

Gene Duplication and Positive Selection Explains Unusual Physiological Roles of the Relaxin Gene in the European Rabbit

José Ignacio Arroyo · Federico G. Hoffmann ·
Juan C. Opazo

Received: 15 September 2011 / Accepted: 25 January 2012 / Published online: 22 February 2012
© Springer Science+Business Media, LLC 2012

Abstract The relaxin gene family is a group of genes involved in different physiological roles, most of them related to reproduction. In vertebrates the genes in this family are located in three separate chromosomal locations, and have been called relaxin family locus (RFL) A, B, and C. Among mammals the RFLA and RFLC are the most conserved as no gene copy-number variation has been observed thus far. The RFLB locus is also conserved on most mammals other than primates, where there are several gene gains and losses. Interestingly, the relaxin gene found on the RFLB locus in the European rabbit has acquired a novel role. In addition to the classical reproductive roles, this gene is expressed in tracheobronchial epithelial cells and its expression has been linked to squamous differentiation. We reconstructed the evolutionary history of the European rabbit RFLB locus using the tools of comparative genomics and molecular evolution. We found that the European rabbit possess a RFLB locus which is unique among mammals in that there are five tandemly arranged relaxin gene copies, which contrast with the single relaxin copy gene found in most mammals. In

addition we also found that the ancestral pre-duplication gene was subject to the action of positive selection, and several amino acid sites were identified under the action of natural selection including the sites B12 and B13 which are part of the receptor recognition and binding site.

Keywords Positive selection · Gene duplication · Relaxin · European rabbit · Gene family evolution

Introduction

Understanding how new biological functions arise represents an area of active research in evolutionary biology. From a genetic standpoint, gene duplication is recognized as a key biological event in generating the raw material for evolution to act on (Ohno 1970). After duplication, the functional and regulatory divergence between the resulting gene copies is thought to contribute to the generation of evolutionary novelties (Ohno 1970; Zhang 2003; Hahn 2009), and positive selection has been recognized as a potentially important evolutionary force in generating these new functions. Morris Goodman (1981, 1982) postulated that in the first stages of functional differentiation the amino acid changes that remodel the molecule could be driven to fixation by positive Darwinian selection, whereas at later stages the previously fixed changes would be preserved by purifying selection so as to maintain the newly acquired physiological role. In this regard the relaxin gene family represents an interesting model system to explore the interplay between gene duplication and functional differentiation in originating a group of genes with a variety of physiological functions (Sherwood 2004; Wilkinson et al. 2005; Arroyo et al. 2012; Hoffmann and Opazo 2011).

Electronic supplementary material The online version of this article (doi:10.1007/s00239-012-9487-2) contains supplementary material, which is available to authorized users.

J. I. Arroyo · J. C. Opazo (✉)
Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
e-mail: jopazo@gmail.com

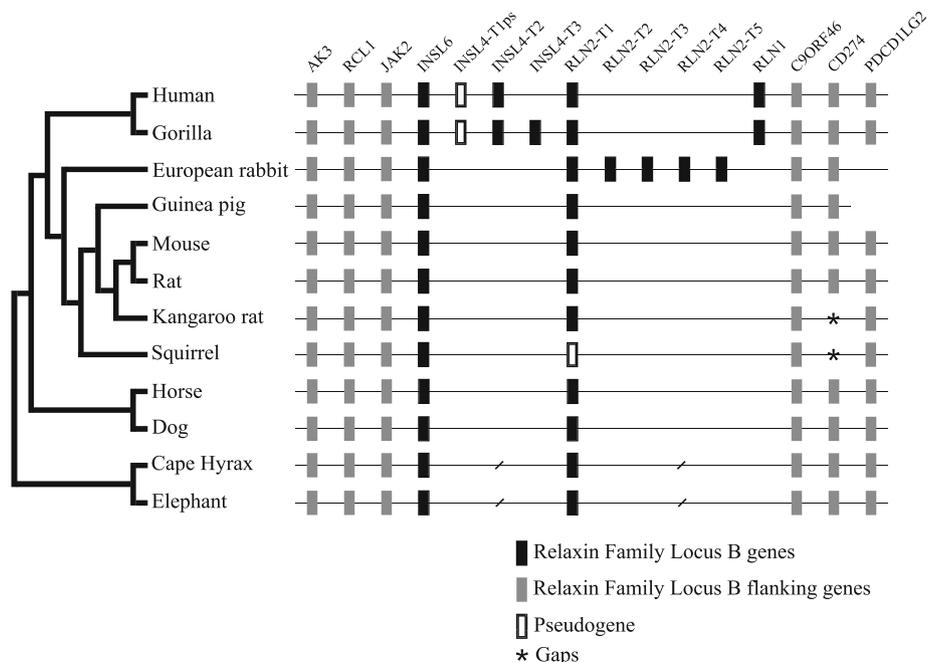
F. G. Hoffmann
Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS, USA

F. G. Hoffmann
Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Starkville, MS, USA

Members of the relaxin gene family are located on three different locations in the genome, which have been called relaxin family locus (RFL) A, B, and C (Park et al. 2008a). The number and nature of genes in these three loci is well conserved in most mammalian groups. The RFLA locus contains an ortholog of the human INSL5 gene, which has remained as a single-copy gene in all species surveyed, and the RFLC locus contains orthologs of the human INSL3 and RLN3 genes, which have also remained as single-copy genes in all species surveyed (Park et al. 2008a, Arroyo et al. 2012, Hoffmann and Opazo 2011). Copy number variation is also minimal in the RFLB locus in most mammalian species with the exception of primates (Fig. 1; Park et al. 2008a; Arroyo et al. 2012; Hoffmann and Opazo 2011). In most mammals the RFLB locus contains two genes: a single copy of a relaxin-like gene, labeled as either RLN or RLN2, and a single copy of an ortholog of the human INSL6 gene, but primates show additional variation in the number and nature of genes in this locus (Fig. 1; Bièche et al. 2003; Park et al. 2008a, b; Arroyo et al. 2012; Hoffmann and Opazo 2011). For example, haplorhines primates, the group that includes tarsiers, New World monkeys, Old World monkeys and apes, also possess a copy of INSL4 at this locus (Bièche et al. 2003; Park et al. 2008a, b; Arroyo et al. 2012), and apes also possess duplicated RLN1 and RLN2 paralogs. Interestingly, phylogenetic reconstructions indicate that both the INSL4 gene of haplorhines and the duplicate RLN1/RLN2 genes of apes, are actually older than what their phyletic distributions would suggest (Arroyo et al. 2012; Hoffmann and Opazo 2011).

From a functional standpoint, the RLN2 genes of primates and the single-copy relaxin gene in the RFLB locus of non-primate mammals have physiological roles mainly associated to the reproductive system, such as softening the connective tissue of the interpubic ligament, growth of mammary gland and nipples, follicle development, and embryo implantation among others (Burger and Sherwood 1995; Graham and Dracy 1953; Hall 1947; Krantz et al. 1950; O’day et al. 1989; Sherwood and O’Byrne 1974; Zhao et al. 2000; Samuel et al. 2007; Shabanpoor et al. 2009). In addition to its role in reproductive physiology, recent studies have also identified non-reproductive functions of these genes associated to fibrosis (Bathgate et al. 2006; Samuel 2005; Samuel and Hewitson 2006; Sherwood 2004), wound healing (Bathgate et al. 2006; Sherwood 2004), cardiac protection (Bani et al. 2005; Dschietzig et al. 2006; Samuel and Hewitson 2006; Samuel et al. 2006), allergic responses (Bani 1997) and cancer (Kamat et al. 2006; Silvertown et al. 2003). Despite the variety and apparent importance of physiological roles in which the product of this gene is involved, it is not indispensable as species of the family Bovidae (e.g., cow and sheep) have secondarily lost it during its evolutionary history (Park et al. 2008a). Because of the functional similarities between the relaxin gene in the RFLB locus from most mammals and the RLN2 gene from apes, the name RLN2 is generally applied to these two genes, although these two genes are not orthologs (Shabanpoor et al. 2009). We will follow this convention to facilitate comparisons with the extensive body of functional studies already published. Thus, the term RLN2 will be used to refer to the RLN2 gene

Fig. 1 Genomic structure of the RFLB in mammals. Stars indicate gaps in genomic coverage, while diagonal slashes indicate that the genes were identified in different genomic pieces. The orientation of the clusters is from 5’ (on the left) to 3’ (on the right). Phylogenetic relationships are based on recent published literature (Hallstrom and Janke 2008, 2010)



of primates, and to the relaxin gene in the RFLB locus from non-primate mammals.

In the European rabbit (*Oryctolagus cuniculus*), in addition to the classical physiological functions in the reproductive system (Fields et al. 1995), the RLN2 gene is also expressed in tracheobronchial epithelial cells and its expression has been linked to squamous differentiation (Jetten et al. 1992). In principle by comparing the sequence and genomic structure of the genes in the RFLB locus of the European rabbit we could gain insights into the changes and evolutionary forces that are responsible for the acquisition of this new physiological function, which appears to be an autapomorphy restricted to the European rabbit lineage. Accordingly, the goals of this study are to apply the tools of molecular evolution and comparative genomics to better understand the evolution of the RFLB locus in the European rabbit. In particular, we want to evaluate the putative role of positive Darwinian selection in the fixation of the specific amino acid replacements that distinguish the European rabbit RLN2 from other mammalian RLN1 and RLN2 genes from the RFLB locus, and make inferences about the sites that are likely to be involved with the acquisition of this new function in this species. Results of our study indicate that the European rabbit possess five tandemly arranged RLN2 gene copies in the RFLB locus, which contrast to the single copy of RLN2 gene found in all mammals other than primates, and indicate that positive Darwinian selection played a significant role in their pre-duplicative history.

Materials and Methods

Data

We obtained genomic DNA sequences for structural genes in the Relaxin Family Locus B for 35 mammals, one bird, one amphibian and two fish species from the Ensembl database (Supplementary Table 1). We identified RLN/INSL-like genes by comparing annotated exon sequences to unannotated genomic sequences, using the program Blast2seq (Tatusova and Madden 1999) available from NCBI (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>). We also included sequences from shorter records based on genomic DNA or cDNA, in order to attain a broad taxonomic coverage. To distinguish among tandemly arrayed genes copies, we index each gene-copy with the symbol T followed by a number that corresponds to the linkage order in the 5′–3′ orientation, thus, the first gene in the cluster is labeled T1, the second T2, and so forth. It is worth reminding the reader that for historical reasons, the name RLN2 is applied to two set of genes that are not 1:1 orthologs, the RLN2 gene of primates and to the single-copy RLN gene in RFLB locus in all other mammals.

Alignment

To explore sensitivity of our phylogenetic analyses to changes in the alignment method 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). The biological accuracy of alignments was assessed using the program MUMSA (Lassmann and Sonnhammer 2005) which ranks the quality of each alignment by calculating a multiple overlap score (MOS) to each of the different alignments, which ranges from 0 to 1, with 1 being the highest quality. We then selected the alignment with the highest MOS score for all ensuing analyses.

Phylogenetic Inference

We estimated phylogenetic relationships among the different RFLB genes in all major groups of mammals (monotremes, marsupials, and placentals), using sequences from chicken (*Gallus gallus*), western clawed frog (*Xenopus tropicalis*), fugu (*Takifugu rubripes*), and medaka (*Oryzias latipes*) as outgroup. We used a maximum likelihood and a Bayesian approach, as implemented in the program Treefinder version October 2008 (Jobb et al. 2004) and Mr.Bayes v3.1.2 (Ronquist and Huelsenbeck 2003), respectively. The best fitting models for each structural domain was estimated separately using the propose model routine from the program Treefinder version October 2008 (Jobb et al. 2004). In the case of maximum likelihood, we estimated the best tree under the selected models, and assessed support for the nodes with 1,000 bootstrap pseudoreplicates. In Bayesian analysis, two simultaneous independent runs were performed for 30×10^6 iterations of a Markov Chain Monte Carlo algorithm, with six simultaneous chains sampling trees every 1,000 generations. Support for the nodes and parameter estimates were derived from a majority rule consensus of the last 15,000 trees sampled after convergence. The average standard deviation of split frequencies remained 0.01 after the burn-in threshold.

Natural Selection Analysis

To investigate the possible role of positive Darwinian selection on the evolution of the multiple copies of the RLN2 genes in the European rabbit, we explored variation in ω , the ratio of the rate of non-synonymous substitutions (d_N) to the rate of synonymous substitutions (d_S), in a maximum likelihood framework using the codeml program from PAML v4.4 (Yang 2007). In brief, if amino acid

replacements are neutral, then d_N and d_S would be very similar and $\omega = d_N/d_S \approx 1$, under purifying selection most amino acid substitutions would be deleterious, and so $\omega < 1$, whereas under positive selection amino acid replacements would be advantageous, and so $\omega > 1$. All these analyses were based on a phylogeny that includes a representative panel of RLN2 genes from all major mammalian lineages (Supplementary Fig. S1). We compared two sets of models, the first set focused on comparing changes in ω ($=d_N/d_S$) along the branches of the tree, and the second set of models focused on comparing changes in ω along the different sites in the alignment between background and foreground sets of branches. These different models were implemented to explore changes in ω relative to the duplication in the European rabbit RLN2 paralogs. In all cases three starting ω values (0.5, 1, and 2) were used to check the existence of multiple local optima. Nested models were compared using the likelihood ratio test (LRT). We first compared the following four branch models (Supplementary Fig. S2): (i) a 1- ω model which fitted a single ω estimate to all branches in the tree; (ii) a 2- ω model, which assigned one ω to all European rabbit branches and a second ω to all non-rabbit branches in the tree; (iii) a 3- ω model, with one ω for the ancestral branch of the European rabbit RLN2 clade, a second ω for the European rabbit RLN2 clade branches, and a third ω for non-rabbit branches; and (iv) a 7- ω model that assigns a separate ω to each European rabbit RLN2 branch, plus a separate ω estimate for non-rabbit branches (Supplementary Fig. S2). We also implemented branch-site models which explore changes in ω for a set of sites in a specific branch of the tree in order to assess changes in their selective regime (Yang and Dos Reis 2011). In this case the ancestral branch of the European rabbit clade was labeled as the foreground branch. We compared the modified model A (Yang et al. 2005; Zhang et al. 2005), in which some sites are allowed to change to an $\omega > 1$ in the foreground branch, with the corresponding null hypothesis of neutral evolution. The Bayes Empirical Bayes (BEB) method was used to identify sites under positive selection (Nielsen and Yang 1998; Yang et al. 2000). Because the branch-site analysis estimates rates of evolution on a codon by codon basis, its implementation is particularly useful in cases when different gene segments evolve at different rates, as is the case with the different domains of the RLN2 genes.

Results and Discussion

Genomic Structure of the RFLB in the European Rabbit

The number and nature of genes located on the RFLB locus is well conserved among most mammals other than primates (Park et al. 2008a, b; Hoffmann and Opazo 2011;

Arroyo et al. 2012). Most species examined are characterized by having an INSL6 gene at the 5' end of the RFLB locus, and a relaxin-like gene at the 3' end, which is generally referred to as RLN2. The RFLB locus is flanked by janus kinase 2 (JAK2), RNA terminal phosphate cyclase-like 1 (RCL1) and adenylate kinase 3 (AK3) on the 5' side, and by the chromosome 9 open reading frame 46 (C9Orf46), CD274 molecule (CD274), and programmed cell death 1 ligand 2 (PDCD1LG2) on the 3' end (Fig. 1; Park et al. 2008a, b). Previous studies on the RFLB locus in the European rabbit suggested it conformed to the general pattern described above (Fields et al. 1995; Jetten et al. 1992). However, our examinations of the RFLB locus in European rabbit reveal the presence of five tandemly arranged RLN2 gene copies, in addition to a single copy of INSL6 at the 5' end of the cluster (Fig. 1). Each of these five tandemly arranged RLN2 genes is characterized by the canonical two exons/one intron gene structure typical of mammalian relaxins, and has a perfectly intact open reading frame. Experimental evidence indicates that these genes are mainly expressed in the uterus, placenta, and tracheobronchial epithelial cells (Jetten et al. 1992; Fields et al. 1995), and utilized by the endometrium, uterine cervix, and mammary glands (Fields and Fields 2005). Thus, the RFLB locus in the European rabbit includes six putatively protein-coding genes, whereas the corresponding locus includes two putatively coding genes in all rodents, which are the order sister to Lagomorpha (Fig. 1). The rabbit RFLB is located on chromosome 1 and extends for 194 kb from the start codon of INSL6 on the 5' end to the stop codon of the RLN2 paralog in the 3' end. Synteny is conserved relative to other mammals, as genes found up- and downstream of the RFLB locus in the European rabbit are orthologous with the genes found in the corresponding locations in most other mammals (Fig. 1).

Thus far, duplications of the ancestral gene located on the RFLB locus of vertebrates have only been reported for mammals. The oldest duplication gave rise to the INSL6 gene in the common ancestor of placental mammals, a gene that has remained as a single-copy gene in all species examined (Fig. 1; Park et al. 2008a, b; Arroyo et al. 2012). A second duplication event gave rise to INSL4, which is only found in primates (Bièche et al. 2003; Hoffmann and Opazo 2011). Even though INSL4 was thought to derive from a primate-specific duplication, phylogenetic reconstructions suggest that INSL4 derives from a duplication in the last common ancestor of Euarchontoglires (Arroyo et al. 2012; Hoffmann and Opazo 2011). There are also several duplications of the RLN2 and INSL4 genes that are more recent in time, and are restricted to primates for the most part (Arroyo et al. 2012). The RLN2 gene duplicated in the last common ancestor of anthropoids (Arroyo et al. 2012) but the two resulting primate duplicates, RLN1 and

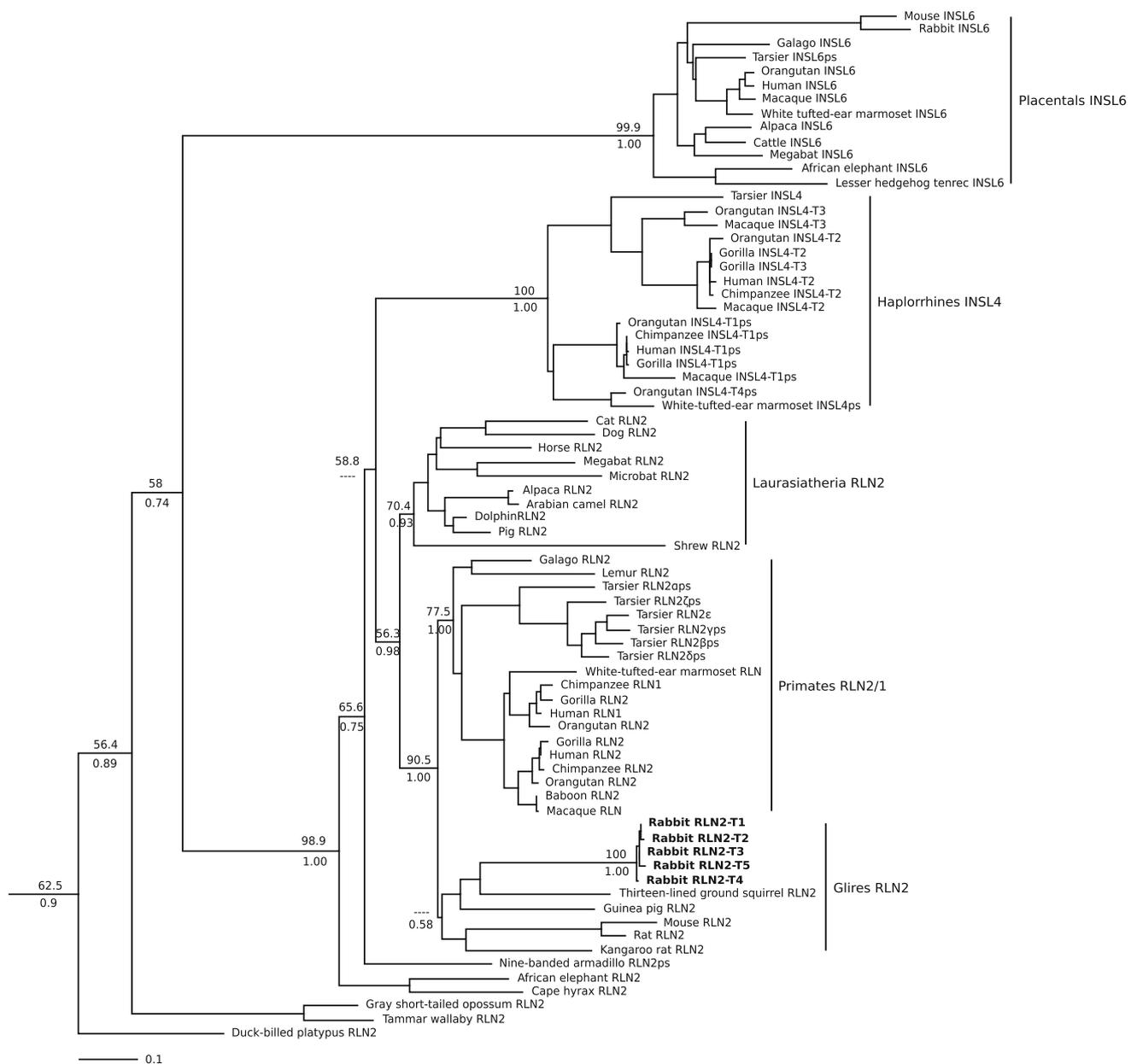


Fig. 2 Maximum likelihood phylogram depicting phylogenetic relationships among the RFLB genes in mammals. The RLN2 paralogous genes identified in the European rabbit are in bold. Number above the nodes correspond to maximum likelihood bootstrap support values, and numbers below the nodes correspond to

Bayesian posterior probabilities. Sequences from the chicken (*Gallus gallus*), western clawed frog (*Xenopus tropicalis*), fugu (*Takifugu rubripes*), and medaka (*Oryzias latipes*) were used as outgroups

RLN2, have only been retained as a pair in hominoids (Arroyo et al. 2012).

Orthologous Relationships of the RFLB Genes in Mammals

We conducted a phylogenetic analysis to resolve orthologous relationships between the European rabbit RLN2 paralogs and the different genes on the RFLB locus in other

mammals, focused on the European rabbit and primates, the groups that possess multiple copies of relaxin-like genes in RFLB. At the deeper level, the phylogenies we obtained are concordant with the evolutionary scenarios proposed for the genes located on the RFLB locus (Fig. 2). The INSL6 clade was recovered as sister to the clade that includes the RLN1, RLN2, and INSL4 genes of placental mammals (Arroyo et al. 2012; Hoffmann and Opazo 2011), as expected for a gene that derives from a duplication that predates the

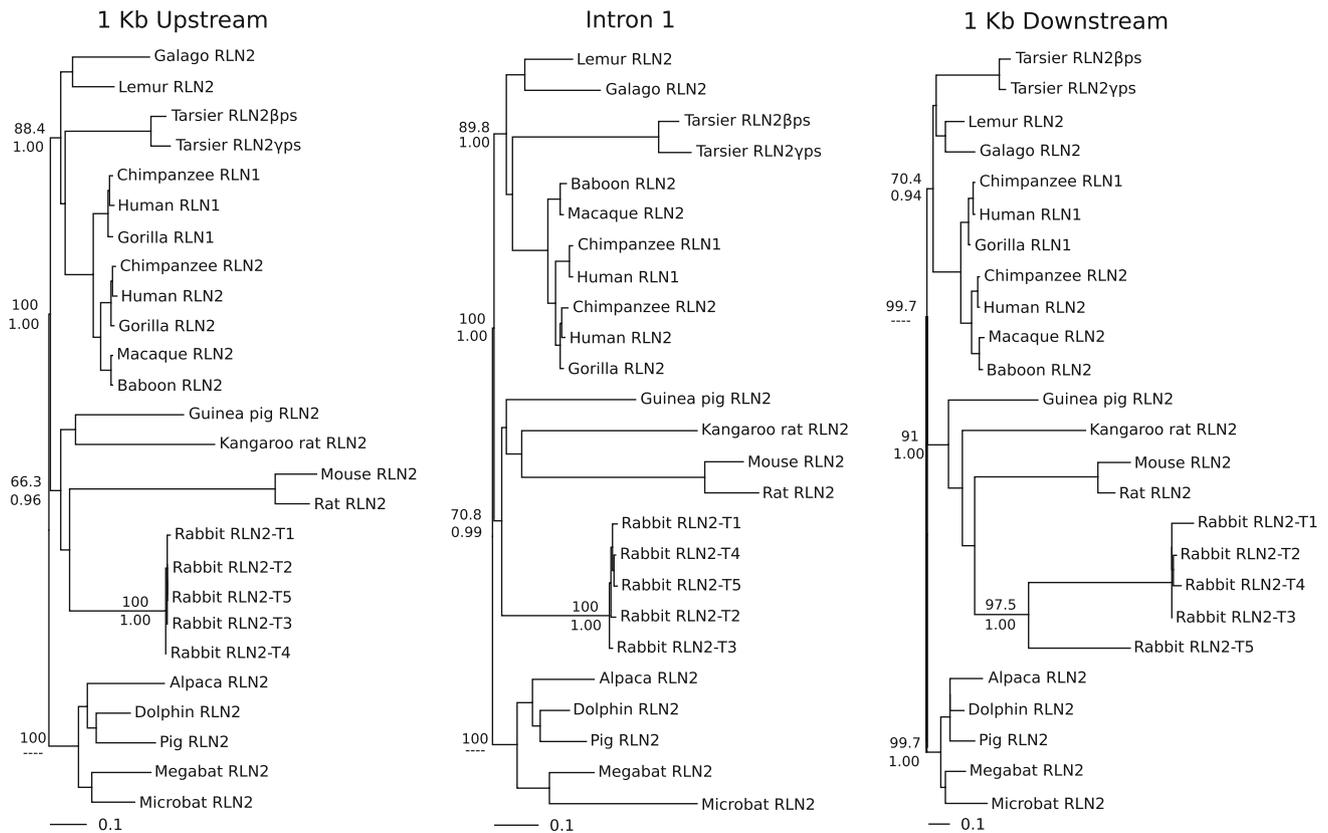


Fig. 3 Maximum likelihood phylogenies depicting relationships among relaxin genes in boreoeutherian mammals based on 1 kb of 5' flanking sequence (*left*), intron 1 (*center*), and 1 kb of 3' flanking

sequence (*right*). Number above the nodes correspond to maximum likelihood bootstrap support values, and numbers below the nodes correspond to Bayesian posterior probabilities

radiation of the group (Fig. 2). The INSL4 clade was embedded within the clade that includes the RLN1 and RLN2 genes (Fig. 2), confirming that it derives from a relatively old duplication of an RLN-like ancestor (Bièche et al. 2003; Arroyo et al. 2012; Hoffmann and Opazo 2011). With regards to the European rabbit RLN2 paralogs, maximum likelihood and Bayesian phylogenies place the five gene copies in a strongly supported monophyletic clade sister to the single-copy RLN2 gene from the ground squirrel, embedded within the clade that includes the RLN2 genes from all other rodents (Fig. 2).

This pattern suggests that the duplications that gave rise to the presence of multiple RLN2 paralogs in the European rabbit are independent of those that gave rise to the duplicate RLN1 and RLN2 genes of apes. However, this could also be due to gene conversion, which can overwrite genetic information and result in phylogenies where paralogous genes from the same species are more similar to each other than they are to their orthologs in closely related species. Because interparalog gene conversion is typically restricted to coding regions, analyzing variation in non-coding regions has the potential to retrieve the true orthologous relationships, even in the presence of gene conversion. Accordingly, we conducted

phylogenetic analyses based on 1 kb of upstream flanking sequence, 1 kb of downstream flanking sequence, and intron 1 for all relaxin genes of Boreoeutherian mammals, including the five tandemly arranged genes of the European rabbit. As shown in Fig. 3, all phylogeny reconstructions recovered the five tandemly arranged RLN2 gene copies of the European rabbit in a monophyletic group with strong support, suggesting that the five tandemly arranged RLN2 genes of the European rabbit are the product of duplications specific to the European rabbit lineage. In agreement with these analysis, gene conversion analyses using GeneConv (Sawyer 1989) also failed to detect any evidence of gene conversion.

Natural Selection at the Molecular Level

Functional studies indicate that the European rabbit RLN2 gene has acquired a novel biological role relative to other mammals (Jetten et al. 1992). Because this functional change coincided with the emergence of the five European rabbit RLN2 paralogs, we explored the potential role of positive Darwinian selection in this process in a maximum likelihood framework. These analyses focused on assessing whether positive selection played a role in the preservation

of the European rabbit RLN2 duplicates, and on exploring when positive selection acted relative to the duplicative history of these paralogs. We started by comparing the d_N/d_S ratio ($=\omega$) among the different RLN2 paralogs of the European rabbit in a pairwise manner using the YN00 method (Yang et al. 2000). Results from this approach are compatible with positive selection in two comparisons, (T5/T2 and T5/T4), whereas in other six comparisons only non-synonymous substitutions were estimated, and in the last two the ω ratio was lower than 1 (Supplementary material). We then incorporated phylogenetic information into the analyses and estimated variation in ω along a tree built with a subset of the sequences included in Fig. 2 that included representatives of all major groups of mammals (Supplementary Fig. S1). Overall, the simplest model, which assigned a single ω for all branches (the 1- ω model), indicates that all RLN1 and RLN2 genes in the study have relatively high ω estimates (Table 1), in line with previous findings from primates (Wilkinson et al. 2005). When we compared variation in ω among different set of branches, we found that a 2- ω model that assumes one ω value for the European rabbit branches and a second one for the non-rabbit branches was significantly better than the 1- ω value for all branches on the tree (Table 1; $2\Delta \ln L = 5.24$, $df = 1$, $P < 0.025$). Under the 2- ω model, European rabbit branches had an ω of 1.37, whereas non-rabbit branches had an ω of 0.71 (Table 1). The 3- and 7- ω models that allowed separate ω estimates were not significantly better than the 2- ω model. In all cases the synonymous and non-synonymous rate estimates obtained for the post-duplication branches were at least an order of magnitude lower than the estimates for the pre-duplication branch obtained by branch models.

We further explored the putative role of positive Darwinian selection in the European rabbit pre-duplication branch using branch-site analyses, which are summarized in Table 1. Here we set the European rabbit pre-duplication branch as foreground, and all other portions of the tree as background. The likelihood ratio test indicates that the model that estimates a class of sites with an ω value of 3.94 (Model A) had a significantly better fit than the null model in which the ω value was fixed to 1 ($2\Delta \ln L = 5.52$, $df = 1$; $P < 0.025$). The branch-site analysis indicates that 27% of the residues switched to a class of sites with an ω value of 3.94 in the European rabbit pre-duplication branch, and the BEB analysis placed 17 codons in this class, three located on the signal peptide, seven on the region encoding for the B chain, one on the region encoding for the C chain, and six located on the region encoding for the A chain (Table 1). Four of these sites, one in each region, are over the 90% posterior probability threshold of the BEB analysis (Table 1). From a biochemical standpoint, two of the putatively selected sites on the B chain map to functionally relevant portions of the protein. The residues in the B12 and B13 positions, which are part of the receptor recognition and binding site (Fig. 4), have been substituted for amino acids with different physical–chemical properties relative to rodents and human. In the case of the B12, glycine, a non-polar amino acid, was replaced by arginine, a positively charged residue, and in the case of B13 arginine was replaced by asparagine, a polar uncharged residue (Fig. 3). Interestingly, all these substitutions are shared by the different European rabbit RLN2 paralogs, which is consistent with the results of PAML analyses, which indicates that positive selection in the European rabbit pre-duplication branch was responsible for the remodeling of this protein.

Table 1 Log likelihood values and parameter estimates under different branch and branch-site models

Model	$\ln L$	Parameter estimates	Positively selected sites
Branch models			
1- ω	-6672.46	$\omega_{\text{all}} = 0.74$	N/A
2- ω	-6669.84	$\omega_{\text{non-lagomorph branches}} = 0.71$; $\omega_{\text{lagomorph branches}} = 1.37$	N/A
3- ω	-6669.83	$\omega_{\text{non-lagomorph branches}} = 0.71$; $\omega_{\text{ancestral lagomorph clade}} = 1.35$; $\omega_{\text{lagomorph crown group}} = **$	N/A
7- ω	-6666.52	$\omega_{\text{non-lagomorph branches}} = 0.71$; $\omega_{\text{ancestral lagomorph branch}} = 1.35$; $\omega_{\text{descendant lagomorph branches}} = **$	N/A
Branch-site models			
Null model A	-6557.92	$p_0 = 0.26$, $p_1 = 0.32$, $p_{2a} = 0.18$, $p_{2b} = 0.24$; $\omega_0 = 0.25$, $\omega_1 = 1$, $\omega_2 = 1$	N/A
Model A	-6555.16	$p_0 = 0.29$, $p_1 = 0.42$, $p_{2a} = 0.11$, $p_{2b} = 0.16$; $\omega_0 = 0.25$, $\omega_1 = 1$, $\omega_2 = 3.94$	SP3, SP15, SP21 ; B1, B7, B12, B13 , B17, B19, B25; C97 ; A4, A5, A7, A13, A18, A21

$\ln L$ likelihood value, p proportion of site class, ω omega value for branches or site classes, *SP* signal peptide, *B* B chain, *C* C chain, *A* A chain. Sites inferred under positive selection at the 90% level are in bold. ** See results in the text

	B Chain			A Chain		
	1	10	20	1	10	20
Human RLN2	DSWMEEV	IKL	CGRELVRAQIAICGMSTWS	QLYSALANKCCHV	GCTKRSLARFC	
European rabbit RLN2-T1	YEWMMENVK	ICRNDFV	RTAIEVCGHVHLE	QFSESLPEECCKY	GCPRYLLMYC	
European rabbit RLN2-T2	YEWMMENVK	ICRNDFV	RTAIEVCGHVHLE	QFSESLPEECCKY	GCPRYLLLYC	
European rabbit RLN2-T3	YEWMMENVK	ICRNDFV	RTAIEVCGHVHLE	QFSESLPEECCKY	GCPRYLLMYC	
European rabbit RLN2-T4	YEWMMENVK	ICRNDFV	RTAIEVCGHVHLE	QFSESLPEKCKY	GCPRYLLMYC	
European rabbit RLN2-T5	YEWMMENVK	ICRNDFV	RTAIEVCGHVHLE	QFSESLPEECCKY	GCPRYLLMYC	

Fig. 4 An alignment of RLN2 amino acid sequences from human and the European rabbit. Positively selected amino acid sites according to the branch-site analyses are shade. The box with discontinuous line denote the receptor recognition and binding site

Taken together, these analyses indicate that positive Darwinian selection has played a strong role in the evolution of the European rabbit RLN2 genes, specially in the pre-duplication branch. Even though our analyses also estimate elevated non-synonymous substitution rates for the European rabbit post-duplication branches, because of the relatively low estimates of d_N and in particular of d_S , the estimates of ω are not reliable for these branches. As a results it is difficult to distinguish between relaxations of purifying selection and positive selection. However, pairwise comparisons between the RLN2-T2 and T5, and between the T4 and T5 paralogs leave open the possibility that positive selection might still be acting in the European rabbit post-duplication branches.

Although the duplication and functional divergence of protein-coding genes has long been recognized as an important process in the origin of biological novelties (Ohno 1970; Zhang 2003; Hahn 2009), the putative role of adaptive changes in promoting these innovations remains unclear. Most of the models that invoke positive selection posit that post-duplication changes that lead to the differentiation of the resulting duplicates are driven by natural selection (Innan and Kondrashov 2010), but our results do not match the expectations of these models. The codon-based models implemented in this study would suggest that there was a burst of positive Darwinian selection in the proto-RLN2 gene, prior to the duplication events that gave rise to all the European rabbit RLN2 paralogs. Bringing together the results of our evolutionary analyses and the functional data from Jetten et al. (1992), it is possible to make an association between the positive selection regime mainly inferred for the ancestral gene of the European rabbit and the gain-of-function observed in this species which is related to cell differentiation in tracheobronchial epithelia. It remains to be seen whether the different rabbit RLN2 paralogs have diverged in their pattern of gene expression, and thus, could have partitioned some of their functional roles among them, and whether the elevated d_N/d_S rates reported for primates (Wilkinson et al. 2005) might be related to currently unknown functional changes.

Acknowledgments This study 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). FGH acknowledges a grant from the National Science Foundation (EPS-0903787).

References

Arroyo JI, Hoffmann FG, Opazo JC (2012) Gene turnover and differential retention in the relaxin/insulin-like gene family in primates. *Molecular Phylogenetics and Evolution* (in press)

Bani D (1997) Relaxin: a pleiotropic hormone. *Gen Pharmacol* 28:13

Bani D, Nistri S, Sacchi TB, Bigazzi M (2005) Basic progress and future therapeutic perspectives of relaxin in ischemic heart disease. *Ann N Y Acad Sci* 1041:423

Bathgate RA, Ivell R, Sanborn BM, Sherwood OD, Summers RJ (2006) International Union of Pharmacology LVII: recommendations for the nomenclature of receptors for relaxin family peptides. *Pharmacol Rev* 58:7

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

Burger LL, Sherwood OD (1995) Evidence that cellular proliferation contributes to relaxin-induced growth of both the vagina and the cervix in the pregnant rat. *Endocrinology* 136:4820

Do CB, Mahabhashyam MS, Brudno M, Batzoglou S (2005) ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res* 15:330

Dschietzig T, Bartsch C, Baumann G, Stangl K (2006) Relaxin-a pleiotropic hormone and its emerging role for experimental and clinical therapeutics. *Pharmacol Ther* 112:38

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinf* 5:113

Fields PA, Fields MJ (2005) Potential binding sites for the relaxin in pregnant rabbits. *Ann NY Acad Sci* 1041:123

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. *Biol Reprod* 53:1139

Goodman M (1981) Decoding the pattern of protein evolution. *Prog Biophys Molec Biol* 37:105

Goodman M (1982) Positive selection causes purifying selection. *Nature* 295:630

Graham EF, Dracy AE (1953) The effect of relaxin and mechanical dilatation of the bovine cervix. *J Dairy Sci* 36:772

Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered* 100:605

- Hall K (1947) The effects of pregnancy and relaxin on the histology of the pubic symphysis in the mouse. *J Endocrinol* 5:174
- Hallstrom BM, Janke A (2008) Resolution among major placental mammal interordinal relationships with genome data imply that speciation influenced their earliest radiations. *BMC Evol Biol* 8:162
- Hallstrom BM, Janke A (2010) Mammalian evolution may not be strictly bifurcating. *Mol Biol Evol* 27:2804
- Hoffmann FG, Opazo JC (2011) Evolution of the relaxin/insulin-like gene family in placental mammals: implications for its early evolution. *J Mol Evol* 72:72
- Innan H, Kondrashov F (2010) The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet* 11:97
- Jetten AM, Bernacki SH, Floyd EE, Saunders NA, Pieniazek J, Lotan R (1992) Expression of a preprorelaxin-like gene during squamous differentiation of rabbit tracheobronchial epithelial cells and its suppression by retinoic acid. *Cell Growth Differ* 3:549
- Jobb G, Av Haeseler, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18
- Kamat AA, Feng S, Agoulnik IU, Kheradmand F, Bogatcheva NV, Coffey D, Sood AK, Agoulnik AI (2006) The role of relaxin in endometrial cancer. *Cancer Biol Ther* 5:71
- Katoh K, Asiminos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* 537:39
- Krantz JC, Bryant HH, Carr CJ (1950) The action of aqueous corpus luteum extract upon uterine activity. *Surg Gynecol Obstet* 90:372
- Lassmann T, Sonnhammer EL (2005) Automatic assessment of alignment quality. *Nucleic Acids Res* 33:7120
- Lassmann T, Frings O, Sonnhammer EL (2009) Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. *Nucleic Acids Res* 37:858
- Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929
- Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302:205
- O'Day MB, Winn RJ, Easter RA, Dziuk PJ, Sherwood OD (1989) Hormonal control of the cervix in pregnant gilts. II. Relaxin promotes changes in the physical properties of the cervix in ovariectomized hormone-treated pregnant gilts. *Endocrinology* 125:3004
- Ohno S (1970) Evolution by gene duplication. Springer-Verlag, New York
- 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
- 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
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572
- Samuel CS (2005) Relaxin: antifibrotic properties and effects in models of disease. *Clin Med Res* 3:241
- Samuel CS, Hewitson TD (2006) Relaxin in cardiovascular and renal disease. *Kidney Int* 69:1498
- Samuel CS, Du XJ, Bathgate RA, Summers RJ (2006) 'Relaxin' the stiffened heart and arteries: the therapeutic potential for relaxin in the treatment of cardiovascular disease. *Pharmacol Ther* 112:529
- Samuel CS, Hewitson TD, Unemori EN, Tang ML (2007) Drugs of the future: the hormone relaxin. *Cell Mol Life Sci* 64:1539
- Sawyer SA (1989) Statistical tests for detecting gene conversion. *Mol Biol Evol* 6:526
- Shabanpoor F, Separovic F, Wade JD (2009) The human insulin superfamily of polypeptide hormones. In: Litwack G (ed) *Insulin and IGFs*. Academic Press, London, pp 1–31
- Sherwood OD (2004) Relaxin's physiological roles and other diverse actions. *Endocr Rev* 25:205
- Sherwood CD, O'Byrne EM (1974) Purification and characterization of porcine relaxin. *Arch Biochem Biophys* 160:185
- Silvertown JD, Summerlee AJ, Klonisch T (2003) Relaxin-like peptides in cancer. *Int J Cancer* 107:513
- Subramanian AR, Kaufmann M, Morgenstern B (2008) DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. *Algorithms Mol Biol* 3:6
- Tatusova TA, Madden TL (1999) BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247
- Wilkinson TN, Speed TP, Tregear GW, Bathgate RAD (2005) Evolution of the relaxin-like peptide family. *BMC Evol Biol* 5:14
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586
- Yang Z, dos Reis M (2011) Statistical properties of the branch-site test of positive selection. *Mol Biol Evol* 28:1217
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431
- Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22:1107
- Zhang J (2003) Evolution by gene duplication: an update. *Trends Ecol Evol* 18:292
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472
- Zhao L, Samuel CS, Tregear GW, Beck F, Wintour EM (2000) Collagen studies in late pregnant relaxin null mice. *Biol Reprod* 63:697