

# Evolution of the Relaxin/Insulin-like Gene Family in Placental Mammals: Implications for Its Early Evolution

Federico G. Hoffmann · Juan C. Opazo

Received: 16 June 2010 / Accepted: 25 October 2010 / Published online: 17 November 2010  
© Springer Science+Business Media, LLC 2010

**Abstract** The relaxin (RLN) and insulin-like (INSL) gene family is a group of genes involved in a variety of physiological roles that includes bone formation, testicular descent, trophoblast development, and cell differentiation. This family appears to have expanded in vertebrates relative to non-vertebrate chordates, but the relative contribution of whole genome duplications (WGDs) and tandem duplications to the observed diversity of genes is still an open question. Results from our comparative analyses favor a model of divergence post vertebrate WGDs in which a single-copy progenitor found in the last common ancestor of vertebrates experienced two rounds of WGDs before the functional differentiation that gave rise to the RLN and INSL genes. One of the resulting paralogs was subsequently lost, resulting in three proto-RLN/INSL genes on three separate chromosomes. Subsequent rounds of tandem gene duplication and divergence originated the set of paralogs found on a given cluster in extant vertebrates. Our study supports the hypothesis that differentiation of the RLN and INSL genes took place independently

in each RLN/INSL cluster after the two WGDs during the evolutionary history of vertebrates. In addition, we show that INSL4 represents a relatively old gene that has been apparently lost independently in all Euarchontoglires other than apes and Old World monkeys, and that RLN2 derives from an ape-specific duplication.

**Keywords** Gene family evolution · Relaxin · Primates · Gene duplication · Whole genome duplication · Genome evolution

## Introduction

Gene duplication has been recognized as an important source of raw material for biological innovations (Ohno 1970). Following gene duplication, the functional and regulatory divergence of the different copies of an ancestral gene may contribute to the generation of evolutionary novelties (Hahn 2009; Lynch 2007; Nei and Rooney 2005; Ohno 1970; Zhang 2003). The relaxins and insulin-like genes provide an interesting example of the evolutionary versatility that can be generated by the differentiation of duplicated genes. In this gene family, the different paralogs have acquired a variety of physiological roles including bone formation, testicular descent, trophoblast development, and cell differentiation (Bathgate et al. 2003; de Pablo and de la Rosa 1995; Reinecke and Collet 1998; Sherwood 2004).

The insulin–relaxin gene superfamily comprises four different kinds of genes: insulin (INS), insulin growth factors (IGFs), relaxins (RLNs), and insulin-like peptides (INSLs) (Chan and Steiner 2000; Nagamatsu et al. 1991; Olinski et al. 2006a; Park et al. 2008a). The number and nature of genes in the insulin–relaxin superfamily in a

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-010-9403-6) contains supplementary material, which is available to authorized users.

---

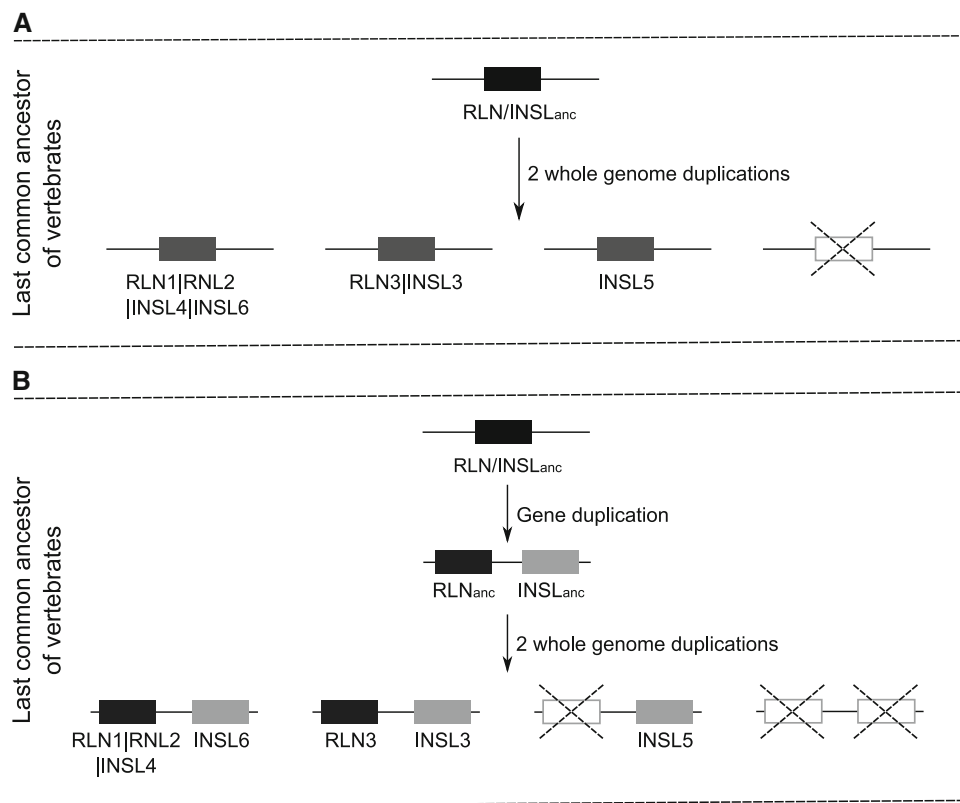
F. G. Hoffmann  
School of Biological Sciences, University of Nebraska,  
Lincoln, NE 68588, USA

F. G. Hoffmann  
Department of Biochemistry and Molecular Biology, Mississippi  
State University, Box 9650, Mississippi State, MS 39762, USA

J. C. Opazo (✉)  
Instituto de Ecología y Evolución, Facultad de Ciencias,  
Universidad Austral de Chile, Casilla 567, Valdivia, Chile  
e-mail: jopazo@gmail.com

species is variable. For example, humans possess a single INS ortholog, 2 IGFs, 3 RLNs, and 4 INSLs in their genomes, whereas mouse possess 2 INS paralogs, 2 IGFs, 2 RLNs, and 3 INSLs (Park et al. 2008a; Wentworth et al. 1986). These genes have been placed into a single superfamily because of structural similarities and the presence of six conserved cysteine residues (Chan and Steiner 2000). The insulin–relaxin superfamily was further classified into two separate families, with INS and IGFs in the first, and RLNs and INSLs in the second. The RLNs and INSLs genes are located in separate clusters, and synteny analyses have shown that these clusters were established early in vertebrate evolution (Olinski et al. 2006a, b; Park et al. 2008a, b). For example, in humans, one cluster includes the RLN1, RLN2, INSL4, and INSL6 genes and is located on chromosome 9, a second cluster includes the RLN3 and INSL3 genes and is located on chromosome 19, and the third cluster includes the INSL5 gene and is located on chromosome 1 (Park et al. 2008a).

Vertebrate genomes have undergone two rounds of whole genome duplications (WGDs) prior to the divergence between cyclostomes and gnathostomes (Dehal and Boore 2005; Kuraku et al. 2009; Meyer and Schartl 1999; Ohno 1970), and these two WGDs have been postulated to be major determinants of the diversification of the insulin–relaxin gene superfamily (Olinski et al. 2006a, b; Park et al. 2008a). The accepted model of evolution of the RLN/INSL family posits that all extant members of this group derive from a single progenitor found in the common ancestor of vertebrates. Through the two successive rounds of WGDs, the single-copy proto-RLN/INSL gene would have given rise to three proto-RLN/INSL paralogs, located on three different chromosomes (Fig. 1a, Olinski et al. 2006b; Park et al. 2008a, b), plus a paralog that was secondarily lost. Subsequent rounds of tandem gene duplication and divergence of each of these three paralogs would eventually give rise to the different clusters found today in extant vertebrates (Hsu 2003; Olinski et al.



**Fig. 1** Alternative hypotheses regarding the evolutionary history of the RLN/INSL-like genes. **a** The divergence post-WGD model. Starting with a single-copy proto-RLN/INSL gene, two successive rounds of whole genome duplications originated three proto-RLN/INSL genes (one was lost) on three different chromosomes. After that, duplications of each of the resulting paralogs originated all of the other paralogs found on the same chromosome. According to this model, paralogs on a chromosome should form monophyletic groups.

**b** The divergence pre-WGD model. In this case, the ancestor of the RLN and INSL genes would have diverged prior to the two rounds of whole genome duplications. Here the two successive rounds of whole genome duplications would have generated the copies of the RLN–INSL ancestral arrangement. Under this scenario, the RLN and INSL genes would form reciprocally monophyletic clades. Our results favor the divergence post-WGD model depicted on **a**

2006b; Park et al. 2008a, b). Under this model, paralogs found on the same cluster should share a most recent common ancestor to the exclusion of paralogs on separate clusters. From a functional standpoint, this would imply that the functional divergence of the different RLN and INSL paralogs occurred after the two rounds of WGD (divergence post-WGD model, Fig. 1a). An alternative view to explain the presence of both RLN and INSL genes on different chromosomes could invoke a tandem duplication predating the WGDs. In this scenario, the ancestors of RLN and INSL genes would have already been present in the ancestral locus (Fig. 1b), and functional divergence would have preceded WGDs (divergence pre-WGD model, Fig. 1b). Thus, the major difference between the two competing models is in the timing of the functional differentiation relative to the two rounds of WGDs in the last common ancestor of extant vertebrates.

On a more recent time scale, tandem duplications have also played an important role in the evolution of the vertebrate RLN and INSL repertoire. Within mammals, for example, apes possess an additional RLN gene, RLN2, with no clear ortholog in any other vertebrate group, and a similar pattern can be observed for the INSL4 gene, which is only found in catarrhine primates, the group that includes apes and Old World monkeys (Bieche et al. 2003; Park et al. 2008a, b). Given that the data at hand suggests that the RLN2 and INSL4 genes derive from lineage-specific tandem duplications, we would expect them to nest within the corresponding primate clade in the corresponding phylogenies: the RLN2 clade should nest within apes sequences, and the INSL4 should nest within catarrhine sequences.

From a phylogenetic standpoint, the divergence post-WGD and divergence pre-WGD models make mutually exclusive topological predictions. For the divergence post-WGD model (Fig. 1a) we would expect paralogs on the same cluster to form monophyletic groups. By contrast, in the divergence pre-WGD model (Fig. 1b) we would expect RLN and INSL paralogs to cluster in separate clades regardless of their genomic location. Similarly, if the RLN2 and INSL4 genes were the result of lineage-specific tandem duplications, they would be expected to group with other paralogs of the corresponding lineage. Accordingly, the main goal of this study are (1) to compare the divergence post-WGD and divergence pre-WGD models, and (2) to assess the relative contribution of WGDs and tandem duplications to the observed diversity of RLN and INSL genes in vertebrates, two questions that were not directly addressed in the previous studies (Good-Avila et al. 2009; Olinski et al. 2006a, b; Park et al. 2008a, b).

## Materials and Methods

### Data

We selected placental mammals as a model system because they have the most diverse repertoire of RLN/INSL genes, and allow us to evaluate competing models explaining the origin of these genes, and address specific questions relative to the gain and loss of RLN/INSL genes in this group. Accordingly, the DNA sequences from structural genes in the RLN/INSL gene family of placental mammals were obtained from the Ensembl database (release 55). In each case, RLN/INSL-like genes were identified by comparing known exon sequences with genomic fragments using the program Blast2 version 2.2 (Tatusova and Madden 1999) available from NCBI (<http://www.ncbi.nlm.nih.gov/blast/bl2seq>). Sequences derived from shorter records based on genomic DNA or cDNA were also included in order to attain a broad and balanced taxonomic coverage of placental mammals (Supplementary Table 1).

To explore the sensitivity of our analyses to changes in the alignment method, nucleotide translated sequences were aligned using Dialign-TX (Subramanian et al. 2008), Kalign2 (Lassmann et al. 2009), the E-INS-i, G-INS-i, and L-INS-i strategies from Mafft v.6 (Katoh et al. 2009), MUSCLE v3.5 (Edgar 2004), Probcons (Do et al. 2005), and Tcoffee (Notredame et al. 2000). Nucleotide alignments were generated using the amino acid alignments as a template with the software PAL2NAL (Suyama et al. 2006). Finally, the biological accuracy of alignments was assessed using the software MUMSA (Lassmann and Sonnhammer 2005), which compares alignment blocks from different alignment strategies to assess the difficulty of an alignment case, and also ranks the quality of each alternative alignment. For each set of alignments of a given set of sequences, MUMSA provides an Average Overlap Score (AOS) which gives a measure of the alignment difficulty that ranges from 0 to 1, with 1 being the least difficult. In addition, it also assigns a Multiple Overlap Score (MOS) score to each of the different alignments, which also ranges from 0 to 1, with 1 being the highest quality.

### Phylogenetic Inference

Phylogenetic relationships among the different RLN/INSL-like DNA sequences in the dataset were estimated using Bayesian and maximum likelihood approaches, as implemented in Mr.Bayes v3.1.2 (Ronquist and Huelsenbeck 2003) and Treefinder version October 2008 (Jobb et al. 2004), respectively. The best fitting models were estimated separately for each gene segment, and also for each codon position within each segment using the “propose model”

routine from Treefinder version October 2008 (Supplementary material 2; Jobb et al. 2004). For the Bayesian analyses, two simultaneous independent runs were performed for  $30 \times 10^6$  iterations of a Markov Chain Monte Carlo algorithm, with five simultaneous chains, sampling every 1000 generations. Support for the nodes and parameter estimates were derived from a majority rule consensus of the last 15,000 trees sampled after convergence. In maximum likelihood, we estimated the best tree for each alignment, and support for the nodes was estimated with 1,000 bootstrap pseudoreplicates.

## Results and Discussion

### Alignment Accuracy

The AOS score for the alignments was 0.55, and the MOS scores for each individual alignment ranged from 0.577 to 0.689, with higher scores denoting higher quality. Based on these scores, we selected the four multiple alignments with the best MUMSA scores (L-INS-i, E-INS-i, G-INS-i, and Probcons), and compared the likelihood scores of the resulting trees. We then selected the tree with the highest likelihood score, which was obtained with the G-INS-i MAFFT alignment strategy, as our best tree. Results obtained with the other three alignment strategies are reported as Supplementary material 3.

### Phylogenetic Analysis

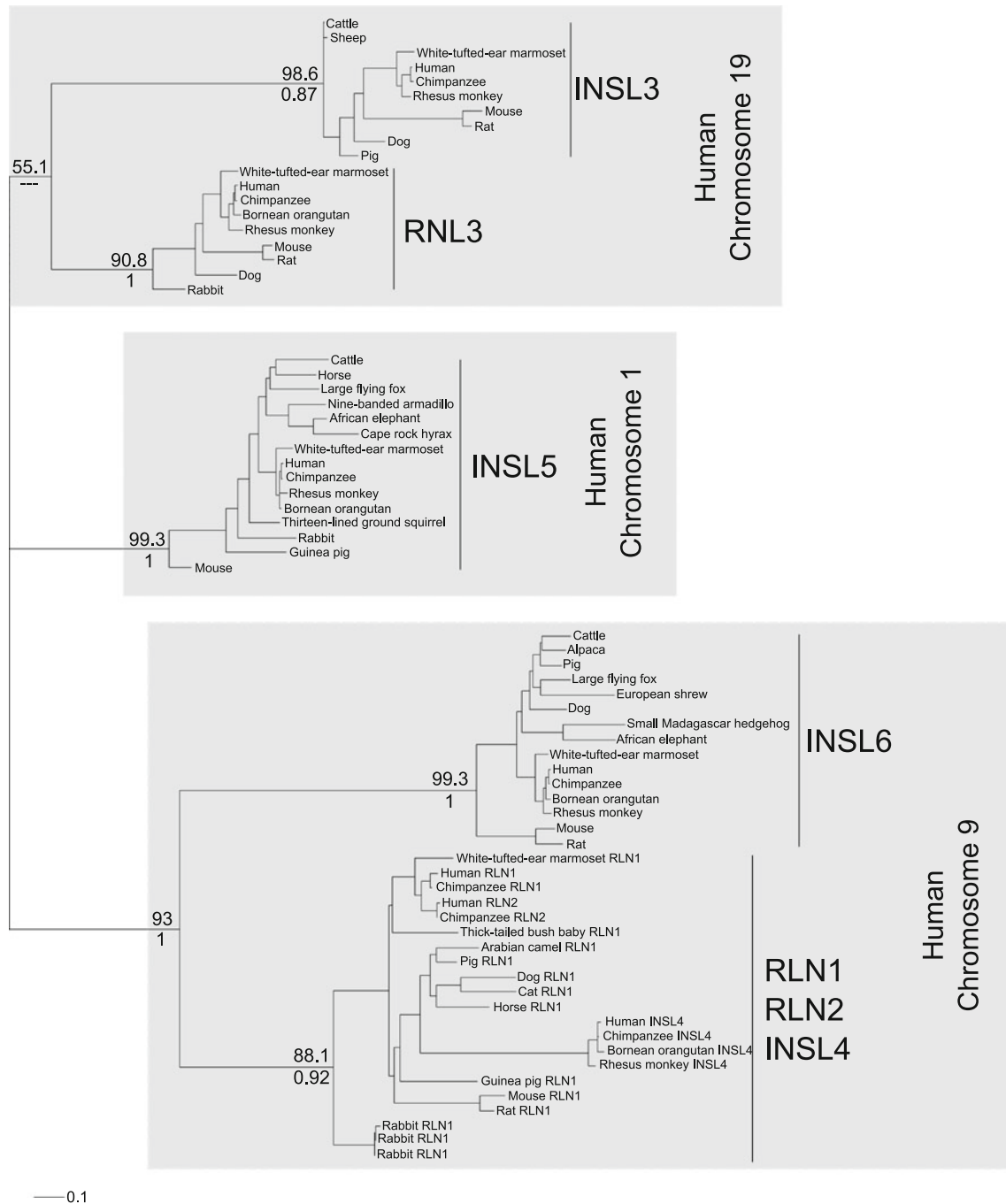
The topology recovered in our analysis is congruent with the divergence post-WGD model (Fig. 1a), as genes found on the same clusters share a most recent common ancestor to the exclusion of paralogs found on separate clusters (Fig. 2). We recovered the monophyly of each of the INSL3, INSL5, RNL3, INSL6 paralogs, plus the monophyly of a clade containing the RLN1, RLN2, and INSL4 genes, with strong support in all phylogenies (Fig. 2). Among these groups, the relationship between the INSL6 and RLN1, RLN2, RLN3, and INSL4 clades is strongly supported, while the relationship between the RNL3 and INSL3 paralogs, is moderately supported (Fig. 2).

To further clarify relationships within the RLN1, RLN2, and INSL4 clade, we performed a second set of analyses restricted to these genes, and added marsupial and platypus sequences as outgroups. Because the RLN2 gene is restricted to apes, represented by human and chimp in our study, it was expected to derive from an ape-specific duplication (Wilkinson et al. 2005; Park et al. 2008a, b). Our phylogenies are consistent with this interpretation: the human and chimp RLN2 orthologs were sister to the human and chimp RLN1 genes (Figs. 2, 3), indicating that

the ape RLN1 and RLN2 genes derives from the duplication of a proto-RLN1 ortholog (Fig. 4), in agreement with previous studies. Similarly, it is generally thought that the duplication that gave rise to INSL4 is an evolutionary innovation specific to the catarrhine lineage (Bieche et al. 2003). According to this scenario, INSL4 would be expected to be sister to the RLN1/RLN2 clade of human and chimp, to the exclusion of the marmoset RLN1 gene. However, our analyses are not compatible with this scenario: the RLN1 gene of the marmoset was placed sister to the RLN1/RLN2 genes of human and chimp with strong support (Fig. 3). Additionally, all RLN sequences from Euarchontoglires, the group that included primates, rodents, and lagomorphs, were recovered as a monophyletic group to the exclusion of the INSL4 clade with strong support (Fig. 3). In all cases the INSL4 clade is embedded within the RLN1 clade with strong support (Figs. 2, 3), suggesting that this gene arose from an RLN, and not from an INSL ancestor (Bieche et al. 2003; Olinski et al. 2006b; Wilkinson et al. 2005), thus, this phylogeny suggests that the INSL4 gene derives from the duplication of an RLN-like gene that predates the radiation of Euarchontoglires (Fig. 3), and that the gene was secondarily lost in all Euarchontoglires other than catarrhine primates. This latter point was also supported by the approximately unbiased topology test (Shimodaira 2002), which rejected the placement of the INSL4 as sister to the catarrhine RLN1 and RLN2 clade ( $P < 0.0001$ ), but not as a sister group of the Euarchontoglires clade ( $P = 0.42$ ).

### Evolution of the RLN/INSL-Like Genes

In the past, different models for the evolution of the RNL/INSL gene family were developed that made specific predictions regarding genealogical relationships among the different paralogs (Hsu 2003; Olinski et al. 2006b; Park et al. 2008a, b). However, the most recent phylogenetic studies did not compare the competing evolutionary scenarios in a phylogenetic framework (Good-Avila et al. 2009; Park et al. 2008a; Wilkinson et al. 2005). According to the divergence post-WGD model of evolution (Fig. 1a), the RLN and INSL paralogs derive from a single RLN/INSL ancestral gene that underwent two successive rounds of WGDs and gave rise to three RLN/INSL genes located on three different genomic locations. Each of these resulting genes would have been the progenitors of the paralogs found on a given cluster (Fig. 1a, Hsu 2003; Olinski et al. 2006b; Park et al. 2008a, b). The alternative divergence pre-WGD model would require the duplication of the single RLN/INSL ancestral gene and ensuing differentiation into a proto-RLN and a proto-INSL gene prior to the two rounds of WGDs. Here, the RLN and INSL ancestral genes would have been already present when the



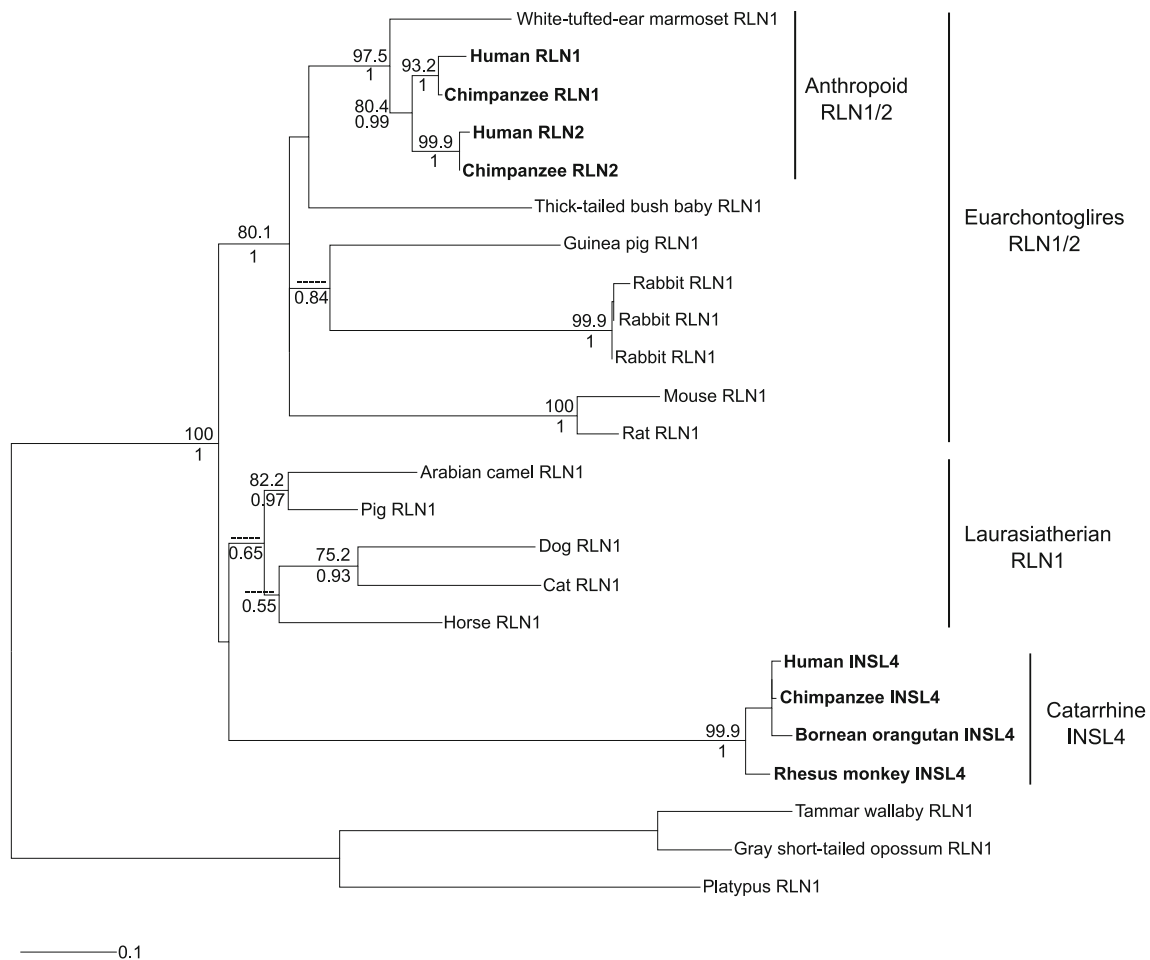
**Fig. 2** Unrooted maximum likelihood phylogeny ( $\ln L = -18013.49$ ) describing phylogenetic relationships among the RLN/INSL-like genes of placental mammals. Values on the nodes denote bootstrap

support values (*above*) and Bayesian posterior probabilities (*below*). The chromosomal location of the human paralogs is indicated with *shading*

WGDs took place (Fig. 1b). The phylogenetic predictions of these two competing models are mutually exclusive and easily recognizable. Under the divergence post-WGD model we would expect genes on the same cluster to be monophyletic, whereas under the divergence pre-WGD model we would expect to find all RLN genes in one clade, and all INSL genes in another. The topology recovered in our ML and Bayesian analyses is congruent with the

divergence post-WGD model, and the approximately unbiased topology test (Shimodaira 2002) was marginally significant in rejecting the divergence pre-WGD model in all cases ( $P = 0.0508$ ).

From an evolutionary perspective, the main implication of this finding is that the differentiation of the RLN and INSL genes occurred independently in the different clusters after the two rounds of WGDs. The initial stages of this



**Fig. 3** Maximum likelihood phylogeny (lnL = -7543.97) describing phylogenetic relationships among RLN1, RLN2, and INSL4 genes of placental mammals. Sequences were aligned using the G-INS-i strategy from Mafft v.6 (Kato et al. 2009). Two marsupial

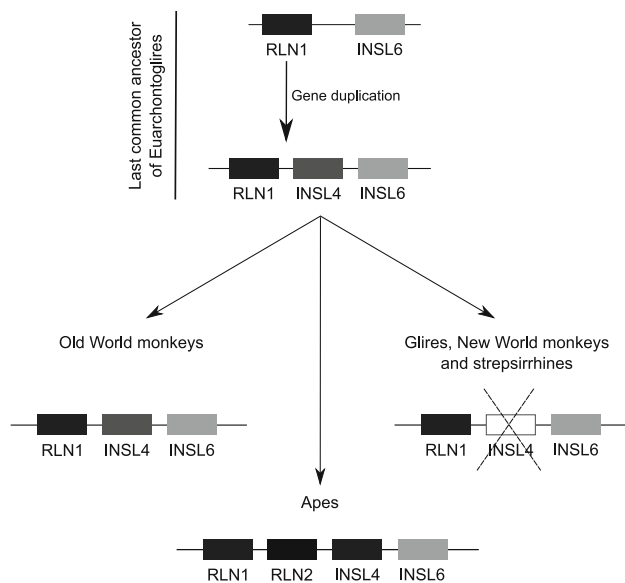
and one monotreme sequences were added to root the tree. Values on the nodes denote bootstrap support values (*above*) and Bayesian posterior probabilities (*below*). Note the position of the INSL4 clade outside the RLN1/RLN2 clade from Euarchontoglires

process would have occurred early in the evolution of vertebrates as orthology among tetrapod and teleost fish members of the family has been well established in previous studies (Good-Avila et al. 2009; Park et al. 2008a).

On a more recent time scale, tandem duplications have also played an important role in the evolution of the vertebrate RLN and INSL repertoire. For example, most mammals possess five relatively old paralogs of this family in their genomes: INSL5, RLN1, INSL6, RLN3, and INSL3 (Park et al. 2008a, b), while primates have two additional, younger genes (INSL4 and RNL2) that are not present in any other placental lineages (Bieche et al. 2003). These genes derive from more recent tandem duplications (Fig. 4). The RLN2 gene apparently originated between 29 and 18 mya (Goodman et al. 1998; Steiper and Young 2009) in the last common ancestor of apes (Figs. 3, 4; Good-Avila et al. 2009; Park et al. 2008a; Wilkinson et al. 2005), whereas the INSL4 gene arose from an RLN ancestor prior to the divergence of Euarchontoglires

(Figs. 2, 3, 4). Outside placental mammals, monotremes and marsupials appear to possess four paralogs (INSL5, RLN1, RLN3, and INSL3; Park et al. 2008a, b).

Variation in the RLN/INSL gene complement is also observed among other vertebrates. There are two RLN and two INSL genes in the western clawed frog (*Xenopus tropicalis*), but only one RLN and one INSL gene in chicken (*Gallus gallus*) (Park et al. 2008a, b). Ray-finned fish underwent an additional round of WGD, and as a result, possess duplicated RLN/INSL clusters and duplicated copies of INSL5 and RLN3 genes relative to tetrapods (Good-Avila et al. 2009; Park et al. 2008a). On the other hand, the RNL1 gene was lost in zebrafish (*Danio rerio*), but is found as a single-copy gene in other ray-finned fish species (Good-Avila et al. 2009; Park et al. 2008a), as is the case with the INSL3 gene, which has been found as a single-copy gene in the zebrafish, the spotted green pufferfish (*Tetraodon nigroviridis*), and the fugu (*Takifugu rubripes*) (Good-Avila et al. 2009).



**Fig. 4** Evolutionary reconstruction of the duplication history of the RLN2 and INSL4 genes in placental mammals. Starting with the RLN1–INSL6 gene arrangement, the proto-RLN1 gene would have gone through two tandem duplication events giving rise to the INSL4 gene in the last common ancestor of Euarchontoglires, and to the RLN2 in the last common ancestor of apes. The INSL4 gene was lost independently in all Euarchontoglires other than catarrhine primates

## Conclusions

This study provides strong support for the divergence post-WGD model of evolution for the vertebrate RLN/INSL family of genes (Fig. 1a). Under the proposed scenario, there would have been a single ancestral proto-RLN/INSL gene prior to the two rounds of WGDs in the last common ancestor of extant vertebrates. The two successive rounds of WGD would have then generated the progenitors of the different RLN–INSL clusters on each chromosome and subsequent duplications within each cluster then gave rise to the present RLN/INSL gene clusters. From a functional standpoint, our study illustrates the interplay between gene duplication and functional differentiation in the generation of biological novelties. In this case, the differentiation of the proto-RLN/INSL genes deriving from the WGDs gave rise to RLN and INSL independently on each separate cluster. Our study also indicates that lineage-specific patterns of duplication, deletion, and retention of genes have played a strong role in shaping the RLN and INSL gene complement in extant species. An example of this process is the INSL4 gene, which at present is only found in catarrhine primates, but appears to derive from a relatively older duplication that predates divergence among Euarchontoglires, and has apparently been secondarily lost in all Euarchontoglires other than catarrhines (Fig. 4).

**Acknowledgments** This work was funded by grants to JCO from the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 11080181), Programa Bicentenario de Ciencia y Tecnología (PSD89), and the Oliver Pearson Award from the American Society of Mammalogists (ASM). The authors also thank Dominique Alò, Amy Runck and Zachary A. Cheverson for critical comments, and Yves Van de Peer and two anonymous reviewers for helpful suggestions on the manuscript.

## References

- Bathgate RAD, Samuel CS, Burazin TCD, Gundlach AL, Tregear GW (2003) Relaxin: new peptides, receptors and novel actions. *Trends Endocrinol Metab* 14:207–213
- Bieche I, Laurent A, Laurendeau I, Duret L, Giovannardi Y, Frenco J-L, Olivi M, Fausser J-L, Evain-Brion D, Vidaud M (2003) Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. *Biol Reprod* 68:1422–1429
- Chan SJ, Steiner DF (2000) Insulin through the ages: phylogeny of a growth promoting and metabolic regulatory hormone. *Am Zool* 40:222–231
- de Pablo F, de la Rosa EJ (1995) The developing CNS: a scenario for the action of proinsulin, insulin and insulin-like growth factors. *Trends Neurosci* 18:143–150
- Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3:314
- Do CB, Mahabhashyam MSP, Brudno M, Batzoglu S (2005) ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res* 15:330–340
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Good-Avila SV, Yegorov S, Harron S, Bogerd J, Glen P, Ozon J, Wilson BC (2009) Relaxin gene family in teleosts: phylogeny, syntenic mapping, selective constraint, and expression analysis. *BMC Evol Biol* 9:293
- Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol* 9:585–598
- Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered* 100:605–617
- Hsu SYT (2003) New insights into the evolution of the relaxin–LGR signaling system. *Trends Endocrinol Metab* 14:303–309
- Jobb G, Haeseler AV, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18
- Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* 537:39–64
- Kuraku S, Meyer A, Kuratani S (2009) Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after. *Mol Biol Evol* 26:47–59
- Lassmann T, Sonnhammer ELL (2005) Automatic assessment of alignment quality. *Nucleic Acids Res* 33:7120–7128
- Lassmann T, Frings O, Sonnhammer ELL (2009) Kalign2: high performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res* 37:858–865
- Lynch M (2007) *The origins of genome architecture*. Sinauer Associates, Sunderland, MA
- Meyer A, Schartl M (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the

- evolution of novel gene functions. *Curr Opin Cell Biol* 11:699–704
- Nagamatsu S, Chan SJ, Falkmer S, Steiner DF (1991) Evolution of the insulin gene superfamily. Sequence of a preproinsulin-like growth factor cDNA from the Atlantic hagfish. *J Biol Chem* 266:2397–2402
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152
- Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205–217
- Ohno S (1970) *Evolution by gene duplication*. Springer-Verlag, New York
- Olinski RP, Dahlberg C, Thorndyke M, Hallbook F (2006a) Three insulin-relaxin-like genes in *Ciona intestinalis*. *Peptides* 27:2535–2546
- Olinski RP, Lundin L-G, Hallbook F (2006b) 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
- Park J-I, Semyonov J, Chang CL, Yi W, Warren W, Hsu SYT (2008a) Origin of INSL3-mediated testicular descent in therian mammals. *Genome Res* 18:974–985
- Park J-I, Semyonov J, Yi W, Chang CL, Hsu SYT (2008b) Regulation of receptor signaling by relaxin A chain motifs: derivation of pan-specific and LGR7-specific human relaxin analogs. *J Biol Chem* 283:32099–32109
- Reinecke M, Collet C (1998) The phylogeny of the insulin-like growth factors. *Int Rev Cytol* 183:1–94
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Sherwood OD (2004) Relaxin's physiological roles and other diverse actions. *Endocr Rev* 25:205–234
- Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508
- Steiper ME, Young NM (2009) *Primates (Primates)*. In: Hedges SB, Kumar S (eds) *The timetree of life*. Oxford University Press, Oxford, pp 482–486
- Subramanian AR, Kaufmann M, Morgenstern B (2008) DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithms Mol Biol* 3:6
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:609–612
- Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247–250
- Wentworth BM, Schaefer IM, Villa-Komaroff L, Chirgwin JM (1986) Characterization of the two nonallelic genes encoding mouse preproinsulin. *J Mol Evol* 23:305–312
- Wilkinson TN, Speed TP, Tregear GW, Bathgate RAD (2005) Evolution of the relaxin-like peptide family. *BMC Evol Biol* 5:14
- Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18:292–298