Molecular Phylogenetics and Evolution 63 (2012) 768–776

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Gene turnover and differential retention in the relaxin/insulin-like gene family in primates

José Ignacio Arroyo [a], Federico G. Hoffmann [b,c], Juan C. Opazo [a,*]

[a] Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile
[b] Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, MS, USA
[c] Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, MS, USA

ARTICLE INFO

Article history:
Received 29 March 2011
Revised 15 February 2012
Accepted 17 February 2012
Available online 1 March 2012

Keywords:
Gene family evolution
Relaxin
Primates
Gene duplication
Insulin-like peptide
Birth-and-death
Differential retention

ABSTRACT

The relaxin/insulin-like gene family is related to the insulin gene family, and includes two separate types of peptides: relaxins (RLNs) and insulin-like peptides (INSLs) that perform a variety of physiological roles including testicular descent, growth and differentiation of the mammary glands, trophoblast development, and cell differentiation. In vertebrates, these genes are found on three separate genomic loci, and in mammals, variation in the number and nature of genes in this family is mostly restricted to the Relaxin Family Locus B. For example, this locus contains a single copy of RLN in platypus and opossum, whereas it contains copies of the INSL6, INSL4, RLN2 and RLN1 genes in human and chimp. The main objective of this research is to characterize changes in the size and membership composition of the RLN/INSL gene family in primates, reconstruct the history of the RLN/INSL genes of primates, and test competing evolutionary scenarios regarding the origin of INSL4 and of the duplicated copies of the RLN gene of apes. Our results show that the relaxin/INSL-like gene family of primates has had a more dynamic evolutionary history than previously thought, including several examples of gene duplications and losses which are consistent with the predictions of the birth-and-death model of gene family evolution. In particular, we found that the differential retention of relatively old paralogs played a key role in shaping the gene complement of this family in primates. Two examples of this phenomenon are the origin of the INSL4 gene of catarrhines (the group that includes Old World monkeys and apes), and of the duplicate RLN1 and RLN2 paralogs of apes. In the case of INSL4, comparative genomics and phylogenetic analyses indicate that the origin of this gene, which was thought to represent a catarrhine-specific evolutionary innovation, is as old as the split between carnivores and primates, which took place approximately 97 million years ago. In addition, in the case of the RLN1 and RLN2 genes of apes our phylogenetic trees and topology tests indicate that the duplication that gave rise to these two genes maps to the last common ancestor of anthropoid primates. All these genomic changes in gene complement, which are particularly prevalent among anthropoid primates, might be linked to the many physiological and anatomical changes found in this group. Given the various roles of members of the RLN/INSL-like gene family in reproductive biology, it might be that changes in this gene family are associated to changes in reproductive traits.

1. Introduction

The relaxin/insulin-like gene family is related to the insulin gene family, and includes two separate types of peptides: relaxins (RLNs) and insulin-like peptides (INSLs) that perform a variety of physiological roles mostly related to reproduction (Bathgate et al., 2003; Sherwood, 2004; Park, Chang and Hsu, 2005). Members of this gene family share a conserved two-exon one-intron structure, and encode for a prohormone that consists of a signal peptide plus three domains (B-, C-, and A-; Hudson et al., 1983). In most of the cases the mature peptide results from the cleavage of the signal peptide and the C-domain (Adham et al., 1993; Chan and Steiner, 2000).

From a genomic standpoint, comparative analyses have revealed that the RLN/INSL-like gene family has greatly expanded in vertebrates relative to non-vertebrate chordates such as sea squirt and amphioxus (Olinski et al., 2006a,b; Park et al., 2008a; Hoffmann and Opazo, 2011). Early in their evolutionary history vertebrates underwent two rounds of whole genome duplication (WGDs) (Dehal and Boore, 2005; Kuraku et al., 2009; Meyer and Schartl, 1999; Ohno, 1970), which have been linked to the initial expansion of this gene family (Olinski et al., 2006a,b; Park et al., 2008a; Hoffmann and Opazo, 2011). The WGDs gave rise to the

* Corresponding author. Fax: +56 63 221344.
E-mail address: jopazo@gmail.com (J.C. Opazo).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.02.011

precursors of each of the three *Relaxin Family Loci* currently found in vertebrates, and subsequent tandem duplications within each *Relaxin Family Loci* generated the gene repertoire found in extant species (Olinski et al., 2006b; Park et al., 2008a; Hoffmann and Opazo, 2011). Synteny comparisons show that the genomic location of these loci is well conserved across vertebrates (Park et al., 2008a,b; Yegorov et al., 2009; Good-Avila et al., 2009). For example, the *Relaxin Family Locus A* (RFLA) is defined by the presence of the INSL5 gene, and is flanked by the TCTEX1D1 and WDR78 genes, and the *Relaxin Family Locus B* (RFLB) is defined by the presence of the descendants of the RLN gene, INSL6 and 4, and RLN2 and 1, and is flanked by the JAK2 and C9orf46 genes. The *Relaxin Family Locus C* (RFLC) is defined by the presence of the INSL3 and RLN3 genes, but unlike the RFLA and RFLB, this locus has become fragmented in vertebrates. In human the two remnants of this ancestral cluster are located ~4 Mb apart on chromosome 19, and one of them contains the RLN3 gene flanked by RFX1 and IL27RA genes, and the other contains the INSL3 gene flanked by the JAK3 and B3GNT3 genes.

Among mammals the number and nature of genes present in the RFLA and RFLC loci is conserved, and copy number variation is mostly restricted to the RFLB locus (Park et al., 2008a,b; Good-Avila et al., 2009). The RFLB of both monotremes and marsupials contains a single RLN gene, probably reflecting the ancestral condition, whereas the RFLB of placental mammals contains additional genes. An INSL6 ortholog is found in all placental mammals, and primates posses also a copy of INSL4, and duplicated RLN1 and RLN2 paralogs in this locus (Bièche et al., 2003; Wilkinson et al., 2005; Park et al., 2008a,b). Contrary to expectations, INSL4 was found to derive from a proto-RLN and not from a proto-INSL ancestor (Bièche et al., 2003; Hoffmann and Opazo, 2011) and has only been found in catarrhine primates (the group that includes Old World monkeys and apes), as a single copy gene. INSL4 is mainly expressed in placenta (Chassin et al., 1995; Bièche et al., 2003), and although the precise function of INSL4 is not well known, it appears to be involved in fetal and placental growth and differentiation as well as bone formation (Laurent et al., 1998; Millar et al., 2005). Because of its restricted phyletic distribution, INSL4 was thought to derive from a duplication in the common ancestor of catarrhine primates (Bièche et al., 2003; Park et al., 2008a,b). However, phylogenetic analyses placed the INSL4 clade outside Euarchontoglires (the group that encompasses primates, rodents, and lagomorphs) suggesting a more ancient origin for this gene (Hoffmann and Opazo, 2011).

The presence of the duplicate RLN1 and RLN2 paralogs in apes, the group that includes gibbons, orangutans, gorillas, chimpanzees, and humans, is the second evolutionary innovation in this gene family that is found among primates (Evans et al., 1994). RLN1 is mainly expressed in human deciduas, placenta trophoblast and the prostate gland, however, its functional significance remains unclear (Hansell et al., 1991). RLN2 is a pleiotropic hormone produced during pregnancy in the corpus luteum that displays a variety of biological functions (Hudson et al., 1984), but has also been detected in other nonreproductive tissues where its physiological role is unknown (Gunnersen et al., 1995). Among apes, chimpanzee and human have retained functional copies of both RLN1 and 2, while either one or the two copies of RLN1 and RLN2 have been inactivated in most other species (Evans et al., 1994). The RLN1 and RLN2 paralogs are thought to derive from a duplication of the single copy RLN gene present in the RFLB of the common ancestor of apes (Wilkinson et al., 2005; Park et al., 2008a,b; Hoffmann and Opazo, 2011). The phylogenetic prediction for this hypothesis is that the RLN1 and RLN2 genes of apes would form a monophyletic group embedded within the single copy RLN genes from the RFLB of other primates.

The main objective of this research is to characterize changes in the size and membership composition of the RLN/INSL gene family in primates, reconstruct their duplicative history, and test competing evolutionary scenarios regarding the origin of INSL4 and of the RLN1 and RLN2 paralogs of apes. Results of our comparative genomic analysis reveal an unsuspected level of copy number variation in the RFLB locus of primates, and indicate that a complex combination of lineage specific duplications and deletions, plus the differential retention of ancestral genes have all played a role in shaping the gene complement of the RLN/INSL-like gene family in primates. Further, our phylogenies also suggest that the duplications giving rise to INSL4 and the RLN1 and RLN2 genes are older than previously thought.

2. Material and methods

2.1. DNA sequence data

We used bioinformatic searches to identify the full complement of structural genes in the RLN/INSL-like gene family in nine primate species covering the major taxonomic groups of primates. We included two strepsirrhines (mouse lemur, *Microcebus murinus*; galago, *Otolemur garnettii*); a tarsier (*Tarsius syrichta*); one New World monkey (marmoset, *Callithrix jacchus*), one Old World monkey (rhesus monkey, *Macaca mulatta*), and four different species of apes (orangutan, *Pongo pygmaeus*; gorilla, *Gorilla gorilla*; chimpanzee, *Pan troglodytes*; and human; *Homo sapiens*). DNA sequences from structural genes were obtained from the Ensembl database. In all cases RLN/INSL-like genes were manually annotated by comparing known exon sequences with genomic fragments using the program Blast2seq (Tatusova and Madden, 1999) available from NCBI (www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi). Sequences derived from shorter records based on genomic DNA or cDNA were also included in order to attain a broad and balanced taxonomic coverage (Supplementary Table S1, Supplementary material). Putatively functional genes were characterized by an open intact reading frame with the canonical two exon/one intron structure typical of vertebrate relaxins, whereas pseudogenes were identifiable because of their high sequence similarity to functional orthologs and the presence of inactivating mutations, and/or the lack of exons.

Because mammalian RLN/INSL-like genes have undergone multiple rounds of duplication that have resulted in the presence of sets of tandemly repeated paralogous gene copies in many species, we index each duplicated gene with the symbol T followed by a number that corresponds to the linkage order in the 5′ to 3′ orientation.

2.2. Sequence alignment

To explore the sensitivity of our analyses to changes in the alignment method, nucleotide sequences were aligned based on their amino acid translation 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 corresponding nucleotide alignments were generated using the amino acid alignments as a template with the software PAL2NAL (Suyama et al., 2006). The biological accuracy of alignments was assessed using the software MUMSA (Lassmann and Sonnhammer, 2005), which compares different alignments for a given set of sequences. For each set of alignments, MUMSA provides an Average Overlap Score (AOS) which is a measure of the alignment difficulty, and ranks the alignments based on the Multiple Overlap Score (MOS) score, a measure of alignment quality.

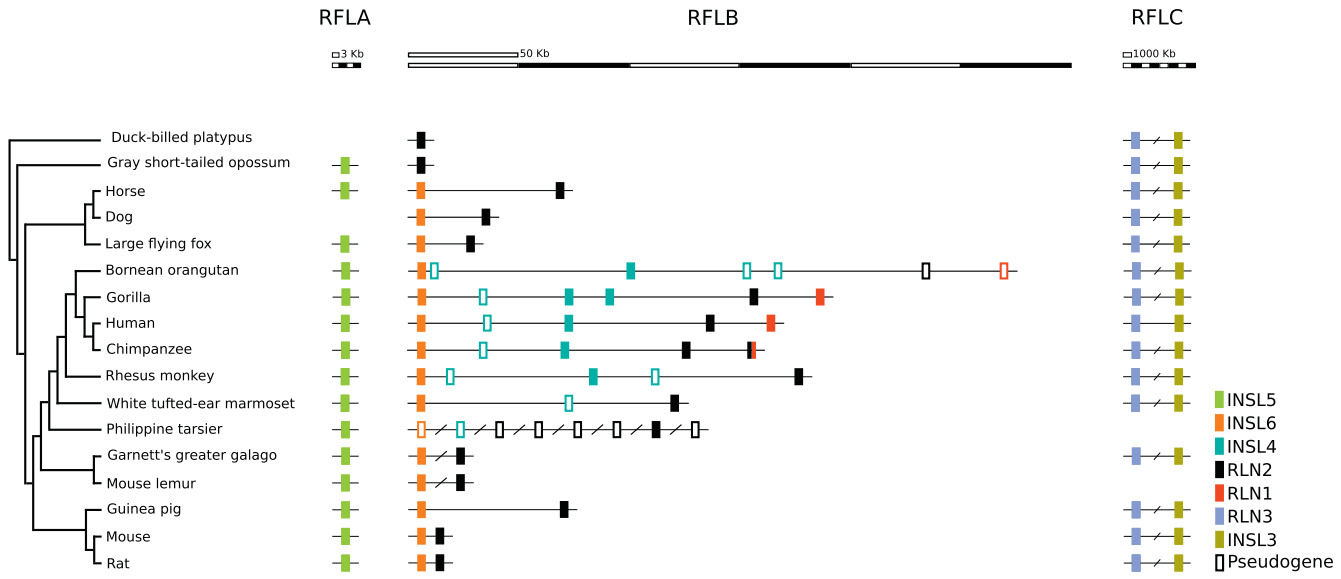

Fig. 1. Genomic structure of the RLN/INSL-like gene family in primates. Diagonal slashes indicate gaps in genomic coverage. Segments containing such gaps are not drawn to scale. The orientation of the clusters is from 5′ (on the left) to 3′ (on the right).

2.3. Phylogenetic inference

Because phylogenetic relationships and orthology among the different paralogs in the RFLA and RFLC loci has been resolved previously (Hoffmann and Opazo, 2011), we focused on reconstructing evolutionary relationships among the mammalian paralogs found in RFLB, with special emphasis on primates. Phylogenetic relationships among the different members of the RLN/INSL gene family 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 different domains of the RLN/INSL genes probably evolve under somewhat different regimes. To accommodate this, each of the domains (signal peptide, and peptides B-, C- and A), were set as independent partitions and were allowed to have and independent model of nucleotide substitution. For each of these partitions, the best fitting model of nucleotide substitution was estimated separately using the "propose model" routine from Treefinder version October 2008 (Supplementary Table S2; Supplementary material; Jobb et al., 2004). For the Bayesian analyses, two simultaneous independent runs were performed for 30×10^6 iterations of a Markov Chain Monte Carlo algorithm, with six 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 using the models of nucleotide substitution previously selected, and support for the nodes was estimated with 1000 bootstrap pseudoreplicates.

3. Results and discussion

3.1. Relaxin/Insulin-like gene repertoire in primates

In order to characterize changes in the RLN/INSL gene complement in primates, we first manually identified the number and nature of the different RLN/INSL-like genes in the genomes of the representatives of the major groups of primates for which genomic information was available. Our sample included apes, Old World monkeys, New World monkeys, tarsiers and strepsirrhines. All primates examined possess a single copy of the INSL5 gene in the RFLA locus (Fig. 1), and single copies of the INSL3 and RLN3 genes in the RFLC locus (Fig. 1). By contrast, the RFLB locus show extensive variation in its gene complement. The number of functional genes in this locus ranges from 1, in the tarsier, to 5 in the gorilla, with differences also observed in the number of pseudogenes present in the different species (Fig. 1). Mouse lemur and galago represent the simplest cluster structure, with 2 functional genes, INSL6 and RLN2, and no pseudogenes, whereas all other primates in the study possess different combinations of genes and pseudogenes in this cluster. In the case of gorilla there are five putatively functional genes and a pseudogene, while in orangutan there are two putatively functional genes and five pseudogenes.

Prior surveys reported that the INSL4 gene was only present among catarrhines as a single copy gene (Bièche et al., 2003; Park et al., 2008a,b; Hoffmann and Opazo, 2011), however, we found that many species posses additional copies of this paralog, which in most cases are clearly identifiable as pseudogenes (Fig. 1). These pseudogenes are characterized by having multiple stop codons, and in all cases other than the INSL4-T3ps of the orangutan and rhesus monkey, the remaining sequence is restricted to the second exon. In addition to finding INSL4 pseudogenes among catarrhines, there are also INSL4-like pseudogenes in the marmoset (*Callithrix jacchus*, a New World Monkey) and the tarsier (*Tarsius syrichta*). These results suggest that the INSL4 gene could be at least traced back to the last common ancestor of haplorhines, the group that includes tarsiers, New World monkeys, Old World monkeys and apes, which is estimated to have occurred between ~77.5 and ~71.1 mya (Steiper and Young, 2009).

In agreement with previous studies, we identified the presence of the duplicated paralogs RLN1 and RLN2 in the RFLB of all apes (Evans et al., 1994; Park et al., 2008a,b; Hoffmann and Opazo, 2011). In the case of the chimpanzee, the 3′ copy, RLN1, appears to be a chimeric gene (Fig. 1), which apparently derives from a gene conversion event in which the first exon of the 5′ copy, RLN2, has overwritten the corresponding fragment of the 3′ copy gene (Evans et al., 1994). The case of the tarsier deserves special attention as 6 copies of the RLN2 gene were found (Fig. 1), however, the order of which could not be established as all of them were annotated in different scaffolds (Supplementary Table S1; Supplementary material). All other primate species posses a single RLN2 gene in the RFLB, (Crawford et al., 1989; Klonisch et al., 2001;

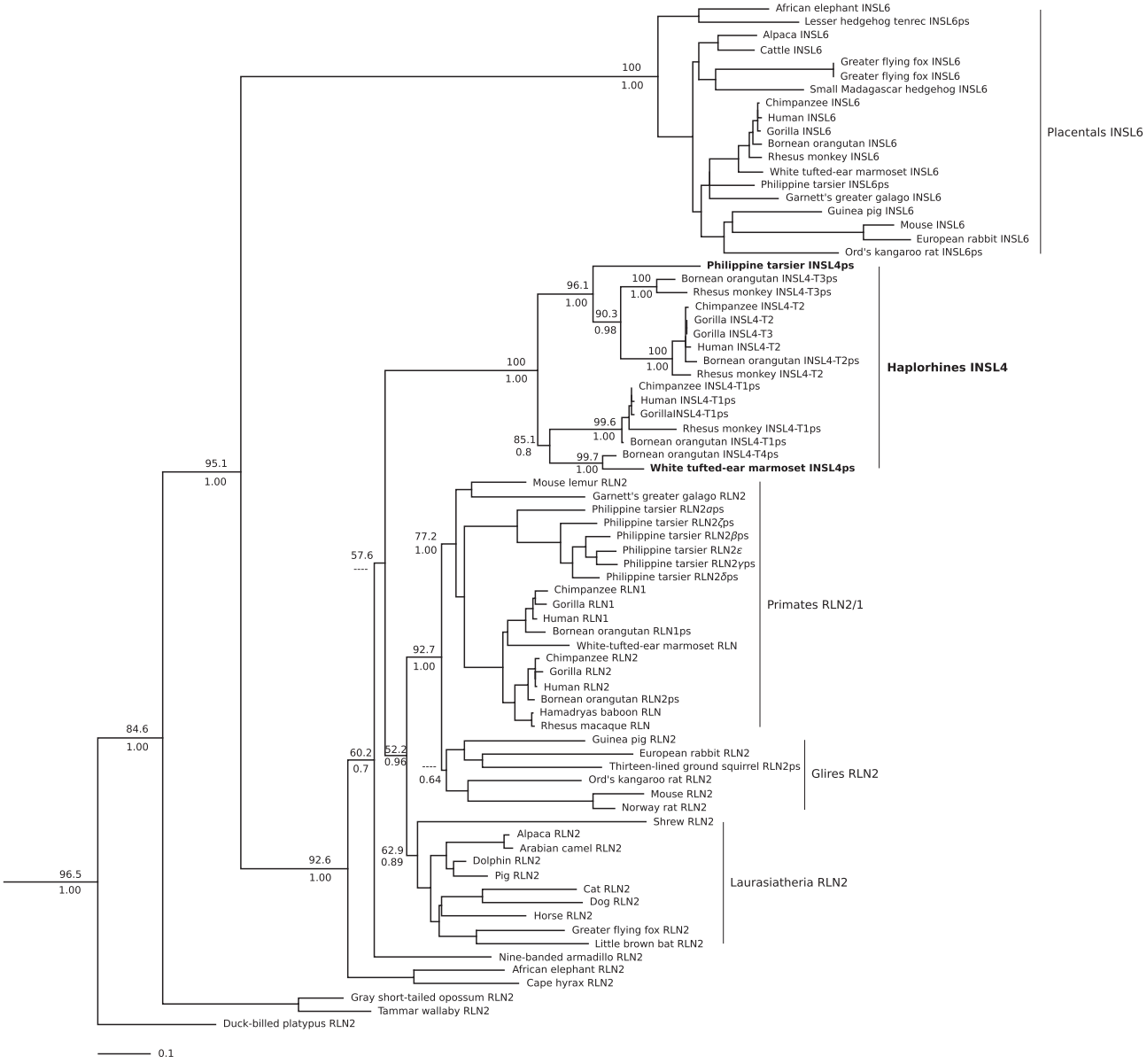

Fig. 2. Maximum likelihood phylogram describing phylogenetic relationships among the RLN/INSL-like genes of tetrapods. The RLN genes from fugu, medaka, frog, chicken, platypus, opossum, and wallaby were used as outgroup. The two INSL4 pseudogenes identified outside the catarrhine primate group are in bold. Numbers above the nodes correspond to maximum likelihood bootstrap support values, and numbers below the nodes correspond to Bayesian posterior probabilities.

Park et al., 2008a,b; Hoffmann and Opazo, 2011). 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 applied to both of them (Shabanpoor et al., 2009). Although these two genes are not orthologs, we will follow this convention to facilitate comparison with the extensive body of functional studies. Thus, the term RLN2 will used to refer to the RLN2 gene of primates, and to the relaxin gene in the RFLB locus from non-primate mammals.

3.2. Phylogeny of the genes in the Relaxin Family Locus B

Because orthology for the genes in the RFLA and RFLC genomic loci has been resolved previously based on both synteny comparisons and phylogenetic reconstructions (Park et al., 2008a,b; Hoffmann and Opazo, 2011), we focused our analyses on the different genes found on the RFLB locus. Our alignment included all RFLB paralogs from primates, plus additional representative sequences from placental mammals, and included RLN2 sequences from fish, frog, chicken, platypus, opossum and wallaby as outgroup sequences. The average overlap score for the alignments is 0.57, and the multiple overlap scores for each individual alignment range from 0.55 to 0.71, 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 Dialign), and compared the likelihood scores of the resulting trees. We then selected the tree with the highest likelihood score, which was obtained with the L-INS-i MAFFT alignment strategy, as our best tree. Results obtained with the other three alignment strategies are reported as Supplementary material.

In all cases the INSL6 and INSL4 clades are placed in strongly supported monophyletic clades (Fig. 2). The INSL6 clade is sister to the clade that includes the INSL4, RLN 1 and RLN2 sequences

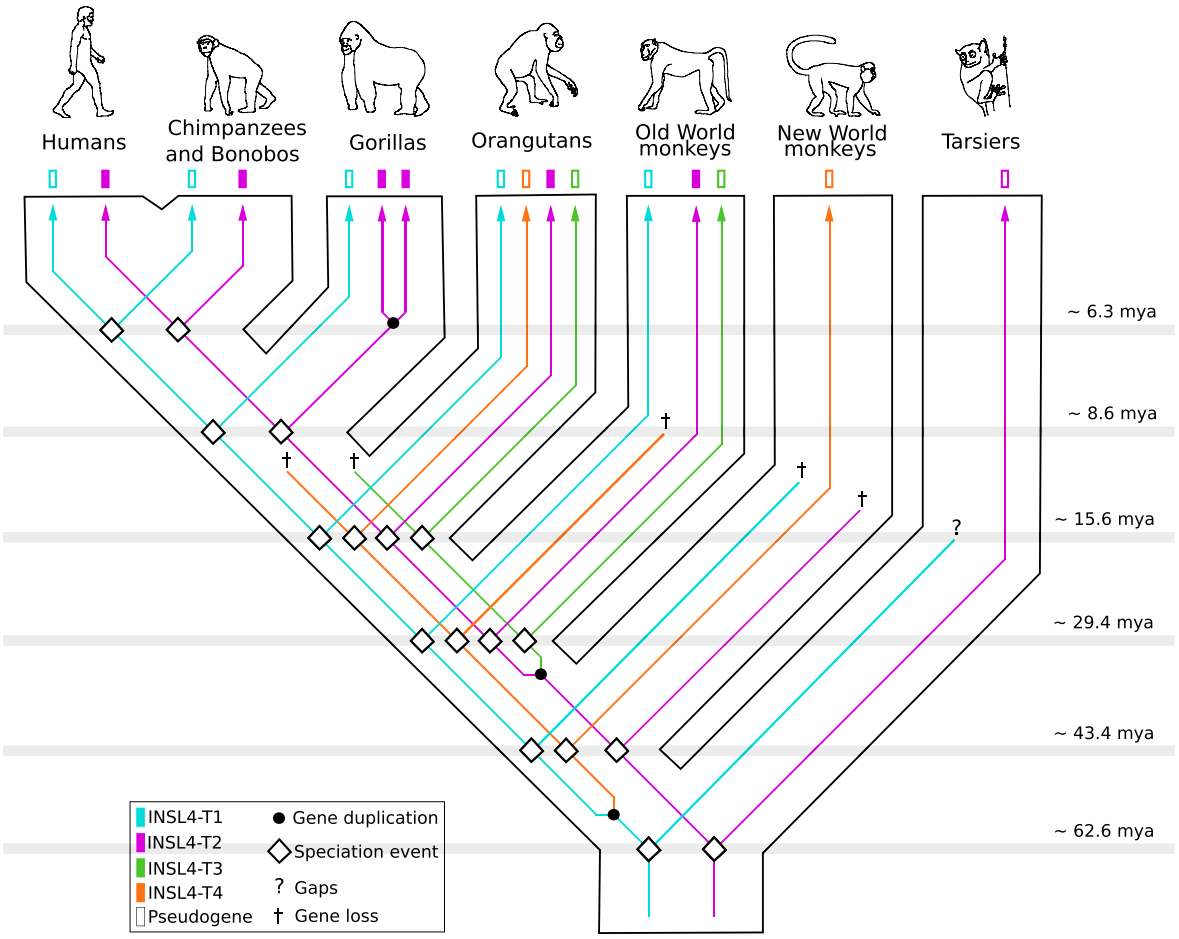

Fig. 3. An evolutionary model for the evolution of the INSL4 gene in primates. The model starts from the last common ancestor of haplorhines, which had duplicate copies of INSL4 gene. The first of these ancestral genes gave rise to the INSL4-T1 pseudogene of catarrhines and to the INSL4-T4 pseudogene of orangutans and New World monkeys. The second ancestral copy of INSL4 became non-functional in tarsiers, was lost in New World monkeys, but underwent an additional duplication in the last common ancestor of catarrhines and gave rise to the INSL4-T2 paralog found in all catarrhines, plus the INSL4-T3 pseudogene of orangutan and Old World monkeys. Finally, in the lineage leading to gorillas the INSL4-T2 gene underwent an additional duplication, and as a result, this is the only species that has two functional copies of INSL4.

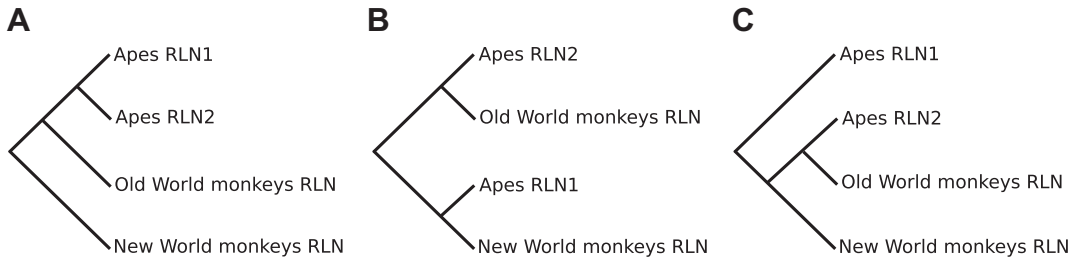

Fig. 4. Schematic representations of alternative hypotheses regarding phylogenetic relationships among the duplicated RLN genes in anthropoid primates. In (A) RLN1 and RLN2 genes arose via duplication of an ancestral RLN gene in the last common ancestor of apes. In (B) the duplication event that gave rise to RLN1 and RLN2 genes predates the radiation of anthropoid primates, although a two gene arrangement was present in the last common ancestor of anthropoid primates, only apes appear to have retained both copies, whereas New and Old World monkeys independently lost complementary gene copies, RLN2 and RLN1 respectively. In (C) the duplication event also predates the radiation of anthropoid primates but this time New and Old World monkeys have independently lost the RLN1 paralog.

from placental mammals (Fig. 2), consistent with the postulated evolutionary origin of INSL6 in the last common ancestor of placental mammals (Park et al., 2008a,b). The INSL4 clade is embedded within RLN1 and RLN2 sequences from other placental mammals confirming that it derives from a RLN-like and not an INSL-like ancestor (Bièche et al., 2003; Hoffmann and Opazo, 2011). In addition, the phylogeny confirms the orthology of all the INSL4 pseudogenes we previously identified.

3.3. Evolutionary history of the INSL4 gene

Because of its restricted phyletic distribution, the INSL4 gene was thought to derive from a duplication specific to the catarrhine lineage (Bièche et al., 2003; Park et al., 2008a,b). However, we found two independent pieces of evidence that are inconsistent with this scenario. First, we identified INSL4 pseudogenes outside catarrhines, in marmoset and tarsier (Figs. 1 and 2). Secondly, the

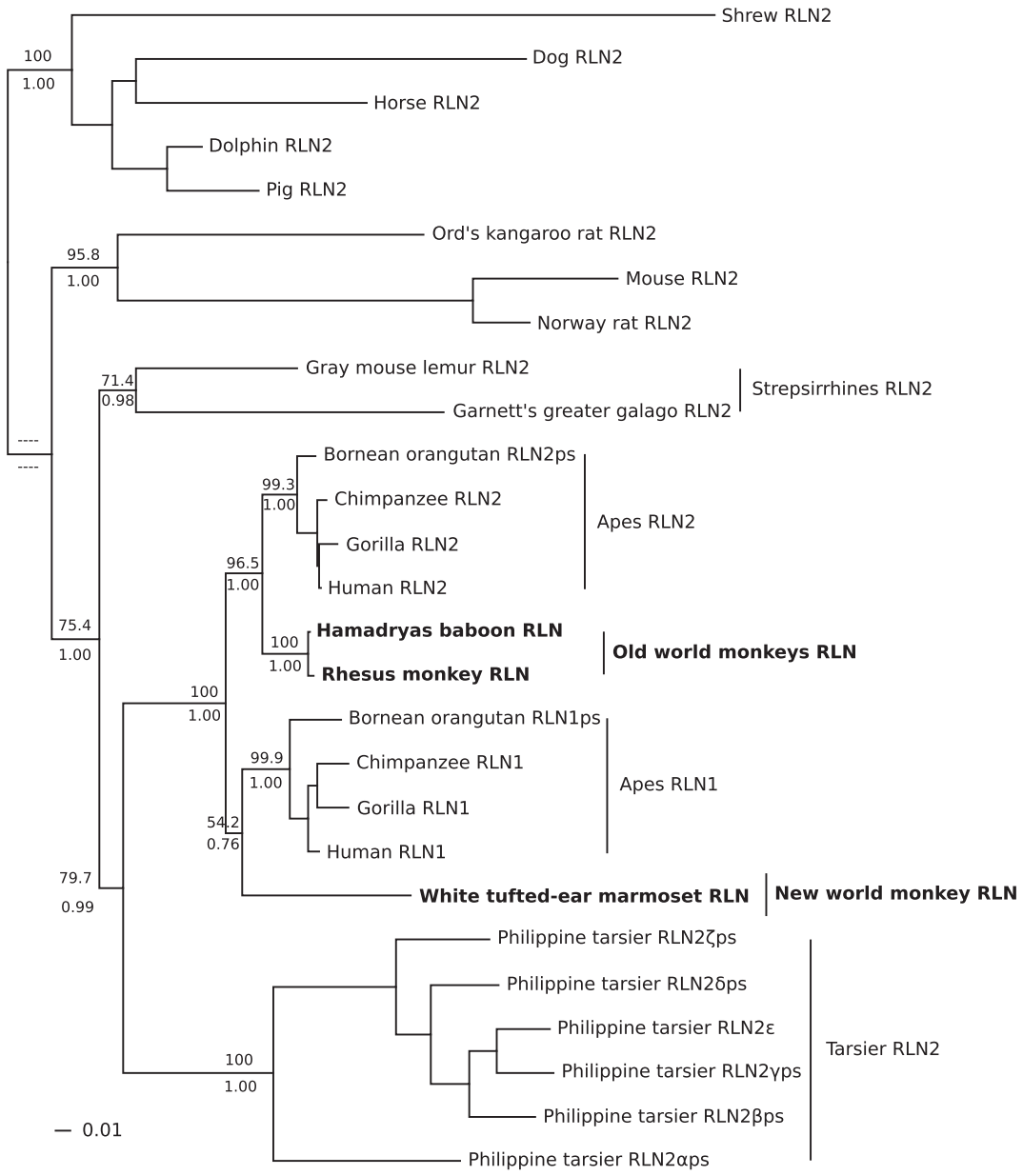

Fig. 5. Maximum likelihood phylogram describing phylogenetic relationships among the RLN1 and RLN2 genes of Boreoeutherian mammals. Sequences were aligned using L-INS-i strategy from Mafft v.6 (Katoh et al., 2009). Numbers above the nodes correspond to maximum likelihood bootstrap support values, and those below the nodes correspond to Bayesian posterior probabilities. Single copy RLN genes found in New and Old World monkeys are in bold.

phylogenetic position of the INSL4 clade is also indicative of an older origin for this gene (Fig. 2). All INSL4 sequences are placed in clade sister to a group composed of all the RLN1 and RLN2 sequences from boreoeutherian mammals (Fig. 2), indicating that the origin of INSL4 predates divergence between euarchontoglires and laurasiatherians. An observation that is statistically supported by the approximately unbiased topology test (Shimodaira, 2002), which rejected the placement of the INSL4 clade as sister to the RLN sequences of any primate group.

Among catarrhines, the INSL4 gene had been found as a single copy gene, with no variation in copy number (Bièche et al., 2003; Park et al., 2008a,b; Hoffmann and Opazo, 2011). However, our survey revealed an unexpected level of copy number variation for INSL4 and identified several pseudogenes outside the catarrhine clade (Fig. 1). Within primates, the different INSL4 paralogs are arranged into five strongly supported clades: (*i*) INSL4-T1ps from catarrhine primates; (*ii*) INSL4-T4ps from orangutan and marmoset; (*iii*) INSL4-T2 from catarrhine primates; (*iv*) INSL4-T3ps from orangutan and rhesus monkey; and (*v*) INSL4ps from tarsier (Fig. 2). The sister group relationship between the first two clades is recovered with strong support, and the tarsier INSL4ps is placed sister to the clade containing the INSL4-T2 genes from catarrhines and the INSL4-T3ps from orangutan and rhesus monkey with strong support too (Fig. 2). This phylogeny indicates that the last common ancestor of haplorhines possessed duplicate INSL4 paralogs (Fig. 3), one of which gave rise to the INSL4-T1 pseudogene of catarrhines and the INSL4-T4 pseudogene of orangutans and New World monkeys (Fig. 3). The second ancestral copy of INSL4 became non-functional in tarsiers, was lost in New World monkeys, but underwent an additional duplication in the last common ancestor of catarrhines and gave rise to the INSL4-T2 paralog found in all catarrhines, plus the INSL4-T3 pseudogene of orangutan and

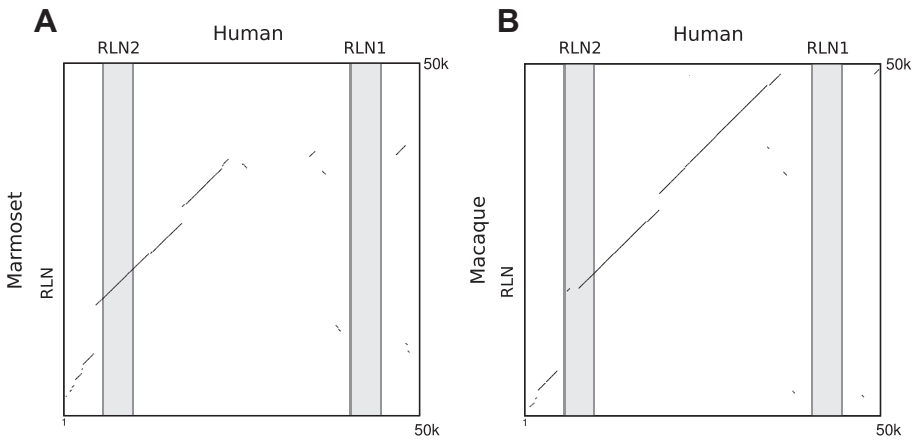

Fig. 6. Dot-plot comparisons between the single copy RLN gene of marmoset (*Callithrix jacchus*, a New World monkey) and macaque (*Macaca mulata*, an Old World monkey), and the RLN1 and RLN2 genes of human (*Homo sapiens*). Dark and light gray vertical lines correspond to exons and introns, respectively. In the case of the human sequence dot plots were based on the complete coding sequence of both RLN genes plus 5 kb of upstream and downstream flanking sequence. In the case of the New World monkeys, we used ~16 kb of upstream flanking sequence and ~30 kb of downstream flanking sequences, and in the case of Old World monkeys we used ~20 kb of upstream flanking sequence and ~30 kb of downstream flanking sequence.

Old World monkeys. Finally, in the lineage leading to gorillas the INSL4-T2 gene underwent an additional duplication, and is the only species that has two functional copies of this gene (Fig. 3).

Certain portions of the genomes, called recombination hotspots, have been shown to be more prone for recombination events than others (Paigen and Petkov, 2010). These locations are characterized by high rates of genetic exchanges that include gene duplications, deletions and translocations, which over enough evolutionary time could be a source of genomic diversity for evolution to act on (Batzer and Deininger, 2002). Because of the prevalence of copy number variation, we expected recombination rate to be higher in RFLB when compared to the two other RFLs. Contrary to our expectations, at least in humans, recombination rate was lowest for the RFLB (1.1 cM/Mb) than either RFLA (1.7 cM/Mb) or RFLC (3.6 cM/Mb). If this pattern holds across primates, our results would suggest that there is the potential for relatively rapid rates of gene gain and loss, even in the absence of high recombination rates.

3.4. Differential retention of RLN genes in anthropoid primates

The evolutionary history of the RLN1-RLN2 clade of genes is also complex. Because they are only found in apes, the duplicated RLN1 and RLN2 paralogs were thought to derive from an ape-specific duplication (Wilkinson et al., 2005; Park et al., 2008a,b; Hoffmann and Opazo, 2011). If this was the case, we would expect to recover a phylogenetic tree in which the single copy RLN gene in the RFLB

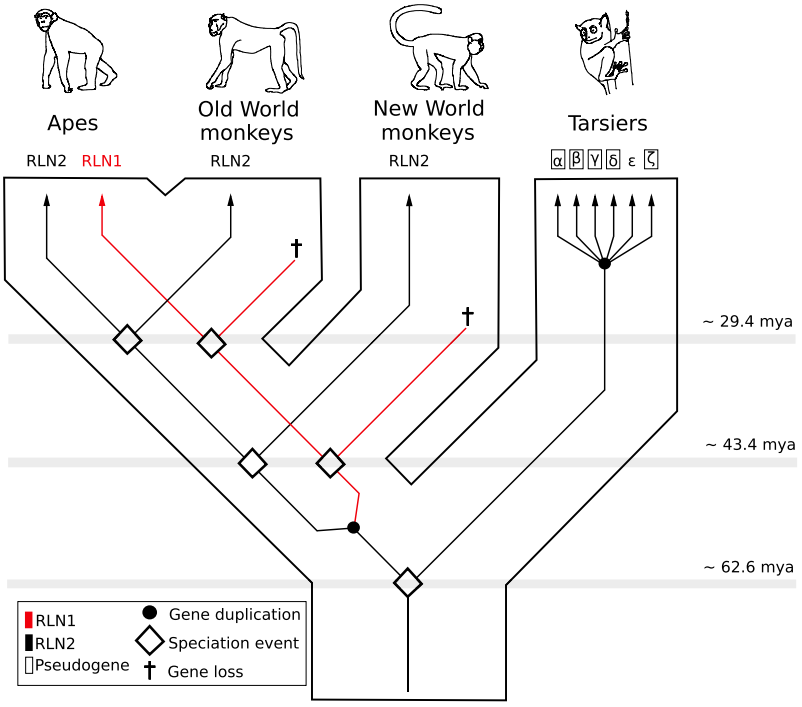

Fig. 7. An evolutionary model for the evolution of the RLN1 and RLN2 genes in anthropoid primates. The model indicates that that the RLN1 and RLN2 paralogs derive from the duplication of a proto-RLN gene in the last common ancestor of anthropoid primates, and not in the last common ancestor of apes as previously postulated. Although a two gene arrangement was present in the last common ancestor of anthropoid primates, only apes appear to have retained both copies, whereas New and Old World monkeys independently lost the RLN1 paralog.

of Old World monkeys was sister to the clade containing reciprocally monophyletic RLN1 and RLN2 sequences of apes. Additionally, the single copy RLN gene found in the RFLB of New World monkeys, which is labeled as RLN2 in Fig. 2, would be expected to fall sister to the catarrhine clade (Fig. 4A). However, our results are not compatible with these predictions: our phylogenies group the single copy RLN gene of marmoset with the RLN1 clade of apes with weak support, and also group the single copy RLN of Old World monkeys with the RLN2 clade of apes with strong support (Fig. 5). This pattern suggests that the duplicative history of the RLN1 and RLN2 paralogs of apes is older than expected, as the phylogeny in Fig. 5 indicates that these genes originated via duplication of a proto-RLN gene in the last common ancestor of anthropoid primates. Thus, the presence of single copy RLN genes in the RFLB of marmoset and Old World monkeys would reflect the independent loss of one of the two ancestral paralogs (Fig. 4B and C). This finding is supported by the approximately unbiased test (Shimodaira, 2002) which rejected the topology that is consistent with an ape specific duplication of the proto-RLN gene (Fig. 4A; $P < 10^{-4}$).

The strongly supported relationship between the single copy RLN of the RFLB of Old World monkeys and the RLN2 clade of apes suggests that Old World monkeys have all retained a copy of the RLN2 paralog (Fig. 5). However, phylogenetic analyses could not conclusively resolve the orthology of the single RLN gene of marmoset. The availability of genomic sequences allowed us to make comparisons involving sequences from non-coding flanking and intronic regions in addition to coding sequences. In principle comparisons of the single copy RLN gene of New and Old World monkeys and the duplicated paralogs of apes should allow us to see regions of similarity between species. Dot-plot comparisons revealed that the single copy RLN gene of the RFLB of New and Old World monkeys are more similar to the RLN2 gene of humans. In addition to the coding sequence, high similarity regions include the intron and flanking sequences (Fig. 6), suggesting that marmoset would have also retained the RLN2 paralog. Thus, our results would indicate that that the RLN1 and RLN2 paralogs derive from the duplication of a proto-RLN gene in the last common ancestor of anthropoid primates, and not in the last common ancestor of apes as previously postulated (Wilkinson et al., 2005; Park et al., 2008a,b; Hoffmann and Opazo, 2011). Although a two gene arrangement was present in the last common ancestor of anthropoid primates, only apes appear to have retained both copies, whereas New and Old World monkeys independently lost the RLN1 paralog (Fig. 7).

To investigate the potential role of positive selection in the retention of the duplicate RLN1 and RLN2 paralogs, we performed branch and branch-site tests using the maximum likelihood codon substitution model as implemented in PAML v4.4 (Yang, 2007). Our results do not reveal evidence of positive Darwinian selection in any of the tests comparing the RLN paralogs from the RFLB of primates (Supplementary Table S3; Supplementary material). Thus, it seems that adaptive changes in the coding sequences of these genes do not appear to have played a strong role in the functional differentiation of these genes. The possibility remains that changes in non-coding sequences were driven by Darwinian selection.

4. Conclusions

In this study we show that the RLN/INSL-like gene family in primates had a more dynamic evolutionary history than previously thought. We identified several gene gains and losses consistent with the predictions of the birth-and-death model of gene family evolution (Nei and Rooney, 2005). We also found that the differential retention of relatively old paralogs played a strong role in shaping the complement of RNL/INLS genes in the RFLB of primates. Specifically, we found that the duplication giving rise to the INSL4 paralogs, and the duplication giving rise to the RLN1 and RLN2 duplicate paralogs from apes are both older than expected. As a result of the combination of lineage-specific duplications and the differential retention of relatively old duplicates (as summarized in Figs. 3 and 7), the repertoire of genes present in the RFLB was much more variable among primates relative to other mammals (Fig. 1). These findings are in good agreement with previous research studies that show extensive genomic changes in anthropoid primates (Maston and Ruvolo, 2002; Grossman et al., 2004; Koike et al., 2007; Hahn et al., 2007; Hoffmann et al., 2008; Than et al., 2009). Given the role of members of the RLN/INSL-like gene family in reproductive biology (Ivell, 1997; Bathgate et al., 2003; Sherwood, 2004; Shabanpoor et al., 2009), we speculate that the observed changes in gene complement in this gene family might be associated to changes in reproductive features, such as longer gestation periods.

Acknowledgments

This paper is dedicated to the memory of Morris Goodman (1925–2010). We thank Wen-Hsiung Li and two anonymous reviewers for helpful comments on the manuscript. 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). FGH acknowledges grant support from the National Science Foundation (EPS-0903787).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2012.02.011.

References

Adham, I.M., Burkhardt, E., Benahmed, M., Engel, W., 1993. Cloning of a cDNA for a novel insulin- like peptide of the testicular Leydig cells. J. Biol. Chem. 268, 26668–26672.

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

Batzer, M.A., Deininger, P.L., 2002. Alu repeats and human genomic diversity. Nat. Rev. Genet. 3, 370–379.

Bièche, I., Laurent, A., Laurendeau, I., Duret, L., Giovangrandi, Y., Frendo, 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, S.J., Steiner, D.F., 2000. Insulin through the ages: phylogeny of a growth promoting and metabolic regulatory hormone. Am. Zool. 40, 222–231.

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.

Crawford, R.J., Hammond, V.E., Roche, P.J., Johnston, P.D., Tregear, G.W., 1989. Structure of rhesus monkey relaxin predicted by analysis of the single-copy rhesus monkey relaxin gene. J. Mol. Endocrinol. 3, 169–174.

Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3, e314.

Do, C.B., Mahabhashyam, M.S., Brudno, M., Batzoglou, S., 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res. 15, 330–340.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Evans, B.A., Fu, P., Tregear, G.W., 1994. Characterization of two relaxin genes in the chimpanzee. J. Endocrinol. 140, 385–392.

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 Evol. Biol. 9, 293.

Grossman, L.I., Wildman, D.E., Schmidt, T.R., Goodman, M., 2004. Accelerated evolution of the electron transport chain in anthropoid primates. Trends Genet. 20, 578–585.

Gunnersen, J.M., Crawford, R.J., Tregear, G.W., 1995. Expression of the relaxin gene in rat tissues. Mol. Cell. Endocrinol. 110, 55–64.
Hahn, M.W., Demuth, J.P., Han, S.G., 2007. Accelerated rate of gene gain and loss in primates. Genetics 177, 1941–1949.
Hansell, D.J., Bryant-Greenwood, G.D., Greenwood, F.C., 1991. Expression of the human relaxin H1 gene in the decidua, trophoblast, and prostate. J. Clin. Endocrinol. Metab. 72, 899–904.
Hoffmann, F.G., Opazo, J.C., 2011. Evolution of the relaxin/insulin-like gene subfamily in placental mammals: implications for its early evolution. J. Mol. Evol. 72, 72–79.
Hoffmann, F.G., Opazo, J.C., Storz, J.F., 2008. Rapid rates of lineage-specific gene duplication and deletion in the alpha-globin gene family. Mol. Biol. Evol. 25, 591–602.
Hudson, P., Haley, J., John, M., Cronk, M., Crawford, R., Haralambidis, J., Tregear, G., Shine, J., Niall, H., 1983. Structure of a genomic clone encoding biologically active human relaxin. Nature 307, 628–631.
Hudson, P., John, M., Crawford, R., Haralambidis, J., Scanlon, D., Gorman, J., Tregear, G., Shine, J., Niall, H., 1984. Relaxin gene expression in human ovaries and the predicted structure of a human preprorelaxin by analysis of cDNA clones. EMBO J. 3, 2333–2339.
Ivell, R., 1997. Biology of the relaxin-like factor (RLF). Rev. Reprod. 2, 133–138.
Jobb, G., von Haeseler, A., 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.
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.
Koike, C., Uddin, M., Wildman, D.E., Gray, E.A., Trucco, M., Starzl, T.E., Goodman, M., 2007. Functionally important glycosyltransferase gain and loss during catarrhine primate emergence. Proc. Natl. Acad. Sci. USA 104, 559–564.
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, E.L., 2005. Automatic assessment of alignment quality. Nucleic Acids Res. 33, 7120–7128.
Lassmann, T., Frings, O., Sonnhammer, E.L., 2009. Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features. Nucleic Acids Res. 37, 858–865.
Laurent, A., Rouillac, C., Delezoide, A.L., Giovangrandi, Y., Vekemans, M., Bellet, D., Abitol, M., Vidaud, M., 1998. Insulin-like 4 (INSL4) gene expression in human embrionic and trophoblastic tissues. Mol. Reprod. Dev. 51, 123–129.
Maston, G.A., Ruvolo, M., 2002. Chorionic gonadotropin has a recent origin within primates and an evolutionary history of selection. Mol. Biol. Evol. 19, 320–335.
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.
Millar, L., Streiner, N., Webster, L., Yamamoto, S., Okabe, R., Kawamata, T., Shimoda, J., Bullesbach, E., Schwabe, C., Bryant-Greenwood, G., 2005. Early placental insulin-like protein (INSL4 or EPIL) in placental and fetal membrane growth. Biol. Reprod. 73, 695–702.
Nei, M., Rooney, A.P., 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39, 121–152.
Notredame, C., Higgins, D.G., 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.
Olinski, R.P., Dahlberg, C., Thorndyke, M., Hallböök, F., 2006a. Three insulin–relaxin-like genes in *Ciona intestinalis*. Peptides 27, 2535–2546.
Olinski, R.P., Lundin, L.G., Hallböök, F., 2006b. Conserved synteny between the *Ciona* genome and human paralogons identifies large duplication events in the molecular evolution of the insulin–relaxin gene family. Mol. Biol. Evol. 23, 10–22.
Paigen, K., Petkov, P., 2010. Mammalian recombination hot spots: properties, control and evolution. Nat. Rev. Genet. 11, 221–233.
Park, J.I., Chang, C.L., Hsu, S.Y., 2005. New Insights into biological roles of relaxin and relaxin-related peptides. Rev. Endocr. Metab. Disord. 6, 291–296.
Park, J.I., Semyonov, J., Chang, C.L., Yi, W., Warren, W., Hsu, S.Y., 2008a. Origin of INSL3- mediated testicular descent in therian mammals. Genome Res. 18, 974–985.
Park, J.I., Semyonov, J., Yi, W., Chang, C.L., Hsu, S.Y., 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.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Shabanpoor, F., Separovic, F., Wade, J.D., 2009. The human insulin superfamily of polypeptide hormones. In: Litwack, G. (Ed.), Insulin and IGFs. Academic Press, pp. 1–31.
Sherwood, O.D., 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, M.E., Young, N.M., 2009. In: Hedges, S.B., Kumar, S. (Eds.), Primates (Primates). Oxford University Press, pp. 482–486.
Subramanian, A.R., 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, W609–612.
Tatusova, T.A., Madden, T.L., 1999. BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. 174, 247–250.
Than, N.G., Romero, R., Goodman, M., Weckle, A., Xing, J., Dong, Z., Xu, Y., Tarquini, F., Szilagyi, A., Gal, P., Hou, Z., Tarca, A.L., Kim, C.J., Kim, J.S., Haidarian, S., Uddin, M., Bohn, H., Benirschke, K., Santolaya-Forgas, J., Grossman, L.I., Erez, O., Hassan, S.S., Zavodszky, P., Papper, Z., Wildman, D.E., 2009. A primate subfamily of galectins expressed at the maternal-fetal interface that promote immune cell death. Proc. Natl. Acad. Sci. USA 106, 9731–9736.
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., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24, 1586–1591.
Yegorov, S., Good-Avila, S.V., Parry, L., Wilson, B.C., 2009. Relaxin family genes in humans and teleosts. Ann. N. Y. Acad. Sci. 1160, 42–44.