Insulin/TOR Signal Transduction Pathway: A Network-Level Analysis of Selective Pressures. *Genome Biology and Evolution*, 3(0):87–101.
- Anisimova, M., Bielawski, J. P., y Yang, Z. (2001). Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. *Molecular Biology and Evolution*, 18(8):1585–92.
- Anisimova, M. y Yang, Z. (2007). Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. *Molecular Biology and Evolution*, 24(5):1219–28.
- Aristide, L., Rosenberger, A. L., Tejedor, M. F., y Perez, S. I. (2015). Modeling lineage

and phenotypic diversification in the New World monkey (Platyrrhini, Primates) radiation. *Molecular Phylogenetics and Evolution*, 82(Pt B):375–85.

- Bajaj, M., Blundell, T. L., Horuk, R., Pitts, J. E., Wood, S. P., Gowan, L. K., Schwabe, C., Wollmer, A., Gliemann, J., y Gammeltoft, S. (1986). Coypu insulin. Primary structure, conformation and biological properties of a hystricomorph rodent insulin. *The Biochemical Journal*, 238(2):345–51.
- Barreiro, L. B. y Quintana-Murci, L. (2010). From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nature Reviews Genetics*, 11(1):17–30.
- Beintema, J. J. y Campagne, R. N. (1987). Molecular evolution of rodent insulins. *Molecular Biology and Evolution*, 4(1):10–8.
- Belfiore, A. (2007). The role of insulin receptor isoforms and hybrid insulin/IGF-I receptors in human cancer. *Current Pharmaceutical Design*, 13(7):671–86.
- Belfiore, A., Frasca, F., Pandini, G., Sciacca, L., y Vigneri, R. (2009). Insulin receptor isoforms and insulin receptor/insulin-like growth factor receptor hybrids in physiology and disease. *Endocrine Reviews*, 30(6):586–623.
- Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J., y Sayers, E. W. (2012). GenBank. Nucleic Acids Research, 40(Database issue):D48–53.
- Bielawski, J. P., Baker, J. L., y Mingrone, J. (2016). Inference of episodic changes in natural selection acting on protein coding sequences via CODEML. En *Current Protocols in Bioinformatics*, volumen 54, pp. 6.15.1–6.15.32. John Wiley & Sons, Inc., Hoboken, NJ, USA.
- Blanga-Kanfi, S., Miranda, H., Penn, O., Pupko, T., DeBry, R. W., y Huchon, D. (2009). Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology*, 9(1):71.
- Blundell, T. L. y Wood, S. P. (1975). Is the evolution of insulin Darwinian or due to selectively neutral mutation? *Nature*, 257(5523):197–203.
- Bolger, A. M., Lohse, M., y Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–20.
- Boucher, J., Tseng, Y.-H., y Kahn, C. R. (2010). Insulin and insulin-like growth factor-1 receptors act as ligand-specific amplitude modulators of a common pathway regulating gene transcription. *The Journal of Biological Chemistry*, 285(22):17235–45.

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., y Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1):421.
- Capella-Gutierrez, S., Silla-Martinez, J. M., y Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–3.
- Carter-Su, C., Schwartz, J., y Argetsinger, L. S. (2016). Growth hormone signaling pathways. *Growth Hormone & IGF Research*, 28:11–15.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, 17(4):540–52.
- Chan, S. J., Episkopou, V., Zeitlin, S., Karathanasis, S. K., MacKrell, A., Steiner, D. F., y Efstratiadis, A. (1984). Guinea pig preproinsulin gene: an evolutionary compromise? *Proceedings of the National Academy of Sciences of the United States of America*, 81(16):5046–50.
- Chan, S. J., Seino, S., Gruppuso, P. A., Schwartz, R., y Steiner, D. F. (1987). A mutation in the B chain coding region is associated with impaired proinsulin conversion in a family with hyperproinsulinemia. *Proceedings of the National Academy of Sciences of the United States of America*, 84(8):2194–7.
- Chen, E. Y., Liao, Y. C., Smith, D. H., Barrera-Saldaña, H. A., Gelinas, R. E., y Seeburg, P. H. (1989). The human growth hormone locus: nucleotide sequence, biology, and evolution. *Genomics*, 4(4):479–97.
- Chen, X. y Tompa, M. (2010). Comparative assessment of methods for aligning multiple genome sequences. *Nature Biotechnology*, 28(6):567–72.
- Chrarif, D. y Lobry, J. R. (2007). SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. En Bastolla, U., Porto, M., Roman, H. E., y Vendruscolo, M., editores, *Structural Approaches to Sequence Evolution*, Biological and Medical Physics, Biomedical Engineering, pp. 207–232. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Churakov, G., Sadasivuni, M. K., Rosenbloom, K. R., Huchon, D., Brosius, J., y Schmitz, J. (2010). Rodent evolution: back to the root. *Molecular Biology and Evolution*, 27(6):1315–26.
- Clemmons, D. R. (2004). The relative roles of growth hormone and IGF-1 in controlling insulin sensitivity. *The Journal of Clinical Investigation*, 113(1):25–7.

Criscuolo, A. y Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new

software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, 10:210.

- Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M. R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., Rothfels, K., Shamovsky, V., Song, H., Williams, M., Birney, E., Hermjakob, H., Stein, L., y D'Eustachio, P. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Research*, 42(Database issue):D472–7.
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Aken, B. L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S. M. J., Spudich, G., Trevanion, S. J., Yates, A., Zerbino, D. R., y Flicek, P. (2014). Ensembl 2015. *Nucleic Acids Research*, 43(Database issue):D662–9.
- Daughaday, W. H., Emanuele, M. A., Brooks, M. H., Barbato, A. L., Kapadia, M., y Rotwein, P. (1988). Synthesis and secretion of insulin-like growth factor II by a leiomyosarcoma with associated hypoglycemia. *The New England Journal of Medicine*, 319(22):1434–40.
- Daughaday, W. H., Trivedi, B., y Baxter, R. C. (1993). Serum "big insulin-like growth factor II" from patients with tumor hypoglycemia lacks normal E-domain O-linked glycosylation, a possible determinant of normal propeptide processing. *Proceedings of the National Academy of Sciences of the United States of America*, 90(12):5823–7.
- Davidsen, P. K., Herbert, J. M., Antczak, P., Clarke, K., Ferrer, E., Peinado, V. I., Gonzalez, C., Roca, J., Egginton, S., Barberá, J. A., y Falciani, F. (2014). A systems biology approach reveals a link between systemic cytokines and skeletal muscle energy metabolism in a rodent smoking model and human COPD. *Genome Medicine*, 6(8):59.
- Davies, K. T. J., Bennett, N. C., Tsagkogeorga, G., Rossiter, S. J., y Faulkes, C. G. (2015). Family wide molecular adaptations to underground life in African mole-rats revealed by phylogenomic analysis. *Molecular Biology and Evolution*, 32(12):3089–107.
- Daza, D. O., Sundstrom, G., Bergqvist, C. A., Duan, C., y Larhammar, D. (2011). Evolution of the insulin-like growth factor binding protein (IGFBP) family. *Endocrinology*, 152(6):2278–89.
- de Magalhães, J. P., Budovsky, A., Lehmann, G., Costa, J., Li, Y., Fraifeld, V., y Church, G. M. (2009). The Human Ageing Genomic Resources: online databases and tools for biogerontologists. *Aging Cell*, 8(1):65–72.

- de Magalhães, J. P. y Costa, J. (2009). A database of vertebrate longevity records and their relation to other life-history traits. *Journal of Evolutionary Biology*, 22(8):1770–4.
- de Magalhães, J. P., Costa, J., y Church, G. M. (2007). An analysis of the relationship between metabolism, developmental schedules, and longevity using phylogenetic independent contrasts. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, 62(2):149–60.
- De Meyts, P. (1994). The structural basis of insulin and insulin-like growth factor-I receptor binding and negative co-operativity, and its relevance to mitogenic versus metabolic signalling. *Diabetologia*, 37(Suppl 2):S135–48.
- De Meyts, P. (2004). Insulin and its receptor: structure, function and evolution. *BioEssays: News* and Reviews in Molecular, Cellular and Developmental Biology, 26(12):1351–62.
- De Meyts, P. (2008). The insulin receptor: a prototype for dimeric, allosteric membrane receptors? *Trends in Biochemical Sciences*, 33(8):376–84.
- de Oliveira, F. B., Molina, E. C., y Marroig, G. (2009). Paleogeography of the South Atlantic: a route for primates and rodents into the New World? En Garber, P. A., Estrada, A., Bicca-Marques, J. C., Heymann, E. W., y Strier, K. B., editores, *South American Primates: Comparative Perspectives in the Study of Behavior, Ecology, and Conservation. Developments in Primatology: Progress and Prospects*, capítulo 3, pp. 55–68. Springer.
- Delport, W., Poon, A. F. Y., Frost, S. D. W., y Kosakovsky Pond, S. L. (2010). Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, 26(19):2455–7.
- Delsuc, F. y Tilak, M.-K. (2015). Naked but not hairless: the pitfalls of analyses of molecular adaptation based on few genome sequence comparisons. *Genome Biology and Evolution*, 7(3):768–74.
- Dufty, A. M., Clobert, J., y Moller, A. P. (2002). Hormones, developmental plasticity and adaptation. *Trends in Ecology and Evolution*, 17(4):190–6.
- Duguay, S. J., Jin, Y., Stein, J., Duguay, A. N., Gardner, P., y Steiner, D. F. (1998). Post-translational processing of the insulin-like growth factor-2 precursor. Analysis of O-glycosylation and endoproteolysis. *The Journal of Biological Chemistry*, 273(29):18443–51.
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–7.
- ElShewy, H. y Luttrell, L. (2009). Insulin-like growth factor-2-mannose-6 phosphate receptors. *Vitamins and Hormones*, 80:667–97.

- Emmons, L. H., Leite, Y. L. R., y Patton, J. (2015). Family Echimyidae. En Patton, J., Pardiñas, U., y D'Elía, G., editores, *Mammals of South America. Volume 2. Rodents*, pp. 887–880. The University of Chicago Press, Chicago.
- Fabre, P.-H., Galewski, T., Tilak, M.-k., y Douzery, E. J. P. (2013). Diversification of South American spiny rats (Echimyidae): a multigene phylogenetic approach. *Zoologica Scripta*, 42(2):117–34.
- Fabre, P.-H., Hautier, L., Dimitrov, D., y Douzery, E. J. P. (2012). A glimpse on the pattern of rodent diversification: a phylogenetic approach. *BMC Evolutionary Biology*, 12(1):88.
- Fabre, P.-H., Rodrigues, A., y Douzery, E. J. P. (2009). Patterns of macroevolution among Primates inferred from a supermatrix of mitochondrial and nuclear DNA. *Molecular Phylogenetics and Evolution*, 53(3):808–25.
- Fabre, P.-H., Upham, N. S., Emmons, L. H., Justy, F., Leite, Y. L. R., Loss, A. C., Orlando, L., Tilak, M.-K., Patterson, B. D., y Douzery, E. J. P. (2017). Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. *Molecular Biology and Evolution*, 34(3):613–33.
- Fang, X., Seim, I., Huang, Z., Gerashchenko, M. V., Xiong, Z., Turanov, A. A., Zhu, Y., Lobanov, A. V., Fan, D., Yim, S. H., Yao, X., Ma, S., Yang, L., Lee, S.-G., Kim, E. B., Bronson, R. T., Šumbera, R., Buffenstein, R., Zhou, X., Krogh, A., Park, T. J., Zhang, G., Wang, J., y Gladyshev, V. N. (2014). Adaptations to a subterranean environment and longevity revealed by the analysis of mole rat genomes. *Cell Reports*, 8(5):1354–64.
- Feyertag, F., Berninsone, P. M., y Alvarez-Ponce, D. (2017). Secreted proteins defy the expression level evolutionary rate anticorrelation. *Molecular Biology and Evolution*, 34(3):692–706.
- Figuet, E., Nabholz, B., Bonneau, M., Mas Carrio, E., Nadachowska-Brzyska, K., Ellegren, H., y Galtier, N. (2016). Life history traits, protein evolution, and the Nearly Neutral Theory in amniotes. *Molecular Biology and Evolution*, 33(6):1517–27.
- Fletcher, W. y Yang, Z. (2010). The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Molecular Biology and Evolution*, 27(10):2257–67.
- Frasca, F., Pandini, G., Scalia, P., Sciacca, L., Mineo, R., Costantino, A., Goldfine, I. D., Belfiore, A., y Vigneri, R. (1999). Insulin receptor isoform A, a newly recognized, high-affinity insulin-like growth factor II receptor in fetal and cancer cells. *Molecular and Cellular Biology*, 19(5):3278–88.
- Gharib, W. H. y Robinson-Rechavi, M. (2013). The branch-site test of positive selection is

surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Molecular Biology and Evolution*, 30(7):1675–86.

- Ghosh, P., Dahms, N. M., y Kornfeld, S. (2003). Mannose 6-phosphate receptors: new twists in the tale. *Nature Reviews Molecular Cell Biology*, 4(3):202–13.
- Giudice, J., Leskow, F. C., Arndt-Jovin, D. J., Jovin, T. M., y Jares-Erijman, E. A. (2011). Differential endocytosis and signaling dynamics of insulin receptor variants IR-A and IR-B. *Journal of Cell Science*, 124(Pt 5):801–11.
- Goldman, N. y Yang, Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution*, 11(5):725–36.
- 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, Y.-X., y Barrera Saldaña, H. A. (2006). Growth hormone locus expands and diverges after the separation of New and Old World Monkeys. *Gene*, 380(1):38–45.
- Gouveia-Oliveira, R., Sackett, P. W., y Pedersen, A. G. (2007). MaxAlign: maximizing usable data in an alignment. *BMC Bioinformatics*, 8(1):312.
- Greenall, S. A., Bentley, J. D., Pearce, L. A., Scoble, J. A., Sparrow, L. G., Bartone, N. A., Xiao, X., Baxter, R. C., Cosgrove, L. J., y Adams, T. E. (2013). Biochemical characterization of individual human glycosylated pro-insulin-like growth factor (IGF)-II and big-IGF-II isoforms associated with cancer. *The Journal of Biological Chemistry*, 288(1):59–68.
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., y Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3):307–21.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., Henschel, R., LeDuc, R. D., Friedman, N., y Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8):1494–512.
- Haig, D., Graham, C., Ogata, E., Asano, T., Iiri, T., Katada, T., Ui, M., Grubb, J., Sly, W., y Nishimoto, I. (1991). Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell*, 64(6):1045–6.
- Haneda, M., Polonsky, K. S., Bergenstal, R. M., Jaspan, J. B., Shoelson, S. E., Blix, P. M., Chan, S. J., Kwok, S. C., Wishner, W. B., y Zeidler, A. (1984). Familial hyperinsulinemia due to

a structurally abnormal insulin. Definition of an emerging new clinical syndrome. *The New England Journal of Medicine*, 310(20):1288–94.

- Hansen, B. F., Danielsen, G. M., Drejer, K., Sorensen, A. R., Wiberg, F. C., Klein, H. H., y Lundemose, A. G. (1996). Sustained signalling from the insulin receptor after stimulation with insulin analogues exhibiting increased mitogenic potency. *Biochemical Journal*, 315(1):271–9.
- Harris, R. A., Tardif, S. D., Vinar, T., Wildman, D. E., Rutherford, J. N., Rogers, J., Worley, K. C., y Aagaard, K. M. (2014). Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proceedings of the National Academy of Sciences of the United States of America*, 111(4):1467–72.
- Harrison, P. W., Jordan, G. E., y Montgomery, S. H. (2014). SWAMP: sliding window alignment masker for PAML. *Evolutionary Bioinformatics Online*, 10:197–204.
- Harrow, J. L., Steward, C. A., Frankish, A., Gilbert, J. G., Gonzalez, J. M., Loveland, J. E., Mudge, J., Sheppard, D., Thomas, M., Trevanion, S., y Wilming, L. G. (2014). The Vertebrate Genome Annotation browser 10 years on. *Nucleic Acids Research*, 42(Database issue):D771–9.
- Hedges, S. B. y Kumar, S. (2009). The timetree of life. Oxford University Press.
- Hedges, S. B., Marin, J., Suleski, M., Paymer, M., y Kumar, S. (2015). Tree of Life Reveals Clock-Like Speciation and Diversification. *Molecular Biology and Evolution*, 32(4):835–45.
- Hernández-Sánchez, C., Mansilla, A., de Pablo, F., y Zardoya, R. (2008). Evolution of the insulin receptor family and receptor isoform expression in vertebrates. *Molecular Biology and Evolution*, 25(6):1043–53.
- Horuk, R., Blundell, T. L., Lazarus, N. R., Neville, R. W., Stone, D., y Wollmer, A. (1980). A monomeric insulin from the porcupine (Hystrix cristata), an Old World hystricomorph. *Nature*, 286(5775):822–4.
- Horuk, R., Goodwin, P., O'connor, K., Neville, R. W., Lazarus, N. R., y Stone, D. (1979). Evolutionary change in the insulin receptors of hystricomorph rodents. *Nature*, 279(5712):439–40.
- Horuk, R., Rodbell, M., Cushman, S. W., y Wardzala, L. J. (1983). Proposed mechanism of insulin-resistant glucose transport in the isolated guinea pig adipocyte. Small intracellular pool of glucose transporters. *The Journal of Biological Chemistry*, 258(12):7425–9.
- Huang, C. G., Eng, J., Pan, Y. C., Hulmes, J. D., y Yalow, R. S. (1986). Guinea pig glucagon differs from other mammalian glucagons. *Diabetes*, 35(5):508–12.

- Hubbard, S. R. y Till, J. H. (2000). Protein tyrosine kinase structure and function. *Annual Review* of *Biochemistry*, 69:373–98.
- Huchon, D., Chevret, P., Jordan, U., Kilpatrick, C. W., Ranwez, V., Jenkins, P. D., Brosius, J., y Schmitz, J. (2007). Multiple molecular evidences for a living mammalian fossil. *Proceedings* of the National Academy of Sciences of the United States of America, 104(18):7495–9.
- Huchon, D. y Douzery, E. J. (2001). From the Old World to the New World: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents. *Molecular Phylogenetics and Evolution*, 20(2):238–51.
- Huchon, D., Madsen, O., Sibbald, M. J. J. B., Ament, K., Stanhope, M. J., Catzeflis, F., de Jong, W. W., y Douzery, E. J. P. (2002). Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. *Molecular Biology and Evolution*, 19(7):1053–65.
- Jameson Kiesling, N. M., Yi, S. V., Xu, K., Gianluca Sperone, F., y Wildman, D. E. (2015). The tempo and mode of New World monkey evolution and biogeography in the context of phylogenomic analysis. *Molecular Phylogenetics and Evolution*, 82(Pt B):386–99.
- Jensen, M. y De Meyts, P. (2009). Molecular mechanisms of differential intracellular signaling from the insulin receptor. *Vitamins and Hormones*, 80:51–75.
- Jensen, M., Palsgaard, J., Borup, R., de Meyts, P., y Schäffer, L. (2008). Activation of the insulin receptor (IR) by insulin and a synthetic peptide has different effects on gene expression in IR-transfected L6 myoblasts. *The Biochemical Journal*, 412(3):435–45.
- Jiggins, F. M. y Kim, K. W. (2007). A screen for immunity genes evolving under positive selection in Drosophila. *Journal of Evolutionary Biology*, 20(3):965–70.
- Jombart, T., Kamvar, Z. N., Schliep, K., Archer, E., y Harris, R. (2015). apex: Phylogenetic Methods for Multiple Gene Data.
- Jordan, G. y Goldman, N. (2012). The effects of alignment error and alignment filtering on the sitewise detection of positive selection. *Molecular Biology and Evolution*, 29(4):1125–39.
- Julenius, K. y Pedersen, A. G. (2006). Protein Evolution Is Faster Outside the Cell. Molecular Biology and Evolution, 23(11):2039–48.
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., y Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(Database issue):D457–62.

- Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J., y McInerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology*, 6(1):29.
- Keightley, M. C. y Fuller, P. J. (1996). Anomalies in the endocrine axes of the guinea pig: relevance to human physiology and disease. *Endocrine Reviews*, 17(1):30–44.
- Kellerer, M., Lammers, R., Ermel, B., Tippmer, S., Vogt, B., Obermaier-Kusser, B., Ullrich, A., y Häring, H. U. (1992). Distinct alpha-subunit structures of human insulin receptor A and B variants determine differences in tyrosine kinase activities. *Biochemistry*, 31(19):4588–96.
- Killian, J. K., Nolan, C. M., Wylie, A. A., Li, T., Vu, T. H., Hoffman, A. R., y Jirtle, R. L. (2001). Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary. *Human Molecular Genetics*, 10(17):1721–8.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., y Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4):R36.
- Kim, J. y Ma, J. (2011). PSAR: measuring multiple sequence alignment reliability by probabilistic sampling. *Nucleic Acids Research*, 39(15):6359–68.
- Kind, K. L., Clifton, P. M., Grant, P. A., Owens, P. C., Sohlstrom, A., Roberts, C. T., Robinson, J. S., y Owens, J. A. (2003). Effect of maternal feed restriction during pregnancy on glucose tolerance in the adult guinea pig. *American Journal of Physiology. Regulatory, Integrative and Comparative Physiology*, 284(1):R140–52.
- King, G. L. y Kahn, C. R. (1981). Non-parallel evolution of metabolic and growth-promoting functions of insulin. *Nature*, 292(5824):644–6.
- King, G. L., Kahn, C. R., y Heldin, C. H. (1983). Sharing of biological effect and receptors between guinea pig insulin and platelet-derived growth factor. *Proceedings of the National Academy of Sciences of the United States of America*, 80(5):1308–12.
- Kosaki, A., Nelson, J., y Webster, N. J. (1998). Identification of intron and exon sequences involved in alternative splicing of insulin receptor pre-mRNA. *The Journal of Biological Chemistry*, 273(17):10331–7.
- Kosakovsky Pond, S. y Muse, S. V. (2005). Site-to-site variation of synonymous substitution rates. *Molecular Biology and Evolution*, 22(12):2375–85.

- Kosakovsky Pond, S. L. y Frost, S. D. W. (2005a). A simple hierarchical approach to modeling distributions of substitution rates. *Molecular Biology and Evolution*, 22(2):223–34.
- Kosakovsky Pond, S. L. y Frost, S. D. W. (2005b). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics*, 21(10):2531–3.
- Kosakovsky Pond, S. L. y Frost, S. D. W. (2005c). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, 22(5):1208–22.
- Kosakovsky Pond, S. L., Frost, S. D. W., y Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, 21(5):676–9.
- Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D. W., Delport, W., y Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution*, 28(11):3033–43.
- Kosiol, C. y Anisimova, M. (2012). Selection on the protein-coding genome. *Methods in Molecular Biology*, 856:113–40.
- Kosiol, C., Holmes, I., y Goldman, N. (2007). An empirical codon model for protein sequence evolution. *Molecular Biology and Evolution*, 24(7):1464–1479.
- Kück, P., Meusemann, K., Dambach, J., Thormann, B., von Reumont, B. M., Wägele, J. W., y Misof, B. (2010). Parametric and non-parametric masking of randomness in sequence alignments can be improved and leads to better resolved trees. *Frontiers in Zoology*, 7(1):10.
- Kumar, P., Henikoff, S., y Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 4(7):1073–81.
- Lackey, D. E. y Olefsky, J. M. (2015). Regulation of metabolism by the innate immune system. *Nature Reviews Endocrinology*, 12(1):15–28.
- Lanfear, R., Kokko, H., y Eyre-Walker, A. (2014). Population size and the rate of evolution. *Trends in Ecology and Evolution*, 29(1):33–41.
- Leibiger, B., Leibiger, I. B., Moede, T., Kemper, S., Kulkarni, R. N., Kahn, C. R., de Vargas, L. M., y Berggren, P. O. (2001). Selective insulin signaling through A and B insulin receptors regulates transcription of insulin and glucokinase genes in pancreatic beta cells. *Molecular Cell*, 7(3):559–70.
- LeRoith, D., Kavsan, V. M., Koval, A. P., y Roberts, C. T. (1993). Phylogeny of the insulin-like

growth factors (IGFs) and receptors: a molecular approach. *Molecular Reproduction and Development*, 35(4):332–8.

- Lessa, E. P., Cook, J. A., D'Elía, G., y Opazo, J. C. (2014). Rodent diversity in South America: transitioning into the genomics era. *Frontiers in Ecology and Evolution*, 2:39.
- Levinovitz, A., Norstedt, G., van den Berg, S., Robinson, I. C., y Ekström, T. J. (1992). Isolation of an insulin-like growth factor II cDNA from guinea pig liver: expression and developmental regulation. *Molecular and Cellular Endocrinology*, 89(1-2):105–10.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., y 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–9.
- Li, Y., Ye, C., Shi, P., Zou, X.-J., Xiao, R., Gong, Y.-Y., y Zhang, Y.-P. (2005). Independent origin of the growth hormone gene family in New World monkeys and Old World monkeys/hominoids. *Journal of Molecular Endocrinology*, 35(2):399–409.
- Liao, B. Y., Weng, M. P., y Zhang, J. (2010). Impact of Extracellularity on the Evolutionary Rate of Mammalian Proteins. *Genome Biology and Evolution*, 2(0):39–43.
- Liao, Y., Smyth, G. K., y Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research*, 41(10):e108.
- Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J., y Russell, R. B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure*, 11(11):1453–9.
- Linnell, J., Groeger, G., y Hassan, A. B. (2001). Real Time kinetics of Insulin-like Growth Factor II (IGF-II) interaction with the IGF-II/Mannose 6-Phosphate receptor. *Journal of Biological Chemistry*, 276(26):23986–91.
- Loss-Oliveira, L., Aguiar, B. O., y Schrago, C. G. (2012). Testing synchrony in historical biogeography: the case of new world primates and hystricognathi rodents. *Evolutionary Bioinformatics Online*, 8:127–37.
- Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. En Methods in Molecular Biology, volumen 1079, pp. 155–70.
- Löytynoja, A. y Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10557–62.

MacManes, M. D. y Lacey, E. A. (2012). The social brain: transcriptome assembly and

characterization of the hippocampus from a social subterranean rodent, the colonial tuco-tuco (Ctenomys sociabilis). *PLoS One*, 7(9):e45524.

- MacManes, M. D., Lacey, E. A., y Lessa, E. P. (2015). Genetics, genomics, and evolutionary themes in caviomorph rodents. En Vassallo, A. I. y Antenucci, D., editores, *Biology of Caviomorph Rodents: Diversity and Evolution*, capítulo 9, pp. 273–294. SAREM Series A, Buenos Aires.
- Madsen, O., Scally, M., Douady, C. J., Kao, D. J., DeBry, R. W., Adkins, R., Amrine, H. M., Stanhope, M. J., de Jong, W. W., y Springer, M. S. (2001). Parallel adaptive radiations in two major clades of placental mammals. *Nature*, 409(6820):610–4.
- Maglott, D., Ostell, J., Pruitt, K. D., y Tatusova, T. (2005). Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Research*, 33(Database issue):D54–8.
- Mallick, S., Gnerre, S., Muller, P., y Reich, D. (2009). The difficulty of avoiding false positives in genome scans for natural selection. *Genome Research*, 19(5):922–33.
- Markova-Raina, P. y Petrov, D. (2011). High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes. *Genome Research*, 21(6):863–74.
- McClain, D. A. (1991). Different ligand affinities of the two human Insulin Receptor splice variants are reflected in parallel changes in sensitivity for Insulin action. *Molecular Endocrinology*, 5(5):734–39.
- McGaugh, S. E., Bronikowski, A. M., Kuo, C.-H., Reding, D. M., Addis, E. A., Flagel, L. E., Janzen, F. J., y Schwartz, T. S. (2015). Rapid molecular evolution across amniotes of the IIS-TOR network. *Proceedings of the National Academy of Sciences of the United States of America*, 112(22):7055–60.
- McNelis, J. C. y Olefsky, J. M. (2014). Macrophages, immunity, and metabolic disease. *Immunity*, 41(1):36–48.
- Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., Goodbla, A., Eizirik, E., Simão, T. L. L., Stadler, T., Rabosky, D. L., Honeycutt, R. L., Flynn, J. J., Ingram, C. M., Steiner, C., Williams, T. L., Robinson, T. J., Burk-Herrick, A., Westerman, M., Ayoub, N. A., Springer, M. S., y Murphy, W. J. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, 334(6055):521–4.
- Mitchell, M. L., Guillemin, R., y Selye, H. (1954). The effect of somatotrophic hormone on the growth of normal and hypophysectomized Guinea pigs. *Endocrinology*, 54(1):111–14.

- Montgelard, C., Forty, E., Arnal, V., y Matthee, C. A. (2008). Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. *BMC Evolutionary Biology*, 8(1):321.
- Morgan, D. O., Edman, J. C., Standring, D. N., Fried, V. A., Smith, M. C., Roth, R. A., y Rutter, W. J. (1987). Insulin-like growth factor II receptor as a multifunctional binding protein. *Nature*, 329(6137):301–7.
- Muggeo, M., Ginsberg, B. H., Roth, J., Neville, D. M., De Meyts, P., y Kahn, C. R. (1979). The insulin receptor in vertebrates is functionally more conserved during evolution than insulin itself. *Endocrinology*, 104(5):1393–402.
- Murphy, W. J., Eizirik, E., Johnson, W. E., Zhang, Y. P., Ryder, O. A., y O'Brien, S. J. (2001). Molecular phylogenetics and the origins of placental mammals. *Nature*, 409(6820):614–8.
- Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D. P., Smith, D. M., Scheffler, K., y Kosakovsky Pond, S. L. (2015). Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, 32(5):1365–71.
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., y Kosakovsky Pond, S. L. (2012). Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genetics*, 8(7):e1002764.
- Muse, S. V. y Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, 11(5):715–24.
- Nenadic, O. y Greenacre, M. (2007). Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package. *Journal of Statistical Software*, 20(3).
- Nielsen, R. y Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 148(3):929–36.
- Nishi, M. y Steiner, D. F. (1990). Cloning of complementary DNAs encoding islet amyloid polypeptide, insulin, and glucagon precursors from a New World rodent, the degu, Octodon degus. *Molecular Endocrinology*, 4(8):1192–8.

Nuzzo, R. (2014). Scientific method: Statistical errors. Nature, 506(7487):150-52.

Oh, Y., Nagalla, S. R., Yamanaka, Y., Kim, H. S., Wilson, E., y Rosenfeld, R. G. (1996). Synthesis and characterization of insulin-like growth factor-binding protein (IGFBP)-7. Recombinant human mac25 protein specifically binds IGF-I and -II. *The Journal of Biological Chemistry*, 271(48):30322–5.

- Ohta, T. (1987). Very slightly deleterious mutations and the molecular clock. *Journal of Molecular Evolution*, 26(1-2):1–6.
- Ohta, T. (1993). Pattern of nucleotide substitutions in growth hormone-prolactin gene family: a paradigm for evolution by gene duplication. *Genetics*, 134(4):1271–6.
- Ojeda, R. A., Novillo, A., y Ojeda, A. A. (2015). Large scale richness patterns, biogeography and ecological diversification in caviomorph rodents. En Vassallo, A. I. y Antenucci, C. D., editores, *Biology of Caviomorph Rodents: Diversity and Evolution*, capítulo 3, pp. 121–138. SAREM Series A, Buenos Aires.
- Olinski, R. P., Lundin, L.-G., y Hallbook, F. (2005). Conserved synteny between the Ciona genome and human paralogons identifies large duplication events in the molecular evolution of the insulin-relaxin gene family. *Molecular Biology and Evolution*, 23(1):10–22.
- Opazo, J. C., Palma, R. E., Melo, F., y Lessa, E. P. (2005). Adaptive evolution of the insulin gene in caviomorph rodents. *Molecular Biology and Evolution*, 22(5):1290–8.
- Opazo, J. C., Soto-Gamboa, M., y Bozinovic, F. (2004). Blood glucose concentration in caviomorph rodents. *Comparative Biochemistry and Physiology. Part A, Molecular and Integrative Physiology*, 137(1):57–64.
- Opazo, J. C., Wildman, D. E., Prychitko, T., Johnson, R. M., y Goodman, M. (2006). Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). *Molecular Phylogenetics and Evolution*, 40(1):274–80.
- Papper, Z., Jameson, N. M., Romero, R., Weckle, A. L., Mittal, P., Benirschke, K., Santolaya-Forgas, J., Uddin, M., Haig, D., Goodman, M., y Wildman, D. E. (2009). Ancient origin of placental expression in the growth hormone genes of anthropoid primates. *Proceedings* of the National Academy of Sciences, 106(40):17083–88.
- Penn, O., Privman, E., Ashkenazy, H., Landan, G., Graur, D., y Pupko, T. (2010). GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Research*, 38(Web Server):W23–8.
- Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M. P. C., Silva, A., O'Brien, S. J., y Pecon-Slattery, J. (2011). A molecular phylogeny of living primates. *PLoS Genetics*, 7(3):e1001342.
- Pérez, M. E. y Pol, D. (2012). Major radiations in the evolution of Cavid rodents: reconciling fossils, ghost lineages, and relaxed molecular clocks. *PLoS One*, 7(10):e48380.

- Perez, S. I., Tejedor, M. F., Novo, N. M., y Aristide, L. (2013). Divergence Times and the Evolutionary Radiation of New World Monkeys (Platyrrhini, Primates): An Analysis of Fossil and Molecular Data. *PLoS One*, 8(6):e68029.
- Petronella, N. y Drouin, G. (2011). Gene conversions in the growth hormone gene family of primates: Stronger homogenizing effects in the Hominidae lineage. *Genomics*, 98(3):173–81.
- Philippon, H., Souvane, A., Brochier-Armanet, C., Perrière, G., Vernot, B., y Durand, D. (2017). IsoSel: protein isoform selector for phylogenetic reconstructions. *PLoS One*, 12(3):e0174250.
- Poux, C., Chevret, P., Huchon, D., de Jong, W. W., y Douzery, E. J. P. (2006). Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Systematic Biology*, 55(2):228–44.
- Premzl, M. (2015). Third party data gene data set of eutherian growth hormone genes. *Genomics Data*, 6:166–9.
- Privman, E., Penn, O., y Pupko, T. (2012). Improving the performance of positive selection inference by filtering unreliable alignment regions. *Molecular Biology and Evolution*, 29(1):1–5.
- Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C. M., Hart, J., Landrum, M. J., McGarvey, K. M., Murphy, M. R., O'Leary, N. A., Pujar, S., Rajput, B., Rangwala, S. H., Riddick, L. D., Shkeda, A., Sun, H., Tamez, P., Tully, R. E., Wallin, C., Webb, D., Weber, J., Wu, W., DiCuccio, M., Kitts, P., Maglott, D. R., Murphy, T. D., y Ostell, J. M. (2014). RefSeq: an update on mammalian reference sequences. *Nucleic Acids Research*, 42(Database issue):D756–63.
- Pupko, T., Pe, I., Shamir, R., y Graur, D. (2000). A Fast Algorithm for Joint Reconstruction of Ancestral Amino Acid Sequences. *Molecular Biology and Evolution*, 17(6):890–6.
- R Core Team (2015). R: A Language and Environment for Statistical Computing.
- Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., y Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1):24–6.
- Rodriguez, M. H. y Jungery, M. (1986). A protein on Plasmodium falciparum-infected erythrocytes functions as a transferrin receptor. *Nature*, 324(6095):388–91.
- Romiguier, J., Ranwez, V., Douzery, E. J. P., y Galtier, N. (2010). Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. *Genome Research*, 20(8):1001–9.

- Rosenberger, A. y Tejedor, M. (2013). The misbegotten: long lineages, long branches and the interrelationships of Aotus, Callicebus and the saki-uakaris. En Veiga, L. M., Barnett, A. A., Ferrari, S. F., y Norconk, M. A., editores, *Evolutionary Biology and Conservation of Titis, Sakis and Uacaris*, capítulo 2. Cambridge.
- Rosenberger, A. L. (1979). Cranial anatomy and implications of Dolichocebus, a late Oligocene ceboid primate. *Nature*, 279(5712):416–8.
- Roth, A. C. J., Gonnet, G. H., y Dessimoz, C. (2008). Algorithm of OMA for large-scale orthology inference. *BMC Bioinformatics*, 9(1):518.
- Roth, R. A. (1988). Structure of the receptor for insulin-like growth factor II: the puzzle amplified. *Science*, 239(4845):1269–71.
- Rotwein, P. (1999). Molecular Biology of IGF-I and IGF-II. En Rosenfeld, R. G. y Roberts,
  C. T., editores, *The IGF System: Molecular Biology, Physiology, and Clinical Applications*, pp. 19–35. Humana Press, Totowa, NJ.
- Sahm, A., Bens, M., Platzer, M., y Szafranski, K. (2017). PosiGene: automated and easy-to-use pipeline for genome-wide detection of positively selected genes. *Nucleic Acids Research*, 45(11):e100.
- Sano, T., Kawata, K., Ohno, S., Yugi, K., Kakuda, H., Kubota, H., Uda, S., Fujii, M., Kunida, K., Hoshino, D., Hatano, A., Ito, Y., Sato, M., Suzuki, Y., y Kuroda, S. (2016). Selective control of up-regulated and down-regulated genes by temporal patterns and doses of insulin. *Science Signaling*, 9(455):ra112.
- Schmidt, B., Kiecke-Siemsen, C., Waheed, A., Braulke, T., y von Figura, K. (1995). Localization of the insulin-like growth factor II binding site to amino acids 1508-1566 in repeat 11 of the mannose 6-phosphate/insulin-like growth factor II receptor. *The Journal of Biological Chemistry*, 270(25):14975–82.
- Schneider, A., Souvorov, A., Sabath, N., Landan, G., Gonnet, G. H., y Graur, D. (2009). Estimates of positive Darwinian selection are inflated by errors in sequencing, annotation, and alignment. *Genome Biology and Evolution*, 1:114–8.
- Schwartz, G. P., Burke, G. T., y Katsoyannis, P. G. (1987). A superactive insulin: B10-aspartic-acid-insulin (human). *Proceedings of the National Academy of Sciences of the United States of America*, 84(18):6408–11.
- Seino, S. y Bell, G. I. (1989). Alternative splicing of human insulin receptor messenger RNA. *Biochemical and Biophysical Research Communications*, 159(1):312–6.

- Seino, S., Steiner, D. F., y Bell, G. I. (1987). Sequence of a New World primate insulin having low biological potency and immunoreactivity. *Proceedings of the National Academy of Sciences of the United States of America*, 84(21):7423–7.
- Sela, I., Ashkenazy, H., Katoh, K., y Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Research*, 43(Web Server):W7–14.
- Sen, S., Talukdar, I., y Webster, N. J. G. (2009). SRp20 and CUG-BP1 modulate Insulin Receptor exon 11 alternative splicing. *Molecular and Cellular Biology*, 29(3):871–80.
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., y Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, 29(1):308–11.
- Shiao, M.-S., Liao, B.-Y., Long, M., y Yu, H.-T. (2008). Adaptive evolution of the insulin two-gene system in mouse. *Genetics*, 178(3):1683–91.
- Siddle, K. (2012). Molecular basis of signaling specificity of insulin and IGF receptors: neglected corners and recent advances. *Frontiers in Endocrinology*, 3:34.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., y Higgins, D. G. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1):539.
- Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M. H., Baldock, R., Barbiera, G., Bardou, P., Beck, T., Blake, A., Bonierbale, M., Brookes, A. J., Bucci, G., Buetti, I., Burge, S., Cabau, C., Carlson, J. W., Chelala, C., Chrysostomou, C., Cittaro, D., Collin, O., Cordova, R., Cutts, R. J., Dassi, E., Di Genova, A., Djari, A., Esposito, A., Estrella, H., Eyras, E., Fernandez-Banet, J., Forbes, S., Free, R. C., Fujisawa, T., Gadaleta, E., Garcia-Manteiga, J. M., Goodstein, D., Gray, K., Guerra-Assunção, J. A., Haggarty, B., Han, D.-J., Han, B. W., Harris, T., Harshbarger, J., Hastings, R. K., Hayes, R. D., Hoede, C., Hu, S., Hu, Z.-L., Hutchins, L., Kan, Z., Kawaji, H., Keliet, A., Kerhornou, A., Kim, S., Kinsella, R., Klopp, C., Kong, L., Lawson, D., Lazarevic, D., Lee, J.-H., Letellier, T., Li, C.-Y., Lio, P., Liu, C.-J., Luo, J., Maass, A., Mariette, J., Maurel, T., Merella, S., Mohamed, A. M., Moreews, F., Nabihoudine, I., Ndegwa, N., Noirot, C., Perez-Llamas, C., Primig, M., Quattrone, A., Quesneville, H., Rambaldi, D., Reecy, J., Riba, M., Rosanoff, S., Saddiq, A. A., Salas, E., Sallou, O., Shepherd, R., Simon, R., Sperling, L., Spooner, W., Staines, D. M., Steinbach, D., Stone, K., Stupka, E., Teague, J. W., Dayem Ullah, A. Z., Wang, J., Ware, D., Wong-Erasmus, M., Youens-Clark, K., Zadissa, A., Zhang, S.-J., y Kasprzyk, A. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Research, 43(Web Server):W589-98.

- Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., y Kosakovsky Pond, S. L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, 32(5):1342–53.
- Soares, M. B., Schon, E., Henderson, A., Karathanasis, S. K., Cate, R., Zeitlin, S., Chirgwin, J., y Efstratiadis, A. (1985). RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Molecular and Cellular Biology*, 5(8):2090–103.
- Sojo, V., Dessimoz, C., Pomiankowski, A., Lane, N., M, C., S, B., Y, Z., D, P., KG, L. R., y R, A. (2016). Membrane Proteins Are Dramatically Less Conserved than Water-Soluble Proteins across the Tree of Life. *Molecular Biology and Evolution*, 33(11):2874–84.
- Song, S., Liu, L., Edwards, S. V., y Wu, S. (2012). Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy of Sciences of the United States of America*, 109(37):14942–7.
- Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., Stadler, T., Steiner, C., Ryder, O. A., Janečka, J. E., Fisher, C. A., y Murphy, W. J. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS One*, 7(11):e49521.
- Springer, M. S., Stanhope, M. J., Madsen, O., y de Jong, W. W. (2004). Molecules consolidate the placental mammal tree. *Trends in Ecology and Evolution*, 19(8):430–8.
- Steiner, D. F. (2004). The proinsulin C-peptide–a multirole model. *Experimental Diabesity Research*, 5(1):7–14.
- Storz, J. F. (2016). Causes of molecular convergence and parallelism in protein evolution. *Nature Reviews Genetics*, 17(4):239–50.
- Suyama, M., Torrents, D., y Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, 34(Web Server):W609–12.
- Talavera, G. y Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology*, 56(4):564–77.
- Thomas, G. W., Hahn, M. W., y Hahn, Y. (2017). The effects of increasing the number of taxa on inferences of molecular convergence. *Genome Biology and Evolution*, 9(1):213–21.
- Tong, Z. Q., Jack, E., Moule, M., Goldfine, I. D., y Yip, C. C. (1994). Rodent insulin receptors are

immunologically different from other mammalian insulin receptors. *General and Comparative Endocrinology*, 94(3):374–81.

- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., Pimentel, H., Salzberg, S. L., Rinn, J. L., y Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7(3):562–78.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., y Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, 28(5):511–15.
- Uhles, S., Moede, T., Leibiger, B., Berggren, P.-O., y Leibiger, I. B. (2003). Isoform-specific insulin receptor signaling involves different plasma membrane domains. *The Journal of Cell Biology*, 163(6):1327–37.
- Upham, N. S. y Patterson, B. D. (2012). Diversification and biogeography of the Neotropical caviomorph lineage Octodontoidea (Rodentia: Hystricognathi). *Molecular Phylogenetics and Evolution*, 63(2):417–29.
- Upham, N. S. y Patterson, B. D. (2015). Evolution of caviomorph rodents: a complete phylogeny and timetree for living genera. En Vassallo, A. I. y Antenucci, D., editores, *Biology of Caviomorph Rodents: Diversity and Evolution*, capítulo 2, pp. 63–120. SAREM Series A, Buenos Aires.
- Vera, F., Zenuto, R. R., y Antenucci, C. D. (2008). Decreased glucose tolerance but normal blood glucose levels in the field in the caviomorph rodent Ctenomys talarum: The role of stress and physical activity. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology*, 151(2):232–8.
- Villanueva-Cañas, J. L., Laurie, S., y Albà, M. M. (2013). Improving genome-wide scans of positive selection by using protein isoforms of similar length. *Genome Biology and Evolution*, 5(2):457–67.
- Vogt, B., Carrascosa, J. M., Ermel, B., Ullrich, A., y Häring, H. U. (1991). The two isotypes of the human insulin receptor (HIR-A and HIR-B) follow different internalization kinetics. *Biochemical and Biophysical Research Communications*, 177(3):1013–8.
- Voloch, C. M., Vilela, J. F., Loss-Oliveira, L., y Schrago, C. G. (2013). Phylogeny and chronology of the major lineages of New World hystricognath rodents: insights on the biogeography of the Eocene/Oligocene arrival of mammals in South America. *BMC Research Notes*, 6(1):160.

Vucetich, M. G., Arnal, M., Deschamps, C. M., Pérez, M. E., y Vieytes, E. C. (2015). A brief

history of caviomorph rodents as told by the fossil record. En Vassallo, A. I. y Antenucci, D., editores, *Biology of Caviomorph Rodents: Diversity and Evolution*, capítulo 1, pp. 11–62. SAREM Series A, Buenos Aires.

- Waddell, P. J., Kishino, H., y Ota, R. (2001). A phylogenetic foundation for comparative mammalian genomics. *Genome Informatics. International Conference on Genome Informatics*, 12:141–54.
- Wallis, M. (2009). New insulin-like growth factor (IGF)-precursor sequences from mammalian genomes: the molecular evolution of IGFs and associated peptides in primates. *Growth Hormone and IGF Research*, 19(1):12–23.
- Wallis, M. (2015). Coevolution of insulin-like growth factors, insulin and their receptors and binding proteins in New World Monkeys. *Growth Hormone and IGF Research*, 25(4):158–67.
- Wallis, O. C. y 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). *Journal of Molecular Evolution*, 63(5):591–601.
- Wallis, O. C., Zhang, Y. P., y Wallis, M. (2001). Molecular evolution of GH in primates: characterisation of the GH genes from slow loris and marmoset defines an episode of rapid evolutionary change. *Journal of Molecular Endocrinology*, 26(3):249–58.
- Wang, S., Wei, W., Zheng, Y., Hou, J., Dou, Y., Zhang, S., Luo, X., y Cai, X. (2012). The role of insulin C-peptide in the coevolution analyses of the insulin signaling pathway: a hint for its functions. *PLoS One*, 7(12):e52847.
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., y Barton, G. J. (2009). Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9):1189–91.
- Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L., y Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution*, 32(3):820–32.
- Whittaker, J., Sørensen, H., Gadsbøll, V. L., y Hinrichsen, J. (2002). Comparison of the functional insulin binding epitopes of the A and B isoforms of the insulin receptor. *The Journal of Biological Chemistry*, 277(49):47380–4.
- Wildman, D. E., Jameson, N. M., Opazo, J. C., y Yi, S. V. (2009). A fully resolved genus level phylogeny of neotropical primates (Platyrrhini). *Molecular Phylogenetics and Evolution*, 53(3):694–702.
- Williams, C., Hoppe, H.-J., Rezgui, D., Strickland, M., Forbes, B. E., Grutzner, F., Frago, S., Ellis, R. Z., Wattana-Amorn, P., Prince, S. N., Zaccheo, O. J., Nolan, C. M., Mungall, A. J., Jones, E. Y., Crump, M. P., y Hassan, A. B. (2012). An exon splice enhancer primes IGF2:IGF2R binding site structure and function evolution. *Science*, 338(6111):1209–13.
- Williams, C., Rezgui, D., Prince, S. N., Zaccheo, O. J., Foulstone, E. J., Forbes, B. E., Norton, R. S., Crosby, J., Hassan, A. B., y Crump, M. P. (2007). Structural insights into the interaction of Insulin-like Growth Factor 2 with IGF2R domain 11. *Structure*, 15(9):1065–78.
- Wilson, D. E. y Reeder, D. M., editores (2005). Mammal Species of the World: A Taxonomic and Geographic Reference, Volume 1. JHU Press, 3rd edición.
- Wong, K. M., Suchard, M. A., y Huelsenbeck, J. P. (2008). Alignment uncertainty and genomic analysis. *Science*, 319(5862):473–6.
- Worley, K. C., Warren, W. C., Rogers, J., Locke, D., Muzny, D. M., Mardis, E. R., Weinstock, G. M., Tardif, S. D., Aagaard, K. M., Walker, J. A., Wang, Q., Warner, C. M., Wildman, D. E., Witherspoon, D. J., Wright, R. A., Wu, Y., Xiao, W., Xing, J., Zdobnov, E. M., Zhu, B., Gibbs, R. A., y Wilson, R. K. (2014). The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics*, 46(8):850–7.
- Wriston, J. C. (1981). Biochemical peculiarities of the guinea pig and some possible examples of convergent evolution. *Journal of Molecular Evolution*, 17(1):1–9.
- Xu, J. y Messina, J. L. (2009). Crosstalk between growth hormone and insulin signaling. *Vitamins and Hormones*, 80:125–53.
- Xu, Y., Goodyer, C. G., Deal, C., y Polychronakos, C. (1993). Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochemical and Biophysical Research Communications*, 197(2):747–54.
- Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5):568–73.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8):1586–91.
- Yang, Z. y Bielawski, J. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology and Evolution*, 15(12):496–503.
- Yang, Z. y Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, 19(6):908–17.

- Yang, Z., Nielsen, R., Goldman, N., y Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, 155(1):431–49.
- Yang, Z., Wong, W. S. W., y Nielsen, R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4):1107–18.
- Zhang, J., Nielsen, R., y Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Molecular Biology and Evolution*, 22(12):2472–9.
- Zimmerman, A. E., Moule, M. L., y Yip, C. C. (1974). Guinea pig insulin. II. Biological activity. *The Journal of Biological Chemistry*, 249(13):4026–9.

## A APÉNDICE: tablas suplementarias

Tabla A	<b>\-1</b> :	Lista	de	ortólogos	para	INS	(insulin	)
---------	--------------	-------	----	-----------	------	-----	----------	---

Especie	Transcripto	Proteína
Homo sapiens	NM_001185098	NP_001172027
Gorilla gorilla	ENSGGOT0000030259	ENSGGOP0000018194
Macaca fascicularis	NM_001284919.1	NP_001271848
Callithrix jacchus	XM_002755667.2	XP_002755713
Saimiri boliviensis	XM_010331147.1	XP_010329449
Aotus nancymaae	XM_012456192.1	XP_012311615
Otolemur garnettii	ENSOGAP00000019490	ENSOGAP00000019490
Tupaia chinensis	ENSTBEP0000006277	ENSTBEP0000006277
Ochotona princeps	XM_004599211.1	XP_004599268
Oryctolagus cuniculus	ENSOCUT0000000242	ENSOCUP0000000211
Ictidomys tridecemlineatus	ENSSTOT0000022326	ENSSTOP0000017025
Mus musculus (Ins1)	ENSMUST0000039652	ENSMUSP00000049095
Mus musculus (Ins2)	ENSMUST0000000220	ENSMUSP0000000220
Nannospalax galili	XM_008839204.1	XP_008837426
Dipodomys ordii	XM_013021025.1	XP_012876479
Fukomys damarensis	KN122776.1	KFO28273
Heterocephalus glaber	JH173076.1	EHB15713
Octodon degus	M57671.1	AAA40590
Chinchilla lanigera	XM_005384189.1	XP_005384246
Cavia porcellus	ENSCPOT0000025967	ENSCPOP00000019016
Bos taurus	ENSBTAT00000017289	ENSBTAP00000017289
Equus caballus	ENSECAT0000026502	ENSECAP0000022114
Canis lupus familiaris	NM_001130093.1	NP_001123565

Especie	Transcripto	Proteína
Homo sapiens	NM_000618	NP_000609
Pongo abelii	XM_002823659.2	XP_002823705.1
Macaca mulatta	ENSMMUT0000003165	ENSMMUP0000002987
Callithrix jacchus	ENSCJAT0000027842	ENSCJAP0000026346
Saimiri boliviensis	XM_003929619.2	XP_003929668.1
Tarsius syrichta	XM_008062221.1	XP_008060412.1
Otolemur garnettii	XM_003783298.2	XP_003783346.1
Tupaia chinensis	XM_006141401.2	XP_006141463.1
Ochotona princeps	XM_004589598.2	XP_004589655.1
Oryctolagus cuniculus	ENSOCUT0000025296	ENSOCUP0000021389
Ictidomys tridecemlineatus	XM_005322353.2	XP_005322410.1
Mus musculus	NM_001111275.2	NP_001104745.1
Nannospalax galili	XM_008832606.1	XP_008830828.1
Dipodomys ordii	XM_013020148.1	XP_012875602.1
Fukomys damarensis	XM_010631349.1	XP_010629651.1
Heterocephalus glaber	XM_004844842.2	XP_004844899.1
Octodon degus	XM_004623847.1	XP_004623904.1
Ctenomys sociabilis	JU555046.1	NA
Chinchilla lanigera	XM_005374571.2	XP_005374628.1
Cavia porcellus	XM_005006331.2	XP_005006388.2
Bos taurus	XM_005206497.3	XP_005206554.1
Equus caballus	NM_001082498.2	NP_001075967.2
Canis lupus familiaris	NM_001313855.1	NP_001300784.1

 Tabla A-2: Lista de ortólogos para IGF1 (insulin-like growth factor 1)

 Tabla A-3: Lista de ortólogos para IGF2 (insulin-like growth factor 2)

Especie	Transcripto	Proteína
Homo sapiens	NM_001127598	NP_001121070
Gorilla gorilla	XM_004050432.1	XP_004050480.1
Macaca mulatta	XM_015113349.1	XP_014968835.1
Callithrix jacchus	ENSCJAT0000036062	ENSCJAP0000034138
Saimiri boliviensis	XM_010331146.1	XP_010329448.1
Aotus nancymaae	XM_012456191.1	XP_012311614.1
Otolemur garnettii	XM_012810504.1	XP_012665958.1
Tupaia chinensis	KB320452.1	ELW71921.1
Ochotona princeps	XM_004599213.2	XP_004599270.1
Oryctolagus cuniculus	NM_001171406.1	NP_001164877.1
Ictidomys tridecemlineatus	XM_005340768.2	XP_005340825.1
Mus musculus	NM_010514.3	NP_034644.2
Nannospalax galili	XM_008839206.1	XP_008837428.1
Dipodomys ordii	XM_013021164.1	XP_012876618.1
Fukomys damarensis	KN122776.1	KFO28272.1
Heterocephalus glaber	JH173076.1	EHB15714.1
Octodon degus	XM_004627081.1	XP_004627138.1
Ctenomys sociabilis	JU555048.1	NA
Chinchilla lanigera	XM_005384193.2	XP_005384250.1
Cavia porcellus	XM_003468138.3	XP_003468186.1
Bos taurus	XM_005227270.3	XP_005227327.1
Equus caballus	NM_001114539.1	NP_001108011.1
Canis lupus familiaris	XM_005631053.2	XP_005631110.1

Especie	Transcripto	Proteína
Homo sapiens	NM_000208	NP_000199
Macaca mulatta	XM_005587740.2	XP_005587797.1
Callithrix jacchus	XM_008987076.1	XP_008985324.1
Saimiri boliviensis	XM_003938873.2	XP_003938922.1
Tarsius syrichta	XM_008070535.1	XP_008068726.1
Microcebus murinus	XM_012761725.1	XP_012617179.1
Tupaia chinensis	XM_014584917.1	XP_014440403.1
Ochotona princeps	XM_004595727.1	XP_004595784.1
Oryctolagus cuniculus	XM_008249177.1	XP_008247399.1
Ictidomys tridecemlineatus	XM_005332099.2	XP_005332156.1
Mus musculus	XM_006508700.1	XP_006508763.1
Nannospalax galili	XM_008844559.1	XP_008842781.1
Dipodomys ordii	XM_013033503.1	XP_012888957.1
Fukomys damarensis	XM_010605588.1	XP_010603890.1
Heterocephalus glaber	XM_004865316.1	XP_004865373.1
Octodon degus	XM_004644903.1	XP_004644960.1
Ctenomys sociabilis	JU555040.1	NA
Chinchilla lanigera	XM_013509165.1	XP_013364619.1
Cavia porcellus	XM_003468370.3	XP_003468418.1
Bos taurus	XM_002688832.5	XP_002688878.3
Equus caballus	XM_014741203.1	XP_014596689.1
Canis lupus familiaris	XM_005633222.1	XP_005633279.1

 Tabla A-4: Lista de ortólogos para INSR (insulin receptor)

 Tabla A-5: Lista de ortólogos para IGF1R (insulin-like growth factor 1 receptor)

Especie	Transcripto	Proteína
Homo sapiens	NM_000875	NP_000866
Gorilla gorilla	XM_004056810.1	XP_004056858.1
Macaca mulatta	NM_001261352.1	NP_001248281.1
Callithrix jacchus	XM_008998317.1	XP_008996565.1
Aotus nancymaae	XM_012438640.1	XP_012294063.1
Saimiri boliviensis	XM_003940530.2	XP_003940579.1
Tarsius syrichta	XM_008055345.1	ENSTSYP0000002092
Microcebus murinus	XM_012764522.1	XP_012619976.1
Tupaia chinensis	XM_014592757.1	XP_014448243.1
Ochotona princeps	XM_004593165.1	XP_004593222.1
Oryctolagus cuniculus	XM_008248788.1	XP_008247010.1
Ictidomys tridecemlineatus	XM_005320094.2	XP_005320151.1
Mus musculus	NM_010513.2	NP_034643.2
Nannospalax galili	XM_008852347.1	XP_008850569.1
Dipodomys ordii	XM_013012046.1	XP_012867500.1
Fukomys damarensis	KM190089.1	AII23384.1
Heterocephalus glaber	XM_004852704.2	XP_004852761.1
Octodon degus	XM_004623262.2	XP_004623319.1
Ctenomys sociabilis	JU555047.1	NA
Chinchilla lanigera	XM_005381615.2	XP_005381672.1
Cavia porcellus	XM_003475268.3	XP_003475316.1
Bos taurus	NM_001244612.1	NP_001231541.1
Equus asinus	XM_014849602.1	XP_014705088.1
Canis lupus familiaris	XM_545828.5	XP_545828.2

Especie	Transcripto	Proteína
Homo sapiens	NM_014215.2	NP_055030.1
Gorilla gorilla	XM_004026992.1	XP_004027041.1
Macaca mulatta	XM_015112355.1	XP_014967841.1
Callithrix jacchus	XM_008984622.1	XP_008982870.1
Aotus nancymaae	XM_012449794.1	XP_012305217.1
Saimiri boliviensis	XM_003937855.2	XP_003937904.1
Tarsius syrichta	XM_008049799.1	XP_008047990.1
Microcebus murinus	XM_012767685.1	XP_012623139.1
Tupaia chinensis	XM_006155229.2	XP_006155291.1
Ochotona princeps	XM_004589235.1	XP_004589292.1
Oryctolagus cuniculus	XM_002715312.2	XP_002715358.1
Ictidomys tridecemlineatus	XM_005331409.2	XP_005331466.1
Mus musculus	NM_011832.2	NP_035962.2
Nannospalax galili	XM_008855747.1	XP_008853969.1
Dipodomys ordii	XM_013030700.1	XP_012886154.1
Fukomys damarensis	XM_010643780.1	XP_010642082.1
Heterocephalus glaber	XM_004871250.1	XP_004871307.1
Octodon degus	XM_004629719.1	XP_004629776.1
Chinchilla lanigera	XM_013508213.1	XP_013363667.1
Cavia porcellus	XM_003475400.2	XP_003475448.1
Bos taurus	NM_001191235.1	NP_001178164.1
Equus asinus	XM_014836935.1	XP_014692421.1
Canis lupus familiaris	XM_547526.3	XP_547526.1

 Tabla A-6:
 Lista de ortólogos para INSRR (insulin receptor-related receptor)

Especie	Transcripto	Proteína
Homo sapiens	NM_000876.3	NP_000867.2
Pan troglodytes	XM_009452338.1	XP_009450613.1
Macaca mulatta	XM_005551504.2	XP_005551561.1
Callithrix jacchus	XM_002806738.3	XP_002806784.1
Aotus nancymaae	XM_012453040.1	XP_012308463.1
Saimiri boliviensis	XM_010344340.1	XP_010342642.1
Tarsius syrichta	XM_008074381.1	XP_008072572.1
Lemur catta	AF339160.1	AAK71866.1
Tupaia chinensis	XM_014589243.1	XP_014444729.1
Ochotona princeps	XM_004595481.1	XP_004595538.1
Oryctolagus cuniculus	AF339158.1	AAK71864.1
Ictidomys tridecemlineatus	XM_005321423.2	XP_005321480.2
Mus musculus	NM_010515.2	NP_034645.2
Nannospalax galili	XM_008833873.1	XP_008832095.1
Dipodomys ordii	XM_013024527.1	XP_012879981.1
Fukomys damarensis	XM_010631164.1	XP_010629466.1
Heterocephalus glaber	NM_001310295.1	NP_001297224.1
Octodon degus	XM_004639438.1	XP_004639495.1
Ctenomys sociabilis	JU555056.1	NA
Chinchilla lanigera	XM_013517350.1	XP_013372804.1
Cavia porcellus	XM_013148819.1	XP_013004273.1
Bos taurus	NM_174352.2	NP_776777.1
Equus asinus	XM_014858321.1	XP_014713807.1
Canis lupus familiaris	NM_001122602.1	NP_001116074.1

 Tabla A-7: Lista de ortólogos para IGF2R (insulin-like growth factor 2 receptor)

 Tabla A-8: Lista de ortólogos para IGFBP1 (insulin-like growth factor binding protein 1)

Especie	Transcripto	Proteína
Homo sapiens	NM_000596.2	NP_000587
Pongo abelii	XM_002817975.2	XP_002818021.1
Macaca mulatta	XM_001085935.3	XP_001085935.2
Callithrix jacchus	XM_002751403.2	XP_002751449.1
Aotus nancymaae	XM_012458556.1	XP_012313979.1
Saimiri boliviensis	XM_003930841.2	XP_003930890.1
Tarsius syrichta	XM_008060766.1	XP_008058957.1
Otolemur garnettii	XM_003797752.2	XP_003797800.1
Tupaia chinensis	XM_006166039.2	XP_006166101.1
Ochotona princeps	XM_004582232.2	XP_004582289.1
Ictidomys tridecemlineatus	XM_005319258.2	XP_005319315.1
Mus musculus	NM_008341.4	NP_032367.3
Nannospalax galili	XM_008842187.1	XP_008840409.1
Dipodomys ordii	XM_013032957.1	XP_012888411.1
Fukomys damarensis	XM_010624910.1	XP_010623212.1
Heterocephalus glaber	XM_004839882.2	XP_004839939.1
Octodon degus	XM_004630353.1	XP_004630410.1
Chinchilla lanigera	XM_005396640.1	XP_005396697.1
Cavia porcellus	XM_013148215.1	XP_013003669.1
Bos taurus	NM_174554.3	NP_776979.2
Equus caballus	NM_001308602.1	NP_001295531.1
Canis lupus familiaris	XM_005629499.2	XP_005629556.1

Especie	Transcripto	Proteína
Homo sapiens	NM_000597.2	NP_000588
Pongo abelii	XM_003780700.2	XP_003780748.1
Macaca fascicularis	XM_005574235.2	XP_005574292.1
Callithrix jacchus	XM_008999459.1	XP_008997707.1
Saimiri boliviensis	XM_003925726.2	XP_003925775.1
Tarsius syrichta	XM_008048149.1	XP_008046340.1
Microcebus murinus	XM_012785185.1	XP_012640639.1
Tupaia chinensis	XM_006157469.2	XP_006157531.2
Ochotona princeps	XM_004576843.2	XP_004576900.1
Oryctolagus cuniculus	XM_002712467.2	XP_002712513.1
Ictidomys tridecemlineatus	XM_005331832.2	XP_005331889.1
Mus musculus	NM_008342.3	NP_032368.2
Nannospalax galili	XM_008825688.1	XP_008823910.1
Dipodomys ordii	XM_013010959.1	XP_012866413.1
Fukomys damarensis	XM_010611979.1	XP_010610281.1
Heterocephalus glaber	XM_004864615.2	XP_004864672.1
Octodon degus	XM_004641349.1	XP_004641406.1
Ctenmys sociabilis	JU555061.1	NA
Chinchilla lanigera	XM_013519997.1	XP_013375451.1
Cavia porcellus	XM_003462081.3	XP_003462129.1
Bos taurus	NM_174555.1	NP_776980.1
Ceratotherium simum	XM_004427770.2	XP_004427827.1
Canis lupus familiaris	XM_005640742.1	XP_005640799.1

 Tabla A-9: Lista de ortólogos para IGFBP2 (insulin-like growth factor binding protein 2)

 Tabla A-10: Lista de ortólogos para IGFBP3 (insulin-like growth factor binding protein 3)

Especie	Transcripto	Proteína
Homo sapiens	NM_000598.4	NP_000589
Gorilla gorilla	XM_004045417.1	XP_004045465.1
Macaca mulatta	NM_001260547.2	NP_001247476.1
Callithrix jacchus	XM_003733559.2	XP_003733607.2
Aotus nancymaae	XM_012458567.1	XP_012313990.1
Tarsius syrichta	ENSTSYT00000010931	ENSTSYP00000010020
Microcebus murinus	XM_012763427.1	XP_012618881.1
Tupaia chinensis	XM_006150682.1	XP_006150744.1
Ochotona princeps	XM_004582495.2	XP_004582552.1
Ictidomys tridecemlineatus	XM_005319260.2	XP_005319317.1
Mus musculus	NM_008343.2	NP_032369.2
Nannospalax galili	XM_008842186.1	XP_008840408.1
Dipodomys ordii	XM_013032940.1	XP_012888394.1
Fukomys damarensis	XM_010625066.1	XP_010623368.1
Heterocephalus glaber	XM_004840542.2	XP_004840599.2
Octodon degus	XM_004630351.1	XP_004630408.1
Chinchilla lanigera	XM_013520639.1	XP_013376093.1
Cavia porcellus	XM_003465779.3	XP_003465827.2
Bos taurus	NM_174556.1	NP_776981.1
Equus przewalskii	XM_008519309.1	XP_008517531.1
Odobenus rosmarus	XM_004409433.1	XP_004409490.1

× 0 .	, 01
Transcripto	Proteína
NM_001552	NP_001543
XM_002827599.2	XP_002827645.1
NM_001260546.2	NP_001247475.1
XM_002748562.2	XP_002748608.1
XM_012452026.1	XP_012307449.1
XM_003942876.2	XP_003942925.2
XM_008068658.1	XP_008066849.1
XM_012765836.1	XP_012621290.1
XM_006156031.2	XP_006156093.1
XM_004591175.2	XP_004591232.1
XM_002719357.2	XP_002719403.1
XM_005321822.2	XP_005321879.1
NM_010517.3	NP_034647.1
XM_008829447.1	XP_008827669.1
XM_013037993.1	XP_012893447.1
KN125323.1	KFO18783.1
XM_004859434.2	XP_004859491.1
XM_004633729.1	XP_004633786.1
JU555062.1	NA
XM_013519360.1	XP_013374814.1
XM_013149634.1	XP_013005088.1
XM_005898040.2	XP_005898102.1
XM_014833768.1	XP_014689254.1
XM_845091.3	XP_850184.2
	Transcripto           NM_001552           XM_002827599.2           NM_001260546.2           XM_002748562.2           XM_003942876.2           XM_008068658.1           XM_012765836.1           XM_006156031.2           XM_004591175.2           XM_005321822.2           NM_010517.3           XM_008829447.1           XM_004859434.2           XM_004633729.1           JU555062.1           XM_013149634.1           XM_005898040.2           XM_014833768.1           XM_014833768.1           XM_014833768.1

 Tabla A-11: Lista de ortólogos para IGFBP4 (insulin-like growth factor binding protein 4)

Especie	Transcripto	Proteína
Homo sapiens	NM_000599	NP_000590
Pongo abelii	XM_009238079.1	XP_009236354.1
Macaca fascicularis	NM_001284032.1	NP_001270961.1
Callithrix jacchus	XM_002749768.2	XP_002749814.1
Saimiri boliviensis	XM_003925559.2	XP_003925608.1
Tarsius syrichta	XM_008050925.1	XP_008049116.1
Otolemur garnettii	XM_003784970.2	XP_003785018.1
Tupaia chinensis	XM_006157446.1	XP_006157508.1
Ochotona princeps	XM_004576842.2	XP_004576899.1
Oryctolagus cuniculus	XM_002712468.2	XP_002712514.1
Ictidomys tridecemlineatus	XM_005331831.2	XP_005331888.1
Mus musculus	NM_010518.2	NP_034648.2
Nannospalax galili	XM_008825689.1	XP_008823911.1
Dipodomys ordii	XM_013009585.1	XP_012865039.1
Fukomys damarensis	XM_010611945.1	XP_010610247.1
Heterocephalus glaber	XM_004864614.2	XP_004864671.1
Octodon degus	XM_004641348.1	XP_004641405.1
Ctenomys sociabilis	JU555063.1	NA
Chinchilla lanigera	XM_005395681.1	XP_005395738.1
Cavia porcellus	XM_003462071.3	XP_003462119.1
Bos taurus	NM_001105327.2	NP_001098797.1
Equus caballus	NM_001308603.1	NP_001295532.1
Canis lupus familiaris	XM_847792.4	XP_852885.1

 Tabla A-12: Lista de ortólogos para IGFBP5 (insulin-like growth factor binding protein 5)

Especie	Transcripto	Proteína
Homo sapiens	NM_002178	NP_002169
Gorilla gorilla	XM_004053209.1	XP_004053257.1
Macaca nemestrina	XM_011728855.1	XP_011727157.1
Callithrix jacchus	XM_002752462.3	XP_002752508.1
Aotus nancymaae	XM_012450304.1	XP_012305727.1
Saimiri boliviensis	XM_003939212.2	XP_003939261.1
Tarsius syrichta	ENSTSYT00000012906	ENSTSYP00000011843
Otolemur garnettii	XM_003790419.2	XP_003790467.1
Tupaia chinensis	XM_006169120.1	XP_006169182.1
Ochotona princeps	XM_004583289.1	XP_004583346.1
Oryctolagus cuniculus	XM_002711012.2	XP_002711058.1
Ictidomys tridecemlineatus	XM_005325010.2	XP_005325067.1
Mus musculus	NM_008344.3	NP_032370.2
Nannospalax galili	XM_008838447.1	XP_008836669.1
Dipodomys ordii	XM_013032471.1	XP_012887925.1
Fukomys damarensis	XM_010609285.1	XP_010607587.1
Heterocephalus glaber	XM_004872712.1	XP_004872769.1
Octodon degus	XM_004637870.1	XP_004637927.1
Ctenomys sociabilis	JU555064.1	NA
Chinchilla lanigera	XM_005397143.2	XP_005397200.1
Cavia porcellus	XM_003476266.3	XP_003476314.1
Bison bison	XM_010857897.1	XP_010856199.1
Equus asinus	XM_014847494.1	XP_014702980.1
Odobenus rosmarus	XM_004406238.2	XP_004406295.1

 Tabla A-13: Lista de ortólogos para IGFBP6 (insulin-like growth factor binding protein 6)

 Tabla A-14:
 Lista de ortólogos para IGFBP7 (insulin-like growth factor binding protein 7)

Especie	Transcripto	Proteína
Homo sapiens	NM_001553.2	NP_001544.1
Pongo abelii	XM_002814758.3	XP_002814804.1
Macaca mulatta	XM_001083041.3	XP_001083041.1
Callithrix jacchus	XM_002745806.3	XP_002745852.1
Aotus nancymaae	XM_012441065.1	XP_012296488.1
Saimiri boliviensis	XM_010344210.1	XP_010342512.1
Tarsius syrichta	XM_008073627.1	XP_008071818.1
Propithecus coquereli	XM_012641113.1	XP_012496567.1
Tupaia chinensis	XM_006167714.2	XP_006167776.2
Oryctolagus cuniculus	XM_008267816.1	XP_008266038.1
Ictidomys tridecemlineatus	XM_005320054.2	XP_005320111.1
Mus musculus	NM_008048.3	NP_032074.3
Nannospalax galili	XM_008833286.1	XP_008831508.1
Dipodomys ordii	XM_013008932.1	XP_012864386.1
Fukomys damarensis	XM_010609528.1	XP_010607830.1
Heterocephalus glaber	XM_004849681.2	XP_004849738.1
Octodon degus	XM_012513856.1	XP_012369310.1
Chinchilla lanigera	XM_013518518.1	XP_013373972.1
Cavia porcellus	XM_003471680.3	XP_003471728.1
Bos taurus	NM_001102300.2	NP_001095770.1
Equus asinus	XM_014857837.1	XP_014713323.1
Felis catus	XM_003985355.3	XP_003985404.3

Especie	Transcripto	Proteína
Homo sapiens	NM_000515.4	NP_000506.2
Gorilla gorilla	XM_004041177.1	XP_004041225.1
Macaca mulatta	NM_001042738.1	NP_001036203.1
Callithrix jacchus	NM_001290346.1	NP_001277275.1
Cebus albifrons	AM235213.1	CAJ84332.1
Saimiri boliviensis	NM_001279978.1	NP_001266907.1
Tarsius syrichta	AF339080.1	AAK62307.1
Microcebus murinus	XM_012747257.1	XP_012602711.1
Tupaia chinensis	XM_006146836.1	XP_006146898.1
Oryctolagus cuniculus	XM_008271805.1	XP_008270027.1
Ochotona princeps	XM_004597248.1	XP_004597305.1
Ictidomys tridecemlineatus	XM_005328271.1	XP_005328328.1
Mus musculus	NM_008117.3	NP_032143.1
Nannospalax galili	XM_008846583.1	XP_008844805.1
Dipodomys ordii	XM_013033276.1	XP_012888730.1
Fukomys damarensis	XM_010616749.1	XP_010615051.1
Heterocephalus glaber	XM_004868633.2	XP_004868690.1
Octodon degus	XM_004630026.2	XP_004630083.2
Ctenomys sociabilis	JU553788.1	NA
Chinchilla lanigera	XM_005402180.2	XP_005402237.1
Cavia porcellus	NM_001172859.1	NP_001166330.1
Bos taurus	NM_180996.1	NP_851339.1
Equus caballus	NM_001081948.1	NP_001075417.1
Canis lupus familiaris	NM_001003168.1	NP_001003168.1

 Tabla A-15: Lista de ortólogos para GH1 (growth hormone 1)

Especie	Transcripto	Proteína
Homo sapiens	NM_001242399.2	NP_001229328.1
Gorilla gorilla	XM_004058902.1	XP_004058950.1
Macaca mulatta	XM_015139959.1	XP_014995445.1
Callithrix jacchus	XM_002745018.3	XP_002745064.1
Aotus nancymaae	XM_012433844.1	XP_012289267.1
Saimiri boliviensis	XM_010336645.1	XP_010334947.1
Tarsius syrichta	XM_008052716.1	XP_008050907.1
Microcebus murinus	XM_012777771.1	XP_012633225.1
Tupaia chinensis	XM_006166964.2	XP_006167026.2
Oryctolagus cuniculus	XM_008262045.1	XP_008260267.1
Ochotona princeps	XM_004583677.2	XP_004583734.1
Ictidomys tridecemlineatus	XM_005325612.2	XP_005325669.1
Mus musculus	NM_010284.3	NP_034414.2
Nannospalax galili	XM_008842104.1	XP_008840326.1
Dipodomys ordii	XM_013018643.1	XP_012874097.1
Fukomys damarensis	XM_010633879.1	XP_010632181.1
Heterocephalus glaber	XM_004848566.2	XP_004848623.1
Octodon degus	XM_012517043.1	XP_012372497.1
Ctenomys sociabilis	JU553790.1	NA
Chinchilla lanigera	XM_013514982.1	XP_013370436.1
Cavia porcellus	XM_013152804.1	XP_013008258.1
Bos taurus	NM_176608.1	NP_788781.1
Equus caballus	XM_001498656.5	XP_001498706.3
Canis lupus familiaris	XM_005618865.2	XP_005618922.1

 Tabla A-16:
 Lista de ortólogos para GHR (growth hormone receptor)

**Tabla A-17:** Lista de ortólogos para GHRH (growth hormone releasing hormone)

Especie	Transcripto	Proteína
Homo sapiens	NM_021081.5	NP_066567.1
Gorilla gorilla	XM_004062114.1	XP_004062162.1
Macaca nemestrina	XM_011766633.1	XP_011764935.1
Callithrix jacchus	XM_008995894.1	XP_008994142.1
Aotus nancymaae	XM_012454925.1	XP_012310348.1
Saimiri boliviensis	XM_003936354.1	XP_003936403.1
Tarsius syrichta	XM_008071312.1	XP_008069503.1
Microcebus murinus	XM_012754949.1	XP_012610403.1
Tupaia chinensis	XM_006144111.2	XP_006144173.1
Oryctolagus cuniculus	XM_008256017.1	XP_008254239.1
Marmota marmota	XM_015504611.1	XP_015360097.1
Mus musculus	NM_010285.2	NP_034415.1
Nannospalax galili	XM_008855675.1	XP_008853897.1
Dipodomys ordii	XM_013010845.1	XP_012866299.1
Fukomys damarensis	XM_010640029.1	XP_010638331.1
Heterocephalus glaber	XM_004875526.1	XP_004875583.1
Octodon degus	XM_012514664.1	XP_012370118.1
Chinchilla lanigera	XM_013513839.1	XP_013369293.1
Cavia porcellus	XM_013150442.1	XP_013005896.1
Bos taurus	NM_178325.1	NP_847895.1
Equus przewalskii	XM_008535251.1	XP_008533473.1
Canis lupus familiaris	NM_001290112.1	NP_001277041.1

Especie	Transcripto	Proteína
Homo sapiens	NM_000823.3	NP_000814.2
Gorilla gorilla	XM_004045265.1	XP_004045313.1
Macaca fascicularis	XM_005549861.2	XP_005549918.1
Callithrix jacchus	XM_009002263.1	XP_009000511.1
Aotus nancymaae	XM_012475147.1	XP_012330570.1
Saimiri boliviensis	XM_010345721.1	XP_010344023.1
Tarsius syrichta	XM_008051428.1	XP_008049619.1
Otolemur garnettii	XM_003788555.1	XP_003788603.1
Tupaia chinensis	XM_006146251.1	XP_006146313.1
Oryctolagus cuniculus	XM_008261609.1	XP_008259831.1
Ochotona princeps	XM_004582395.2	XP_004582452.2
Ictidomys tridecemlineatus	XM_005319170.1	XP_005319227.1
Mus musculus	NM_001003685.3	NP_001003685.2
Nannospalax galili	XM_008827524.1	XP_008825746.1
Dipodomys ordii	XM_013035128.1	XP_012890582.1
Fukomys damarensis	XM_010628333.1	XP_010626635.1
Heterocephalus glaber	XM_004856289.1	XP_004856346.1
Octodon degus	XM_004626527.1	XP_004626584.1
Chinchilla lanigera	XM_005377509.2	XP_005377566.1
Cavia porcellus	XM_013150538.1	XP_013005992.1
Bos taurus	XM_010804292.2	XP_010802594.1
Equus caballus	XM_001916546.3	XP_001916581.1
Canis lupus familiaris	XM_539502.3	XP_539502.3

Tabla A-18: Lista de ortólogos para GHRHR (growth hormone releasing hormone receptor)

**Tabla A-19:** Modelos de sustitución nucleotídica. Se presentan los modelos con mejor ajuste para cada<br/>gen según el valor de AICc. Los cálculos fueron realizados con Modelgeneartor v0.8.5<br/>(Keane *et al.*, 2006)

Gen	Modelo	AICc
INS	HKY+G	6112,8112
IGF1	K81uf+I+G	5414,8118
IGF2	TVM+G	3377,87139
INSR	GTR+I+G	56199,6067
IGF1R	GTR+I+G	50466,5398
INSRR	GTR+I+G	50078,6184
IGF2R	GTR+I+G	64558,3475
IGFBP1	GTR+G	15245,4586
IGFBP2	TrN+G	10133,3785
IGFBP3	GTR+I+G	4297,7259
IGFBP4	GTR+I+G	3825,3508
IGFBP5	TIM+I+G	4032,2742
IGFBP6	GTR+I+G	6009,75088
IGFBP7	GTR+I+G	2957,2371
GH1	HKY+G	5219,6549
GHR	HKY+I+G	14270,0909
GHRH	TVM+G	5028,3623
GHRHR	GTR+G	20399,0099

				) ) )	no han																
Taxa	<b>A1</b>	C1	G1	T1	A2	C2	G2	<b>T2</b>	A3	C3	G3	T3	A	C	IJ	L	TOTAL	GC1	GC2	GC3	GC
Homo	2531	2394	3123	1850	3030	2219	2169	2480	1329	3725	3059	1785	6890	8338	8351	6115	29694	19	15	23	56
Gorilla	2164	2074	2701	1573	2577	1934	1896	2106	1164	3181	2608	1560	5905	7189	7205	5239	25538	19	15	23	56
Macaca	2540	2387	3116	1857	3029	2223	2170	2478	1332	3727	3073	1768	6901	8337	8359	6103	29700	19	15	23	56
Callithrix	2559	2393	3077	1845	3000	2230	2176	2468	1307	3674	3039	1854	6866	8297	8292	6167	29622	18	15	23	56
Saimiri	2460	2303	2992	1778	2891	2149	2082	2411	1281	3531	2942	1779	6632	7983	8016	5968	28599	19	15	23	56
Aotus	2046	1876	2465	1455	2359	1796	1728	1959	1101	2864	2361	1516	5506	6536	6554	4930	23526	18	15	22	56
Tarsier	2313	2105	2717	1629	2678	1986	1906	2194	1165	3318	2713	1568	6156	7409	7336	5391	26292	18	15	23	56
Otolemur	2484	2423	3134	1796	2987	2229	2175	2446	1161	3932	3241	1503	6632	8584	8550	5745	29511	19	15	24	58
Tupaia	2453	2337	3050	1780	2910	2193	2122	2395	1076	4054	3124	1366	6439	8584	8296	5541	28860	19	15	25	58
Ochotona	2323	2418	3035	1745	2850	2169	2121	2381	943	4056	3163	1359	6116	8643	8319	5485	28563	19	15	25	59
Oryctolagus	2309	2337	3029	1722	2816	2147	2071	2363	831	4184	3296	1086	5956	8668	8396	5171	28191	19	15	27	61
Ictidomys	2500	2396	3144	1835	2992	2235	2185	2463	1187	3896	3173	1619	6679	8527	8502	5917	29625	19	15	24	57
Mus	2594	2328	3077	1883	3023	2215	2139	2505	1467	3517	2845	2053	7084	8060	8061	6441	29646	18	15	21	54
Nannospalax	2493	2265	3032	1856	2925	2179	2088	2454	1410	3383	2801	2052	6828	7827	7921	6362	28938	18	15	21	54
Dipodomys	2414	2307	3002	1794	2926	2112	2073	2406	1215	3666	2980	1656	6555	8085	8055	5856	28551	19	15	23	57
Fukomys	2509	2326	3037	1812	2973	2169	2112	2430	1238	3727	3043	1676	6720	8222	8192	5918	29052	18	15	23	56
Heterocephalus	2535	2405	3110	1846	3013	2225	2180	2478	1282	3795	3123	1696	6830	8425	8413	6020	29688	19	15	23	57
Octodon	2511	2405	3093	1820	3010	2199	2157	2463	1267	3788	3104	1670	6788	8392	8354	5953	29487	19	15	23	57
Ctenomys	1906	1646	2281	1367	2301	1596	1530	1773	987	2636	2208	1369	5194	5878	6019	4509	21600	18	14	22	55
Chinchilla	2552	2379	3106	1840	3024	2208	2167	2478	1267	3803	3133	1674	6843	8390	8406	5992	29631	19	15	23	57
Cavia	2530	2391	3102	1826	3033	2210	2142	2464	1355	3744	3019	1731	6918	8345	8263	6021	29547	19	15	23	56
Bos	2464	2452	3151	1811	2974	2237	2191	2476	1065	4150	3289	1374	6503	8839	8631	5661	29634	19	15	25	59
Equus	2495	2415	3136	1843	3009	2237	2171	2472	1172	3931	3166	1620	6676	8583	8473	5935	29667	19	15	24	57
Canis	2521	2395	3135	1836	2986	2221	2197	2483	1182	3809	3182	1714	6689	8425	8514	6033	29661	19	15	24	57

**Tabla A-20:** Contenido de bases y G+C (%) para los diferentes *taxa* estudiados (suma de todas las secuencias por *taxa*).

<b>Tabla A-21:</b> Us	so de an	ninoáci	dos pai	ra los d	iferente	s taxa	estudia	dos (su	ma de t	odas la:	s secue	ncias p	or taxa	·							
Taxa	Α	С	D	F	F	G	Η	Ι	К	L	Μ	Z	Р	Q	R	S	Т	V	W	Υ	Total
Homo	589	421	489	693	365	739	235	402	493	918	182	408	607	372	562	779	550	613	141	340	8686
Gorilla	532	374	416	582	306	648	198	336	424	787	154	332	527	336	480	667	481	523	121	288	8512
Macaca	588	418	488	691	367	735	234	401	490	915	181	412	607	372	572	783	551	614	139	342	0066
Callithrix	577	419	499	667	351	723	226	398	480	925	183	410	620	382	583	788	558	611	138	336	9874
Saimiri	570	396	486	651	341	686	210	389	453	904	178	397	586	364	576	745	540	599	132	330	9533
Aotus	482	331	408	506	281	578	177	327	379	735	125	323	478	299	449	630	464	491	112	267	7842
Tarsier	504	352	449	587	326	637	203	373	451	793	162	370	558	310	516	889	510	540	127	308	8764
Otolemur	614	417	498	678	351	733	235	380	484	915	189	387	614	370	587	745	554	611	140	335	9837
Tupaia	604	403	475	667	363	700	224	374	471	885	169	390	589	356	573	762	549	604	135	327	9620
Ochotona	606	410	500	647	355	707	224	360	436	918	173	372	601	355	576	722	531	575	137	316	9521
Oryctolagus	611	396	481	648	354	689	208	354	439	891	164	374	582	354	565	725	518	600	132	312	9397
Ictidomys	620	414	486	689	366	749	237	394	482	921	182	386	603	378	580	769	546	600	139	334	9875
Mus	581	425	497	681	367	713	224	420	500	915	198	402	598	376	564	761	574	605	138	343	9882
Nannospalax	591	402	490	644	368	706	218	390	482	902	193	399	588	356	544	761	538	601	137	336	9646
Dipodomys	580	412	471	659	354	697	224	372	480	905	180	393	567	368	548	729	521	595	131	331	9517
Fukomys	585	398	498	655	356	697	223	383	500	806	181	394	580	372	559	774	548	602	140	331	9684
Heterocephalus	598	418	493	681	362	730	235	386	497	932	190	395	605	375	575	776	564	809	139	337	9896
Octodon	885	422	485	676	356	714	231	371	492	915	191	404	584	387	574	771	565	630	138	335	9829
Ctenomys	414	300	394	514	274	503	160	273	388	622	148	312	414	273	409	591	408	456	87	260	7200
Chinchilla	599	424	491	682	358	721	226	385	499	922	200	407	580	385	572	784	558	613	137	334	9877
Cavia	601	414	485	989	362	721	236	383	503	925	185	401	585	386	556	775	564	609	136	336	9849
Bos	629	417	486	675	354	738	238	388	476	937	174	398	808	361	593	762	546	623	135	340	9878
Equus	610	415	488	692	361	734	229	392	493	926	181	408	627	364	589	768	527	612	138	335	6886
Canis	605	419	497	674	361	744	228	403	485	927	177	398	611	361	587	771	544	615	137	343	9887

Prueba de sele	ección positiva					
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	48	4,5896	8,0797	-3030,4276	3,8975	0,0484
Nulo ( $\omega^+ = 1$ )	47	4,5207	7,9609	-3032,3789		
dN/dS para la	s clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,5313	0,1167	0,2886	0,0634		
$\omega_{background}$	0,0787	1,0000	0,0787	1,0000		
$\omega^+$	0,0787	1,0000	2,0176	2,0176		
Sitios con BEE	$B>0,95$ para $\omega^+$	->1 en el linaje	blanco:			
9 P 0,980*	27 N 0,985*	41 L 0,997**	45 E 0,998**	97 S 0,983*		
22 A 0,963*	28 Q 0,991**	42 V 0,991**	53 K 1,000**	101 L 0,992**		
26 V 0,990**	34 H 0,990*	44 G 0,990*	70 G 0,983*	102 Y 0,989*		

 Tabla A-22: Modelo de ramas y sitios (CODEML) para INS seleccionando el clado Caviomorpha como linaje blanco.

Tabla A-23: Modelo de ramas y sitios (	(CODEML) para	IGF2 seleccionando	la rama	ancestral	de
Platyrrhini como linaje blanc	co.				

Prueba de selec	Prueba de selección positiva								
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue			
$\overline{\omega^+}$	48	4,6206	4,1702	-3409,0371	10,3759	0,0013			
Nulo ( $\omega^+ = 1$ )	47	4,5726	4,1211	-3414,2251					
dN/dS para las	clases de sitio	,							
clase de sitio	0	1	2a	2b					
proporción	0,7185	0,2535	0,0207	0,0073					
$\omega_{background}$	0,0562	1,0000	0,0562	1,0000					
$\omega^+$	0,0562	1,0000	25,5621	25,5621					
Sitios con BEB	> 0,95 para a	$\mathbf{p}^+ > 1$ en el li	naje blanco:						
129 G 0,998**									

<sup>136</sup> A 0,995\*\*

136 A 0,976\*

Prueba de seleco	ción positiva					
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\omega^+$	48	4,5871	4,1284	-3413,2107	3,6221	0,0570
Nulo ( $\omega^+ = 1$ )	47	4,5741	4,1225	-3415,0218		
dN/dS para las	clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,7151	0,2531	0,0235	0,0083		
$\omega_{background}$	0,0559	1,0000	0,0559	1,0000		
$\omega^+$	0,0559	1,0000	5,3933	5,3933		
Sitios con BEB >	> 0,95 <b>para</b> ω	$^+ > 1$ en el lin	aje blanco:			
129 G 0,988*						

**Tabla A-24:** Modelo de ramas y sitios (CODEML) para IGF2 seleccionando el clado Platyrrhini como linaje blanco.

NOTA: Para los sitios bajo selección positiva se indica: posición del codón; aminoácido en la secuencia de *H. sapiens* usada como referencia; probabilidad posterior por BEB.

Tabla A-25: Modelo de ramas y sitios (CODEML)	) para INSR	seleccionando	la rama	ancestral	de
Caviomorpha como linaje blanco.					

Prueba de selece	Prueba de selección positiva								
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue			
$\overline{\omega^+}$	46	3,2386	5,1860	-26331,7236	3,1860	0,0743			
Nulo ( $\omega^+ = 1$ )	45	3,2397	5,1827	-26333,3166					
dN/dS para las	clases de sitio	,							
clase de sitio	0	1	2a	2b					
proporción	0,9548	0,0432	0,0019	0,0001					
$\omega_{background}$	0,0242	1,0000	0,0242	1,0000					
$\omega^+$	0,0242	1,0000	17,9400	17,9400					
Sitios con BEB >	> 0,95 <b>para</b> a	$p^+ > 1$ en el li	naje blanco:						

185 I 0,962\*

Prueba de sele	ección positiva					
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	50	3,5224	4,1844	-23630,7092	11,3255	0,0008
<b>Nulo</b> ( $\omega^+ = 1$ )	49	3,5022	4,1731	-23636,3719		
dN/dS para la	as clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,8740	0,0314	0,0913	0,0033		
$\omega_{background}$	0,0210	1,0000	0,0210	1,0000		
$\omega^+$	0,0210	1,0000	3,4362	3,4362		
Sitios con BEI	$B>0,95$ para $\omega^+$	> 1 en el linaje l	olanco:			
24 S 0,994**	293 S 0,997**	315 G 0,996**	558 P 0,961*			
280 D 0,996**	294 E 0,992**	320 Y 0,995**	694 G 0,997**			
291 S 0,998**	295 G 0,994**	442 K 0,998**				

**Tabla A-26:** Modelo de ramas y sitios (CODEML) para IGF1R seleccionando la rama ancestral de Platyrrhini como linaje blanco.

Prueba de selec	Prueba de selección positiva								
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue			
$\overline{\omega^+}$	48	4,5816	4,0975	-24077,8653	2,8645	0,0906			
Nulo ( $\omega^+ = 1$ )	47	4,5786	4,0960	-24079,2975					
dN/dS para las	clases de sitio								
clase de sitio	0	1	2a	2b					
proporción	0,8933	0,0859	0,0190	0,0018					
$\omega_{background}$	0,0395	1,0000	0,0395	1,0000					
$\omega^+$ .	0,0395	1,0000	2,7946	2,7946					
Sitios con BEB	> 0,95 <b>para</b> ω	$^+ > 1$ en el lir	naje blanco:						
710 Q 0,979*									
718 H 0,975*									

 Tabla A-27: Modelo de ramas y sitios (CODEML) para INSRR seleccionando el clado Platyrrhini como linaje blanco.

NOTA: Para los sitios bajo selección positiva se indica: posición del codón; aminoácido en la secuencia de *H. sapiens* usada como referencia; probabilidad posterior por BEB.

1008 T 0,977\*

Prueba de selección positiva								
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue		
$\overline{\omega^+}$	50	3,2047	6,5959	-62658,8805	7,9656	0,0048		
Nulo ( $\omega^+ = 1$ )	49	3,2031	6,5930	-62662,8633				
dN/dS para las	clases de sit	tio						
clase de sitio	0	1	2a	2b				
proporción	854	142	3	1				
$\omega_{background}$	77	1000	77	1000				
000000000000000000000000000000000000000			01000	21000				

**Tabla A-28:** Modelo de ramas y sitios (CODEML) para IGF2R seleccionando la rama ancestral de Platyrrhini como linaje blanco.

2023 G 0,967\*

NOTA: Para los sitios bajo selección positiva se indica: posición del codón; aminoácido en la secuencia de *H. sapiens* usada como referencia; probabilidad posterior por BEB.

Prueba de sele	ección positiva					
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	41	3,5633	8,9313	-7426,9828	20,6683	<0,0001
Nulo ( $\omega^+ = 1$ )	40	3,4775	8,7174	-7437,3169		
dN/dS para la	s clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,4209	0,1333	0,3385	0,1072		
$\omega_{background}$	0,1187	1,0000	0,1187	1,0000		
$\omega^+$	0,1187	1,0000	5,8759	5,8759		
Sitios con BEB	$B > 0,95$ para $\omega$	$^+ > 1$ en el linaje	blanco:			
40 N 0,998**	74 K 1,000**	122 R 0,975*	158 S 0,975*	201 P 0,998**		
42 P 1,000**	85 R 0,987*	128 M 0,999**	177 A 0,968*	,		
69 A 0,989*	113 S 0,989*	138 L 0,998**	184 E 0,974*			

**Tabla A-29:** Modelo de ramas y sitios (CODEML) para IGFBP1 seleccionando la rama ancestral de Platyrrhini como linaje blanco.

Prueba de sel	ección positiva					
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	41	3,5717	8,9187	-7411,5058	15,8802	0,0001
Nulo ( $\omega^+ = 1$ )	40	3,5117	8,8135	-7419,4459		
dN/dS para la	as clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,4459	0,1578	0,2927	0,1036		
$\omega_{background}$	0,1048	1,0000	0,1048	1,0000		
$\omega^+$ .	0,1048	1,0000	2,8630	2,8630		
Sitios con BEI	$B > 0,95$ para $\omega$	$^+ > 1$ en el linaje	e blanco:			
40 L 0,984*	74 A 0,998**	113 S 0,998**	122 L 1,000**	128 M 0,992**	195 K 0,973*	
42 P 0,996**	98 T 0,971*	117 T 1,000**	125 F 0,973*	138 L 0,992**	201 S 0,986*	
70 A 0,981*	99 R 0,982*	118 E 0,973*	126 H 0,988*	150 A 0,970*	234 S 1,000**	

**Tabla A-30:** Modelo de ramas y sitios (CODEML) para IGFBP1 seleccionando el clado Platyrrhini como linaje blanco.

Tabla A-31: Modelo de ramas y s	os (CODEML) para IGFBP1 seleccionando el clado Caviomorpha
como linaje blanco.	

Prueba de seleco	Prueba de selección positiva								
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue			
$\overline{\omega^+}$	41	3,4923	8,7973	-7472,4468	2,9111	0,0880			
Nulo ( $\omega^+ = 1$ )	40	3,4841	8,7671	-7473,9024					
dN/dS para las	clases de sitio								
clase de sitio	0	1	2a	2b					
proporción	0,7473	0,2438	0,0068	0,0022					
$\omega_{background}$	0,1399	1,0000	0,1399	1,0000					
$\omega^+$	0,1399	1,0000	6,3711	6,3711					
Sitios con BEB >	> 0,95 <b>para</b> <i>w</i>	$^+ > 1$ en el lin	aje blanco:						
7 A 0,968*									

Prueba de selecc	ción positiva					
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	48	3,5150	4,5119	-4839,2496	2,8843	0,0894
Nulo ( $\omega^+ = 1$ )	47	3,4752	4,4647	-4840,6918		
dN/dS para las	clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,8107	0,0616	0,1187	0,0090		
$\omega_{background}$	0,0712	1,0000	0,0712	1,0000		
$\omega^+$	0,0712	1,0000	3,4911	3,4911		
Sitios con BEB >	> 0,95 <b>para</b> ω	$^+ > 1$ en el lin	aje blanco:			
134 L 0,996**						
139 K 0,978*						

**Tabla A-32:** Modelo de ramas y sitios (CODEML) para IGFBP2 seleccionando la rama ancestral de Platyrrhini como linaje blanco.

Tabla A-33: Modelo de ramas y sitios (CODEM	L) para IGFBP4 seleccionando la rama ancestral de
Platyrrhini como linaje blanco.	

Prueba de selección positiva							
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue	
$\overline{\omega^+}$	48	6,2037	3,6138	-3702,7358	5,6568	0,0174	
Nulo ( $\omega^+ = 1$ )	47	6,0479	3,5416	-3705,5642			
dN/dS para las	clases de sitio						
clase de sitio	0	1	2a	2b			
proporción	0,8068	0,0516	0,1331	0,0085			
$\omega_{background}$	0,0264	1,0000	0,0264	1,0000			
$\omega^+$	0,0264	1,0000	7,2375	7,2375			
Sitios con BEB >	> 0,95 <b>para</b> ω	$^+ > 1$ en el lin	aje blanco:				
20 E 0,999**							

164 Q 0,991\*\* 193 A 0,950\*

Prueba de selección positiva							
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue	
$\overline{\omega^+}$	50	3,6588	6,0055	-5939,9215	3,1510	0,0759	
<b>Nulo</b> ( $\omega^+ = 1$ )	49	3,6532	5,9935	-5941,4970			
dN/dS para las	clases de sitio						
clase de sitio	0	1	2a	2b			
proporción	0,6933	0,2933	0,0094	0,0040			
$\omega_{background}$	0,1178	1,0000	0,1178	1,0000			

**Tabla A-34:** Modelo de ramas y sitios (CODEML) para IGFBP6 seleccionando la rama ancestral de Caviomorpha como linaje blanco.

NOTA: NA: no aplicable. No se infirieron sitios bajo selección positiva según un valor de BEB > 0,95.

Tabla A-35: Modelo de ramas y sitios (Co	DEML) para	GHRH se	eleccionando	la rama	ancestral	de
Platyrrhini como linaje blanco						

Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	46	4,9811	6,2654	-2509,7298	10,3710	0,0013
Nulo ( $\omega^+ = 1$ )	45	4,8673	6,1053	-2514,9153		
dN/dS para las	clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,7526	0,1603	0,0718	0,0153		
$\omega_{background}$	0,2145	1,0000	0,2145	1,0000		
$\omega^+$	0,2145	1,0000	64,0947	64,0947		
Sitios con BEB	> 0,95 para a	$\mathbf{p}^+ > 1$ en el li	naje blanco:			

105 S 0,998\*\*

NA

Tabla A-36: Modelo de ramas y sitios (CO	ODEML) para GHRH seleccionando el clado Platyrrhini como
linaje blanco.	

Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	46	4 8978	6 1412	-2514 6216	3 3883	0.0657
Nulo ( $\omega^+ = 1$ )	45	4,8721	6,1138	-2516,3158	5,5005	0,0007
dN/dS para las	clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,7974	0,1683	0,0284	0,0060		
$\omega_{background}$	0,2199	1,0000	0,2199	1,0000		
$\omega^+$	0,2199	1,0000	11,4519	11,4519		
Sitios con BEB >	> 0.95 <b>para</b> @	$^+ > 1$ en el lir	naie blanco:			

105 S 0,951\*

NOTA: Para los sitios bajo selección positiva se indica: posición del codón; aminoácido en la secuencia de *H. sapiens* usada como referencia; probabilidad posterior por BEB.

Tabla A-37: Modelo de ramas y sitios (CODI	EML) para GHRHR seleccionando la rama ancestral de
Platyrrhini como linaje blanco.	

Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	48	3,9298	5,8670	-9951,0971	6,9287	0,0085
Nulo ( $\omega^+ = 1$ )	47	3,9299	5,7139	-9954,5614		
dN/dS para las	clases de sitio	,				
clase de sitio	0	1	2a	2b		
proporción	0,8044	0,1927	0,0024	0,0006		
$\omega_{background}$	0,0831	1,0000	0,0831	1,0000		
$\omega^+$	0.0831	1.0000	531,7692	531,7692		

205 S 0,968\*

**Tabla A-38:** Modelo de ramas y sitios (CODEML) para GHRHR seleccionando el clado Platyrrhini como linaje blanco.

Prueba de sele	cción positiva					
Modelo	NP	Ts/Tv	TL	lnL	LRT	pvalue
$\overline{\omega^+}$	48	3,9566	5,7558	-9932,4410	10,4562	0,0012
Nulo ( $\omega^+ = 1$ )	47	3,9360	5,7377	-9937,6691		
dN/dS para las	s clases de sitio					
clase de sitio	0	1	2a	2b		
proporción	0,7607	0,1814	0,0468	0,0112		
$\omega_{background}$	0,0764	1,0000	0,0764	1,0000		
$\omega^+$ °	0,0764	1,0000	4,0045	4,0045		
Sitios con BEB	$>$ 0,95 para $\omega^+$ $>$	1 en el linaje bla	nco:			
128 I 0,987*	256 W 0,963*	267 E 0,983*				
205 S 0,959*	259 C 0,979*					
249 P 0,994**	260 K 0,998**					

INS

Tabla A-39: Resultados de aBSREL para los genes INS, IGF1 e IGF2.

Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección	
1	35	81%	2%	0	
2	8	19%	98%	0	
Modelo	lnL	np	AICc		
branch-wise $\omega$ (MG94)	-3013,85	100	6236,09		
branch-site $\omega$	-2973,67	116	6190,69		
Sin ramas con $\omega > 1$ y <i>pvalue</i> corregido por Holm-Bonferroni $< 0, 1$					

IGF1 Clases de  $\omega$ % TL # bajo selección # ramas % ramas 39 26% 1 91% 0 2 4 9% 74% 0 Modelo lnL AICc np -2561,13 100 5328,18 *branch-wise* ω (MG94) -2544,64 branch-site  $\omega$ 108 5312,18 Sin ramas con  $\omega > 1$  y *pvalue* corregido por Holm-Bonferroni < 0, 1

IGF2				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	35	81%	5%	0
2	8	19%	95%	0
Modelo	lnL	np	AICc	
branch-wise $\omega$ (MG94)	-3384,41	100	6973,85	
branch-site $\omega$	-3350,54	116	6939,87	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
ancestro Platyrrhini	8,51	0,2189*	0,0050	$\omega_1 = 0,40(97\%); \omega_2 = 35,20(3\%)$

NOTA: Se consideraron bajo selección aquellas ramas con  $\omega > 1$  que presentaron significancia para un  $\alpha \le 0, 1$ , luego de la corrección de Holm-Bonferroni por pruebas múltiples. En este sentido, IGF2 no presentó ramas con señal de selección positiva significativa, pero igual se muestran las características de la rama ancestral de Platyrrhini debido a su concordancia con los resultados de BSREL (Tabla 3.6). TL: largo del árbol.

118

INSR				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	22	54%	38%	0
2	19	46%	62%	1
Modelo	lnL	np	AICc	
<i>branch-wise</i> ω (MG94)	-26323,57	96	52839,77	
branch-site $\omega$	-26212,26	134	52693,76	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
ancestro Octodon + Ctenomys	10,62	0,0716	0,0017	$\omega_1 = 0,05(98\%); \omega_2 = 9,12(2\%)$
IGF1R				
Clases de <i>w</i>	# ramas	% ramas	% TL	# bajo selección
1	34	76%	8%	0
2	11	24%	92%	2
Modelo	lnL	np	AICc	
<i>branch-wise</i> ω (MG94)	-23560,47	104	47329,6	
branch-site $\omega$	-23432,86	126	47118,69	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
Oryctolagus	34,46	«0,0001	«0,0001	$\omega_1 = 0,02(99\%); \omega_2 = 3330(1\%)$
ancestro Platyrrhini	18,31	0,0016	«0,0001	$\omega_1 = 0,25(99\%); \omega_2 = 32,20(1\%)$
INSRR				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	32	74%	63%	0
2	11	26%	37%	1
Modelo	lnL	np	AICc	
branch-wise $\omega$ (MG94)	-24070,63	100	48341,95	
branch-site $\omega$	-24010,71	122	48266,43	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
Callithrix	13,17	0,0208	0,0005	$\omega_1 = 0, 19(100\%); \omega_2 = 448(0\%)$

 Tabla A-40:
 Resultados de aBSREL para los genes INSR, IGF1R e INSRR.

 Tabla A-41: Resultados de aBSREL para los genes IGF2R e IGFBP1.

IGF2R				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	15	33%	6%	0
2	30	67%	94%	7
Modelo	lnL	np	AICc	
<i>branch-wise</i> $\omega$ (MG94)	-63413,42	104	127035,22	
branch-site $\omega$	-63020,62	164	126370,16	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
Pan	70,56	«s0,0001	«0,0001	$\omega_1 = 0,28(99\%); \omega_2 = 1720(1\%)$
Lemur	24,83	0,0001	«0,0001	$\omega_1 = 0,15(99\%); \omega_2 = 149(1\%)$
Ictidomys	15,63	0,0600	0,0001	$\omega_1 = 0.07(99\%); \omega_2 = 931(1\%)$
Saimiri	12,81	0,0243	0,0006	$\omega_1 = 0,25(100\%); \omega_2 = 94,50(0\%)$
ancestro Homo + Pan	11,12	0,0558	0,0013	$\omega_1 = 0,17(100\%); \omega_2 = 46,90(0\%)$
Callithrix	10,04	0,0940	0,0023	$\omega_1 = 0, 10(100\%); \omega_2 = 10000(0\%)$
ancestro Laurasiatheria	9,91	0,0977	0,0024	$\omega_1 = 0, 11(99\%); \omega_2 = 130(1\%)$
IGFBP1				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	31	76%	10%	0
2	10	24%	90%	1
Modelo	lnL	np	AICc	
branch-wise $\omega$ (MG94)	-7445,72	96	15086,94	
branch-site ω	-7397,66	116	15032,45	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
ancestro Platyrrhini	12,32	0,0304	0,0007	$\omega_1 = 1,00(91\%); \omega_2 = 20,80(9\%)$

IGFBP2				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	36	84%	3%	0
2	7	16%	97%	4
Modelo	lnL	np	AICc	
branch-wise $\omega$ (MG94)	-4763,92	100	9731,47	
branch-site omega	-4679,59	114	9591,89	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
Fukomys	35,17	«0,0001	«0,0001	$\omega_1 = 0,00(95\%); \omega_2 = 10000(5\%)$
Tarsius	27,56	«0,0001	«0,0001	$\omega_1 = 0,01(92\%); \omega_2 = 48,80(8\%)$
ancestro <i>Mus</i> + <i>Nannospalax</i>	11,85	0,0384	0,0009	$\omega_1 = 0.09(97\%); \omega_2 = 10000(3\%)$
ancestro Lagomorpha	11,28	0,0501	0,0012	$\omega_1 = 0,45(98\%); \omega_2 = 137(2\%)$
IGFBP3				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	32	82%	1%	0
2	7	18%	99%	0
Modelo	lnL	np	AICc	
branch-wise ω (MG94)	-4170,66	92	8529,44	
branch-site $\boldsymbol{\omega}$	-4144,82	106	8507,13	
Sin ramas con $\omega > 1$ y pval	ue corregid	o por Holm-Bonfer	<b>roni</b> < 0, 1	
IGFBP4				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	40	93%	5%	0
2	3	7%	95%	0
Modelo	lnL	np	AICc	
branch-wise $\omega$ (MG94)	-3639,45	100	7482,75	
branch-site $\omega$	-3629,34	106	7475,01	

Tabla A-42: Resultados de aBSREL para los genes IGFBP2, IGFBP3 e IGFBP4.

Sin ramas con  $\omega > 1$  y *pvalue* corregido por Holm-Bonferroni < 0, 1

IGFBP5				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	40	93%	57%	0
2	3	7%	43%	0
Modelo	lnL	np	AICc	
<i>branch-wise</i> $\omega$ (MG94)	-3779,52	100	7762,34	
branch-site $\omega$	-3765,28	106	7746,28	
Sin ramas con $\omega > 1$ y	<i>y pvalue</i> cor	regido por Holm-B	onferroni <	0,1
IGFBP6				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	40	89%	10%	0
2	5	11%	90%	1
Modelo	lnL	np	AICc	
<i>branch-wise</i> $\omega$ (MG94)	-5947,40	104	12106,74	
branch-site $\omega$	-5928,60	114	12089,94	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
Octodon	12,45	0,0312	0,0007	$\omega_1 = 0,43(99\%); \omega_2 = 10000(1\%)$
IGFBP7				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	38	93%	68%	0
2	3	7%	32%	0
Modelo	lnL	np	AICc	
<i>branch-wise</i> ω (MG94)	-2834,94	96	5866,26	
branch-site $\omega$	-2823,96	102	5856,86	
<b>Sin ramas con</b> $\omega > 1$ y	<i>y pvalue</i> cor	regido por Holm-B	onferroni <	0,1

 Tabla A-43:
 Resultados de aBSREL para los genes IGFBP5, IGFBP6 e IGFBP7.

 Tabla A-44: Resultados de aBSREL para los genes GH1, GHR, GHRH y GHRHR.

GH1				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	39	87%	71%	0
2	6	13%	29%	0
Modelo	lnL	np	AICc	
<i>branch-wise</i> $\omega$ (MG94)	-5035,15	104	10282,63	
branch-site $\omega$	-5013,97	116	10265,32	
Sin ramas con $\omega > 1$ y	<i>pvalue</i> corr	egido por Holm-Bo	nferroni < (	),1
GHR				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	35	78%	48%	0
2	10	22%	52%	0
Modelo	lnL	np	AICc	
branch-wise $\omega$ (MG94)	-14135,08	104	28479,62	
branch-site ω	-14098,22	124	28446,51	
Sin ramas con $\omega > 1$ y	<i>pvalue</i> corr	egido por Holm-Bo	nferroni < (	), 1
GHRH				
Clases de $\omega$	# ramas	% ramas	% TL	# bajo selección
1	38	93%	0%	0
2	3	7%	100%	0
Modelo	lnL	np	AICc	
branch-wise $\omega$ (MG94)	-2459,62	96	5119,58	
branch-site ω	-2445,74	102	5104,92	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
Mus	10,98	0.0598	0,0014	$\omega_1 = 0.302(91\%); \omega_2 = 10000(9\%)$
ancestro Platyrrhini	9,14	0,1476*	0,0036	$\omega_1 = 1,00(98\%); \omega_2 = 173(2\%)$
GHRHR				
Clases de <i>w</i>	# ramas	% ramas	% TL	# bajo selección
1	32	74%	19%	0
2	11	26%	81%	1
Modelo	lnL	np	AICc	
<i>branch-wise</i> ω (MG94)	-9983,20	100	20168,52	
branch-site $\omega$	-9912,02	122	20071,19	
<b>Ramas con</b> $\omega > 1$	LRT	<i>pvalue</i> corregido	pvalue	distribución de $\omega$
Callithrix	49,40	«0,0001	«0,0001	$\omega_1 = 0,15(96\%); \omega_2 = 151(4\%)$

NOTA: Se consideraron bajo selección aquellas ramas con  $\omega > 1$  que presentaron significancia para un  $\alpha \le 0, 1$ , luego de la corrección de Holm-Bonferroni por pruebas múltiples. En este sentido, en GHRH la rama ancestral de Platyrrhini no presentó señal de selección positiva significativa, pero igual se muestra para enfatizar la concordancia con los resultados de BSREL (Tabla 3.6). TL: largo del árbol.

 Tabla A-45:
 Resultados de BUSTED para los genes IGF2, INSR, IGF1R, INSRR e IGF2R.

IGF2: ancestro Platyr	rhini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-3394,62	67	6925,50	0,0140	0,32 (81%)	0,46 (16%)	32,10 (39
(ramas del background)					0,05 (81%)	0,45 (11%)	1,34 (8%)
modelo con $\omega_3 = 1$	-3398,91	66	6932,00		0,00 (59%)	0,00 (3%)	1,00 (38%
INSR: clado Caviomo	rpha						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-26317,93	65	52766,15	0,0940	0,03 (89%)	0,06 (10%)	6,28 (1%)
(ramas del background)					0,02 (95%)	0,46 (5%)	9,14 (0%)
modelo con $\omega_3 = 1$	-26320,3	64	52768,87		0,00 (91%)	0,00 (3%)	1,00 (6%)
IGF1R: ancestro Platy	rrhini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-23579,79	69	47297,87	«0,0001	0,21 (83%)	0,51 (16%)	33,80 (1%
(ramas del background)					0,02 (97%)	0,81 (2%)	49,50 (1%
modelo con $\omega_3 = 1$	-23588,82	68	47313,93		0,00 (67%)	0,00 (2.0%)	1,00 (31%
INSRR: clado Platyrrl	nini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-24068,68	67	48271,68	0,0150	0,05 (92%)	1,00 (8%)	229 (0%)
(ramas del background)					0,04 (93%)	0,59 (7%)	11,00 (0%
modelo con $\omega_3 = 1$	-24072,89	66	48278,08		0,00 (85%)	0,93 (0.0%)	1,00 (15%
IGF2R: ancestro Platy	rrhini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-63164,55	69	126467,27	0,0240	0,13 (87%)	0,49 (12%)	18,40 (1%
(ramas del background)					0,04 (86%)	1,00 (14%)	805 (0%)
modelo con $\omega_3 = 1$	-63168,29	68	126472,74		0,00 (57%)	0,00 (19%)	1,00 (24%
IGF2R: clado Platyrrh	nini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-63165,11	69	126468,39	«0,0001	0,16 (95%)	0,17 (5%)	41,00 (0%
(ramas del background)	,		·		0,04 (86%)	1,00 (14%)	1260 (0%
modelo con $\omega_3 = 1$	-63176,57	68	126489,31		0,04 (0%)	0,00 (81%)	1,00 (19%

NOTA: Se muestran solo los resultados para las configuraciones de ramas blanco que resultaron bajo selección positiva para  $\alpha \leq 0, 1$ .

IGFDF1: allestro Flat	lyrrinn						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-63164,55	69	126467,27	0,0240	0,13 (87%)	0,49 (12%)	18,40 (1%
modelo sin restricciones	-7455,41	65	15042,43	0,0010	1,00 (63%)	1,00 (6%)	7,16 (31%
(ramas del <i>background</i> )					0,05 (72%)	0,65 (27%)	11,70 (1%
modelo con $\omega_3 = 1$	-7462,76	64	15055,07		1,00 (5%)	0,95 (0%)	1,00 (95%
IGFBP1: clado Platyri	rhini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-63164,55	69	126467,27	0,0240	0,13 (87%)	0,49 (12%)	18,40 (1%
modelo sin restricciones	-7459,83	65	15051,27	0,0150	0,97 (78%)	0,99 (18%)	16,00 (4%
(ramas del background)					0,06 (73%)	0,66 (26%)	11,20 (1%
modelo con $\omega_3 = 1$	-7464,00	64	15057,56		1,00 (6%)	0,94 (0%)	1,00 (94%
IGFBP2: ancestro Hys	tricognathi						
Modelo	lnL	NP	AICc	pvalue	$\boldsymbol{\omega}_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-63164,55	69	126467,27	0,0240	0,13 (87%)	0,49 (12%)	18,40 (1%
modelo sin restricciones	-4785,68	67	9706,98	0,0380	0,19 (82%)	0,19 (17%)	2300 (1%)
(ramas del background)					0,09 (96%)	0,51 (3%)	17,10 (1%
modelo con $\omega_3 = 1$	-4788,95	66	9711,48		0,23 (100%)	0,24 (0%)	1,00 (0%)
IGFBP6: clado Cavion	norpha						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-63164,55	69	126467,27	0,0240	0,13 (87%)	0,49 (12%)	18,40 (1%
modelo sin restricciones	-5962,06	69	12063,85	0,0070	0,14 (95%)	1,00 (5%)	233 (0%)
(ramas del background)					0,27 (96%)	0,27 (3%)	6,82 (1%)
modelo con $\omega_3 = 1$	-5966,96	68	12071,61		0,00 (80%)	0,93 (0%)	1,00 (20%

 Tabla A-46:
 Resultados de BUSTED para los genes IGFBP1, IGFBP2 e IGFBP6.

NOTA: Se muestran solo los resultados para las configuraciones de ramas blanco que resultaron bajo selección positiva para  $\alpha \leq 0, 1$ .

GHRH: ancestro Platy	rrhini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-2475,64	65	5085,08	0,0110	1,00 (88%)	1,00 (10%)	196 (2%)
(ramas del background)					0,228 (70%)	0,23 (20%)	2,16 (10%)
modelo con $\omega_3 = 1$	-2480,14	64	5091,94		1,00 (7%)	0,92 (0.0%)	1,00 (93%)
GHRH: clado Platyrrh	nini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-2478,34	65	5090,47	0,0910	0,52 (92%)	0,54 (8%)	93,90 (0%)
(ramas del background)					0,23 (66%)	0,23 (24%)	2,22 (10%)
modelo con $\omega_3 = 1$	-2480,73	64	5093,14		0,00 (1%)	0,00 (50%)	1,00 (49%)
GHRHR: ancestro Pla	tyrrhini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-9989,22	67	20113,40	0,0520	0,25 (94%)	0,27 (5%)	10000 (1%
(ramas del background)					0,09 (86%)	0.931 (14%)	64,40 (0%)
modelo con $\omega_3 = 1$	-9992,18	66	20117,28		1,00 (13%)	0.00 (74%)	1,00 (13%)
GHRHR: clado Platyr	rhini						
Modelo	lnL	NP	AICc	pvalue	$\omega_1(p_1)$	$\omega_2(p_2)$	$\omega_3(p_3)$
modelo sin restricciones	-9986,87	67	20108,70	«0,0001	0,00 (77%)	1,00 (22%)	384 (1%)
(ramas del background)	,			*	0,06 (78%)	0,49 (16%)	1,45 (6%)
modelo con $\omega_3 = 1$	-10001,58	66	20136,09		0,00 (76%)	0,94 (0%)	1,00 (24%)

Tabla A-47: Resultados de BUSTED para los genes GHRH y GHRHR.

NOTA: Se muestran solo los resultados para las configuraciones de ramas blanco que resultaron bajo selección positiva para  $\alpha \leq 0, 1$ .

Tabla A-48:	Prueba de	e relajación	de la s	selección	en Plat	yrrhini	comparado	con los	restantes	linajes	de
	Euarchon	ta como ref	erenci	a (RELA)	X).						

				· · · · ·			
Gen	Coeficiente k	pvalue	LR	$\omega$ Platyrrhini	ωREF	AICc nulo	AICc alternativo
INS	0,61	0,0686	3,32	0,21; 0,59; 4,18	0,07; 0,42; 10,6	5282,52	5281,37
IGF1	0,71	0,1845	1,76	0,15; 0,15; 1,00	0,07; 0,07; 1,00	4823,79	4824,15
IGF2	1,35	0,4438	0,59	0,09; 0,16; 8,98	0,17; 0,26; 5,09	5892,82	5894,33
INSR	0,21	«0,0001	66,10	0,02; 0,69; 1,47	$\approx$ 0; 0,17; 6,20	47090,14	47026,05
IGF1R	0,35	«0,0001	65,88	0,08; 0,75; 26,1	$\approx$ 0; 0,45; 10.000	42614,08	42550,21
INSRR	0,84	0,0491	3,87	0,12; 0,37; 153	0,08; 0,31; 389	42824,10	42822,24
IGF2R	0,59	0,0118	6,34	0,01; 0,94; 58,90	$\approx$ 0; 0,90; 1.000	107861,83	107857,50
IGFBP1	0	«0,0001	31,45	1,00; 1,00; 1,00	0,02; 0,79; 10,20	12758,38	12728,99
IGFBP2	0,23	«0,0001	28,40	0,13; 0,30; 1,19	$\approx 0; 0,01; 2,12$	8640,52	8614,19
IGFBP3	0,63	0,0793	3,08	0,17; 0,20; 2,13	0,06; 0,08; 3,35	7318,60	7317,61
IGFBP4	0,18	«0,0001	24,90	0,05; 0,66; 1,29	$\approx$ 0; 0,11; 3,98	7517,06	7494,23
IGFBP5	0,43	«0,0001	22,64	0,18; 0,20; 20,1	0,02; 0,03; 1.000	7086,75	7066,18
IGFBP6	0,91	0,8072	0,06	0,15; 0,91; 1,08	0,13; 0,90; 1,09	10794,46	10796,48
IGFBP7	0,85	0,7040	0,14	0,07; 0,08; 4,69	0,05; 0,05; 6,20	5330,55	5332,49
GH1	NA	NA	NA	NA	NA	NA	NA
GHR	0,83	0,5343	0,39	0,27; 0,66; 3,93	0,21; 0,60; 5,23	25565,82	25567,46
GHRH	0,66	0,5002	0,45	0,49; 0,49; 211	0,34; 0,34; 3.330	4292,78	4294,49
GHRHR	21,57	0,0054	7,75	$\approx$ 0; 1,00; 390	0,09; 1,00; 1,32	17557,21	17551,5

NOTA: NA: no aplica; LR: cociente de verosimilitud entre los modelos nulo y alternativo;  $\omega$  REF: valores de  $\omega$  en los euarcontos no platirrinos usado como referencia. Los valores de  $\omega$  en Platyrrhini y REF se indican como  $\omega_1$ ;  $\omega_2$ ;  $\omega_3$ .

## B APÉNDICE: figuras suplementarias




**Figura B-2:** *Timetree*, cladograma calibrado según el tiempo geológico de los procesos de diversificación, para los *taxa* estudiados. La filogenia se obtuvo a partir de la base de datos TimeTree (http://www.timetree.org; Hedges y Kumar, 2009; Hedges *et al.*, 2015).





Figura B-3: Análisis de correspondencia de la composición de codones de los genes y taxa estudiados. Los triángulos azules indican los codones mientras que los círculos rojos representan las especies, algunas de las cuales están etiquetadas. ANA: A. nancymaae; BOV: B. taurus; CAR: Carnivora; CJA: C. jacchus; CLA: C. lanigera; CPO: C. porcellus; CSO: C. sociabilis; FDA: F. damarensis; GGO: G. gorilla; HSA: H. sapiens; MMU: M. mulatta; MUS: M. musculus; NGA: N. galili; OCU: O. cuniculus; OGA: O. garnettii; OPR: O. princeps; PER: Perissodactyla; TBE: T. chinensis; TSY: T. syrichta.



Figura B-4: Análisis de correspondencia del uso de aminoácidos de los genes y taxa estudiados. Los triángulos azules indican los aminoácidos mientras que los círculos rojos representan las especies. ANA: A. nancymaae; BOV: B. taurus; CAR: Carnivora; CJA: C. jacchus; CLA: C. lanigera; CPO: C. porcellus; CSO: C. sociabilis; DOR: D. ordii; FDA: F. damarensis; GGO: G. gorilla; HGL: H. glaber; HSA: H. sapiens; MMU: M. mulatta; MUS: M. musculus; NGA: N. galili; OCU: O. cuniculus; ODE: O. degus; OGA: O. garnettii; OPR: O. princeps; PER: Perissodactyla; SBO: S. boliviensis; TBE: T. chinensis; TSY: T. syrichta.



Figura B-5: Análisis de correspondencia de todas las secuencias de codones. Los símbolos y colores identifican las secuencias pertenecientes al mismo *taxa*. Para mejor claridad se indican *taxa* de orden superior. EUARCHONTA-NP: euarcontos no Platyrrhini; GLIRES-NH: glires no Hystricognathi; OUTGROUP: Laurasiatheria.

Jalview 2.8.0b1	5F1R/IGF1R_codones_291_295.fasta	ulate Web Service 291 293 - 295	850 860 870 880 890 290	caacatoctcagecedagecedectecedegecertete saes <mark>sd</mark> segev	CCAACATCCTCAGCGCCGAAAGCAGCGACTCCGAGGGGGTTTGTG SAES <mark>SD</mark> SEGFV	CCAACATCCTCAGTGCCGAGGCAGCGAGGGGTTTCGTG SAESSDSEGFV	CAGCATCGTCAGCTCCGAGAACTCGGAGAACAACAACAAGTTTGTG SSENSENNKFV	CCAGCATCGTCAGCTCTGAGAACTCGGAGAACAACAACAAGTTCGTG SSENSENNKFV	cagcatcetcagctetgagaactegagagagagagagattege		CCAACATCCCCAGCGAGGAGCAGCGAGGGGGGTTCGTG SAESSDSEGFV	CCAACATCACCAGCGCCGAGGCGGCGACTCTGAAGGCTTCGTG SAESSDSEGFV	:CAACATCCTCAGCGTCGACGAAGGCGAGGGCTTCGTC SVDEGEGFV	CCAACATCCCTAACGCCGATGGCGGCGGACTCCCGAGGGCTTTGTC NADGGDSEGFV	CCAACATCCTCAGGGGGGGGGGGGGGGGGGGGGGGGGGG	CCAACATCCCCAACGCTGAGGAGGAGGAGGGATGGCTTCGTT NAESSDSDGFV	CCAACATCCAGAGTGCTGAGGAGCACTGACTCCTATGGCTTCGTT SAESTDSYGFV	CCAACATCCCCAGGGGGGGGGGGGGGGGGGGGGGGGGGG	CCAACATCCCCCAACGCTGAGGCAGCGAGGGCTTCGTC NAESSDSEGFV				CCAACATCCCCCAATGCCGAGGGGGGGGGCTCTGGGGGGGTTTGTC NAESSDSEGFV					CCAACATCCCCAGCGCCGAGGGGGCGGCTTCGGGGGGGGG	
Jalvi	e Tools <u>V</u> amsas Help Window home/natalia/Documents/INSULINA/5_IGF1R/IGF1R_codones_291_295.fasta	File Edit Select View Format Colour Calculate Web Service	840 850 860 87	<i>ното</i> твасттствсесса́асатсстсавсесвавае	Gorilla TGACTTCTGCGCCAACATCCTCAGCGCCGAAAGC	Macaca TGACTTCTGCGCCAACATCCTCAGTGCCGAGAGC	Saimiri TCAGTTCTGTGCCAGCATCGTCAGCTCCGAGAAC	Aotus TCAATTCTGTGCCAGCATCGTCAGCTCTGAGAAC	Callthrix TCAGTTCTGTGCCAGCATCGTCAGCTCTGAGAAC	Tarsius XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	Microcebus CGACTTCTGCGCCAACATCCCCAGCGCCGAGAGC	Tupaia CGACTTCTGCGCCAACATCACCAGCGCCGAGAGC	Ochotone TGACTTCTGCGCCAACATCCTCAGCGTCGACGAA	Oryctolagus CGACTTCTGCGCCAACATCCCTAACGCCGATGGC	Ictidomys CGACTTCTGTGCCAACATCCTCAGCGCTGAGAGC	Mus CGATTTCTGCGCCAACATCCCCCAACGCTGAGAGC	Nannospalax AGATTTCTGTGCCAACATCCAGAGTGCTGAGAGC	Dipodomys CGACGCCTGCGCCCAACATCCCCCAGCGCCGAGAGC	Fukomys GGACTTCTGCGCCAACATCCCCCAACGCTGAGGGC	Hererocephakus GGACTTCTGCGCCCAACATCCCCCAATGCTGAGGGC	Uctodon IGACI I CIGCGCCAACAI CCCAAAI GCCGAGAGC	Ctenomys CGACTTCTGCGCGCCAACATCCCCCAACGCCGAGAGC	Chinchilla CGACTTCTGCGCCCAACATCCCCCAATGCCGAGAGC	Cava AGACTTCIGCGCCAACATCCCCCAATGCTGAGGC	000040-00040000-4044000-0-0-0-0400-80064	194949191919191919191919191919191919191	Consensus	CGACTTCTGCGCCAACATCCCCAGCGCCGAGAGC	

presentan codones sinónimos para serina, pero la sustitución AGC - TCG involucra al menos tres pasos, siendo al menos uno de ellos un cambio no sinónimo. esas posiciones. Si bien el modelo de ramas y sitios implementado en CODEML detectó selección positiva en cuatro codones muy cercanos, observando el alineamiento se descartan errores en alguna de las secuencias o en la hipóteiss de homología. En el caso del codon 291, catarrinos y platirrinos

Figura B-7: Topología de referencia con largos de rama obtenidos a partir de 12 genes mitocondriales de la hebra pesada. Un alineamiento construido a partir de los 12 genes codificantes de proteínas de la hebra mitocondrial pesada, con exclusión de la tercera posición de los codones, se utilizó para calcular en nucleótido. Los números indican nodos que se utilizan para describir ramas internas en la Tabla 3.8. PhyML el largo de cada rama de la topología. Estos largos indican la tasa de cambio por rama, encontrándose la escala en unidades de sustituciones por



		. 1	ĻO			2	0				3	0				
Homo	HSQGI	FTSD	ÝSKY	LD	SRI	RAQ	DF	VQ	ŴL	ΜN	ТК	RN	IRN	1N	A	ΚR
Rattus	HSQGI	FTSD	YSKY	LD	SRI	RAQ	DF	VQ	WL	ΜN	ТΚ	RN	IRN	IN	I A I	ΚR
Cricetulus	HSQGI	FTSD	YSKY	LD	SRI	RAQ	DF	VQ	WL	ΜN	ТΚ	RN	IRN	IN	A	ΚR
Fukomys	HSQGI	FTSD	YSKY	LD	SRI	RAQ	DF	VQ	WL	ΜN	ТΚ	RN	IRN	IN	A	ΚR
Heterocephalus	HSQGI	FTSD	YSKY	LD	SRI	RAQ	DF	VQ	WL	ΜN	ТΚ	RN	IRN	IN	A	ΚR
Octodon	HSQGT	FTSD	Y S K F	LD	TRI	RAQ	DF	LD	WL	ΚN	ТΚ	RN	IRN	ΙE	A	ΚR
Chinchilla	HSQGI	FTSD	YSKH	LD	SR	Υ <mark>Α</mark> Ω	EF	VQ	WL	Μ <mark>Ν</mark>	ТΚ	RN	IRN	IN	A	ΚR
Cavia	HSQGI	FTSD	YSKY	LD	SR	RAQ	QF	LΚ	WL	LΝ	VΚ	RN	IRN	IN	A	ΚR
Bos	HSQGT	FTSD	YSKY	LD	SRI	RAC	DF	VQ	WL	MN	ТК	RN	KN	IN	A	ΚR

**Figura B-8:** Alineamiento de algunas secuencias aminoacídicas del péptido glucagón. Los sitios están coloreados según el porcentaje de identidad. Se puede observar que todas las sustituciones, excepto una, están concentradas en Caviomorpha. Además, los diferentes caviomorfos presentan diferentes cambios aminoacídicos. Adaptado de Nishi y Steiner (1990).



como "brush-tailed rat" corresponde a O. degus.

región homóloga al inicio del exón 11 en los restantes mamíferos incluidos en el alineamiento. La secuencia identificada en UCSC Genome Browser

	scaffold 42			
			*	
	14.470 kb	14.480 kb	14.490 kb	Þ
LUNG_Tophat: bam Coverage	trot- ol			•
LUNG_Tophat: bam				Þ
MUSCLE_Tophat.bam Coverag				•
MUSCLE_Tophat.bam				
LUNG_Rsubread.bam Coverag	501 · 0			•
LUNG_Rsubread barn				•
MUSCLE_Rsubread.barn Cover ge	4 (o - 5/0)			•
MUSCLE_Rsubread barn				Þ
LUNG_Cufflinks.gtf		CUFF.580.1		4
MUSCLE_Cufflinks.gtf		CUFF.523.1		Þ

Figura B-10: Visualización en IGV (Robinson et al., 2011) del scaffold\_42 de C. porcellus (cavPor3) para la región que contiene al INSR (se observan 11 de los 21 exones del gen). Se muestra el gráfico de cobertura y parte de los alineamientos obtenidos con Tophat (Kim et al., 2013) y Rsubread (Liao et al., 2013), de reads provenientes del secuenciamiento de dos muestras transcriptómicas, una de pulmón y otra de músculo esquelético. En la parte inferior se muestran los modelos de los transcriptos ensamblados por Cufflinks (Trapnell et al., 2010) para ambas muestras. Las líneas verticales rojas están centradas alrededor de la región que es la homóloga a la región del cromosoma 19 de H. sapiens que contiene al exón 11 de 36 nucleótidos (scaffold\_42 14.479.982-14.480.021 en cavPor3). A pesar del alto nivel de transcripción e independientemente de la herramienta utilizada, no hay reads en ninguna de la muestras que mapeen a la región homóloga al exón 11 del INSR humano. En concordancia, el exón no aparece incluído en los modelos de los transcriptos ensamblados por Cufflinks.