



Universidad Austral de Chile

Facultad de Ciencias

Escuela de Ciencias

PROFESOR PATROCINANTE

Dr. Juan C. Opazo

Instituto de Ecología y Evolución

Facultad de Ciencias

EVOLUCIÓN DE LA FAMILIA MULTIGÉNICA DE LAS RELAXINAS EN

MAMÍFEROS

Tesis de grado presentada como parte de los requisitos para optar al grado de *Licenciado en Ciencias Biológicas*

JOSÉ IGNACIO ARROYO GONZÁLEZ

Valdivia Chile

2010

AGRADECIMIENTOS

Esta tesis se realizó en el Laboratorio de Genómica Evolutiva del Instituto de Ecología y Evolución, y fue subsidiada por el proyecto *Evolution of the insulin gene family in mammals: a genomic approach* (FONDECYT N° 11080181). Agradezco la tutoría del profesor director de esta tesis; Dr. Juan C. Opazo, y de los profesores examinadores; Dra. Leyla Cárdenas y Dr. Christian Figueroa.

ÍNDICE DE CONTENIDOS

RESUMEN	1
ABSTRACT	2
1 INTRODUCCIÓN	3
1.1 Evolución por duplicación génica	3
1.2 Modelos de evolución de familias multigénicas	5
1.3 Familia multigénica de las relaxinas en mamíferos como modelo de estudio	8
1.3.2 Estructura y función	9
1.3.3 Evolución	9
1.4 Hipótesis	11
1.5 Objetivo general	11
1.5.2 Objetivos específicos	11
2 MÉTODOS	12
2.1 Secuencias de ADN	12

2.2 Análisis filogenéticos	12
2.3 Selección natural a nivel molecular	13
2.3.1 Prueba de rama	13
2.3.2 Prueba de rama-sitio	13

3 RESULTADOS

3.1 Evolución de la familia multigénica de las relaxinas en mamíferos	15
3.1.1 Estructura genómica de la familia multigénica de las relaxinas en mamíferos	15
3.1.2 Análisis filogenético de la familia multigénica de las relaxinas en mamíferos	15
3.2 Evolución del cluster génico RFLB mamíferos eucariontes	16
3.2.1 Evolución del cluster génico RFLB en el conejo	16
3.2.1.2 Prueba de rama	17
3.2.1.3 Prueba de rama-sitio	18
3.2.2 Evolución del cluster génico RFLB en primates	19
3.2.2.2 Análisis filogenético	20

4 DISCUSIÓN

4.1 Evolución de la familia multigénica de las relaxinas en mamíferos	22
4.1.2 Modelo de evolución de la familia multigénica de las relaxinas en mamíferos	22
4.1.3 Nomenclatura de los genes de la familia multigénica de las relaxinas	23

	iii
4.2 Evolución del cluster RFLB mamíferos eucariontes	24
4.2.1 Evolución del cluster RFLB en el conejo	24
4.2.2 Evolución del cluster RFLB en primates	26
4.2.2.1 Retención diferencial del gen INSL4 en tarsos y primates catarrinos	26
4.2.2.2 Retención diferencial de los genes RLN1 y 2 en monos del nuevo y viejo mundo	28
4.2.2.3 Evolución mosaico de los genes RLN del chimpancé	29
CONCLUSIÓN	30
ANEXO	54
BIBLIOGRAFÍA	58

ÍNDICE DE TABLAS

Tabla 1	Valores de los parámetros estimados en los modelos de rama.	32
Tabla 2	Pruebas de razón de verosimilitud de los modelos de rama.	32
Tabla 3	Valores de los parámetros estimados en los modelos de rama sitio A y A neutral.	32
Tabla 4	Prueba de razón de verosimilitud de los modelos de rama-sitio.	32
Tabla A1	Números de acceso de los genes de la familia multigénica de las relaxinas usadas en el estudio	32

ÍNDICE DE FIGURAS

Figura 1	Duplicación génica por entrecruzamiento desigual.	33
Figura 2	Relaciones de homología entre genes.	33
Figura 3	Modelo de evolución divergente.	34
Figura 4	Modelo de evolución concertada.	34
Figura 5	Modelo de evolución mosaico.	35
Figura 6	Modelo de evolución nacimiento y muerte.	35
Figura 7	Modelo de evolución perdida y retención diferencial.	36
Figura 8	Familia multigénica de las relaxinas en humanos.	36
Figura 9	Modelo de evolución de la familia multigénica de las relaxinas.	37
Figura 10	Modelo de evolución del cluster génico RFLB.	38
Figura 11	Estructura genómica de la familia multigénica de las relaxinas mamíferos.	39

Figura 12	Filograma de máxima verosimilitud de la familia multigénica de las relaxinas en mamíferos.	40
Figura 13	Topología del árbol de la familia multigenica de las relaxinas según el modelo aceptado actualmente.	41
Figura 14	Estructura genómica del cluster génico RFLB en el conejo.	41
Figura 15	Representación grafica de los modelos anidados implementados en la prueba de rama.	42
Figura 16	Representación grafica del modelo implementado en la prueba de ramasitio.	43
Figura 17	Proporción de sitios en cada clase y omega para cada clase del gen RLN2 del conejo.	43
Figura 18	Sitios bajo selección en los péptidos B y A de la proteína codificada por el gen RLN2 del conejo.	44
Figura 19	Estructura genómica del cluster RFLB en primates.	45
Figura 20	Alineamiento de secuencias de los genes RLN1 y RLN2 de chimapancé.	46

Figura 21	Similitud de secuencia de los exones de los genes RLN1 y RLN2 de chimpancé.	46
Figura 22	Alineamiento de secuencias de los genes RLN1 y RLN2 de humano y chimpancé.	47
Figura 23	Topología del árbol de los genes INSL4, RLN2 y RLN1 en primates de acuerdo con el modelo aceptado actualmente.	48
Figura 24	Filograma de máxima verosimilitud de los genes INSL4, RLN2 y RLN1 en mamíferos.	49
Figura 25	Filograma de máxima verosimilitud de los genes INSL4 en primates haplorrinos.	50
Figura 26	Nomenclatura de los genes de la familia multigénica de las relaxinas.	50
Figura 27	Motivo de interacción ligando-receptor del gen RLN2.	51
Figura 28	Neofuncionalización del gen RLN2 en el conejo.	51
Figura 29	Evolución del cluster RFLB en primates.	52
Figura 30	Evolución mosaico de los genes RLN1 y RLN2 en el chimpancé.	53

ABREVIATURAS

BLAST	Basic local alignment search tool
pb	Par de base
IGF	Factor de crecimiento insulínico
INS	Insulina
INSL	Péptido de Leydig
Kb	Kilo base
LRT	Likelihood Ratio-Test
MAFFT	Multiple Alignment with Fast Fourier Transforms
NCBI	National Center for Biotechnology Information
NF	Neofuncionalización
PAML	Phylogenetic Analysis by Maximum Likelihood
RFL	Relaxin family locus
RLN	Relaxina
SF	Subfuncionalización

RESUMEN

La familia multigénica de las relaxinas está compuesta por las relaxinas (RLNs) y los péptidos de Leydig (INSLs). Los genes de esta familia cumplen roles fisiológicos asociados a la reproducción. La estructura de esta familia sólo ha sido descrita en especies modelos, y el estudio de su evolución es bastante limitado. En este sentido, las bases de datos genómicas abren hoy una oportunidad única para el estudio evolutivo de familias multigénicas. Lo anterior es más evidente en mamíferos, en donde la gran cantidad de genomas disponibles permiten hacer inferencias más precisas debido a la densidad de muestreo en su filogenia.

En este contexto el objetivo de esta tesis es estudiar el patrón evolutivo de la familia multigénica de las relaxinas en mamíferos. Para lograr este objetivo, se caracterizaron los miembros de esta familia en las especies de mamíferos para las cuales hay bases de datos genómicas, y se reconstruyeron las relaciones de ancestralidad de los mismos. Por otro lado, se investigó también el potencial rol de la selección natural en la evolución de este grupo de genes.

Los resultados de esta tesis sugieren que la evolución de la familia multigénica de las relaxinas podría ser explicada principalmente por un modelo divergente. Sin embargo, otros procesos tales como nacimiento y muerte en el clúster RFLB en mamíferos eucariontes, retención diferencial de los genes RLN del clúster RFLB en monos del nuevo y viejo mundo, retención diferencial del gen INSL4 en primates haplorrinos y evolución mosaico en el gen RLN2/1 en el chimpancé, también darían cuenta de la evolución de este grupo de genes.

En conclusión, esta tesis muestra la interacción de distintos procesos evolutivos que dan cuenta de la evolución de sistemas genéticos, muchos de los cuales pueden variar tanto en el tiempo como en el espacio.

ABSTRACT

The relaxin gene family comprises two different kinds of genes: relaxines (RLNs) and Leydig peptides (INSLs). In this gene family, the different paralogs have acquired a variety of physiological roles including bone formation, testicular descent, trophoblast development and cell differentiation. The structure of this family only has been described in model species, and the study of its evolution is rather limited. In this sense, genomic databases provide an unique opportunity to study the evolution of multigene families, specially in mammals, where the amount of genomes available is larger in comparison with other vertebrate groups, allowing more precise conclusions.

In this context, the objective of this thesis is to study the evolutionary pattern of the relaxin gene family in mammals. To achieve this goal, members of this gene family were characterized in mammalian species for which genomic information is available. Gene phylogenies were used to infer orthologous relationships, and the potential role of natural selection in this group of genes was also investigated.

Results of this study suggest that the evolution of the relaxin gene family could be explained mainly by a divergent model. However, processes such as birth-and-death in euarchontoglires, differential retention of the RLN genes in new and old world monkeys, differential retention of INSL4 gene in haplorrhine primates, and mosaic evolution of RLN2/1 gene in chimpanzee, would also account for the evolution of this group of genes.

In conclusion, this thesis shows the interaction of different evolutionary processes that account for the evolution of genetic systems, many of which can vary both in time and space.

INTRODUCCIÓN

1.1 Evolución por duplicación génica

La importancia de las duplicaciones génicas como fuente promotora de innovaciones biológicas fue inicialmente sugerida a principios de la década de los 30 por Haldane y Müller (Haldane 1932; Müller 1935), sin embargo no fue sino hasta inicios de la década de los 70, con la publicación del libro *Evolution by gene duplication* por Susumo Ohno (Ohno 1970), en que la real dimensión de este fenómeno fue entendida. En la década de los 90 el desarrollo de las tecnologías de secuenciación y la bioinformática han hecho posible analizar los genomas en su totalidad, lo cual ha proveído evidencia irrefutable acerca de la abundancia de duplicados en los genomas (Zhang 2003).

Las duplicaciones génicas son eventos en los cuales un segmento de ADN se origina a partir de otro preexistente. Estas duplicaciones de pequeña escala pueden originarse por mecanismos como el entrecruzamiento desigual, el cual consiste en el intercambio de información entre cromosomas homólogos que no están correctamente alineados durante el proceso de meiosis. Como resultado de lo anterior se originan segmentos repetidos en tándem en un cromosoma homólogo y una deleción en el otro (Nei & Rooney 2005) (figura 1) .

Es posible establecer distintas relaciones de homología entre los genes que se originan a partir de una duplicación génica, entre los cuales la paralogía y ortología son los más importantes. La primera se define como una relación de homología entre genes de una misma especie que se han originado por un evento de duplicación génica. Por otro lado la ortología se define como una relación de homología entre genes de diferentes especies, originadas por un

evento de especiación (Gogarten & Olendzenski 1999; Thornton & DeSalle 2000) (figura 2).

Los genes que se originan a partir de una duplicación génica pueden seguir distintas trayectorias evolutivas. Una primera alternativa sugiere que los genes podrían conservar la función del gen parental, lo cual permitiría al organismo producir una mayor cantidad del producto génico, lo cual a su vez y bajo ciertas circunstancias podría representar una ventaja. La mantención de las copias con la misma secuencia puede ser explicado por la acción de la selección purificadora y/o la conversión génica (Zhang 2003).

Una segunda posibilidad es el modelo de subfuncionalización (SF), el cual sugiere que la función que era ejercida por el gen parental es asumida ahora por ambos duplicados, lo cual puede evidenciarse como cambios en los patrones espaciales o temporales de expresión (Zhang 2003). Un ejemplo de lo anterior lo constituye la familia multigénica de las β -globinas en mamíferos en donde los distintos parálogos se expresan diferencialmente durante la ontogenia (Hardison 1988).

Una tercera posibilidad es la neofuncionalización (NF), que sugiere que uno de los duplicados adquiriría un rol fisiológico distinto al del gen parental (Zhang 2003). Uno de los mecanismos que explicarían este modelo es la aceleración de la fijación de mutaciones ventajosas, ya sea en la región codificante o reguladora del gen (Rodríguez-Trelles, Tarrío, & Ayala 2003).

Estos dos últimos procesos subfuncionalización (SF) y neofuncionalización (NF) han sido integrados en un nuevo modelo que sugiere que los genes duplicados podrían experimentar una rápida SF y posteriormente una prolongada NF, modelo que ha sido denominado

subneofuncionalización (He & Zhang 2004).

Finalmente vale la pena destacar que mas allá de los modelos antes mencionados, la mayor parte de los genes duplicados pierde su función debido a la ocurrencia de mutaciones deletéreas que inactivan el gen, tales como las mutaciones que alteran el marco de lectura, las mutaciones sin sentido o las deleciones, generando así pseudogenes (Zhang 2003).

1.2 Modelos de evolución de familias multigénicas

Los genes originados por duplicación génica se agrupan en unidades conocidas como familias multigénicas, las cuales se define como un grupo de genes que descienden de un gen ancestral común, y por lo tanto tienen secuencias de ADN y funciones similares (Nei & Rooney 2005). Los miembros de una familia multigénica pueden estar dispuestos en tándem o también dispersos en distintos cromosomas. Diversos modelos han sido propuestos para explicar la evolución de las familias multigénicas, entre ellos los más citados son los modelos de evolución divergente, concertada, mosaico, nacimiento y muerte, y retención diferencial (Nei y Rooney 2005; Yonekura et al. 1987; Zhang, Dyer & Rosenberg 2000).

El modelo de evolución divergente sugiere que los genes evolucionarían de forma independiente, en relación a la fijación de mutaciones (figura 3a), lo cual no implica necesariamente que los genes adquirirían nuevas funciones (Ota & Nei 1994). Como consecuencia de este proceso los genes ortólogos tendrían una mayor similitud de secuencias entre sí que los genes parálogos, de tal manera que al reconstruir una filogenia la topología recuperaría grupos monofiléticos de genes ortólogos (Carson & Scherer 2009) (figura 3b). Un

ejemplo de evolución divergente lo constituye la familia génica V_H en vertebrados (immunoglobulin heavy-chain variable-region) (Ota & Nei 1994).

El modelo de evolución concertada sugiere que los genes no evolucionarían de manera independiente, sino que interactuarían a través de procesos de conversión génica (Nei & Rooney 2005) (figura 4a). Este mecanismo consiste en un proceso de recombinación homóloga no recíproca entre dos regiones del genoma (Li 1997). La consecuencia de este proceso lleva a que los genes parálogos tendrán una mayor similitud de secuencias entre sí que los genes ortólogos, de tal manera que la topología obtenida al reconstruir una filogenia estaría formada por clados de genes parálogos especie específicos (Carson & Scherer 2009; Nei & Rooney 2005) (figura 4b). Un ejemplo de evolución concertada lo constituye los genes de expresión adulta de la familia multigénica de las β -globinas en mamíferos (Opazo et al. 2008).

El modelo de evolución mosaico describe un escenario en el cual la evolución de los genes ocurriría por una combinación de ambos modelos descritos previamente, es decir, que una región (exon, dominio) de los genes evolucionaría de manera divergente, mientras que la otra lo hace de manera concertada generando así genes quiméricos (figura 5a). Los genes quiméricos son genes que combinan distintos módulos funcionales de distintos genes parentales (Jones & Begun 2005; Long et al. 2003a; Long et al. 2003b). Las reconstrucciones filogenéticas de genes que evolucionan según este patrón producen topologías en las cuales pareciera que han ocurrido duplicaciones múltiples (Teshima & Innan 2004). Por lo anterior para detectar este patrón se deben reconstruir filogenias para distintas regiones del gen, de modo que se obtendrían topologías alternativas para las regiones que evolucionan de manera divergente y concertada, en las cuales los genes se recuperarían en clados de genes ortólogos y parálogos respectivamente

(Carson & Scherer 2009)(figura 5b). Un ejemplo de evolución mosaico lo constituye el gen quimérico β/δ -globina en el elefante africano, el cual se originó por entrecruzamiento desigual entre los genes β y δ -globina en esa especie (Opazo et al. 2009).

El modelo de evolución por nacimiento y muerte describe un escenario en el cual los miembros de una familia multigénica se duplican, se inactivan o se pierden aleatoriamente en diferentes grupos, por lo que comparativamente el número de copia de genes variaría en los diferentes linajes (Hughes & Nei 1989; Nei et al. 1997; Piontkivka et al. 2002) (figura 6). Un ejemplo de evolución por nacimiento y muerte es la familia génica MHC (Major Histocompatibility Complex) en mamíferos (Hughes & Nei 1989; Nei, Gu & Sitnikova 1997).

El modelo de retención diferencial sugiere que existe una pérdida y retención diferencial de genes en diferentes clados (figura 7). Este proceso puede ser estocástico, de modo que un grupo cualquiera de genes puede ser aleatoriamente retenido en un clado. Por otra parte, la pérdida y retención diferencial de un grupo específico de genes podría ocurrir por selección natural, de modo que la retención en el genoma de un grupo particular de genes podría ser ventajoso (Force et al. 1999; Hughes 1994). Un ejemplo de lo anterior es la pérdida y retención diferencial de genes del clúster de las β -globinas en la radiación de mamíferos euterios (Opazo et al. 2008).

Si bien los modelos de evolución de familias multigénicas describen escenarios que individualizan el patrón por los cuales un grupo de genes que descienden de un ancestro común evolucionan, la realidad muestra que estos modelos no son excluyentes, de tal manera que la evolución de una familia multigénica puede ser explicada por procesos mixtos, que pueden

cambiar tanto en el tiempo como en espacio. Por ejemplo, la familia multigénica de las β -globinas en mamíferos exhibe patrones evolutivos que se ajustan a los modelos de evolución concertada, en mosaico, nacimiento y muerte y retención diferencial (Opazo et.al. 2008).

1.3 Familia multigénica de las relaxinas en mamíferos como modelo de estudio

La familia multigénica de las relaxinas forma parte de la súperfamilia de la insulina, la cual a su vez ha sido dividida en dos familias, una de ellas formada por la insulina (INS) y los factores de crecimiento insulínicos (IGFs), y la otra, por las relaxinas (RLNs) y los péptidos de Leydig (INSLs) (Steiner et al. 1985; Chan & Steiner 2000). Estos genes han sido agrupados dentro de una súperfamilia basado en la conservación de seis residuos de cisteína, así como en la similitud de la estructura terciaria de las proteínas (Blundell & Humbel 1980; Conlon 2001).

Los genes de la familia multigénica de las relaxinas en vertebrados se encuentran distribuidos en loci ubicados en tres cromosomas distintos, loci que han sido denominados en la literatura como *relaxin family locus A, B y C* (RFLA, B y C) (Park et al. 2008). El número de genes de esta familia varían ampliamente en vertebrados. En aves, por ejemplo, sólo dos genes han sido descritos (RLN2 y INSL5), mientras que en humanos siete genes han sido identificados: INSL5 en el cromosoma 1, INSL6, INSL4, RLN2 y RLN1 en el cromosoma 9, y RLN3 e INSL3 en el cromosoma 19 (figura 8) (Hsu 1999; Wilkinson et al. 2005; Park et al. 2008; Good-Avila et al. 2009).

1.3.2 Estructura y función

La estructura primaria de las proteínas de los miembros de esta familia tienen una baja similitud de secuencia, excepto por la conservación de tres residuos de cisteína y un residuo de glicina en el péptido B (Shabanpoor, Separovic & Wade 2009). Los genes de esta familia son expresados como un preprohormona formada por cuatro dominios funcionales: un péptido señal, un péptido B, un péptido C, y un péptido A. Esta preprohormona sufre una serie de modificaciones pre y post-traduccionales, de modo que la estructura cuaternaria de las proteínas está formada por un péptido heterodimérico con dos dominios estructurales: A y B (Hsu et al. 2005; Steiner et al. 1985; Wilkinson et al. 2005).

Los genes de esta familia tienen roles fisiológicos asociados a la reproducción. Las relaxinas son hormonas pleiotrópicas que realizan una amplia variedad de funciones como promover el crecimiento y ablandamiento del cuello uterino, suavizar el tejido conectivo de los ligamentos interpúbicos, crecimiento de las glándulas mamarias, entre otros (Bathgate et al. 2003; Hsu 2003; Sherwood 2004). Mientras que para los péptidos de Leydig han sido descritas diferentes funciones relacionadas con desarrollo del trofoblasto, descenso testicular, regulación en la formación de hueso y contractibilidad del intestino (Chan & Steiner 2000).

1.3.3 Evolución

El modelo actualmente aceptado sobre la evolución de esta familia sugiere un escenario donde habría ocurrido una duplicación del gen precursor de la súperfamilia en el ancestro común de los olfactores (urocordados y vertebrados) dando origen a los ancestros de las dos familias

mencionadas previamente: INS/IGF y INSL/RLN (figura 9) (Olinski, Lundin, & Hallbook 2006). Posterior a esto habrían ocurrido dos rondas de duplicación genómica en el ancestro común de los vertebrados (Kasahara 2007; Dehal & Boore 2005; Kuraku et al. 2009), originando así cuatro genes indiferenciados en cuatro cromosomas distintos. Uno de estos se habría perdido, quedando de esta forma tres genes indiferenciados en tres cromosomas distintos, los cuales corresponderían a precursores de los clústeres RFLA, B y C (Olinski, Lundin, & Hallbook 2006) (figura 9).

Procesos de nacimiento y muerte de genes habrían moldeado la estructura de esta familia en las especies actuales. En el clúster RFLC los genes RLN3 e INSL3 se habrían originado a partir de una duplicación segmental de 5-7 Mb que habría ocurrido después de la separación de los teleosteos y tetrápodos (Park et al. 2008). En el clúster RFLB el gen INSL6 se habría originado en el ancestro común de los mamíferos placentados (Park et al. 2008) (figura 10a). Los genes INSL4 y RLN1 se habrían originado durante la radiación de los primates, el gen INSL4 se habría originado a partir de un gen RLN en el ancestro común de los catarrinos (monos del viejo mundo y homínidos) (Bièche et al. 2003), mientras que el gen RLN1 se habría originado en el ancestro común de los homínidos (Klonisch et al. 2001) (figura 10b). En el clúster RFLA el gen INSL5 se habría mantenido como gen de copia única.

La divergencia de los RLNs y INSLs habría ocurrido de manera independiente en estos tres linajes cromosómicos en el ancestro común de los vertebrados dado que las relaciones de ortología entre los principales grupos de vertebrados han sido mostradas en trabajos anteriores (Park et al. 2008; Olinski, Lundin, & Hallbook 2006).

1.4 Hipótesis

La baja similitud de secuencia de los miembros de la familia multigénica de las relaxinas en mamíferos se explicaría según un modelo de evolución divergente.

Esta hipótesis, por un lado explicaría la diversidad de funciones de los miembros de la familia, y por otro, predice que la topología del árbol de esta familia multigénica estaría formada por clados de genes ortólogos.

1.5 Objetivo general

- Estudiar el(los) patrón(es) evolutivo(s) de la familia multigénica de las relaxinas en mamíferos.

1.5.2 Objetivos específicos

- Identificar y caracterizar los genes estructurales de la familia multigénica de las relaxinas en mamíferos
- Caracterizar la estructura genómica, inferir relaciones de ortología, y estudiar procesos de selección natural a nivel molecular
- Inferir los mecanismos moleculares que expliquen la dinámica evolutiva de esta familia multigénica.

MÉTODOS

2.1 Secuencias de ADN

Las secuencias de los miembros de la familia multigénica de las relaxinas en distintas especies de mamíferos fueron obtenidas desde las bases de datos públicas del National Center for Biotechnology Center (NCBI) y de la unión del Wellcome Trust con el European Bioinformatics Institute (ENSEMBL). Los genes de la familia de las relaxinas fueron identificados mediante la comparación de exones conocidos con los fragmentos genómicos usando el programa BLASTN versión 2.2.24 (Zhang et al. 2000) disponible en el NCBI. Secuencias extraídas de registros individuales de cADN y/o ADN genómico fueron también incorporadas para obtener una cobertura taxonómica mas amplia (tabla A1).

2.2 Análisis filogenéticos

En los análisis filogenéticos de la familia multigénica de las relaxinas se utilizaron secuencias nucleotídicas de genes estructurales de distintas especies de mamíferos placentados. Se alinearon las secuencias de nucleótidos usando la estrategia G-INS-i del programa MAFFT versión 6 (Kato 2008). Los modelos de evolución molecular para los distintos dominios funcionales del gen fueron obtenidos usando la rutina *propose model* del programa Treefinder versión Octubre 2008 (Jobb et al. 2004). En el análisis de máxima verosimilitud se obtuvo el mejor árbol, y el soporte de las ramas se evaluó a través de un procedimiento bootstrap de 1.000 pseudoréplicas, según un consenso de la regla de la mayoría del 50%. Ambos análisis fueron realizados con Treefinder versión Octubre 2008 (Jobb et al. 2004).

2.3 Selección natural a nivel molecular

2.3.1 Prueba de ramas

Para estudiar el potencial rol de la selección natural a nivel molecular en los diferentes linajes de mamíferos se usaron modelos de rama (*branch models*). Estos modelos estiman diferentes razones de ω (dN/dS) en distintas ramas de la filogenia, de modo que es posible inferir procesos de selección natural en linajes particulares dentro de un árbol (Yang 1998; Yang & Nielsen 1998). Para hacer estos análisis se usó el programa PAML versión 4.3 (Yang 2007), específicamente el módulo CODEML. Todos los modelos fueron corridos tres veces con tres valores iniciales de omega distintos para evitar el problema de los óptimos locales. Los modelos fueron comparados a través de una prueba de razón de verosimilitud *-likelihood-ratio test* (LRT)-.

2.3.2 Prueba de rama-sitio

Se estudió el rol de la selección natural a nivel molecular sobre codones particulares del gen usando los modelos de rama y sitio (*branch-site models*). En estos modelos es posible estimar diferentes razones de ω (dN/dS) en codones particulares del gen, en ramas específicas de la filogenia. Específicamente, se usó el modelo A (Yang & Nielsen 2002; Zhang, Nielsen & Yang 2005), y el método Bayes Empirical Bayes (BEB) para estimar probabilidades posteriores de pertenencia a las distintas clases de sitios (Yang, Wong & Nielsen 2005), tal como está implementado en el módulo CODEML del programa PAML versión 4.3 (Yang 2007). La filogenia de los mamíferos usada en las pruebas de rama y de rama-sitio se construyó según un

consenso de estudios recientes (Madsen 2009; Murphy & Eizirik 2009; Springer et al. 2009; Springer & Krajewski 2009; Steiper & Young 2009).

RESULTADOS

3.1 Evolución de la familia multigénica de las relaxinas en mamíferos

3.1.1 Estructura genómica de la familia multigénica de las relaxinas en mamíferos

La estructura genómica de la familia multigénica de las relaxinas en mamíferos, está formada en la mayoría de los casos por un gen INSL5 en el clúster RFLA, los genes INSL6 y RLN2 en el clúster RFLB y los genes RLN3 e INSL3 en el clúster RFLC. Desviaciones de este patrón son posibles de apreciar y se deberían esencialmente a dinámicas de nacimiento y muerte de genes linaje específico (figura 11).

Dentro de los mamíferos, en el clado de los euarcontoglires (e.j. primates, conejos, ratones), es donde se verifica variación en cuanto al número de copia de genes, la cual está circunscrita principalmente al clúster RFLB (figura 11). En este grupo, y al igual que en los otros grupos de mamíferos los clústeres RFLA y RFLC se encuentran altamente conservados (figura 11).

Dentro de euarcontoglires, es en el clúster RFLB de primates, en donde se verifica la mayor variación en el número de copia de genes, en el cual además de los genes RLN2 e INSL6 presente en la mayoría de los mamíferos, es posible identificar el gen INSL4 en tarseros y primates catarrinos (monos del viejo mundo y homínidos) y el RLN1 en homínidos (figura 11).

3.1.2 Análisis filogenético de la familia multigénica de las relaxinas en mamíferos

Los análisis recuperaron la monofilia de los paralogones de la familia de las relaxinas

(figura 12). En el clado RFLC se recuperó la relación de hermandad de los clados RLN3 e INSL3, mientras que en el clado RFLB se recuperó la relación de hermandad de los clados formados por el gen INSL6 de mamíferos placentados y el clado formado por los genes INSL4, RLN2 y RLN1 (figura 12), ambos con fuerte soporte de ramas. Por otro lado el clado RFLA también se recuperó como un grupo monofilético.

La tricotomía del árbol y las relaciones de hermandad entre los miembros de los clústeres RFLC y B se ajustan al modelo de evolución sugerido para esta familia (Olinski, Lundin, & Hallbook 2005). De acuerdo a éste, la diversificación habría ocurrido por duplicaciones de pequeña escala independientemente en cada cromosoma, prediciendo así la monofilia de genes que están en un mismo cromosoma (figura 13) (Olinski, Lundin, & Hallbook 2005).

3.2 Evolución del clúster génico RFLB en mamíferos eucariontes

3.2.1 Evolución del clúster génico RFLB en el conejo

La estructura del clúster RFLB del conejo (*Oryctolagus cuniculus*) representa una excepción dentro de los mamíferos. En esta especie, el clúster RFLB está formado por un gen INSL6 al extremo 5' y por cinco genes RLN2 dispuestos en tándem, los que han sido caracterizados como RLN2-T1, 2, 3, 4, y 5 (figura 14), desde el extremo 5' al 3'. Todos estos genes fueron identificados en el mismo fragmento genómico de ~195 Kb y conservan el patrón de sintenia observado en mamíferos, en el cual el gen JAK2 se encuentra al extremo 5' del clúster, y el gen C9orf46 al extremo 3'. De acuerdo con la verificación del marco de lectura, todos codifican para proteínas, por lo que serían potencialmente funcionales, aunque se requeriría de estudios funcionales para probarlo.

3.2.1.2 Prueba de rama

Para investigar el posible rol de la selección natural en la evolución de los duplicados del gen RLN2 en el conejo, se estudió la heterogeneidad del parámetro ω (dN/dS) a través de la implementación de cuatro modelos anidados que permiten explorar la variación de este parámetro en las distintas ramas de la filogenia. Los modelos implementados se muestran en la figura 15, y el valor de los parámetros estimados se resumen en la tabla 1.

El modelo que estimó un solo ω para todas las ramas del árbol arrojó un valor de 0.73, mientras que el modelo de 2ω estimó un $\omega_0 = 0.71$ para los linajes de mamíferos no lagomorfos y un $\omega_1 = 1.34$ para el clado del conejo. El modelo 3ω estimó $\omega_0 = 0.71$ para los linajes de mamíferos no lagomorfos, un $\omega_1 = 1.3$ para la rama ancestral del conejo, y un $\omega_2 = 1.52$ para las ramas descendientes. Finalmente el modelo 7ω , el cual estimó omegas independientes para los linajes de mamíferos no lagomorfos, para la rama ancestral del clado del conejo, así como para cada una de las descendientes, sugiere que existen diferentes valores de ω (tabla 1).

Los modelos anteriores fueron comparados a través de una prueba de razón de verosimilitud (LRT). En la primera prueba de hipótesis ($H_0: 1\omega$ vs $H_1: 2\omega$) el modelo de 2 omegas tuvo un ajuste significativamente mejor que el modelo de un omega ($2\Delta\ln L = 4.72$; $df = 1$; $P=0.03$), indicando de esta manera que el clado de los RLNs del conejo tendrían un valor de omega ($\omega_2 = 1.33$) distinto al del resto de los mamíferos ($\omega_1 = 0.71$). En la segunda prueba de hipótesis ($H_0: 2\omega$ vs $H_1: 3\omega$) el LRT reveló que el modelo de tres omegas no tiene un ajuste significativamente mejor que el de dos ($2\Delta\ln L = 0.03$; $df=1$; $P=0.86$), situación similar ocurrió en la comparación entre el modelo de dos y siete omegas ($H_0: 2\omega$ vs $H_1: 7\omega$) ($2\Delta\ln L = 6.6$;

df=5; P=0.25) (tabla 2).

Estos resultados muestran que el modelo de 2ω es el que presenta un mejor ajuste entre todos los modelos implementados, sugiriendo así que la retención de los genes parálogos RLN2 en el genoma del conejo habría ocurrido por un régimen de selección positiva.

3.2.1.3 Prueba de rama-sitio

Se estudió la heterogeneidad del parámetro ω a través de los sitios del gen en el clado de los genes parálogos RLN2 del conejo (*foreground branch*) (figura 16), a través del modelo de rama-sitio A (Yang & Nielsen 2002; Yang et al. 2002; Zhang et al. 2005). Este modelo se comparó a través de una LRT con un modelo A neutral, en el cual se fija el valor de ω_2 a 1. El valor de los parámetros estimados para los respectivos modelos se resumen en la tabla 3, mientras que las proporciones de las clases de sitio y ω de cada sitio se resumen en la figura 17.

El modelo A sugiere que del total de sitios del gen, un 30.86 % tienen un omega promedio de 0.25, lo que indicaría un régimen de selección purificadora. Por otro lado un 43.75 % de los sitios tiene un omega promedio ~ 1 , lo que indicaría un régimen de neutralidad. Finalmente para un 25.39 % de los sitios fue estimado un omega promedio de 4.46, por lo cual indicaría un régimen de selección positiva.

La prueba de hipótesis (H_0 : Modelo A neutral vs H_1 : Modelo A) reveló que el modelo A tuvo un ajuste significativamente mejor que el modelo A neutral ($2\Delta l = 5.79$; $df = 1$; $P < 0.025$) (tabla 4), sugiriendo así la existencia de un régimen de selección positiva, e identificando sitios particulares bajo este régimen. Los resultados del modelo A sugieren un total de 7 sitios bajo selección en el péptido B, y 6 sitios en el péptido A (figura 18).

3.2.2 Evolución del clúster génico RFLB en primates

La estructura genómica del clúster RFLB varía en los distintos grupos de primates, por un lado los primates estrepsirrinos (e.g. lémures, gálagos) tienen la estructura típica de mamíferos formada por los genes INSL6 y RLN2 (figura 19). En tarseros y catarrinos, además de los dos genes antes mencionados el clúster cuenta con el gen INSL4, el cual en monos del nuevo mundo está presente como un pseudogen (figura 19). Los homínidos cuentan con el repertorio más amplio ya que tienen dos copias del gen RLN, el gen RLN2 al extremo 5' y el RLN1 al extremo 3' (figura 19). El gen INSL6 se encuentra presente como gen de copia única en todos los primates, sin embargo en tarseros este gen habría sido inactivado por sustituciones sin sentido.

La identificación de un gen INSL4 funcional en los tarseros y de un pseudogen en monos del nuevo mundo, sugiere que este gen no es exclusivo de catarrinos, tal como había sido sugerido en la literatura (Bièche et. al. 2003). En todos los grupos de primates catarrinos estudiados se identificó una copia funcional del gen INSL4, excepto en el gorila, en el cual se identificaron dos copias (figura 19). Además de los genes funcionales se identificaron pseudogenes tanto al extremo 5' como al 3' del gen funcional (figura 19).

En homínidos el clúster RFLB está formado por dos copias funcionales del gen RLN, sin embargo es posible observar algunas desviaciones de este patrón. En el orangután ambas copias serían pseudogenes, los cuales se habrían inactivado por sustituciones sin sentido (figura 19).

Por otro lado en el chimpancé el RLN del extremo 3' sería un gen quimérico RLN2/1. Esta hipótesis se sugiere a partir del análisis de la similitud de secuencia de los genes RLN en esta especie (figura 20). La similitud de secuencia del primer exón de estos genes es de 100%,

mientras que para el segundo exón es de 87% (figura 21). Por otro lado el primer exón del gen RLN1 en el chimpancé tiene el motivo característico del primer exón del gen RLN2: DSWMEE, el cual en humanos y otros primates distingue la copia 5' de la 3' (Hudson et al. 1983; Evans et al. 1994) (figura 22).

Finalmente en tarseros se identificaron seis copias de este gen en diferentes *scaffolds*, las cuales fueron caracterizadas como α , β , γ , δ , ϵ , y ζ desde el *scaffold* de menor al de mayor denominación (tabla A1), de las cuales solo una de ellas (RLN1 ϵ) codifica para proteínas (figura 19).

3.2.2.2 Análisis filogenético

En la literatura el gen INSL4 ha sido descrito como una innovación de primates catarrinos, por lo que en la filogenia se debería recuperar el clado INSL4 anidado dentro del clado de primates (figura 23). Sin embargo en contraste con esta expectativa el clado INSL4 se recuperó como grupo hermano de los boreoeuterios (figura 24), lo cual sugiere un origen más antiguo para este gen y una retención diferencial en primates haplorrinos. No obstante, el nodo que define esta relación de hermandad entre mamíferos eucariontes y laurasiaterios no tiene soporte por lo que no es posible establecer de manera precisa el origen de este gen.

Por otro lado, las relaciones filogenéticas del clado INSL4 sugieren la existencia de dos grupos de genes, en donde uno de ellos estaría formado por todos los pseudogenes situados al extremo 5', mientras que el otro estaría formado por los genes codificantes y otros pseudogenes situado al extremo 3' (figura 25).

En la estructura del clúster RFLB en homínidos hay dos copias del gen RLN, por lo que

el gen RLN1 ha sido descrito como una innovación evolutiva de este grupo. De acuerdo con esto en la filogenia se debería recuperar el único representante RLN de monos del viejo mundo como grupo hermano del clado de los genes RLN1 y RLN2 de homínidos (figura 23). Sin embargo en contraste con esta expectativa el gen RLN de copia única de los monos del viejo mundo se recupera como grupo hermano del clado RLN2 de homínidos, mientras que el gen RLN de copia única de los monos del nuevo mundo se recupera como grupo hermano del clado RLN1 de homínidos (figura 24).

Las relaciones de hermandad entre estos genes sugieren que la duplicación que originó los duplicados RLN habría ocurrido no en el ancestro común de los homínidos sino que en el de los primates antropoideos. Bajo este escenario, y posterior a su duplicación el gen RLN del extremo 5' se habría perdido en monos del nuevo mundo y habría sido retenido en los catarrinos. Por otro lado el gen del extremo 3' se habría perdido en monos del viejo mundo y habría sido retenido en monos del nuevo mundo y homínidos.

DISCUSIÓN

4.1 Evolución de la familia multigénica de las relaxinas en mamíferos

El análisis de la estructura genómica de la familia multigénica de las relaxinas en mamíferos reveló que los clústers RFLA, B y C están conservados en la mayoría de los linajes, excepto por algunas duplicaciones, pseudogenizaciones y/o deleciones linaje específicas. La estructura del clúster RFLB en euarcontogliares varía en los diferentes linajes, desde casos como el del conejo, en el cual hay seis genes funcionales hasta el de la rata canguro donde es posible encontrar solo uno (figura 11).

El análisis filogenético recuperó tres clados los cuales corresponden a los tres paralogones RFLA, B y C. Dentro de estos clados se recuperaron las relaciones de hermandad de los genes RLN3 e INSL3 en el paralogón RFLC; INSL6, INSL4, RLN2 y RLN1 en el paralogón RFLB, y el gen de copia única INSL5 en el paralogón RFLA (figura 12). De acuerdo con estas relaciones de hermandad la evolución de esta familia se explicaría por un modelo divergente, en el cual la diferenciación de los genes RLN e INSL habría ocurrido de manera independiente en cada paralogón.

4.1.2 Modelo de evolución de la familia multigénica de las relaxinas en mamíferos

La evolución de la familia multigénica de las relaxinas ha ocurrido por varios mecanismos. Por un lado la topología del árbol de esta familia, la cual recupera los genes ortólogos como grupos monofiléticos, sugiere que la evolución de esta familia ha ocurrido principalmente por un proceso de evolución divergente (figura 12). Por otro lado, procesos

aleatorios de nacimiento y muerte en diferentes linajes de mamíferos, como el clúster RFLB de mamíferos eucariontes, y en especial en primates también explicarían la evolución de esta familia multigénica.

4.1.3 Nomenclatura de los genes de la familia multigénica de las relaxinas

La nomenclatura de las relaxinas y los péptidos de Leydig ha sido establecida de acuerdo al orden temporal en que han sido identificados, y han sido caracterizados como RLN o INSL en relación a la similitud de secuencia con estos genes (Shabanpoor, Separovic & Wade 2009). Sin embargo esta nomenclatura no considera la estructura genómica ni las relaciones de homología de estos genes, conduciendo así a errores. Por ejemplo el gen INSL4 fue caracterizado como un INSL (Chassin et al. 1995), sin embargo se originó a partir de un RLN y no a partir de un INSL; por otro lado el gen RLN3 también fue nombrado como INSL7, a pesar de que es un RLN (Bathgate et al. 2002).

La manera de nombrar a los miembros de una familia multigénica es aquella en la cual se consideren las relaciones evolutivas, la distribución en el genoma, y el orden de los genes en el cromosoma. De este modo los genes deberían nombrarse de acuerdo con sus relaciones de ancestralidad, desde el cromosoma de menor a mayor denominación y desde el extremo 5' al 3' (Aguileta Bielawski & Yang 2006; Wain et al. 2002).

De acuerdo con esto y tomando como ejemplo al humano, se sugiere la siguiente nomenclatura para los miembros de la familia multigénica de las relaxinas: el gen INSL5 situado en cromosoma 1, debería ser nombrado como INSL1, el gen INSL6 situado en el cromosoma 9 como INSL2 y el INSL3 en el cromosoma 19 mantendría su nombre. Los genes RLN repetidos

en tándem en el cromosoma 9, se nombrarían como RLN1-T1, RLN1-T2 y RLN1-T3 desde el extremo 5' al 3'. Finalmente el gen RLN3 situado en el cromosoma 19 debería ser nombrado como RLN2 (figura 26).

4.2 Evolución del clúster RFLB en mamíferos eucariontes

4.2.1 Evolución del clúster RFLB en el conejo

La estructura genómica del clúster RFLB en el conejo está formada por cinco genes parálogos RLN2 funcionales, lo cual representa una excepción dentro de los mamíferos. La falta de información en otras especies de lagomorfos impide determinar si este patrón es exclusivo de esta especie, o de un nivel taxonómico superior. Los análisis de selección a nivel molecular sugieren que los genes parálogos RLN2 en el conejo han estado sometidos a un régimen de selección positiva. A esta conclusión se ha llegado tanto a través de la implementación de modelos de ramas, así como de rama-sitio.

Los sitios que participan en la interacción ligando-receptor de la proteína RLN2 y el receptor LGR7, se encuentran en el péptido B y tienen la estructura primaria característica: -Arg¹³-X-X-X-Arg¹⁷-X-X-Ile²⁰- (Hsu et al. 2002) (figura 27a). En mamíferos este motivo está altamente conservado, sin embargo en el conejo la primera arginina de este motivo-Arg¹³- ha sufrido una sustitución por asparagina -Asn¹³-. Esta sustitución cambia las propiedades fisicoquímicas del residuo, ya que la arginina es un aminoácido básico mientras que la asparagina es polar (figura 27b). Esta sustitución podría tener alguna consecuencia en la interacción ligando-receptor y por ende en la actividad de la proteína, debido a que las interacciones moleculares están determinadas por las propiedades electrostáticas de los residuos

de los aminoácidos.

Los análisis teóricos sugieren que un gen bajo selección positiva podría estar adquiriendo una nueva función. Sin embargo en algunos caso un valor de omega mayor que 1 no significa necesariamente que ha ocurrido un cambio funcional del gen. Esto ha sido ya mostrado en otras familias multigénicas, cuyos genes tienen un valor de omega indicativo de selección positiva, sin embargo los cambios parecen no afectar la función de manera apreciable (Golding & Dean 1998).

Por otra parte, un cambio funcional de un gen puede ocurrir incluso en casos donde el valor de omega es menor que 1. Esto se debe a que la función de una proteína puede cambiar por una o unas pocas sustituciones de aminoácidos que son claves en la determinación de la función (ej. estructura terciaria, actividad, etc.), pero que no aumentan el promedio de omega por sobre 1 (Suzuki & Nei 2004). Particularmente, es interesante el ejemplo de los genes M/LWS. Las proteínas codificadas por estos genes son pigmentos visuales que absorben luz de diferentes longitudes de onda, de modo que sólo unas pocas sustituciones son responsables de la amplia variabilidad de la actividad de estas proteínas en los vertebrados. No obstante, ni los modelos de máxima verosimilitud (Nielsen y Yang 1998; Yang 1997; Yang et al. 2000; Yang y Nielsen 2002) ni aquellos basados en parsimonia (Suzuki y Gojorobi 1999; Suzuki, Gojorobi y Nei 2001) sugieren procesos selectivos en estos genes (Suzuki & Nei 2004).

Debido a estas restricciones, los trabajos en la actualidad además de estudios *in silico*, utilizan aproximaciones experimentales para entender las consecuencias funcionales de los cambios de aminoácidos. Un ejemplo de esto es el estudio de Klonisch et al. (2001), quienes estudiaron las consecuencias funcionales de sustituciones en el péptido B del gen RLN2 que

participan en la interacción ligando-receptor en primates estrepsirrinos (e.g. lémures, gálagos), y que de acuerdo a los autores habrían originado una molécula única en este grupo de mamíferos.

Los estudios *in vivo* sobre la función del gen RLN2 en el conejo sugieren, como en otros mamíferos (Bathgate et al. 2003; Hsu 2003), un rol asociado a la reproducción, habiendo sido detectada su expresión en tejidos como la placenta y el útero (Fields et al. 1995). Sin embargo, en esta especie se ha encontrado evidencia de un rol fuera del sistema reproductor. A este respecto Jetten et al. (1999) realizaron un estudio en el cual se detectó la expresión de este gen en células epiteliales traqueobronquiales, y sugirieron que podría cumplir un rol durante la diferenciación escamosa.

En resumen los resultados de este estudio sugieren que la retención de los genes RLN2 en el genoma del conejo habría ocurrido por selección positiva. Por otro lado, los estudios *in vivo* de Jetten et al., sobre la expresión de este gen, sugieren un rol en la diferenciación celular, lo cual representa el primer y único reporte de una función de las relaxinas en mamíferos fuera del sistema reproductor. Estos antecedentes sugieren un probable escenario de neofuncionalización, en el cual post duplicación un(os) gen(es) habrían asumido un rol en la diferenciación celular (figura 28).

4.2.2 Evolución del clúster RFLB en primates

4.2.2.1 Retención diferencial del gen INSL4 en tarseros y primates catarrinos

Los resultados de este estudio sugieren que la evolución del clúster RFLB en primates habría ocurrido por un proceso de nacimiento y muerte de genes. Un hallazgo interesante resultó ser la existencia de pseudogenes del gen INSL4 en el extremo 5' del gen funcional en todos los

primates catarrinos y en el extremo 3' en algunos otros linajes.

En estudios anteriores sobre la evolución de la familia de las relaxinas en primates se ha sugerido que el gen INSL4 se habría originado en el ancestro común de los primates catarrinos (Bièche et al. 2003), sin embargo los análisis de la estructura genómica de este trabajo muestran que este gen también está presente en tarseros y en monos del nuevo mundo.

Desde un punto de vista filogenético, si el gen INSL4 fuese una innovación de los primates catarrinos, el clado de los genes INSL4 debiese ser recuperado dentro del clado de los primates (figura 23), sin embargo, los resultados de este estudio no soportan dicha predicción. En los análisis de este estudio el clado INSL4 fue recuperado como grupo hermano de los mamíferos boreoeuterios, sin embargo esta relación no tiene soporte (figura 24). Por otro lado, la monofilia de los genes RLN de primates tiene un buen soporte de rama (figura 24), por lo que es poco probable que el clado INSL4 hubiese podido ser recuperado dentro de ese clado, descartándose así el carácter de novedad evolutiva de primates catarrinos. Por otro lado, es importante destacar que la monofilia de los clados RLN de euarcontoglires y laurasiateria también están bien soportadas (figura 24), por lo que lo más probable es que este gen se haya originado en el ancestro común del primer grupo y se haya perdido en todos los grupos excepto en primates haplorrinos.

Para establecer fehacientemente que este gen tiene un origen más antiguo se debería identificar un gen INSL4 en especies de mamíferos que no sean primates catarrinos, lo cual no es factible siempre, ya que este gen podría haber sido eliminado del genoma, tal como ha ocurrido con otros genes de esta familia en otros linajes (Southey et al. 2009) o haber sufrido un proceso de pseudogenización tal que hoy no sea posible reconocer rastro alguno de las secuencias (Li

1997). En este sentido el hallazgo de un gen INSL4 funcional en tarseros y un pseudogen en monos del nuevo mundo representa evidencia directa no solamente de un origen más antiguo de este gen, sino que además de un proceso de retención diferencial en primates haplorrinos (figura 29).

4.2.2.2 Retención diferencial de los genes RLN1 y 2 en monos del nuevo y viejo mundo

Los análisis de la estructura genómica del clúster RFLB en primates muestran la existencia de dos copias del gen RLN en homínidos, observación que ha hecho sugerir a otros autores que este gen se habría originado en el ancestro común de este grupo (Klonisch et al. 2001).

De acuerdo con esta hipótesis en la filogenia se debería recuperar el clado RLN1 de homínidos como grupo hermano del clado RLN2 del mismo grupo, con exclusión del gen RLN de los monos del viejo mundo. A su vez, el gen RLN de los monos del nuevo mundo debería ser recobrado como grupo hermano del clado anterior (figura 23). En contraste con esta expectativa el gen RLN de los monos del viejo mundo fue recuperado como grupo hermano del clado RLN2 de homínidos, mientras que el de los monos del nuevo mundo se recuperó como grupo hermano del clado RLN1 de homínidos (figura 24). Esta topología sugiere que la duplicación habría ocurrido en el ancestro común de los primates antropoideos (platirrinos y catarrinos). De acuerdo con este escenario el gen localizado al extremo 5' habría sido retenido en monos del viejo mundo y perdido en monos del nuevo mundo, mientras que el gen del extremo 3' habría sido retenido en monos del nuevo mundo y perdido en monos del viejo mundo. En contraste, en homínidos se encontrarían ambos genes presentes (figura 29).

4.2..2.3 Evolución mosaico en los genes RLN del chimpancé

El análisis de la similitud de secuencia de los genes RLN1 y 2 en el chimpancé muestran que el primer exón de los genes RLN tiene un 100 % de similitud de secuencia entre ellos (figura 20 y 21). Esta similitud de secuencia se explica por un proceso de evolución concertada que abarcaría el primer exón de estos genes. El mecanismo molecular responsable de esta interacción sería un proceso de recombinación homóloga entre ambas regiones del gen (figura 30), sin embargo es difícil inferir si se trata de una interacción recíproca o unilateral. Al respecto, Evans, Fu & Tregear (1994) ya habían advertido esta similitud de secuencia en el primer exón de estos genes cuando los caracterizaron por primera vez en 1994, y sugirieron que este patrón podría ser explicado por un proceso de conversión génica en el cual el primer exón del gen del extremo 5' estaría convirtiendo al primer exón del gen del extremo 3'. No obstante lo anterior su estudio no proveyó ninguna evidencia a favor de esta hipótesis, por lo que alternativamente podría tratarse de un entrecruzamiento desigual.

El mecanismo molecular de recombinación responsable de la evolución concertada entre estos exones podría determinarse a través de un análisis de las secuencias flanqueantes y del intrón, sin embargo esta información no está disponible en las bases de datos públicas por lo que por ahora no es posible determinar cuál es el mecanismo molecular responsable de la interacción de ambos genes.

CONCLUSIÓN

De modo general los resultados de esta investigación muestran que la evolución de la familia multigénica de las relaxinas en mamíferos ha ocurrido por múltiples mecanismos. El patrón global de evolución de esta familia génica se explicaría principalmente por un proceso de evolución divergente, sin embargo, procesos estocásticos de nacimiento y muerte de genes en diferentes linajes, especialmente en el clúster génico RFLB en euarcontoglíres también darían cuenta de su evolución.

En el clúster RFLB del conejo se encontró un patrón único de duplicación linaje específico de genes RLN, los cuales habrían sido retenidos en el genoma por selección positiva. En el clúster RFLB de primates también se verifican procesos de nacimiento y muerte, lo cual sugiere un escenario diferente al que actualmente es aceptado en relación a la evolución del clúster RFLB. Por un lado, el gen INSL4 no sería una novedad evolutiva en primates catarrinos, sino más bien tendría un origen más antiguo que el actualmente sugerido, siendo retenido diferencialmente en primates haplorrinos. Por otro lado, la duplicación génica que originó al gen RLN1 no habría ocurrido en el ancestro común de los homínidos, sino que habría tenido lugar en el ancestro común de los antropoideos. En este sentido, el gen localizado al extremo 5' del cromosoma habría sido retenido en monos del viejo mundo y habría sido perdido en monos del nuevo mundo, mientras que por otro, el gen del extremo 3' habría sido retenido en monos del nuevo mundo y perdido en monos del viejo mundo.

En conclusión, esta tesis muestra que los genomas son sistemas dinámicos que evolucionan en el tiempo a través de la interacción de distintos mecanismos moleculares tales como duplicaciones génicas en tándem de pequeña escala, pseudogenizaciones y deleciones,

recombinación homóloga, muchos de los cuales pueden variar tanto en el tiempo como en el espacio.

Tabla 1. Valores de los parámetros estimados en los modelos de rama.

Modelo	ω^a	$\ln L^b$
1 ω	$\omega_0=0.734$	-6615.6
2 ω	$\omega_0=0.71, \omega_1=1.335,$	-6613.232
3 ω	$\omega_0=0.71, \omega_1=1.304, \omega_2=1.522$	-6613.215
7 ω	$\omega_0=0.71, \omega_1=1.304, \omega_2=0/0.0037, \omega_3=0/0.01, \omega_4=0.0001 \omega_5=0/0.014, \omega_6=0.167$	-6609.904

^adN/dS, ^bLogaritmo natural de verosimilitud.

Tabla 2. Pruebas de razón de verosimilitud de los modelos de rama.

Prueba de hipótesis	$2\Delta \ln L^a$	gl ^b	P ^c
H ₀ : 1 ω vs H ₁ : 2 ω	4.736	1	< 0.05
H ₀ : 2 ω vs H ₁ : 3 ω	0.034	1	> 0.05
H ₀ : 2 ω vs H ₁ : 7 ω	6.655	5	> 0.05

^aRazón de verosimilitud, ^bGrados de libertad, ^cProbabilidad de chi-cuadrado.

Tabla 3. Valores de los parámetros estimados en los modelos rama-sitio A y A neutral.

Modelo	Parámetros (p^a, ω^b)	$\ln L^c$	Sitios ^d
A	$p_0=0.308, p_1=0.437, p_2=0.105, p_3=0.148;$ $\omega_0=0.258, \omega_1=1, \omega_2=4.461$	-6501.6	B: 1, 4, 9, 10, 14, 16, 22 A: 2, 3, 5, 11, 16, 19
A neutral	$p_0=0.243, p_1=0.329, p_2=0.181, p_3=0.245;$ $\omega_0=0.261, \omega_1=1, \omega_2=1$	-6504.498	

^aProporción de sitios, ^b $\omega=dN/dS$, ^cLogaritmo natural de verosimilitud, ^dSitios bajo selección positiva

Tabla 4. Prueba de razón de verosimilitud de los modelos de rama-sitio.

Prueba de hipótesis	$2\Delta \ln L^a$	gl ^b	P ^c
H ₀ : Modelo A vs H ₁ : Modelo A, $\omega_2=1$	5.796	1	< 0.025

^aRazón de verosimilitud, ^bGrados de libertad, ^cProbabilidad de chi-cuadrado

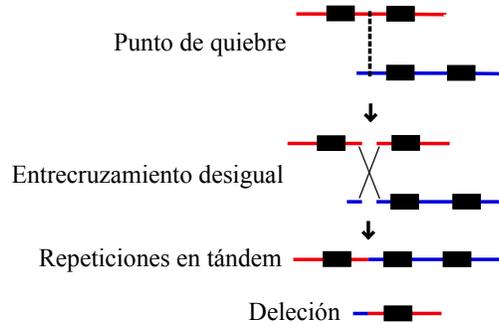


Figura 1. Duplicación génica por entrecruzamiento desigual. El entrecruzamiento desigual ocurre durante el proceso de recombinación homóloga, generalmente entre cromátidas que no están correctamente apareadas, de modo que la recombinación es asimétrica, produciendo segmentos repetidos en tándem en un cromátida y una delección en la otra.

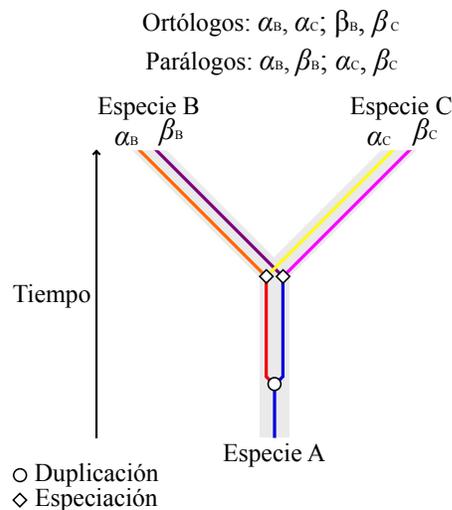


Figura 2. Relaciones de homología entre genes. El modelo describe las relaciones de homología entre genes que se originan a partir de un gen ancestral común. Los distintos colores representan la divergencia de los genes producto de los eventos de duplicación y especiación.

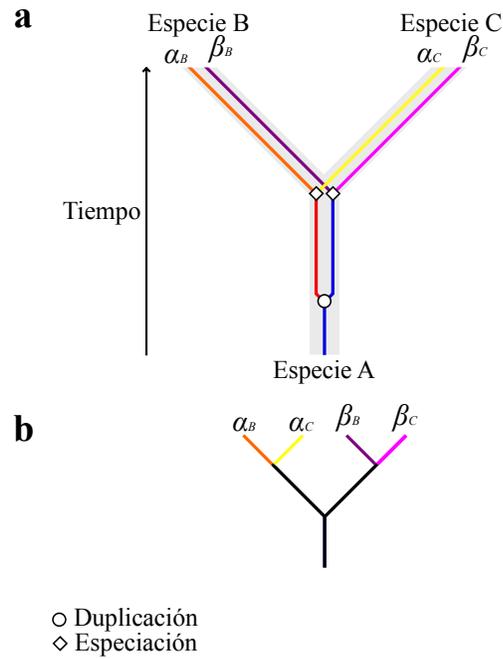


Figura 3. Modelo de evolución divergente. (a) representación gráfica del modelo, (b) predicción filogenética del modelo. Los colores representan divergencia.

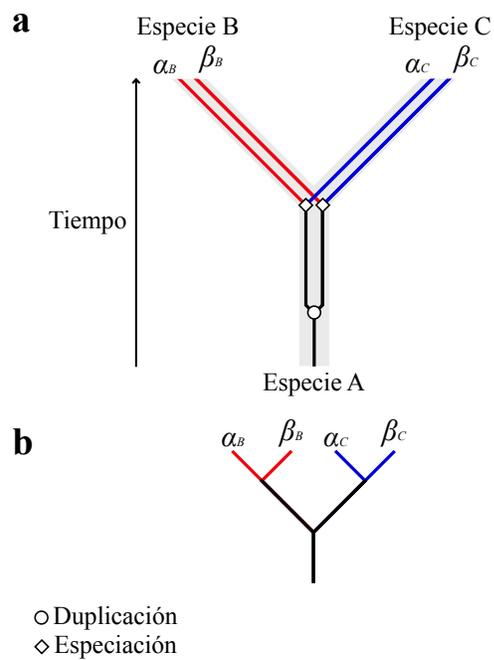


Figura 4. Modelo de evolución concertada. (a) representación gráfica del modelo, (b) predicción filogenética del modelo. Los colores representan divergencia.

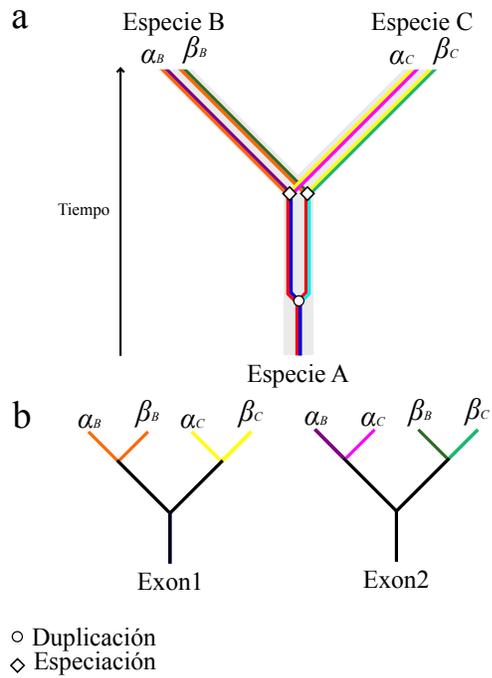


Figura 5. Modelo de evolución mosaico. (a) representación gráfica del modelo, (b) predicción filogenética del modelo. Los colores representan divergencia.

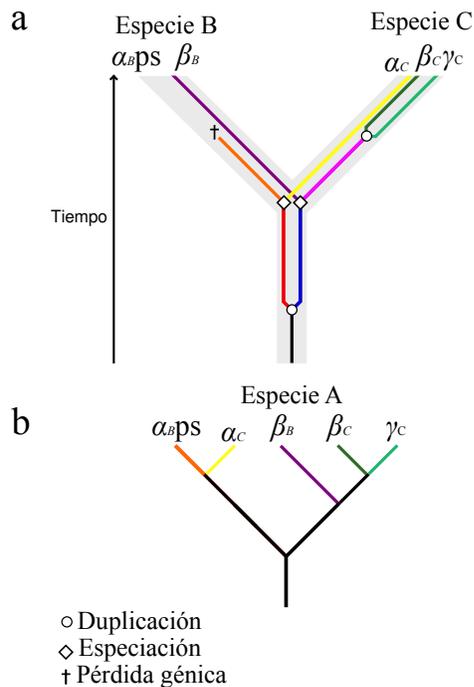


Figura 6. Modelo de evolución nacimiento y muerte. (a) representación gráfica del modelo, (b) predicción filogenética del modelo. Los colores representan divergencia.

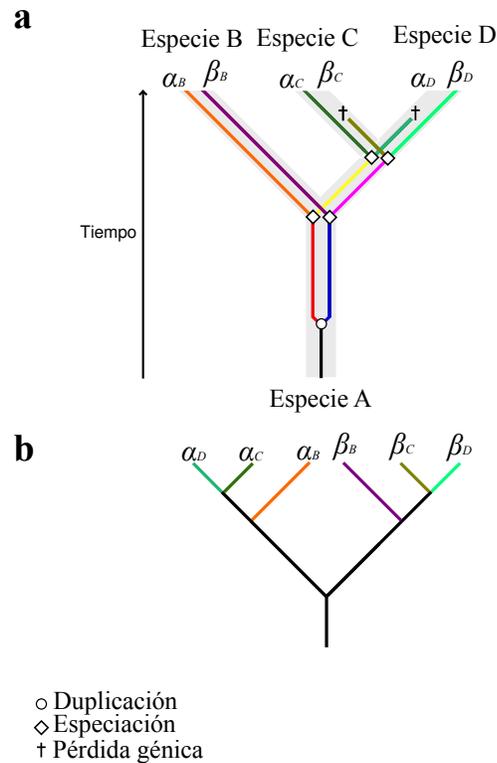


Figura 7. Modelo de evolución retención diferencial. (a) representación gráfica del modelo, **(b)** predicción filogenética del modelo. Los colores representan divergencia.

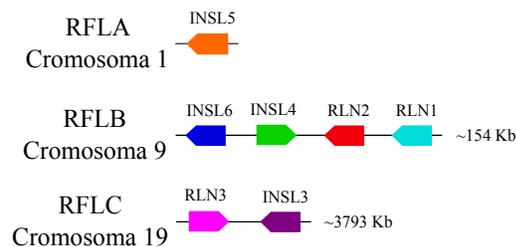


Figura 8. Familia multigénica de las relaxinas en humanos. La familia multigénica de las relaxinas en humanos está distribuida en tres cromosomas distintos: 1, 9 y 19. En el primer loci; RFLA, está el gen INSL5; en el segundo RFLB, están los genes INSL6, INSL4, RLN2 y RLN1; en el tercer loci RFLC están los genes RLN3 e INSL3. La punta del recuadro indica la orientación de los genes en el cromosoma.

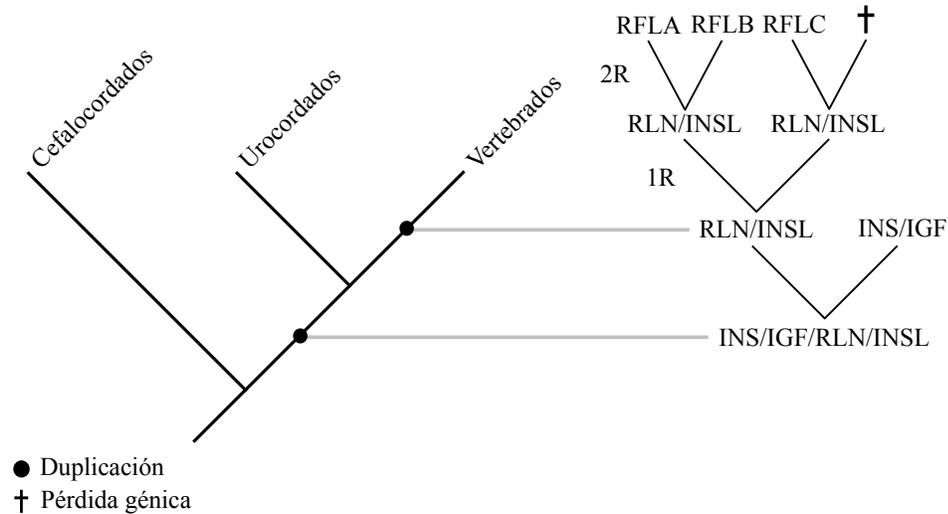


Figura 9. Modelo de evolución de la familia multigénica de las relaxinas. De acuerdo con este modelo en el ancestro común de los olfactores (urocordados y vertebrados) a partir de un gen ancestral INS/IGF/RLN/INSL habría ocurrido la duplicación génica que originó los genes ancestrales de la familia de las relaxinas (RLNs e INSLs) y la familia de la insulina y los factores de crecimiento insulínicos (INSs y IGFs). Luego en el ancestro común hipotético de los vertebrados habrían ocurrido dos rondas de duplicación genómica, originando así cuatro genes RLN/INSL, de los cuales uno de ellos se perdió, mientras que los otros tres corresponderían a los genes ancestrales de los clústeres génicos RFLA, B y C (Olinski, Lundin & Hallbook 2006).

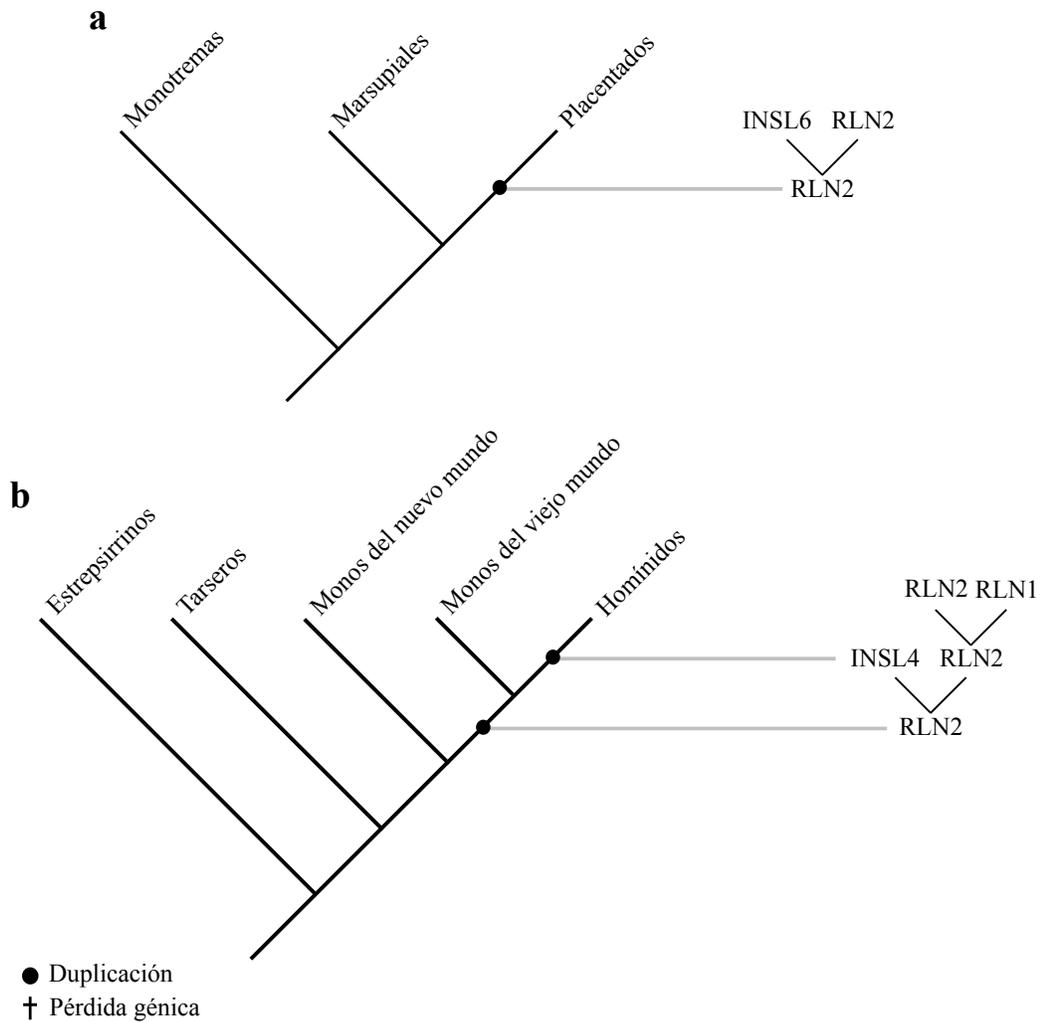


Figura 10. Modelo de evolución del clúster génico RFLB. (a) El clúster génico RFLB se habría diversificado a través de duplicaciones segmentales de pequeña escala; en el ancestro común de los mamíferos placentados el gen INSL6 se habría originado a partir del gen RLN2. **(b)** En primates el gen INSL4 se habría originado a partir de una duplicación del gen RLN2 en el ancestro común de los primates catarrinos (monos del viejo mundo y homínidos), y por otro lado el gen RLN1 se habría originado también a partir del RLN2 en el ancestro común de los homínidos (Olinski, Lundin & Hallbook 2006).

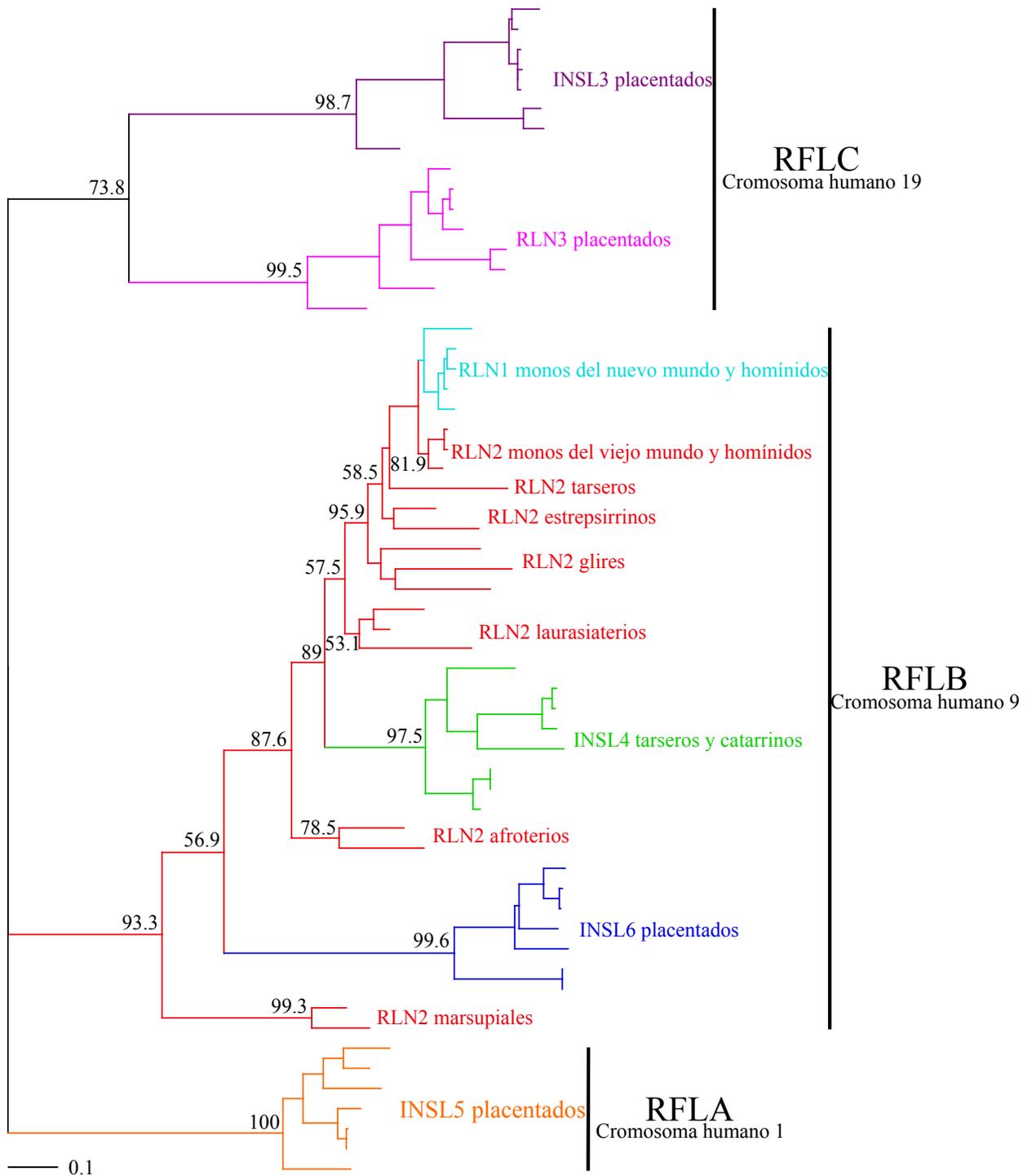


Figura 12. Filograma de máxima verosimilitud de la familia multigénica de las relaxinas en mamíferos. Los valores sobre los nodos corresponde al soporte bootstrap.

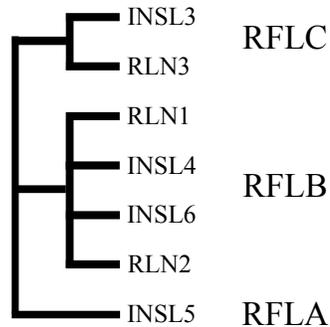


Figura 13. Topología del árbol de la familia multigénica de las relaxinas según modelo aceptado actualmente. De acuerdo con este modelo se debería recuperar una tricotomía, en la cual cada clado estaría formado por los genes de los clústeres RFLA, B y C (Olinski, Lundin & Hallbook 2006).



Figura 14. Estructura genómica del clúster RFLB en el conejo. El clúster génico RFLB del conejo conserva el patrón de sintenia observado en mamíferos, en donde al extremo 5' está el gen JAK2, y al extremo 3' el gen C9orf46.

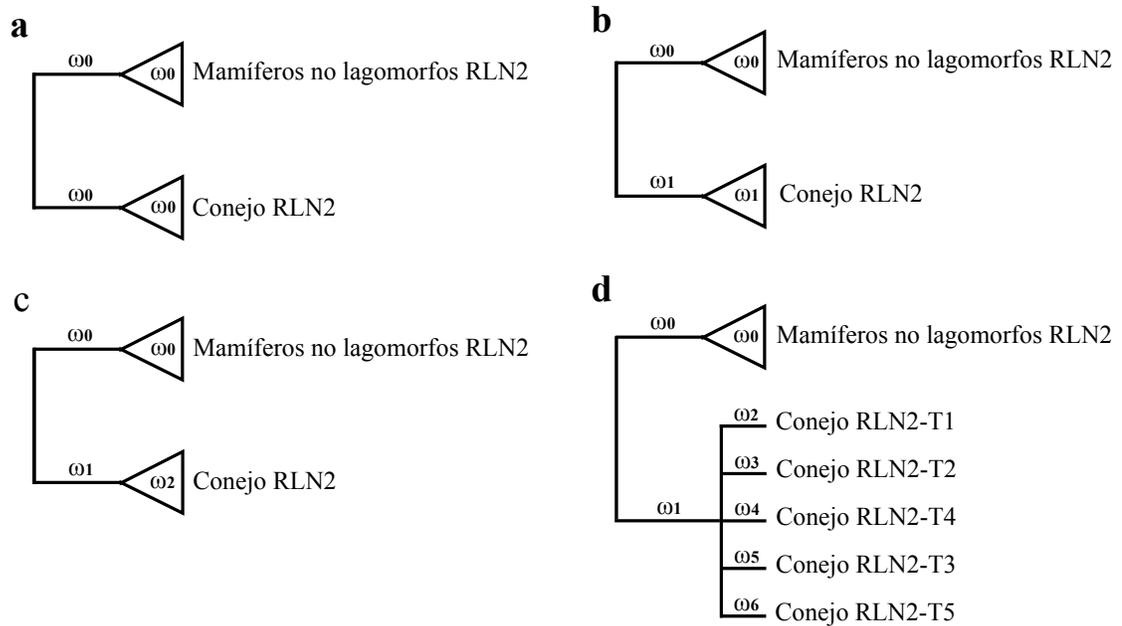


Figura 15. Representación gráfica de los modelos anidados implementados en la prueba de rama. (a) en el modelo de 1ω se estimó un valor de omega para todos los linajes de mamíferos (ω_0). **(b)** en el modelo de 2ω se estimó un valor de omega para los linajes de mamíferos no lagomorfos (ω_0), y un valor de omega para el clado del conejo (rama ancestral y descendientes) (ω_1). **(c)** en el modelo de 3ω , se estimó un valor de omega para para los mamíferos no lagomorfos (ω_0), para la rama ancestral del clado del conejo (ω_1) y para las ramas descendientes (ω_2). **(d)** finalmente, en el modelo de 7ω se estimó un valor de omega para los mamíferos no lagomorfos (ω_0), para la rama ancestral del clado del conejo (ω_1) y un valor de omega para cada una de las ramas descendientes (ω_2 , ω_3 , ω_4 , ω_5 y ω_6).

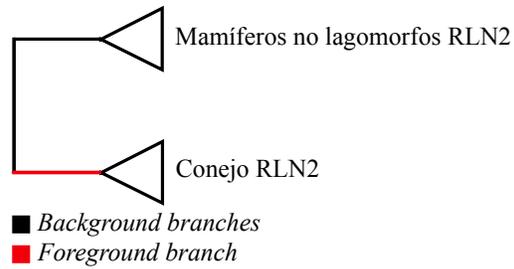


Figura 16. Representación grafica del modelo implementado en la prueba de rama-sitio. Las ramas de mamíferos no lagomorfos fueron consideradas como background branches mientras que la rama del conejo como foreground branch.

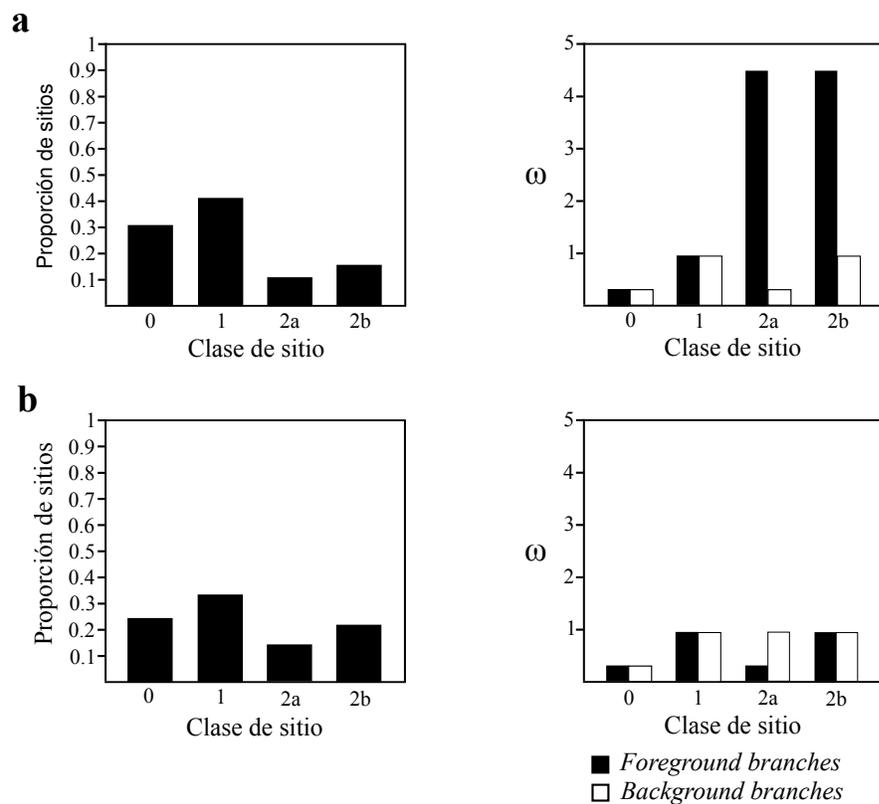


Figura 17. Proporción de sitios en cada clase y omega para cada clase del gen RLN2 del conejo. (a) Modelo A, (b) Modelo A neutral, $\omega_2=1$.

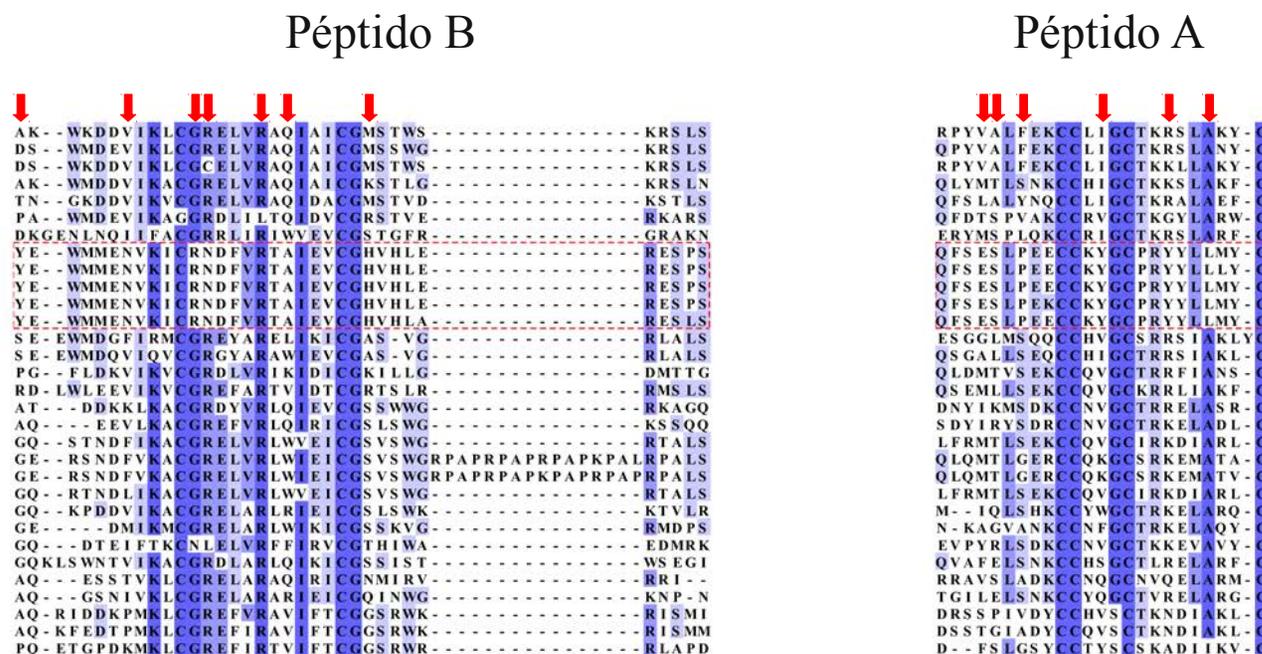


Figura 18. Sitios bajo selección en los péptidos B y A de la proteína codificada por el gen RLN2 del conejo. El alineamiento múltiple de secuencias de los péptidos muestra todas las secuencias de mamíferos usados en el análisis. El color representa el grado de conservación del sitio. Los genes RLN2 del conejo se delimitan por las líneas rojas punteadas. Los sitios seleccionados en los péptidos B y A se indican con una flecha.

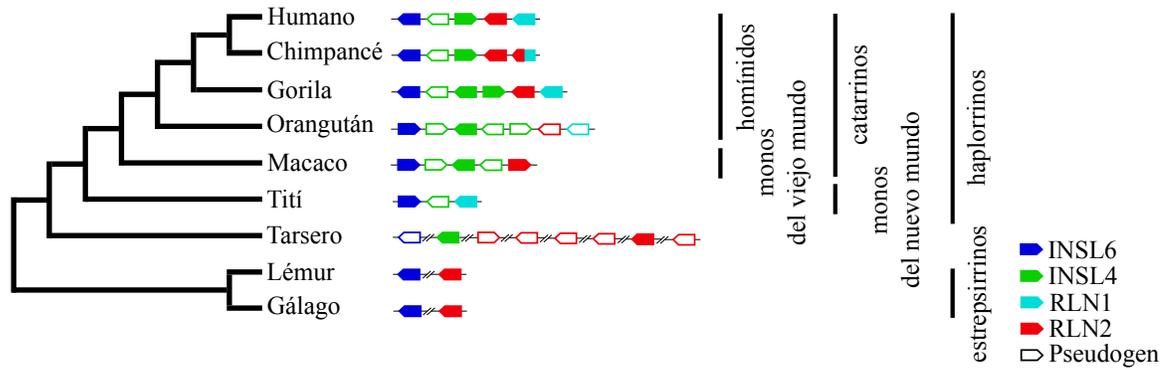


Figura 19. Estructura genómica del cluster RFLB en primates. Las líneas transversales indican que los genes fueron identificados en diferentes fragmentos genómicos.

a

```

Humano RLN1      MPRLFLFHLLLEFCLLLNQFSRAVAAKWKDDVIKLCGRELVRAQIAICGMSTWSKRSLSQE 60
Chimpancé RLN1  MPRLFFFHLLGVCLLLNQFSRAVADSWMDEVIKLCGRELVRAQIAICGMSSWGKRSLSQE 60
                *****:*** .***** . * *:*****:*.*****

Humano RLN1      DAPQTPRPVAEIVPSFINKDTETIIIMLEFIANLPPELKAALSERQPSLPPELQQYVPALK 120
Chimpancé RLN1  DAPQTPRPVAEIVPSFINKDTETIIIMLEFIANLPPELKAALSERQPSLPPELQQYVPALK 120
                *****

Humano RLN1      DNSLSFEEFKLIRNRQSEAADSNPSELKYLGLDTHSQKRRPYVALFEKCLIGCTKRS 180
Chimpancé RLN1  DNSLSFEEFKLIRNRQSEAADSNPSELKYLGLDTHSQKRRQPYVALFEKCLIGCTKRS 180
                *****:*****

Humano RLN1      LAKYC 185
Chimpancé RLN1  LANYC 185
                **:**

```

b

```

Humano RLN2      MPRLFFFHLLGVCLLLNQFSRAVADSWMDEVIKLCGRELVRAQIAICGMSTWSKRSLSQE 60
Chimpancé RLN2  MPRLFFFHLLGVCLLLNQFSRAVADSWMDEVIKLCGRELVRAQIAICGMSSWGKRSLSQE 60
                *****:*****:*.*****

Humano RLN2      DAPQTPRPVAEIVPSFINKDTETINMMSEFVANLPQELKLTLSMQPALPQLQQHVPVLK 120
Chimpancé RLN2  DAPQTPRPVAEIVPSFINKDTETINMMSEFVANLPQELKLTLSMQPALPQLQQYVPVLK 120
                *****:*****

Humano RLN2      DSSLLFEEFKLIRNRQSEAADSSPSELKYLGLDTHSRKKRQLYSALANKCCHVGCTKRS 180
Chimpancé RLN2  DSSLLFEEFKLIRNRQSEAADSSPSELKYLGLDTHSRKKRQLYSALANKCCHVGCTKRS 180
                *****

Humano RLN2      LARFC 185
Chimpancé RLN2  LARFC 185
                *****

```

Figura 22. Alineamiento de secuencias de los genes RLN1 y RLN2 de humano y chimpancé.

(a) RLN1, (b) RLN2. El color rojo representa el primer exón y el negro el segundo. El motivo DSWMDE (indicado en el recuadro) está presente en el primer exón del gen RLN2 en homínidos, sin embargo en el chimpancé este motivo también está presente en el primer exón del gen RLN1.

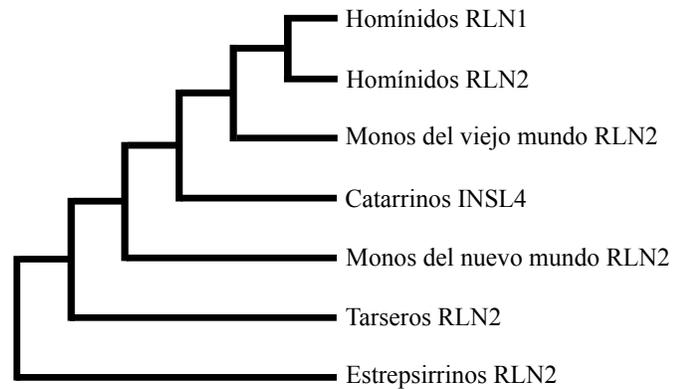


Figura 23. Topología del árbol de los genes INSL4, RLN2 y RLN1 en primates de acuerdo al modelo aceptado actualmente. De acuerdo con este modelo el gen INSL4 sería una innovación de los catarrinos, por lo que el clado INSL4 de los catarrinos debería recuperarse como grupo hermano de los genes RLN de los catarrinos (monos del viejo mundo y homínidos). Por otro lado, el gen RLN1 sería una innovación de los homínidos, por lo que el clado RLN1 de los homínidos debería recuperarse como grupo hermano del clado RLN2 del mismo grupo, mientras que el gen RLN2 de monos del viejo mundo debería recuperarse como grupo hermano de ambos grupos (Bièche et al. 2003; Klonish et al. 2001).

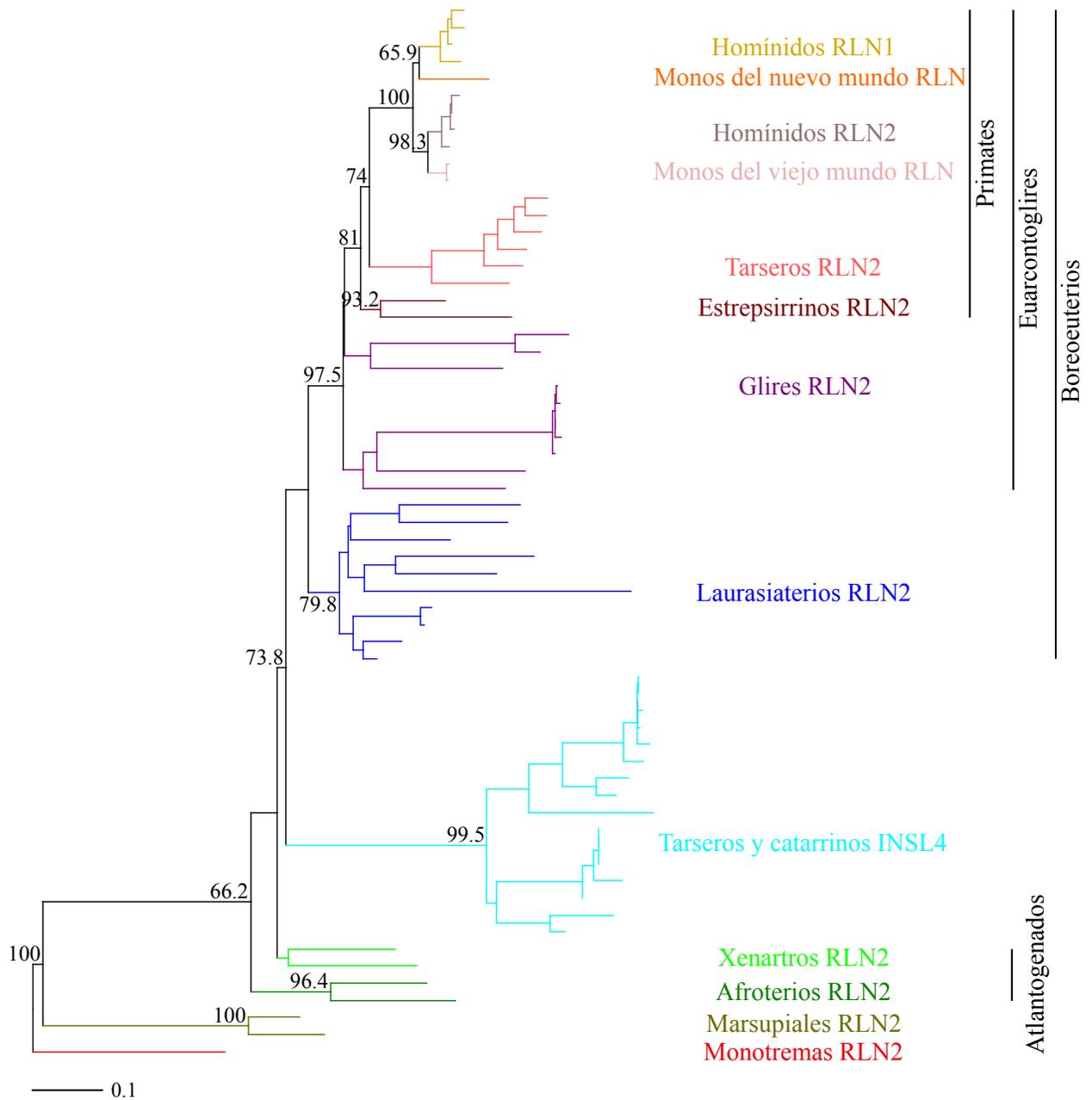


Figura 24. Filograma de máxima verosimilitud de los genes INSL4, RLN2 y RLN1 del cluster RFLB en mamíferos. Los valores sobre los nodos corresponden al soporte bootstrap.

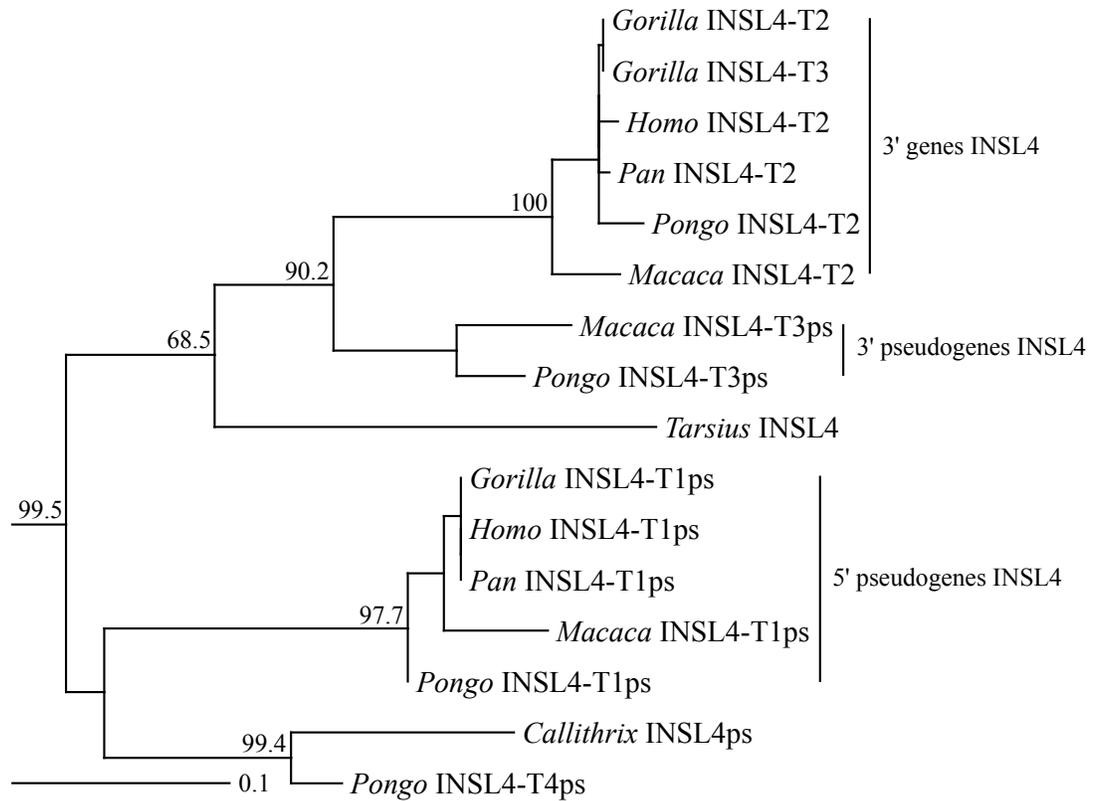


Figura 25. Filograma de maxima verosimilitud de los genes INSL4 en primates haplorrinos. Los valores sobre los nodos corresponden al soporte bootstrap.

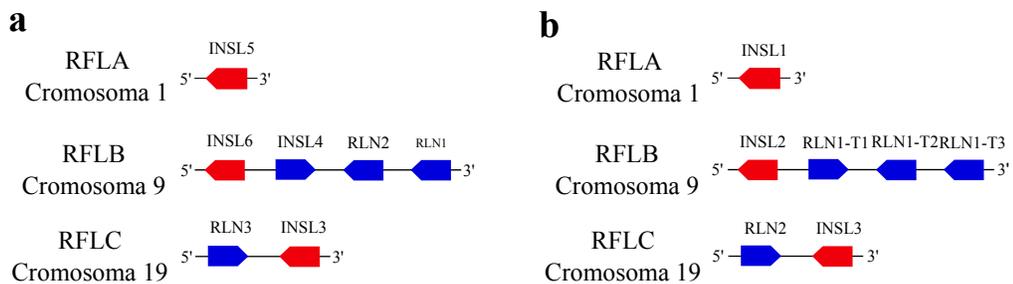


Figura 26. Nomenclatura de los genes de la familia mutigénica de las relaxinas. (a) nomenclatura actual, (b) nomenclatura sugerida. Las relaxinas se grafican en color azul y los péptidos de Leydig en color rojo.

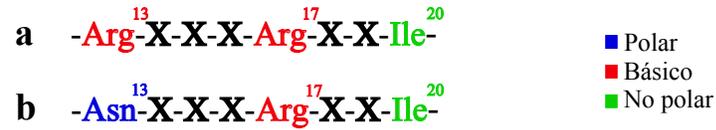


Figura 27. Motivo de interacción ligando-receptor del gen RLN2. (a) motivo en la mayoría de los mamíferos **(b)** motivo en el conejo. La sustitución de una arginina por una asparagina cambia las propiedades fisicoquímicas del residuo.

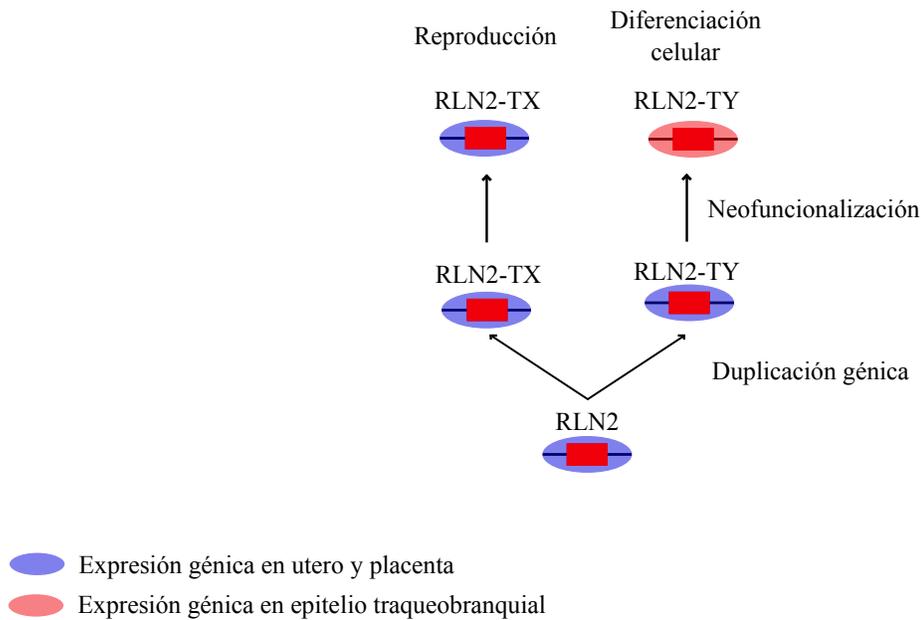


Figura 28. Neofuncionalización del gen RLN2 en el conejo. De acuerdo con esta hipótesis, una vez ocurrida la duplicación del gen RLN2, un(os) gen(es) habrían seguido cumpliendo un rol en el sistema reproductor (RLN-TX), mientras que otro(s) gen(es) habría(n) neofuncionalizado para cumplir un rol en la diferenciación celular epitelial (RLN-TY). Esta hipótesis se sugiere a partir de los estudios *in vivo* de la expresión génica del INSL4 en esta especie, los cuales sugieren que este gen se expresa diferencialmente en el útero y placenta, y también en el epitelio traqueobranquial (Fields et al. 1995; Jetten et al. 1999).

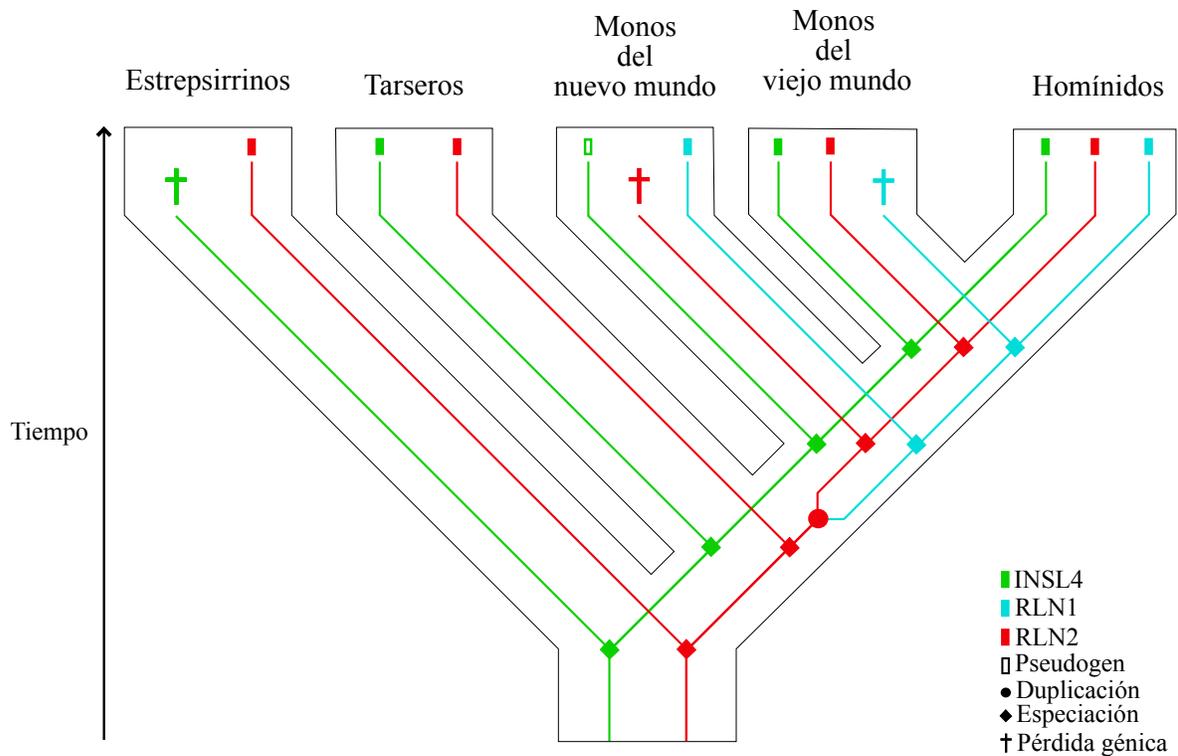


Figura 29. Evolución del cluster RFLB en primates. De acuerdo con este modelo el gen INSL4 se habría originado antes de la radiación de los primates, sin embargo habría sido retenido por tarseros, monos del viejo mundo y homínidos. Por otro lado, la duplicación del gen RLN habría ocurrido en el ancestro común de los antropoideos, sin embargo el gen del extremo 5' se habría perdido en monos del nuevo mundo y habría sido retenido en monos del viejo mundo, mientras que el gen del extremo 3' habría sido retenido en monos del nuevo mundo y se habría perdido en monos del viejo mundo. En homínidos se habrían retenido ambos genes RLN.

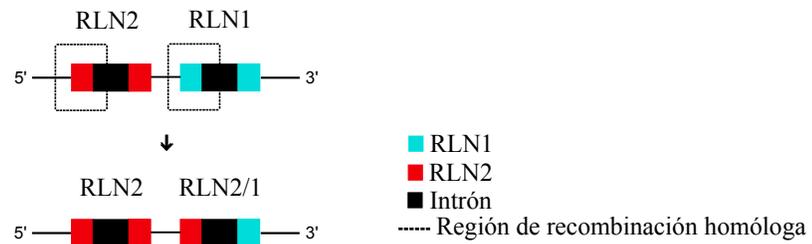


Figura 30. Evolución en mosaico de los genes RLN1 y RLN2 en el chimpancé. La región de recombinación homóloga abarcaría sólo el primer exón, por lo que el primer exón de ambos genes evolucionaría de manera concertada, mientras que el segundo exón de estos genes evolucionaría de manera divergente.

ANEXO A

A.1. Secuencias de ADN

En la tabla A1 se resumen los números de acceso de los genes usado en este trabajo.

Tabla A1. Números de acceso de los genes de la familia multigénica de las relaxinas usadas en el estudio.

Sub clase	Super orden	Orden	Nombre común	Nombre científico	INSL5	
Eutheria	Euarchontoglires	Primates	Humano	<i>Homo sapiens</i>	ENSG00000172410	
			Chimpancé	<i>Pan troglodytes</i>	ENSPTRG00000000835	
			Gorila	<i>Gorilla gorilla</i>	ENSGGOG00000027684	
			Orangután	<i>Pongo pygmaeus</i>	ENSPPYG00000001263	
			Macaco	<i>Macaca mulatta</i>	ENSMMUG00000000640	
			Babuino	<i>Papio hamadryas</i>	Contig50831:4663-7819	
			Tití	<i>Callithrix jacchus</i>	ENSCJAG00000017226	
			Tarsero	<i>Tarsius syrichta</i>	ENSTSYG00000007415	
			Gálago	<i>Otolemur garnettii</i>	ENSOGAG00000006628	
			Lemur	<i>Microcebus murinus</i>	ENSMICG00000004436	
		Lagomorpha	Conejo	<i>Oryctolagus cuniculus</i>	ENSOCUG00000004666	
			Pica	<i>Ochotona princeps</i>	*	
		Rodentia	Rodentia	Cuy	<i>Cavia porcellus</i>	ENSCPOG00000011735
				Ratón	<i>Mus musculus</i>	ENSMUSG00000066090
Rata	<i>Rattus norvegicus</i>			ENSRNOG00000037916		

		Rata canguro	<i>Dipodomys ordii</i>	*	
		Ardilla	<i>Spermophilus tridecemlineatus</i>	ENSSTOG00000016330	
Laurasiatheria	Soricomorpha	Musaraña	<i>Sorex araneus</i>	*	
	Erinaceomorpha	Erizo de tierra	<i>Erinaceus europaeus</i>	*	
	Cetacea	Delfín	<i>Tursiops truncatus</i>	ENSTTRG00000009424	
	Artiodactyla	Vaca	<i>Bos taurus</i>	ENSBTAG00000003850	
		Cerdo	<i>Sus scrofa</i>	*	
		Camello	<i>Camelus dromedarius</i>	*	
		Alpaca	<i>Vicugna pacos</i>	ENSVPAG00000000916	
	Chiroptera	Muriciélago	<i>Pteropus vampyrus</i>	ENSPVAG00000017649	
		Muriciélago	<i>Myotis lucifugus</i>	*	
	Perisodactyla	Caballo	<i>Equus caballus</i>	ENSECAG00000024174	
	Carnivora	Perro	<i>Canis familiaris</i>	*	
		Gato	<i>Felis catus</i>	ENSFCAG00000005503	
Xenarthra	Pilosa	Perezoso	<i>Choloepus hoffmanni</i>	ENSCHOG00000009863	
	Cingulata	Armadillo	<i>Dasyus novemcinctus</i>	ENSDNOG00000015104	
Afrotheria	Afrosoricida	Tenrec	<i>Echinops telfairi</i>	ENSETEG00000006593	
	Hyracoidea	Damán	<i>Procavia capensis</i>	ENSPCAG00000010668	
	Proboscidea	Elefante	<i>Loxodonta africana</i>	ENSLAFG00000015826	
Metatheria	Australidelphia	Diprotodontia	Walabí	<i>Macropus eugenii</i>	ENSMEUG00000014118
	Ameridelphia	Didelphimorphia	Zarigüeya	<i>Monodelphis domestica</i>	*
Prototheria	Monotremata	Ornitorrinco	<i>Ornithorhynchus anatinus</i>	*	

INSL6	INSL4	RLN2	RLN1	RLN3	INSL3
5163863-5339873	5163863-5339873	5163863-5339873	5163863-5339873	ENSPTRG00000010576	NC_000019
5282897-5450516	5282897-5450516	NM_001122774	NM_001118912	NC_006486	NC_006486
5187934-5390917	5187934-5390917	5187934-5390917	5187934-5390917	Chr19:14337468-14340227	Chr19:18224268-18224463
57547568-57803411	57547568-57803411	57547568-57803411	57547568-57803411	ENSPPYG00000009632	Chr19:18260412-18254035
ENSMMUG00000010856	71832351-72032294	71832351-72032294	-	ENSMMUG00000013537	ENSMMUG00000013537
71832351-72032294					
ENSCJAG00000011358	106904656-107050272		106904656-107050272	ENSCJAG00000004965	ENSCJAG00000033575
ENSTSYG00000003808	ENSTSYG00000007147	ENSTSYG00000013076 ENSTSYG00000006242 ENSTSYG00000001306 ENSTSYG00000000722 ENSTSYG00000006196 ENSTSYG00000007074	-	*	*
ENSOGAG00000012255	-	ENSOGAG00000009500	-	ENSOGAG00000002457	*
*	-	ENSMICG0000000426	-		*
1062025-1256369	-	1062025-1256369	-	ENSOCUG00000013451	*
ENSOPRG00000000023	-		-	ENSOPRG00000007508	*
ENSCPOG00000002694	-	ENSCPOG00000001365	-	ENSCPOG00000013564	ENSCPOG00000019362
ENSMUSG000000050957	-	ENSMUSG00000039097	-	ENSMUSG00000045232	ENSMUSG00000079019
ENSRNOG00000015868	-		-	ENSRNOG00000005911	ENSRNOG00000018757
ENSDORG00000003743	-	ENSDORG00000000108	-	ENSDORG00000013264	*
ENSSTOG00000016298	-	ENSSTOG00000012644	-	*	*
ENSSARG00000004569	-	ENSSARG00000004245 ENSSARG00000013075 ENSSARG00000010511	-	ENSSARG00000011917	*
ENSEEUG00000015434	-		-	ENSEEUG00000012184	*
*	-	ENSTTRG00000011584	-	*	*
ENSBTAG00000006651	-		-	ENSBTAG00000038437	ENSBTAG00000025775
ENSSSCG00000005214	-	ENSSSCG00000005216 ENSSSCG00000005213	-	ENSSSCG00000013765	ENSSSCG00000013887
*	-		-	*	*

ENSVPAG0000009194	-	ENSVPAG0000008523 ENSVPAG0000008535	-	*	*
ENSPVAG0000000251	-	ENSPVAG0000000254	-	ENSPVAG00000017551	ENSPVAG00000024188
ENSMLUG0000007982	-	ENSMLUG0000009569	-	ENSMLUG00000013013	*
*	-	ENSECAG00000013020	-	ENSECAG00000014897	ENSECAG00000016450
ENSCAFG00000002113	-	ENSCAFG00000002115	-		ENSCAFG00000015187
ENSFCAG00000005197	-	ENSFCAG00000011471	-	ENSFCAG00000009229 ENSFCAG00000003722	*
*	-	ENSCHOG00000013606	-	*	ENSCHOG00000005444
ENSDNOG00000001668	-	ENSDNOG00000025594 ENSDNOG00000025720	-	ENSDNOG00000012940	*
ENSETEG00000006684	-		-	*	*
*	-	ENSPCAG00000013639	-	*	ENSPCAG00000018886
ENSLAFG00000007479	-	ENSLAFG00000017411	-	*	ENSLAFG00000025675
-	-		-	ENSMEUG00000009042	*
-	-		-	*	*

*gaps en las secuencias genómicas, - los genes no están presentes en estas especies.

BIBLIOGRAFÍA

Aguileta, G., Bielawski, J.P. & Yang, Z. (2006) . Proposed standard nomenclature for the α - and β -globin gene families. *Genes Genet. Syst.* 81:367-371.

Bathgate, R. A., Samuel, C. S., Burazin, T. C., Layfield, S., Claasz, A. A., Reytomas, I. G., Dawson, N. F., Zhao, C., Bond, C., Summers, R. J., Parry, L. J., & Wade, J. D., et al. (2002). Human relaxin gene 3 (H3) and the equivalent mouse relaxin (M3) gene. Novel members of the relaxin peptide family. *J. Biol. Chem.* 277:1148–1157.

Bathgate, R. A., C. S. Samuel, T. C. Burazin, A. L. Gundlach, & G. W. Tregear. (2003). Relaxin: new peptides, receptors and novel actions. *Trends Endocrinol. Metab.* 14:207-213.

Bièche, I., A. Laurent, I. Laurendeau, L. Duret, Y. Giovangrandi, J. L.Frendo, M. Olivi, J. L. Fausser, D. Evain-Brion, & M. Vidaud. (2003). Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. *Biol. Reprod.* 68:1422–1429.

Blundell, T. L., & R. E. Humbel.(1980). Hormone families: pancreatic hormones and homologous growth factors. *Nature* 287:781-787.

Carson, A.R. & Scherer, S.W. (2009). Identifying concerted evolution and gene conversion in mammalian gene pairs lasting over 100 million years. *BMC Evol. Biol.* 9:156.

Chan, J. C., & D. F. Steiner.(2000). Insulin through the ages: phylogeny of the growth promoting and metabolic regulatory hormone. *American Zoologist* 40:213-222.

Chan, S. J., Q. P. Cao, & D. F. Steiner.(1990). Evolution of the insulin superfamily: cloning of a hybrid insulin/insulin-like growth factor cDNA from amphioxus. *PNAS* 87:9319-9323.

Chassin, D., Laurent, A., Janneau, J. L., Berger, R., & Bellet, D. (1995). Cloning of a new member of the insulin gene superfamily (INSL4) expressed in human placenta. *Genomics* 29:465–470.

Conlon, J. M.(2001). Evolution of the insulin molecule: insights into structure-activity and phylogenetic relationships. *Peptides* 22:1183-1193.

Dehal, P. & Boore, J.L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3(10): e314.

Evans, B.A., Fu, P. & Tregear, W. (1994). Characterization of two relaxin genes in the Chimpanzee. *Journal of Endocrinology* 140:385-392.

Fields, P., Kondo, S., Tashima, L., Bryant-Greenwood, G. & Greenwood, F. (1995). Expression of SQ10 (a Preprorelaxin-like gene) in the pregnant rabbit placenta and uterus. *Biology of Reproduction* 53: 1139-1145.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y. & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545.

Gogarten, J.P. & Olendzenski, L. (1999). Orthologs, paralogs and genome comparisons. *Curr Opin Genet. Dev.* 9:630-636.

Golding, G.B. & Dean, A.M. (1998). The structural basis of molecular adaptation. *Mol. Biol. Evol.* 15:335-369.

Good- Avila, S.V., Yegorov, S., Harron, S., Bogerd, J., Glen, P., Ozon, J. & Wilson, B.C. (2009). Relaxin gene family in teleosts: phylogeny, syntenic mapping, selective constraint, and expression analysis. *BMC Evolutionary Biology* 9:293.

Haldane, J.B.S. (1932). *The causes of evolution*. Longmans and Green. London.

Hardison, R.(1988). Hemoglobin from bacteria to man: Evolution of different patterns of gene expression. *J. Exp. Biol.* 201:1099-1117.

He, X. & Zhang, J. (2004). Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in gene duplicate gene evolution. *Genetics* 169:1157-1164.

Hudson, P., Haley, J., John, M., Cronk, M., Mcrawford, R., Haralambidis, J., Tregear, G., Shine, J. & Niall H (1983). Structure of a genomic clone encoding biologically active human relaxin. *Nature* 301:628-631.

Hughes, A.L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* 256:119 –124.

Hughes, A.L. & Nei, M. (1989). Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. *Mol. Biol. and Evol.* 6:559–79.

Hsu, S. Y. (2003). New insights into the evolution of the relaxin-LGR signaling system. *Trends Endocrinol. Metab.* 14:303-309.

Hsu, S. Y., Nakabayashi, K., Nishi, S., Kumagi, J., Kudo, M., Sherwood, O. D. & Hseuh, A. J. W. (2002). Activation of Orphan Receptors by the Hormone Relaxin. *Science* 295, 671-674.

Hsu, S. Y., Semyonov, J., Park, J. I., & Chang, C. L. (2005). Evolution of the signaling system in relaxin-family peptides. *Ann. N.Y. Acad. Sci.* 1041, 520–529.

Jetten, A.M., Bernacki, S.H., Floyd, E.E., Saunders, N.A., Pieniazek, J. & Lotan, R. (1992). Expression of preprorelaxin-like gene during squamous differentiation of rabbit tracheobronchial epithelial cells and its suppression by retinoic acid. *Cell Growth and Differentiation* 3:549-556.

Jobb, G. (2008). TREEFINDER version of October 2008. Munich, Germany. Distributed by the author at www.treefinder.de.

Kasahara (2007). The 2R hypothesis: an update. *Curr. Opin. in Immunol.* 19: 547–552.

Katoh, T. (2008). Recent developments in the MAFFT multiple sequence alignment program.

Briefings in Bioinformatics 9:286-298.

Klonisch, T., Froehlich, C., Tetens, F., Fischer, B. & Hombach-Klonisch, S. (2001). Molecular remodeling of members of the relaxin family during primate evolution. *Mol. Biol. Evol.* 18:393-403.

Kuraku, S., Meyer, A., Kuratani, S. (2008). Timing of Genome Duplications Relative to the Origin of the Vertebrates: Did Cyclostomes Diverge before or after?. *Mol. Biol. Evol.* 26:47-59.

Madsen, O. (2009) Mammals (Mammalia). *The Timetree of Life*. Oxford University Press, pp. 459–461.

Muller, H.J. (1935). The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. *Genetics* 17:237-252.

Murphy, W.J. & Eizirik, E. (2009) Placental mammals (Eutheria). *The Timetree of Life*. Oxford University Press, pp. 471–474.

Nei, M., Gu, X. & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *PNAS* 94:7799–806.

Nei, M., Rogozin, I.B., & Piontkivska, H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *PNAS* 97:10866–10871.

Nei, M. & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 39:121-152.

Nielsen, R., & Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.

Ohno, S. (1970). *Evolution by gene duplication*. Springer-Verlag.

Olinski, R. P., L. G. Lundin, & F. Hallbook. (2005). Genome duplication-driven evolution of gene families: insights from the formation of the insulin family. *Ann. N. Y. Acad. Sci.*1040:426-428.

Olinski, R. P., L. G. Lundin, & F. Hallbook.(2006). Conserved synteny between the Ciona genome and human paralogs identifies large duplication events in the molecular evolution of

the insulin-relaxin gene family. *Mol. Biol. Evol.* 23:10-22.

Opazo, J.C., F.G., Hoffmann & J.F., Storz (2008). Differential loss of embryonic globin genes during the radiation of placental mammals. *PNAS* 105:12950-12955.

Opazo, J.C., Sloan, A.M., Campbell, K.L. & Storz, J.F. (2009). Origin and ascendancy of a chimeric fusion gene: the β/δ globin gene in paenungulate mammals. *Mol. Biol. Evol.* 26:1469-1478.

Ota, T. & Nei, M. (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin V_H gene family. *Mol. Biol. Evol.* 11:469-482.

Park, J-II, Semyonov, J., Chang, J.C., Yi, W., Warren, W. & Hsu, S.Y.T. (2008). Origin of INSL3-mediated testicular descent in therian mammals. *Genome Research* 18:974-985.

Piontkivska, H., Rooney, A. P. & Nei, M. (2002). Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol. Biol. Evol.* 19:689–697.

Rodríguez-Trelles, F., Tarrío, R. & Ayala, F.J. (2003). Convergent neofunctionalization by

positive Darwinian selection after ancient recurrent duplications of the xanthine dehydrogenase gene. PNAS 100:13413-13417.

Ronquist, F. & Huelsenbeck, J.P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

Shabanpoor F, Separovic F & Wade J. (2009). The human insulin superfamily of polypeptide hormones. *Vitamins and Hormones* 80:1-22.

Sherwood, O. D. 2004. Relaxin's physiological roles and other diverse actions. *Endocr. Rev.* 25:205-234.

Steiner, D. F., S. J. Chan, J. M. Welsh, & S. C. Kwok.(1985). Structure and evolution of the insulin gene. *Annu. Rev. Genet.* 19:463-484.

Suzuki, Y., T. Gojobori, & M. Nei. (2001). ADAPTSITE: detecting natural selection at single amino acid sites. *Bioinformatics* 17:660–661.

Suzuki, Y., & M. Nei. (2004). False-positive selection identified by ML-based methods:

examples from the Sig1 gene of the diatom *Thalassiosira weissflogii* and the tax gene of a human T-cell lymphotropic virus. *Mol. Biol. Evol.* 21:914–921.

Suzuki, Y., & T. Gojobori. (1999). A method for detecting positive selection at single amino acid sites. *Mol. Biol. Evol.* 16:1315–1328.

Southey, B.R., Rodriguez-Zas, S.L. & Sweedler, J.V. (2009). Characterization of the prohormone complement in cattle using genomic libraries and cleavage prediction approaches. *BMC Genomics* 10:228.

Springer, M.S. & Krajewski, C.W. (2009) Monotremes (Prototheria). *The Timetree of Life*. Oxford University Press, pp. 462–465.

Springer, M.S., Krajewski, C.W. & Meredith, R.W. (2009) Marsupials (Metatheria). *The Timetree of Life*. Oxford University Press, pp. 466–470.

Steiner, D. F., S. J. Chan, J. M. Welsh, & S. C. Kwok. 1985. Structure and evolution of the insulin gene. *Annu Rev Genet* 19:463-484.

Steiper, M.E., & Young, N.M. (2009) Primates (Primates). *The Timetree of Life*. Oxford

University Press, pp. 482-486.

Thornton, J.W. & DeSalle, R. (2000). Gene family evolution and homology: genomics meets phylogenetics. *Annu. Rev. Genomics Hum. Genet.* 01:41–73.

Wain HM Bruford EA Lovering RC Lush MJ Wright MW Povey S. (2002). Guidelines for human gene nomenclature. *Genomics* 79:4.

Wilkinson, T. N., Speed, T. P., Tregear, G. W., & Bathgate, R. A. (2005). Evolution of the relaxin-like peptide family. *BMC Evol. Biol.* 5:14.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* 13:555–556.

Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15:568-573.

Yang, Z. (2007). PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586-1591.

Yang, Z. & R. Nielsen. (1998). Synonymous and nonsynonymous rate variation in nuclear genes

in mammals. *Journal of Molecular Evolution* 46:409-418.

Yang, Z. & R. Nielsen. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17:32-43.

Yang, Z. & R. Nielsen. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* 19:908-917.

Yang, Z., R. Nielsen, N. Goldman, & A.-M. K. Pedersen. (2000). Codon-substitution models for heterogeneous selection pressure at amino acids. *Genetics* 155:431-449.

Yang, Z., W.S.W., Wong & R. Nielsen. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* 22:1107-1118.

Yokoyama, S. & Takenata, N. (2005). Statistical and molecular analyses of evolutionary significance of red-green color vision and color blindness in vertebrates. *Mol. Biol. Evol.* 22:968-975.

Yonekura, H., Nata, K., Watanabe, T., Kurashina, Y., Yamamoto, H. & Okamoto, H. (1988). Mosaic evolution of the prepropancreatic polypeptide. *The Journal of Biological Chemistry*

263:2990-2997.

Zhang, J., Dyer, K.D. & Rosemberg, H.R. (2000). Evolution of the rodent eosinophil-associated Rnase gene family by rapid gene sorting and positive selection. PNAS 97:4071-4076.

Zhang, J. (2003). Evolution by gene duplication: un update. Trends in Ecology and Evolution 18:292-298.

Zhang, J., R. Nielsen & Z. Yang (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22:2472-2479.

Zhang, Z., Schwartz, S., Wagner, L. & Miller, W (2000). A greedy algorithm for aligning DNA sequences. J Comput Biol 7: 203-214.