

## Washington University School of Medicine Digital Commons@Becker

---

### Open Access Publications

---

2008

# Origin of INSL3-mediated testicular descent in therian mammals

Jae-Il Park

*Stanford University School of Medicine*

Jenia Semyonov

*Stanford University School of Medicine*

Chia Lin Chang

*Chang Gung University School of Medicine*

Wei Yi

*Stanford University School of Medicine*

Wesley Warren

*Washington Univeristy School of Medicine*

*See next page for additional authors*

Follow this and additional works at: [http://digitalcommons.wustl.edu/open\\_access\\_pubs](http://digitalcommons.wustl.edu/open_access_pubs)

---

### Recommended Citation

Park, Jae-Il; Semyonov, Jenia; Chang, Chia Lin; Yi, Wei; Warren, Wesley; and Hsu, Sheau Yu Teddy, "Origin of INSL3-mediated testicular descent in therian mammals." *Genome Research*.18, 974-985. (2008).

[http://digitalcommons.wustl.edu/open\\_access\\_pubs/1982](http://digitalcommons.wustl.edu/open_access_pubs/1982)

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact [engeszer@wustl.edu](mailto:engeszer@wustl.edu).

---

**Authors**

Jae-Il Park, Jenia Semyonov, Chia Lin Chang, Wei Yi, Wesley Warren, and Sheau Yu Teddy Hsu



## Origin of INSL3-mediated testicular descent in therian mammals

Jae-Il Park, Jenia Semyonov, Chia Lin Chang, et al.

*Genome Res.* 2008 18: 974-985 originally published online May 7, 2008

Access the most recent version at doi:[10.1101/gr.7119108](https://doi.org/10.1101/gr.7119108)

---

### Supplemental Material

<http://genome.cshlp.org/content/suppl/2008/05/08/gr.7119108.DC1.html>

### Related Content

**Retroposed SNOfall A mammalian-wide comparison of platypus snoRNAs**

Jürgen Schmitz, Anja Zemann, Gennady Churakov, et al.

[Genome Res. June , 2008 18: 1005-1010](#) **Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes**

Frédéric Veyrunes, Paul D. Waters, Pat Miethke, et al.

[Genome Res. June , 2008 18: 965-973](#) **Defensins and the convergent evolution of platypus and reptile venom genes**

Camilla M. Whittington, Anthony T. Papenfuss, Paramjit Bansal, et al.

[Genome Res. June , 2008 18: 986-994](#) **Conservation of small RNA pathways in platypus**

Elizabeth P. Murchison, Pouya Kheradpour, Ravi Sachidanandam, et al.

[Genome Res. June , 2008 18: 995-1004](#)

### References

This article cites 54 articles, 25 of which can be accessed free at:

<http://genome.cshlp.org/content/18/6/974.full.html#ref-list-1>

Articles cited in:

<http://genome.cshlp.org/content/18/6/974.full.html#related-urls>

### Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see

<http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:

<http://genome.cshlp.org/subscriptions>

# Origin of INSL3-mediated testicular descent in therian mammals

Jae-Il Park,<sup>1</sup> Jenia Semyonov,<sup>1</sup> Chia Lin Chang,<sup>2</sup> Wei Yi,<sup>1</sup> Wesley Warren,<sup>3</sup> and Sheau Yu Teddy Hsu<sup>1,4</sup>

<sup>1</sup>Division of Reproductive Biology, Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, California 94305-5317, USA; <sup>2</sup>Chang Gung University School of Medicine, and Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital, Tao-Yuan 333, Taiwan; <sup>3</sup>Genome Sequencing Center, Washington University School of Medicine, St. Louis, Missouri 63108, USA

Testicular descent is a unique physiological adaptation found in therian mammals allowing optimal spermatogenesis below core body temperature. Recent studies show that INSL3, produced by Leydig cells, and its receptor LGR8 (RXFP2) are essential for mediating the transabdominal phase of testicular descent during early development. However, the origin and genetic basis for this physiological adaptation is not clear. Using syntenic mapping and the functional characterization of contemporary and resurrected relaxin family hormones, we show that derivation of INSL3-mediated testicular descent involved the duplication of an ancestral *RLN3*-like gene that encodes an indiscriminate ligand for LGR7 (RXFP1) and LGR8. This event was followed by acquisition of the LGR7-selective characteristics by a daughter gene (*RLN3*) prior to the evolution of the common ancestor of monotremes, marsupials, and placentals. A subsequent mutation of the other daughter gene (*INSL3*) occurred before the emergence of therian mammals, which then led to the derivation of the reciprocal LGR8-specific characteristics of INSL3. The stepwise evolution of these independent signaling pathways through gene duplication and subsequent divergence is consistent with Darwinian theory of selection and adaptation, and the temporal proximity suggests an association between these genetic events and the concurrent evolution of testicular descent in ancestral therian mammals.

[Supplemental material is available online at [www.genome.org](http://www.genome.org).]

One of the most interesting questions in biology concerns how a novel physiological process evolves in mammals and what are the underlying genetic mechanisms. Recently, it was shown that transabdominal testicular descent is under the direct influence of the INSL3/LGR8 (RXFP2) signaling pathway and is dependent on proper growth of the gubernaculum primordia and the caudal genitor-inguinal ligament (Nef and Parada 1999; Zimmermann et al. 1999; Bogatcheva et al. 2003; Kamat et al. 2004; Wilhelm and Koopman 2006). Curiously, the so-called gubernaculum of Hunter, which extends from the caudal pole of the mesonephric remnants into the inguinal abdominal wall, is common to placentals and marsupials but is absent in monotremata mammals and non-mammalian vertebrates (Griffiths et al. 1993; van der Schoot 1996; Werdelin and Nilsson 1999; Coveney et al. 2002; Mackay et al. 2004; Wilhelm and Koopman 2006). To date, how testicular descent, and the LGR8-specific INSL3 signaling pathway that underlies this therian mammal-specific adaptation, evolved is not clear.

The generation and subsequent fixation of duplicated genes with previously nonexistent characteristics has been shown to be a principal source of new genes and is associated with great leaps in evolution (Long et al. 2003; Chen et al. 2007). By comparative genomic analyses and point mutations, footprints of molecular adaptations through gene duplication and the subsequent divergence of daughter genes have been found in a handful of duplicated enzymes and receptors involving visual reception, oxygen

sensing, steroid hormone signaling, and digestive tract RNA degradation (Yokoyama 2002; Zhang et al. 2002; Long et al. 2003; Thornton 2004; Bridgham et al. 2006). These studies on the cause-effect relationships of gene duplication and biochemical adaptation have focused on proteins that interact with external environmental signals (oxygen for hemoglobins and photons for rhodopsin) or with substrates and ligands not encoded by a gene (foreign RNA for ribonucleases and steroids for steroid receptors). Thus far, no clear case has been made on the cause-effect relationship between the duplication of a polypeptide ligand-mediated signaling network and the evolution of a mammal-specific physiological process (Taylor and Raes 2004; Hughes 2005).

Recently, we identified and characterized relaxin family peptides (human relaxin II [RLN2], INSL3, and relaxin3 [RLN3]) as the cognate ligands for two orphan G protein-coupled receptors (GPCRs), LGR7 (RXFP1) and LGR8 (Hsu et al. 2002; Kumagai et al. 2002; Sudo et al. 2002). Whereas RLN3 and INSL3 are selective agonists for LGR7 and LGR8, respectively, RLN2 activates both LGR7 and LGR8. In addition to signaling through LGR7, RLN3 was shown to be capable of activating two phylogenetically distant GPCRs, GPCR135 (RXFP3) and GPCR142 (RXFP4), but the physiological importance of these interactions remains to be investigated (Liu et al. 2003a,b). Among these ligand-receptor pairs, RLN2/LGR7 signaling is important for the regulation of a subset of therian mammal-specific reproductive processes including parturition through a softened cervix and the development of nipples in females (Zhao et al. 1999; Feng et al. 2005). Disruption of the *relaxin* (*Rln1*) or *Lgr7* gene in mice led to defects in mammary gland and nipple development and an impaired par-

#### <sup>4</sup>Corresponding author.

E-mail [teddyhsu@stanford.edu](mailto:teddyhsu@stanford.edu); fax (650) 725-7102.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.7119108>.

ture process (Zhao et al. 1999; Kamat et al. 2004; Krajnc-Franken et al. 2004). On the other hand, it was shown that expression of *RLN3* and *LGR7* overlaps in several brain regions and that *RLN3/LGR7* signaling could be involved in stress and/or memory regulation (Burazin et al. 2005; Ma et al. 2005; Tanaka et al. 2005). In contrast, *INSL3/LGR8* signaling was found to be essential for the mediation of testicular descent. Deletion of the *Insl3* or *Lgr8* gene prevents gubernaculum growth and causes bilateral cryptorchidism in male mice (Nef and Parada 1999; Zimmermann et al. 1999; Bogatcheva et al. 2003), whereas overexpression of *INSL3* results in ovary descent to the base of the abdominal cavity due to formation of male-like gubernaculum structures (Adham et al. 2002). These data established that the physiological role of *INSL3* is mediated exclusively by *LGR8* and that the development of this specific signaling pathway is critical to the evolution of testicular descent in placental mammals. However, it is not clear how *INSL3/LGR8*-mediated testicular descent evolved in placental and marsupial mammals. Here, based on syntenic mapping and the functional characterization of contemporary and resurrected relaxin family hormones, we identify *INSL3* as a signature protein of therian mammals and describe each of the three critical molecular steps associated with the evolution of testicular descent in therian mammals: (1) gene duplication, (2) diversification of the functional components in one of the daughter genes (*RLN3*), and (3) independent and complementary diversification of the other daughter gene (*INSL3*).

## Results

### Relaxin family genes in vertebrates evolved from three independent chromosomal loci in a common ancestor of vertebrates

In humans, there are seven relaxin family genes; however, the number of their orthologs varies greatly among vertebrates (Hsu 1999; Wilkinson et al. 2005b). Several studies that relied on DNA sequence analysis to delineate the evolutionary relationships of relaxin family genes suggested that the diverging inventory of relaxin family genes in different classes of vertebrates could be a result of lineage-specific and genome-wide duplications (Wilkinson et al. 2005b,c). As evidenced by analysis of synonymous ( $d_s$ ) and nonsynonymous ( $d_n$ ) substitution rates, *RLN1*, *RLN2*, *INSL4*, and *INSL6* genes, clustered on a 170-kb span of human chromosome 9p24 in tandem (Fig. 1A), were found to be under strong positive selection (Wilkinson et al. 2005b), suggesting that these genes are likely derived from lineage-specific segmental duplications as a result of nonallelic homologous recombinations. In contrast, *RLN3* was constrained by purifying selection (Wilkinson et al. 2005b). In addition, these studies indicated that *INSL3* shares a close relatedness to *RLN1/RLN2/INSL4/INSL6* as compared with *RLN3* and *INSL5*. However, the exact evolutionary path of relaxin family genes in vertebrates remained to be clarified because of heterogeneities in the evolutionary rate as a result of repeated gene duplications and subsequent divergence (Thornton and Kolaczowski 2005; Wilkinson et al. 2005b,c).

To resolve this conundrum in the phylogenetic reconstruction of relaxin family genes, we analyzed syntenic loci encoding these genes among vertebrates. Mapping of syntenic loci indicated that these genes evolved from three independent loci including (1) *INSL5* or relaxin family locus A (*RFLA*) corresponding to human chromosome 1p31; (2) *RLN1/RLN2/INSL4/INSL6* or relaxin family locus B (*RFLB*) corresponding to human chromosome

9p24; and (3) *RLN3/INSL3* or relaxin family locus C (*RFLC*) corresponding to human chromosome 19p13 in the most recent common ancestor of vertebrates (Fig. 1A). This result is in direct contrast to the inference based on phylogenetic tree analyses that separated tandem-duplicated paralogs at the same locus onto far-flung branches in the tree. For example, an analysis based on 65 relaxin family peptides from 17 species using the maximum likelihood method indicated the presence of six major branches that diverged deep in vertebrate lineage, and orthologous genes derived from each of these three loci were split into at least two distant branches (Fig. 1B; Supplemental Table 1).

Among these three loci, *RFLA* remains a single gene locus from teleosts to humans (Figs. 1A, 2A). However, unlike tetrapods that encode a single *INSL5* gene, the genomes of teleosts including zebrafish (*Danio rerio*) and pufferfish (*Takifugu rubripes* and *Tetraodon nigroviridis*) encode a pair of co-orthologs on the two syntenic *RFLA* regions derived from whole-genome duplication (WGD) that occurred before the divergence of teleosts and osteoglossomorphs more than 230–350 million years ago (Mya) (Jailon et al. 2004; Crollius and Weissenbach 2005). In contrast, *RFLB* in noneutherian species contains only a single orthologous gene whereas *RFLB* in hominids encodes up to four paralogs (one for each ortholog of human *RLN1*, *RLN2*, *INSL4*, and *INSL6*) (Figs. 1A, 2A). Among these genes, *RLN1* and *INSL4* evolved after the emergence of primates. On the other hand, two paralogous relaxin family genes were identified on *RFLC* (*RFLCI* and *RFLCII*) that are syntenic to human *RLN3* and *INSL3*, respectively, in clawed frog and all mammalian species analyzed (Figs. 1A, 2A). In teleosts, similar to *RFLA*, one co-ortholog was found on each of the two WGD-derived syntenic *RFLC* regions of pufferfish. Thus, the most parsimonious evolutionary path for relaxin family genes in vertebrates originated in three separate genes (*AncRFLA*, *AncRFLB*, and *AncRFLC*) in the most recent common ancestor of tetrapods and teleosts >450 Mya (Fig. 2A).

### INSL3 is a therian mammal-specific signature protein

In addition to revealing the origin of relaxin family genes in vertebrates, syntenic mapping analyses indicated that the neighboring *RLN3* (or *RFLCI*) and *INSL3* (or *RFLCII*) on human chromosome 19p13 were likely derived from a segmental duplication encompassing ~5–7 megabases of a genomic fragment that included ancestors of *EMR*, *RAB*, *MYO*, *PDE4*, *RAB3*, *JUN*, *KLF*, *CALR*, and *ILR* family genes, and that duplication of the *AncRFLC* gene predates the radiation of placentals, marsupials, and monotremes but postdates the separation of teleosts and tetrapods (Fig. 2A,B).

Importantly, we found that only therian mammals contain a pair of orthologous *INSL3* and *RLN3* genes on *RFLC* (Fig. 2B). The two *RFLC* genes in platypus and clawed frog (*RFLCI* and *RFLCII*) encode peptides with a greater similarity to *RLN3* from therian mammals (Fig. 2; Supplemental Table 1). These findings indicate that *INSL3* emerged from an *RLN3*-like ancestor gene on *RFLC*. Because monotremata mammals are similar to reptiles and other non-mammalian tetrapods in having testes that remain close to the caudal pole of the kidneys throughout life (Fig. 2A), but without a gubernaculum structure or testicular descent (van der Schoot 1996), testicular descent likely emerged in therian mammals in parallel with evolution of the therian mammal-specific *INSL3* following separation with monotremata mammals.

Given the finding that *RLN3* and *INSL3* preferentially interact with *LGR7* and *LGR8*, respectively, it is not obvious how the

differential interaction between these two pairs of ligands and receptors could have evolved after duplication of the *AncRFLC* gene. This enigma is complicated by the asymmetric divergence in their sequences; *RLN3* is constrained by strong purifying selection and retains the ancestral sequence features found in species of basal taxonomy, whereas *INSL3*, similar to *RLN1/RLN2/INSL4/INSL6*, constitutes many radical substitutions (Wilkinson et al. 2005c). Because LGR7 and LGR8 originated before the evolution of vertebrates, and highly conserved orthologous LGR7 and LGR8 are found in all classes of vertebrates (>75% amino acid similarity) is shared by human and teleost LGR7 and LGR8 orthologs), with the exception of chicken, which lacks an *RFLC* gene (Hsu et al. 2000; Wilkinson et al. 2005a; Semyonov et al. 2008) (Fig. 1A), we hypothesized that *INSL3* and *RLN3* could have evolved through a neofunctionalization process in which each of the duplicated daughter genes on *RFLC* evolved new fea-

lution of vertebrates, and highly conserved orthologous LGR7 and LGR8 are found in all classes of vertebrates (>75% amino acid similarity) is shared by human and teleost LGR7 and LGR8 orthologs), with the exception of chicken, which lacks an *RFLC* gene (Hsu et al. 2000; Wilkinson et al. 2005a; Semyonov et al. 2008) (Fig. 1A), we hypothesized that *INSL3* and *RLN3* could have evolved through a neofunctionalization process in which each of the duplicated daughter genes on *RFLC* evolved new fea-

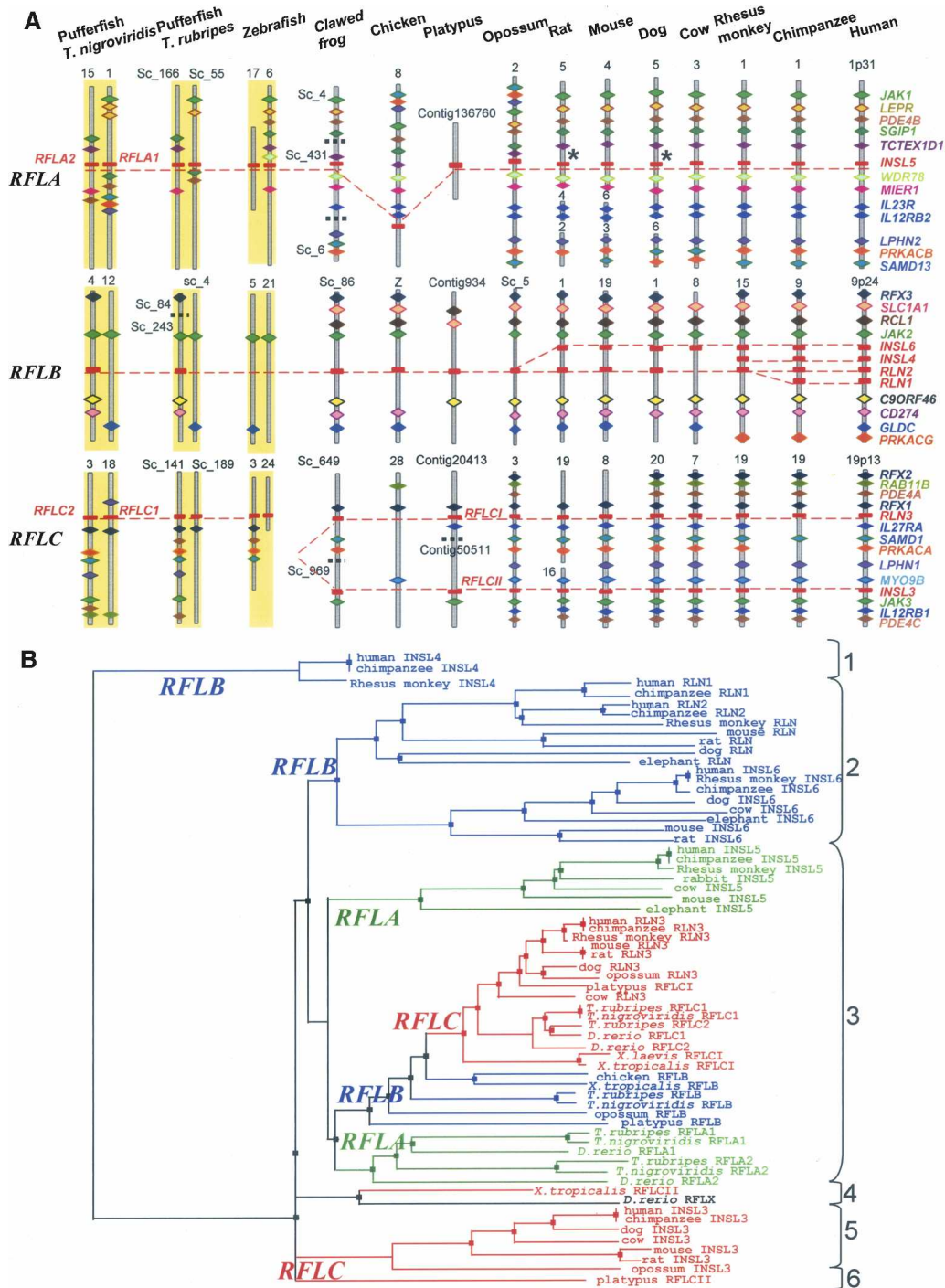


Figure 1. (Legend on next page)

tures that were derived from the original functional characteristics of the parent gene *AncRFLC* (Force et al. 1999; Postlethwait et al. 2004).

### LGR7-specific RLN3 and LGR8-specific INSL3 evolved from a bifunctional ancestral ligand, and the platypus does not contain an LGR8-specific ligand

To trace the molecular footprints underlying the evolution of the LGR7-specific RLN3 and LGR8-specific INSL3 signaling pathways, we sought to analyze the functional characteristics of the ancestral peptide on *RFLC* as well as orthologs from zebrafish and platypus based on LGR7 and LGR8 signaling. From gene resurrection analyses using the maximum likelihood method (Yang et al. 1995; Yang 1997), we generated a series of peptides representing the ancestral RFLC peptides of the common ancestor of vertebrates. Sequence comparison showed that predicted ancestral sequences derived from pooled sequences of non-mammalian species (Supplemental Fig. 1, top) or pools of representative ancestral sequences for each available subclass of vertebrates are in agreement with each other (Supplemental Fig. 1, bottom). The major differences are at the N-terminus of the B chain, which is subjected to alternative processing in different species. Based on these findings, we then generated a recombinant ancestral RFLC peptide (AncRFLC) from the common ancestor of tetrapods and teleosts as well as orthologs from *RFLC* of zebrafish (zRFLC1 and zRFLC2) and platypus (pRFLCI and pRFLCII) using expression constructs in which the B chain was tagged with a Myc epitope and a 6-histidine epitope (Fig. 3A; Supplemental Fig. 2).

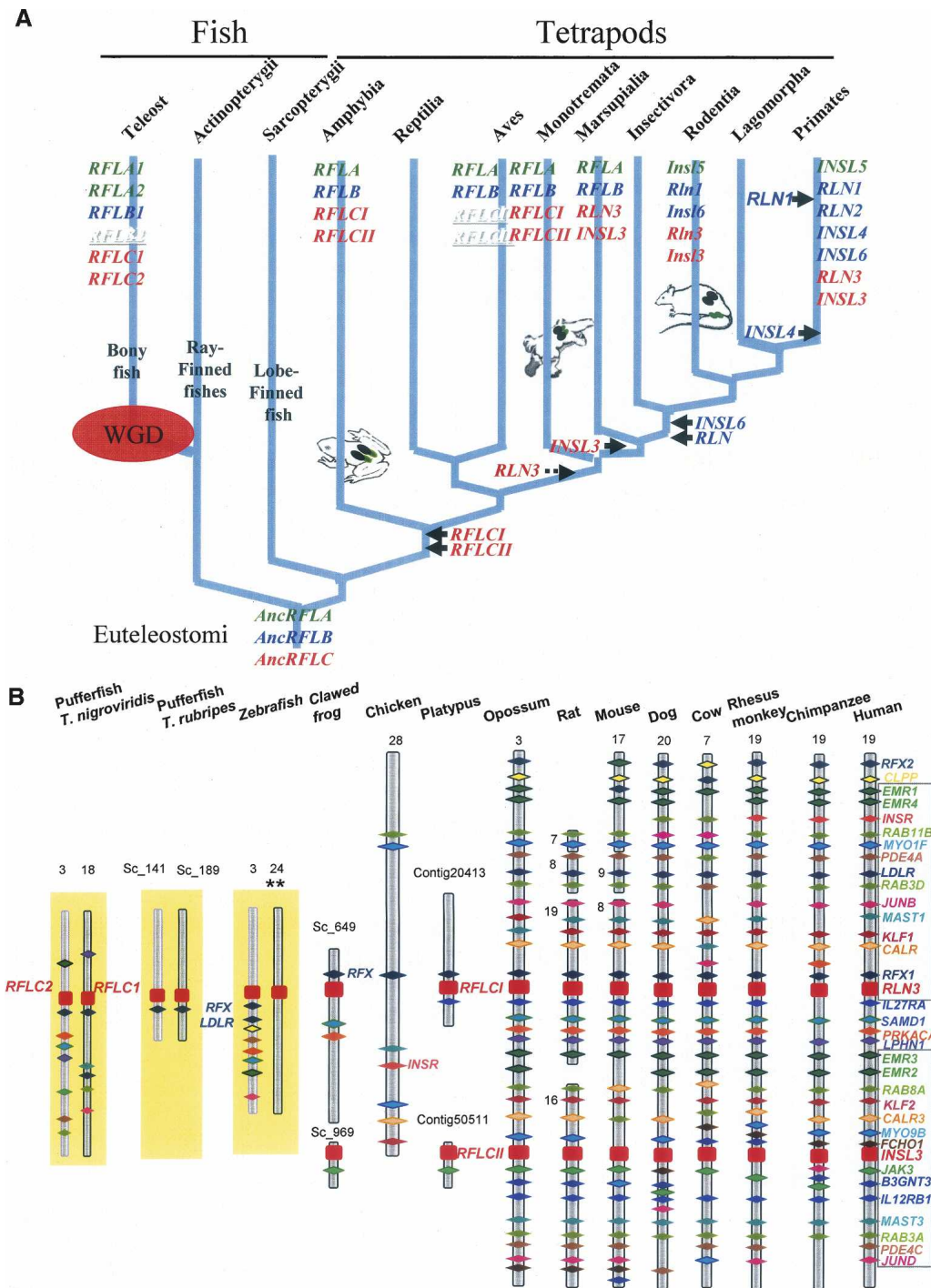
Unlike human RLN3 and INSL3 that activate only LGR7 and LGR8, respectively (Fig. 3B) (Kumagai et al. 2002; Sudo et al. 2002), analysis of receptor-activation activities showed that the AncRFLC peptide and the zebrafish RFLC1 and RFLC2 peptides are capable of robustly activating both LGR7 and LGR8 from human (Fig. 3C,D; see Supplemental Fig. 2 for Western blotting analyses; see Supplemental Table 2 for pEC50). In contrast, the platypus RFLCI peptide, which corresponds to and clusters with RLN3 of therian mammals in a single branch (Fig. 1A,B), exhibited LGR7-specific characteristics, whereas the RFLCII peptide of

platypus, which is syntenic to INSL3, retains ancestral features and activates both LGR7 and LGR8 (Fig. 3E). Therefore, unlike the selective RLN3 and INSL3 genes in humans, the ancestral *RFLC* gene encoded a bifunctional peptide capable of interacting with both LGR7 and LGR8, and the platypus does not contain an LGR8-specific ligand.

### Point mutation of a critical residue in the receptor-binding B chain converts the resurrected AncRFLC ligand from a bifunctional hormone to an LGR8-specific ligand

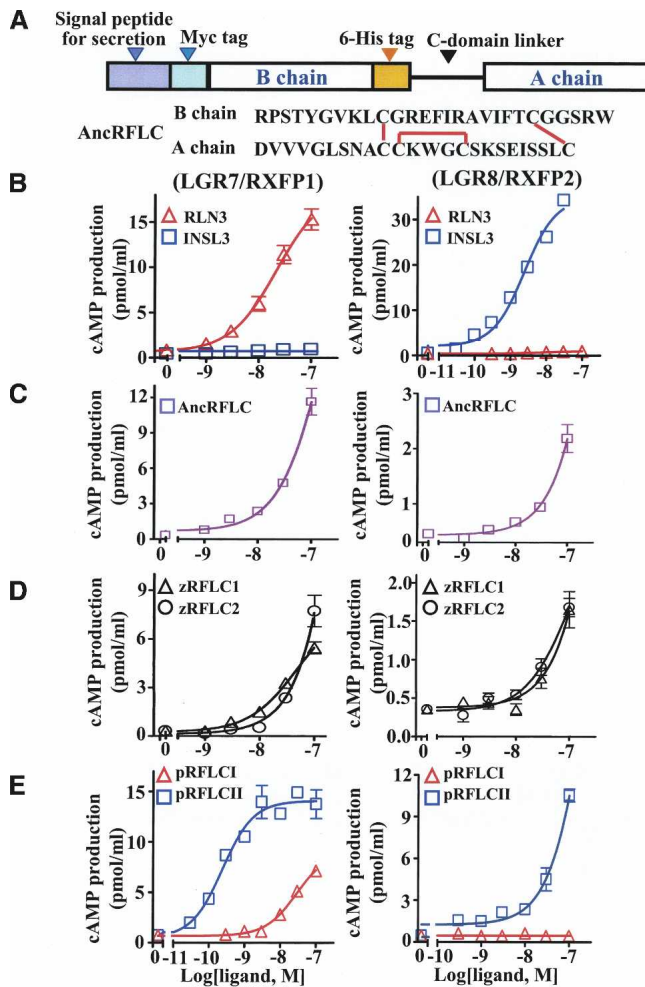
To understand the mechanistic basis for functional divergence of the two bifunctional daughter genes on *RFLC* of ancestral tetrapods, we sought to identify radical substitutions that are responsible for the functional characteristics of the newly emerged INSL3. Because the B chain of relaxin family peptides represents the major functional domain for interaction with LGR7 and LGR8 (Bullesbach and Schwabe 1991; Rosengren et al. 2006b), we focused the analyses on alternative residues on the B chains of INSL3 and RLN3, which share less than 40% amino acid identity in the mature regions. Comparison of primary sequences and tertiary structures showed that five residues on the B chain (R<sup>B12</sup>, E<sup>B13</sup>, F<sup>B20</sup>, T<sup>B21</sup>, and S<sup>B25</sup>) constitute radical substitutions, and three of them (R<sup>B12</sup>, F<sup>B20</sup>, and S<sup>B25</sup>) are on the molecular surface (Fig. 4A). To pinpoint the substitution(s) responsible for the LGR8-specific characteristics of INSL3, we first analyzed mutants of human RLN3 carrying one of each of the three radical substitutions found on the INSL3 surface. Functional analyses showed that substitution of F<sup>B20</sup> and S<sup>B25</sup> with residues found in corresponding positions in INSL3 had a negligible effect on RLN3 bioactivity (Fig. 4B; Supplemental Table 2). Similar to the wild-type peptide, RLN3 F<sup>B20</sup>R and RLN3 S<sup>B25</sup>P mutants function as selective ligands and stimulate cAMP production in cells expressing LGR7 but not LGR8 (Fig. 4B). In contrast, the Arg to His substitution in the RLN3 R<sup>B12</sup>H mutant abolished the LGR7-activation activity of RLN3 (Fig. 4B, left), and surprisingly, this substitution allowed the RLN3 R<sup>B12</sup>H mutant to stimulate cAMP

**Figure 1.** Evolution of relaxin family genes in vertebrates. (A) Syntenic mapping of relaxin family gene loci in human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), Rhesus monkey (*Macaca mulatta*), cow (*Bos taurus*), dog (*Canis familiaris*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), gray short-tailed opossum (*Monodelphis domestica*), platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), clawed frog (*Xenopus tropicalis*), zebrafish (*Danio rerio*), and two pufferfish (*Takifugu rubripes* and *Tetraodon nigroviridis*). The genomes of mammals from human to the egg-laying monotreme platypus encode four to seven paralogous genes. One ortholog of *INSL5* on *RFLA* was identified in all mammals analyzed. In contrast, *RFLB* contains a single gene in marsupial opossum and monotreme platypus but up to four paralogs in human and chimpanzee (one for each human ortholog *RLN1*, *RLN2*, *INSL4*, and *INSL6*). On the other hand, orthologs for *RLN3* and *INSL3* on *RFLC* were identified in all mammalian species analyzed, except platypus in which the two relaxin family genes on *RFLC* (*RFLCI* and *RFLCII*) exhibited great similarity to *RLN3* from other mammals. In contrast, chicken encodes only two relaxin family genes syntenic to *RFLA* and *RFLB* in mammals, respectively. Although genes neighboring *RFLC* could readily be identified on syntenic chicken chromosome 28, no relaxin family gene was detected in this region. On the other hand, the genome of clawed frog encodes four relaxin family genes. Among these paralogs, one each was found to correspond to *RFLA* and *RFLB*, respectively. Although the positions of contigs containing the third and fourth genes have not yet been determined, syntenic mapping indicated that they likely represent counterparts of *RFLCI* and *RFLCII* in platypus, respectively. The genomes of teleosts including zebrafish and pufferfish encode five copies of relaxin family genes that share close sequence relatedness to mammalian *RLN3* (>85% sequence identity at the B chain of mature peptides) (Wilkinson et al. 2005b). Two pairs of these teleost genes were derived from whole-genome duplication (WGD) that occurred before the divergence of teleosts and osteoglossomorphs and correspond to the tetrapod counterpart on *RFLA* and *RFLC*, respectively (Jaillon et al. 2004; Crollius and Weissenbach 2005). On the other hand, only one gene syntenic to *RFLB* was identified, likely representing the remaining member of the WGD-derived ancestral genes on *RFLB*. Relaxin family genes are indicated by red rectangles, whereas neighboring genes are indicated by diamonds. Orthologous genes in different species are identified by color. Horizontal black boxes on chromosome fragments indicate positions where no overlapping contigs are available in the draft genome sequences. The chromosomal numbers or the genomic contig numbers are indicated at the top of the schematic representation of each genomic fragment. The WGD-derived syntenic chromosomal regions in teleosts are indicated by yellow background. (\*) Pseudogene. (B) Phylogenetic analysis of 65 relaxin family peptides from 17 species of vertebrates based on maximum likelihood method. Species analyzed included human (*H. sapiens*), chimpanzee (*P. troglodytes*), Rhesus monkey (*M. mulatta*), rat (*R. norvegicus*), mouse (*M. musculus*), rabbit (*O. cuniculus*), dog (*C. familiaris*), cow (*B. taurus*), elephant (*L. africana*), gray short-tailed opossum (*M. domestica*), platypus (*O. anatinus*), chicken (*G. gallus*), two clawed frogs (*X. tropicalis* and *X. laevis*), zebrafish (*D. rerio*), and two pufferfish (*T. rubripes* and *T. nigroviridis*) (accession nos. are provided in Supplemental Table 4). Genes located on chromosome loci syntenic to *RFLA*, *RFLB*, and *RFLC* are indicated by green, blue, and red letters, respectively. The six major branches separated deep in the phylogenetic tree are indicated on the right. The zebrafish sequence without a locus assignment is indicated by black letters.



**Figure 2.** The evolutionary path of relaxin family genes. (A) Syntenic mapping indicated that seven human relaxin family genes evolved from three separate ancestral genes (*AncRFLA*, *AncRFLB*, and *AncRFLC*) in the common ancestor of tetrapods and teleosts. These ancestral genes gave rise to five relaxin family genes in pufferfish (*RFLA1*, *RFLA2*, *RFLB*, *RFLC1*, and *RFLC2*) as a result of teleost-specific WGD. One of the WGD-derived *RFLB* daughter genes (*RFLB2*) has been lost during teleost evolution. In contrast, a segmental duplication at *RFLC* led to the generation of *RFLC1* and *RFLCII* genes in tetrapods. Whereas *RFLC1* and *RFLCII* were lost in avian species, these two genes became *RLN3* and *INSL3*, respectively, in therian mammals. Arrows indicate the putative time of emergence of different relaxin family genes during vertebrate evolution. All genes are color-coded to indicate their origins. Lost genes are indicated by embossed letters. Positions of kidneys and testes in representative vertebrates are indicated by black balls and green balls, respectively. (B) Syntenic mapping of *RFLC* in vertebrates. The orthologs of human *RLN3* and *INSL3* on *RFLC* of different vertebrates are indicated by red rectangles whereas neighboring genes are indicated by diamonds. Orthologous genes in syntenic chromosomal regions of different species are identified by color. The chromosomal numbers or the genomic contig numbers are indicated at the top or the left of the schematic representation of each genomic fragment. WGD-derived syntenic chromosomal regions in teleosts are indicated by yellow background. Genomic fragments derived from segmental duplication of the region neighboring *RFLC* are indicated by rectangles on the right of the human chromosome 19p13 fragment. (\*\*\*) The co-orthologous relationship of the zebrafish *RFLC2* gene on chromosome 24 to the pufferfish counterparts was determined based on sequence similarity analysis.





**Figure 3.** RLN3 and INSL3 evolved from a bifunctional ligand. (A) Schematic presentation of the expression constructs encoding Myc- and 6-histidine-tagged relaxin family peptides linked with a mini-C domain. Mature B and A chain sequences of AncRFLC are shown (*bottom*). Two inter-chain and one intra-chain disulfide bridges are indicated by red lines. (B–E) LGR7 (RXFP1)- and LGR8 (RXFP2)-activation activities of human RLN3 and INSL3 (B), a resurrected ancestral RFLC peptide (AncRFLC) (C), two zebrafish RFLC peptides (zRFLC1 and zRFLC2) (D), and two platypus RFLC paralogs (pRFLCI and pRFLCII) (E). Mean  $\pm$  SEM,  $N = 4$ .

production in LGR8-expressing cells (Fig. 4B, right). Thus, the  $R^{B12}$  to  $H^{B12}$  substitution effectively converted RLN3 into an analog with receptor specificity identical to that of INSL3. This conversion in receptor specificity is specific, as an alanine substitution at the same position in the RLN3  $R^{B12}A$  mutant ablates the LGR7-activation activity of RLN3 but without any gain of interaction with LGR8 (Fig. 4B). We confirmed this shift in receptor specificity using receptor-binding assays with  $^{125}I$ -labeled tracers. As shown in Figure 4C, RLN3 and INSL3 compete for labeled RLN2 binding to LGR7 and labeled INSL3 binding to LGR8, respectively (see Supplemental Table 3 for  $pIC_{50}$  values). Unlike wild-type RLN3 that selectively competes for labeled RLN2 binding to LGR7, the RLN3  $R^{B12}H$  mutant competes with labeled INSL3 for binding to LGR8 (Fig. 4C, right), but not with labeled RLN2 for binding to LGR7, on transfected cells (Fig. 4C, left).

In addition, to investigate whether the gain-of-function RLN3  $R^{B12}H$  mutant is capable of activating native receptors in

the gubernaculum, we performed primary cultures of gubernaculum cord cells from 7-d-old neonatal rats and tested the effect of the RLN3  $R^{B12}H$  mutant on gubernaculum function (Kumagai et al. 2002). Consistent with studies of recombinant receptors, treatment of gubernaculum cultures with the INSL3 or RLN3  $R^{B12}H$  mutant increased cAMP production dose dependently (Fig. 4D). In contrast, treatment of the RLN3 or the RLN3  $R^{B12}A$  mutant had no effect on cAMP production in the gubernaculum cells.

Taken together, these studies demonstrated that the  $R^{B12}$  to  $H^{B12}$  mutation represents the critical substitution that occurred during the evolution of INSL3 from the AncRFLC gene. In agreement with this hypothesis, we found that introduction of the receptor-specificity-transforming  $R^{B12}$  to  $H^{B12}$  mutation into the AncRFLC and zebrafish RFLC peptides abolished the LGR7-, but not LGR8-, activation activities of these bifunctional ligands (Fig. 5A). Likewise, functional analyses of opossum INSL3, which represents the most diverging INSL3 ortholog in therian mammals but contains the hallmark  $H^{B12}$  in the receptor-binding B chain, showed that it activates only LGR8 (Fig. 5B).

### Restoration of lost LGR7-activation activity in human INSL3 by an $H^{B12}$ to $R^{B12}$ mutation

Apart from the inference that LGR8-specific characteristics of INSL3 are attributable to the  $R^{B12}$  to  $H^{B12}$  mutation, these data predicted that a reversal of this critical substitution in the present-day INSL3 gene could restore the lost function of the parent gene. To investigate this hypothesis, we introduced the  $R^{B12}$  residue found in the AncRFLC peptide into human INSL3 (Fig. 5C). As predicted, substitution of the  $H^{B12}$  with an Arg in the INSL3  $H^{B12}R$  mutant led to restoration of the lost LGR7-activation activity (Fig. 5C, left) without changing the LGR8-activation activity (Fig. 5C, right). In contrast, substitution with an alanine in the same position had negligible effects on the receptor-activating characteristics of INSL3 (Fig. 5C).

Given these results, the most parsimonious course for the evolution of LGR8-specific INSL3 in therian mammals is that the ancestral gene on RFLC encoded a bifunctional ligand, and an Arg to His replacement (a result of two nucleotide changes, CGG to CAT; Supplemental Fig. 3) in one of the duplicated daughter genes eliminated part of the bioactivity (LGR7-activation activity) of the parent gene.

### Reciprocal loss of LGR8-activation activity in RLN3 occurred prior to the separation of therian and monotremata mammals

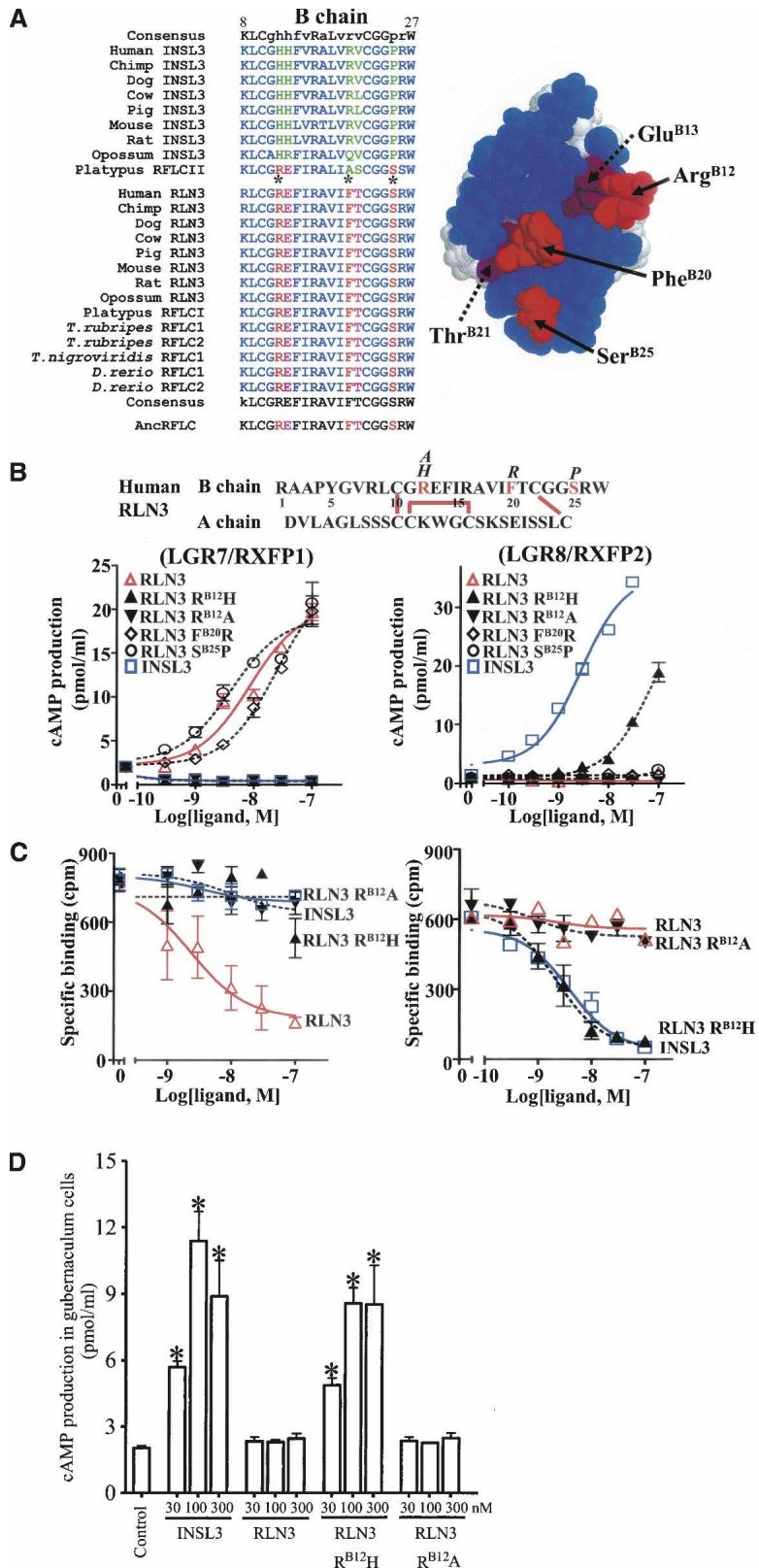
Because the two duplicated daughter genes on RFLC encode bifunctional ligands, the evolution of an LGR7-specific RLN3 in present-day therian mammals likely involved independent mutations. Sequence comparison of mammalian RLN3 and RFLC peptides from non-mammalian vertebrates showed that two pairs of residues at the N terminus of the A chain, A3,4 and A8,9, are on the molecule surface and constitute radical substitutions between the mammalian RLN3 and those from basal taxonomy (Fig. 6A). To pinpoint mutations that are responsible for the shift to LGR7-specific characteristics of mammalian RLN3, we generated mutant AncRFLC peptides carrying substitutions at each of these two pairs of residues (AncRFLC  $VV^{A3,4}LA$  and AncRFLC  $NA^{A8,9}SS$ ; Fig. 6B). Analysis of these mutants showed that point mutations of the N-terminal, but not the C-terminal, pair of residues at the A chain converted the bifunctional AncRFLC peptide into an LGR7-selective agonist (Fig. 6B). To demonstrate further the critical role played by this pair of mutations, we studied RLN3

mutants with reversal substitutions at A3,4 (RLN3 LA<sup>A3,4</sup>VV) and A8,9 (RLN3 SS<sup>A8,9</sup>NA) positions (Fig. 6C). Contrary to studies of AncRFLC mutants, functional analyses showed the RLN3 SS<sup>A8,9</sup>NA mutant remains an LGR7-specific ligand, whereas the RLN3 LA<sup>A3,4</sup>VV mutant exhibits a gain-of-function activity on LGR8 activation. These data are consistent with studies showing the platypus RFLCI peptide, which contains the LGR7-specific "LA pair" hallmark at the A3,4 positions, activates LGR7, but not LGR8 (Figs. 3E, 6A). Therefore, subsequent to the duplication of *AncRFLC*, one of the daughter genes, *RFLCI*, acquired the LGR7-specific characteristics as a result of the VV<sup>A3,4</sup> to LA<sup>A3,4</sup> mutations in the A chain, and this event occurred before the evolution of proto-, meta-, and eutherian mammals.

## Discussion

By mapping of the syntenic chromosomal regions and functional analyses of orthologous relaxin family peptides across vertebrate phyla, we recapitulated the evolutionary footprints of a therian mammal-specific signaling pathway that is crucial to the positioning of the testis from the mesonephric position to the scrotum during embryonic and neonatal development in therian mammals. We showed that the process started with a gene duplication event followed by reciprocal loss of part of the parental functions in the two daughter genes at two separate geological times (Fig. 7A). The stepwise evolution of these independent signaling pathways through gene duplication and subsequent divergence is consistent with Darwinian theory of selection and adaptation, and it suggests an association between the radical substitution of a few key residues of a pair of *RLN3*-like ancestral genes (*RFLCI* and *RFLCII*) and the concurrent evolution of testicular descent in the common ancestor of therian mammals.

These data demonstrate that subsequent to initial duplication each of the two daughter genes on *RFLC* (*RFLCI* and *RFLCII*) lost part of its parental function to increase the signal-to-noise ratio of the LGR7 and LGR8 signaling pathways mediated by them (Fig. 7B). The reciprocal loss of part of the parental function in a temporally separated manner is necessary as, at each point, one of the daughter genes can afford to lose part of



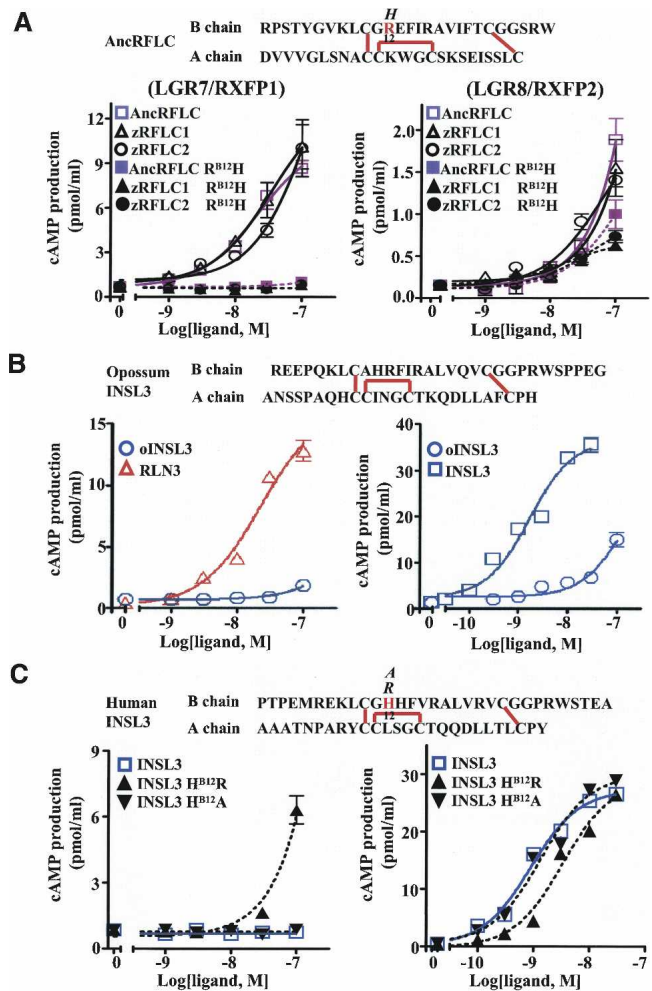
**Figure 4.** (Legend on next page)

its function because the other daughter gene retains it. In addition, the molecular footprints as described here indicated that *INSL3* and *RLN3* evolved in a manner similar to the “divergent resolution” model that was proposed to illustrate the separation of different copies of a duplicated gene in allopatric populations during sympatric evolution (Lynch and Conery 2000; Taylor et al. 2001). In this scenario, segregation of the intercellular signaling pathways of the two daughter genes in individual organisms is selected (Fig. 7B), instead of one of the two copies of duplicated genes in reproductively separated populations. Thus, restoration of the lost functions in the *RLN3* LA<sup>A3,4</sup>VV and *INSL3* H<sup>B12</sup>R mutant peptides effectively reversed two critical steps of this ancient “divergent resolution” event that occurred more than 140–180 Mya (Woodburne et al. 2003). Whereas it is generally agreed that newly duplicated genes evolve with rapid changes and many changes in the new genes are needed for their fixation, the discrete interactions between a ligand and a receptor may allow new hormonal functions to be selected and retained with only one or a few changes in these molecules (Long et al. 2003; Bridgham et al. 2006). In addition, we envision the evolution of tissue-specific *RLN3/LGR7* and *INSL3/LGR8* signaling pathways in thierian mammals could be accompanied by a concurrent diversification in the *cis*-regulatory elements of these ligand and receptor genes.

It has been hypothesized that evolution of transabdominal testicular descent in thierian mammals was driven by natural selection for increased aerobic activity associated with endothermy and avoided potential impairment of spermatogenesis when the core body temperature reached a threshold of 34–35°C in primitive mammals (Cowles 1965; Bennett and Ruben 1979). As the transition of the *RFLCII* gene to *INSL3* in ancestral thierian mammals (140–180 Mya) occurred much later than the evolution of endothermy in the ancestor of mammals (>220 Mya) (Woodburne et al. 2003; Gillooly et al. 2006), genetic events described here could be associated with a high degree of fitness and reproductive success and contributed to the radiation and dominance of thierian mammals in the last 140 million years. Whereas it remains unclear how the male reproductive system in avian species adapted to endothermy, the absence of an RFLC peptide in chicken suggests that alternative mechanisms evolved in avian ancestors to allow optimal spermatogenesis at a high body temperature. Therefore, our study emphasizes the importance of integrating anatomical, physiological, molecular, and phylogenetic data in illustrating the origin and mechanistic bases of gene selection and physiological adaptation.

Furthermore, our investigation provides keys to a fundamental question: How could a novel complex physiological trait or a novel cell type have evolved in a metazoan organism? Because cell type-specific signaling that dictates the proliferation and differentiation of a cell represents one of the most important hallmarks in cell type classification, the duplication of a ligand or its cognate cell surface receptor provided just such a rare opportunity for the generation of a novel intercellular signaling pathway without incurring changes in the fundamental architec-

**Figure 4.** R<sup>B12</sup>H substitution in the B chain of *RLN3* converts LGR7-specific *RLN3* into an LGR8-specific ligand. (A) Identification of radical substitutions in the *INSL3* B chain. Five radical substitutions in *INSL3* (green letters). Spatial positions of these critical residues are shown (right). In the structure model, residues in the B and A chains are indicated by blue and white space balls, respectively. (B) LGR7 (RXFP1)- and LGR8 (RXFP2)-activation activity of human *RLN3*, *INSL3*, *RLN3* R<sup>B12</sup>H, *RLN3* R<sup>B12</sup>A, *RLN3* F<sup>B20</sup>R, and *RLN3* S<sup>B25</sup>P peptides (mean ± SEM, N = 4). (C) Competitive LGR7- and LGR8-binding analysis of *RLN3*, *INSL3*, *RLN3* R<sup>B12</sup>H, and *RLN3* R<sup>B12</sup>A peptides (mean ± SEM, N = 3). (D) Stimulation of cAMP production in cultured rat gubernaculum cells by *INSL3* and *RLN3* R<sup>B12</sup>H peptides (mean ± SEM, N = 4). (\*) Significantly different from controls (P < 0.01).



**Figure 5.** Ablation of LGR7-activation activity of AncRFLC and restoration of lost LGR7-activation activity in human *INSL3* by point mutation. (A–C) LGR7 (RXFP1)- and LGR8 (RXFP2)-activation activity of AncRFLC R<sup>B12</sup>H, zRFLC1 R<sup>B12</sup>H, and zRFLC2 R<sup>B12</sup>H peptides (A), opossum *INSL3* peptide (oINSL3) (B), and *INSL3* H<sup>B12</sup>R and *INSL3* H<sup>B12</sup>A peptides (C). Mean ± SEM, N = 4.

ture of the cellular components. Our data suggest that the evolution of *INSL3* together with the selective expression of *LGR8* in an ancestral pre-gubernaculum cell type allows the emergence of an exclusive signaling pathway for regulating testicular positioning in thierian mammals. Therefore, by considering the *LGR8*-specific *INSL3* signaling network as an adaptive structure rather than a signaling pathway, our study provides a paradigm for studying how a lineage-specific physiological trait or a new cell type, such as the gubernaculum cell, could have emerged, as well as how duplications of ligand and receptor genes contributed to physiological complexity during the evolution of metazoans.

## Methods

### Identification of relaxin family genes

We analyzed the genomic DNA of human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), Rhesus monkey (*Macaca mulatta*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), rabbit (*Oryctolagus cuniculus*), dog (*Canis familiaris*), cow (*Bos taurus*), pig (*Sus scrofa*), elephant (*Loxodonta africana*), gray short-tailed opossum (*Mono delphis domestica*), platypus (*Ornithorhynchus anatinus*), chicken (*Gallus gallus*), clawed frog (*Xenopus tropicalis*), zebrafish, (*D. rerio*), and two pufferfish (*T. rubripes* and *T. nigroviridis*) and identified relaxin family genes by a series of reciprocal sequence comparisons using the BLAST server. Because our analyses allowed the detection of the most divergent *INSL4* or even the functionally unrelated insulin, we ruled out the possibility that additional relaxin family genes exist but were not detected in the genomes of studied species. The evolutionary relationships of relaxin family genes were determined using the ClustalW program (<http://www.ebi.ac.uk/clustalw>). The chromosomal localization of relaxin family genes in different species was determined based on syntenic maps from the Genoscope database (<http://www.genoscope.cns.fr>) and Ensembl's BioMart data mining tool (<http://www.ensembl.org/multi/martview>), followed by verification using the UCSC Genome Bioinformatics web server (<http://genome.ucsc.edu/cgi-bin/hgBlat>).

### Resurrection of ancestral RFLC peptides

Analysis of orthologous *RFLC* genes as a group indicated that, similar to phylogenetic studies of relaxin family genes as a group, the evolution of *RFLC* genes in tetrapods mimicked "a radiation event compressed in time" that cannot be resolved by phylogenetic analyses (Thornton 2004; Rokas et al. 2005). We therefore predicted the ancestral *RFLC* gene sequences using (1) pools of orthologs from non-mammalian species that do not encode an *INSL3* ortholog on *RFLC* to avoid heterotachy incurred by the lineage-specific gene duplication event and subsequent divergence (Thornton and Kolaczowski 2005) and (2) pools of resurrected ancestral or representative peptide sequences for the different subclasses of vertebrates analyzed (Supplemental Fig. 1). Reconstruction was conducted using *codeml(aaml)* in the PAML program package (<http://abacus.gene.ucl.ac.uk/software/paml.html>; Yang 1997).

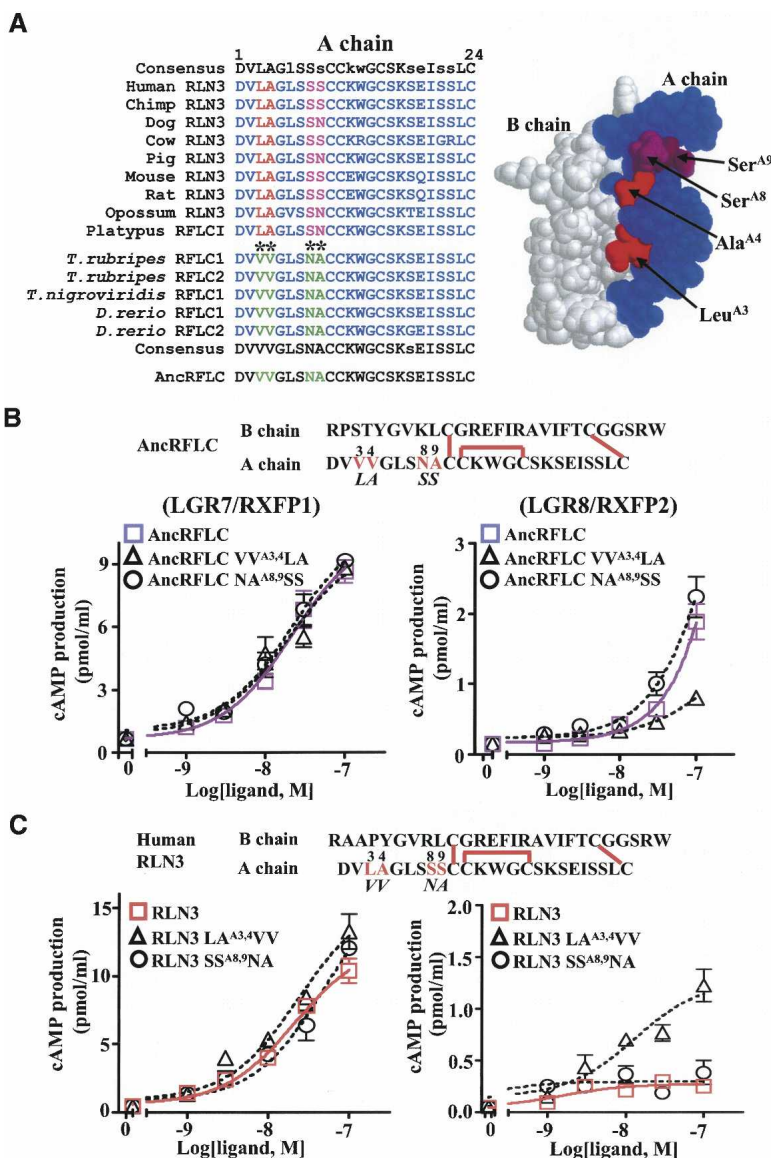
### Comparative structure modeling

Comparative protein modeling was performed with the SWISS-MODEL server

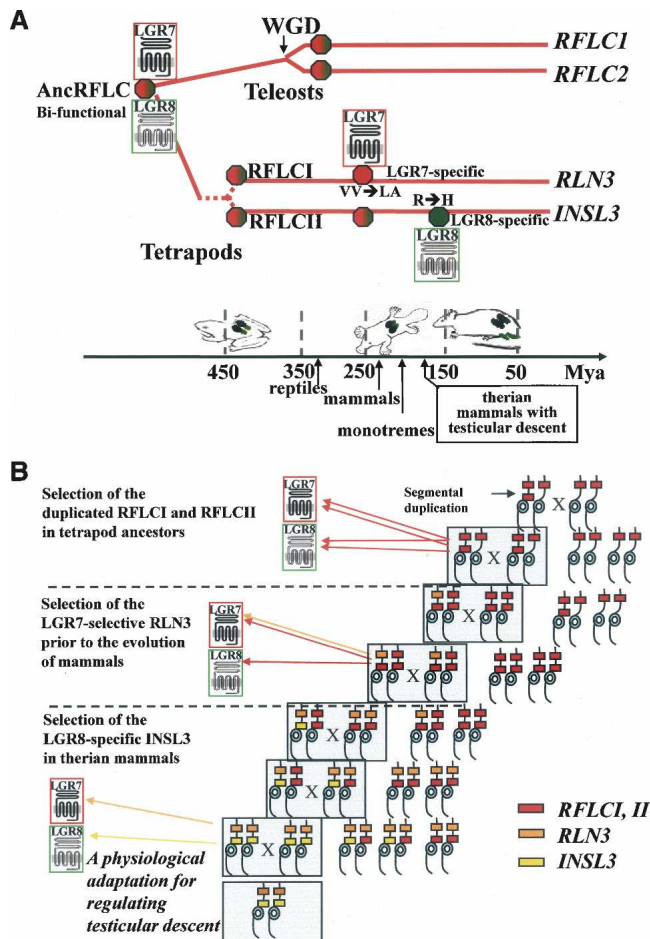
(<http://swissmodel.expasy.org//SWISS-MODEL.html>) using experimentally determined structures for human *RLN3* as the template (Protein Data Bank 2FHW) (Rosengren et al. 2006a).

### Expression of relaxin family peptides

To generate recombinant *RLN3* and *INSL3* from different species, we subcloned cDNAs by overlapping PCR with gene-specific primers. Complementary DNA sequences encoding the mature B and A chains were appended with a signal peptide for secretion sequence from the prolactin precursor at the N terminus. In addition, the B chain was tagged with an N-terminal Myc epitope and a C-terminal 6-histidine tag. The B and A chains were connected by a mini-C domain linker of eight amino acids (SLSQEDAL)



**Figure 6.** Pair of critical substitutions at the A chain ablates LGR8-activation activity of AncRFLC. (A) Sequence comparison and identification of radical substitutions in the mammalian *RLN3* A chain. Two pairs of radical substitutions in mammalian *RLN3* and platypus *RFLC1* (asterisks). Spatial positions of these critical residues are shown (right). In the structure model, residues belong to the B and A chains are indicated by white and blue space balls, respectively. (B, C) LGR7 (RXFP1)- and LGR8 (RXFP2)-activation activities of AncRFLC, AncRFLC VV<sup>A3,4</sup>LA, and AncRFLC NA<sup>A8,9</sup>SS peptides (B), and *RLN3* LA<sup>A3,4</sup>VV and *RLN3* SS<sup>A8,9</sup>NA peptides (C). Mean  $\pm$  SEM,  $N = 4$ .



**Figure 7.** Schematic representation of the evolution trajectory of *RLN3* and *INSL3*. (A) *RLN3* and *INSL3* of therian mammals evolved via gene duplication, followed by emergence of an LGR7 (RXFP1)-specific daughter gene (*RFLC1*) in the common ancestor of mammals, and an LGR8 (RXFP2)-specific daughter gene (*RFLCII*) in the common ancestor of therian mammals. LGR7- and LGR8-specific ligands are indicated by a red and a green hexagonal symbol, respectively. Bifunctional ligands are indicated by hexagonal symbols with a mix of red and green colors. Positions of kidneys and testes in representative vertebrates are indicated by black balls and green balls, respectively. (B) Our results support the hypothesis that natural selection selects signaling pathways with a high signal-to-noise ratio over nonselective ones following the duplication of a ligand gene. A high degree of fitness associated with select mutations can then lead to reproductive success and the selection of these copies. In our example, the two daughter genes, *RLN3* and *INSL3*, were selected through natural selection in a stepwise “divergent resolution” manner (Taylor et al. 2001). Starting with a segmental duplication at *RFLC*, the LGR7-selective *RLN3* and the LGR8-specific *INSL3* were fixed in the population after neofunctionalization but occurred at two disparate geological time points. Red bars on a chromatid represent *RFLC1* and *RFLCII* daughter genes generated by a segmental duplication event in an ancestral tetrapod. Orange bars on a chromatid represent *RLN3* after the acquisition of the LGR7-selective characteristics. Gold bars on a chromatid represent *INSL3* following the R → H mutation. Genotypes selected by evolution are indicated by rectangular boxes. Ligand specificities for LGR7 and LGR8 are indicated by arrows.

flanked by dibasic convertase cleavage sites and subcloned into the expression vector, pcDNA3.1 Zeo. The tagging with a Myc and a 6-histidine epitope allowed efficient quantification and purification of recombinant peptides without affecting the bioactivity of the relaxin family peptides. For efficient processing and

the secretion of mature peptides, cells were routinely cotransfected with the select expression construct and a one-tenth aliquot of a convertase expression vector (Marriott et al. 1992). Four days after transfection, conditioned media were collected and filtered to remove cell debris. Recombinant peptides were purified using nickel affinity chromatography and quantified based on Western blotting analysis using an anti-Myc antibody (Sigma-Aldrich), followed by densitometry analysis of the immunofluorescent signals. For Western blotting analysis, peptides were resolved using 18% SDS-PAGE and analyzed as previously described (Hsu et al. 2002; Kumagai et al. 2002).

### Receptor-activation and receptor-binding analyses

Bioactivity of the relaxin family peptides was determined based on stimulation of the adenylate cyclase activity in HEK293T cells expressing recombinant LGR7 or LGR8 as previously described (Hsu et al. 2002; Kumagai et al. 2002). Because orthologous LGR7 and LGR8 from vertebrates share high polypeptide sequence similarity (>85% similarity among mammalian orthologs and >75% similarity between those of teleosts and humans) and chimerization of LGR7 and LGR8 does not affect the signaling characteristics of these receptors (Sudo et al. 2002), we tested the functional characteristics of different peptides using recombinant human receptors.

To determine the binding characteristics of wild-type and mutant peptides, aliquots of HEK293T cells expressing LGR7 or LGR8 were incubated with increasing doses of purified peptides in the presence of 100 pM of <sup>125</sup>I-labeled RLN2 or INSL3 tracer (Phoenix Pharmaceuticals) at room temperature for 2 h. After washing, radioactivity bound to the cells was measured using a gamma counter. Total binding was determined in the absence, and nonspecific binding in the presence, of 100 nmol of unlabeled RLN2 or INSL3.

### Gubernaculum cell culture

Gubernaculum cells were isolated by modifying an earlier method (Kumagai et al. 2002). Tissues were removed from 1-wk-old rats and dissociated at 37°C in DMEM/F12 with 0.1% collagenase. Cell debris was removed by passage through a sterile filter, and cells were collected by centrifugation. After culture for 48 h in a 5% CO<sub>2</sub> incubator at 37°C, cells were washed and treated with or without hormones in quadruplicate. After 12 h of incubation, total cAMP was measured as previously described (Hsu et al. 2002; Kumagai et al. 2002).

### Statistical analysis

Receptor-activation and receptor-binding activity curves generated in nonlinear regression analyses (Prism; GraphPad Software) were evaluated relative to the samples without hormone treatment. In all cases, data are reported as the mean ± SEM of assays in triplicate or quadruplicate. Statistically significant responses ( $P < 0.01$ ) were determined for each stimulated response to the average nonspecific response from controls using ANOVA and Student's *t*-test.

### Acknowledgments

We thank Caren Spencer, Cynthia Klein, and Rami Rauch for editorial and technical assistance. We thank Drs. Aaron J.W. Hsueh, Anita Payne, and Marco Conti (Department of Obstetrics/Gynecology, Stanford University) for comments on the manuscript. We also thank Drs. Linda Giudice (Department of Obstetrics, Gynecology and Reproductive Sciences, UCSF) and Jonathan S. Berek (Department of Obstetrics/Gynecology, Stanford Univer-

sity) for the encouragement. C.L.C. also thanks Dr. Yung-Kuei Soong (Department of Obstetrics/Gynecology, Chang Gung Memorial Hospital, Taiwan) for the support and encouragement. J.S. is supported by the Bioinformatics Core of the SCCPRR Center at the Stanford University School of Medicine (NIH HD31398). The authors also acknowledge the support of NIH awards (R21 HD47606 and RO1 DK70652 to S.Y.H.) and a March of Dimes research grant (to S.Y.H.).

## References

- Adham, I.M., Steding, G., Thamm, T., Bullesbach, E.E., Schwabe, C., Paprotta, I., and Engel, W. 2002. The overexpression of the *insl3* in female mice causes descent of the ovaries. *Mol. Endocrinol.* **16**: 244–252.
- Bennett, A.F. and Ruben, J.A. 1979. Endothermy and activity in vertebrates. *Science* **206**: 649–654.
- Bogatcheva, N.V., Truong, A., Feng, S., Engel, W., Adham, I.M., and Agoulnik, A.I. 2003. GREAT/LGR8 is the only receptor for insulin-like 3 peptide. *Mol. Endocrinol.* **17**: 2639–2646.
- Bridgman, J.T., Carroll, S.M., and Thornton, J.W. 2006. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**: 97–101.
- Bullesbach, E.E. and Schwabe, C. 1991. Total synthesis of human relaxin and human relaxin derivatives by solid-phase peptide synthesis and site-directed chain combination. *J. Biol. Chem.* **266**: 10754–10761.
- Burazin, T.C., Johnson, K.J., Ma, S., Bathgate, R.A., Tregear, G.W., and Gundlach, A.L. 2005. Localization of LGR7 (relaxin receptor) mRNA and protein in rat forebrain: Correlation with relaxin binding site distribution. *Ann. N.Y. Acad. Sci.* **1041**: 205–210.
- Chen, S.T., Cheng, H.C., Barbash, D.A., and Yang, H.P. 2007. Evolution of *hydra*, a recently evolved testis-expressed gene with nine alternative first exons in *Drosophila melanogaster*. *PLoS Genet.* **3**: e107. doi: 10.1371/journal.pgen.0030107.
- Coveney, D., Shaw, G., Hutson, J.M., and Renfree, M.B. 2002. The development of the gubernaculum and inguinal closure in the marsupial *Macropus eugenii*. *J. Anat.* **201**: 239–256.
- Cowles, R.B. 1965. Hyperthermia, aspermia, mutation rates and evolution. *Q. Rev. Biol.* **40**: 341–367.
- Crollius, H.R. and Weissenbach, J. 2005. Fish genomics and biology. *Genome Res.* **15**: 1675–1682.
- Feng, S., Bogatcheva, N.V., Kamat, A.A., and Agoulnik, A.I. 2005. Genetic targeting of relaxin and *insl3* signaling in mice. *Ann. N.Y. Acad. Sci.* **1041**: 82–90.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531–1545.
- Gillooly, J.F., Allen, A.P., and Charnov, E.L. 2006. Dinosaur fossils predict body temperatures. *PLoS Biol.* **4**: e248. doi: 10.1371/journal.pbio.0040248.
- Griffiths, A.L., Renfree, M.B., Shaw, G., Watts, L.M., and Hutson, J.M. 1993. The tammar wallaby (*Macropus eugenii*) and the Sprague-Dawley rat: Comparative anatomy and physiology of inguinoscrotal testicular descent. *J. Anat.* **183**: 441–450.
- Hsu, S.Y. 1999. Cloning of two novel mammalian paralogs of relaxin/insulin family proteins and their expression in testis and kidney. *Mol. Endocrinol.* **13**: 2163–2174.
- Hsu, S.Y., Kudo, M., Chen, T., Nakabayashi, K., Bhalla, A., van der Spek, P.J., van Duin, M., and Hsueh, A.J. 2000. The three subfamilies of leucine-rich repeat-containing G protein-coupled receptors (LGR): Identification of LGR6 and LGR7 and the signaling mechanism for LGR7. *Mol. Endocrinol.* **14**: 1257–1271.
- Hsu, S.Y., Nakabayashi, K., Nishi, S., Kumagai, J., Kudo, M., Sherwood, O.D., and Hsueh, A.J. 2002. Activation of orphan receptors by the hormone relaxin. *Science* **295**: 671–674.
- Hughes, A.L. 2005. Gene duplication and the origin of novel proteins. *Proc. Natl. Acad. Sci.* **102**: 8791–8792.
- Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* **431**: 946–957.
- Kamat, A.A., Feng, S., Bogatcheva, N.V., Truong, A., Bishop, C.E., and Agoulnik, A.I. 2004. Genetic targeting of relaxin and insulin-like factor 3 receptors in mice. *Endocrinology* **145**: 4712–4720.
- Krajnc-Franken, M.A., van Disseldorp, A.J., Koenders, J.E., Mosselman, S., van Duin, M., and Gossen, J.A. 2004. Impaired nipple development and parturition in LGR7 knockout mice. *Mol. Cell. Biol.* **24**: 687–696.
- Kumagai, J., Hsu, S.Y., Matsumi, H., Roh, J.S., Fu, P., Wade, J.D., Bathgate, R.A., and Hsueh, A.J. 2002. *INSL3*/Leydig insulin-like peptide activates the LGR8 receptor important in testis descent. *J. Biol. Chem.* **277**: 31283–31286.
- Liu, C., Chen, J., Sutton, S., Roland, B., Kuei, C., Farmer, N., Sillard, R., and Lovenberg, T.W. 2003a. Identification of relaxin-3/*INSL7* as a ligand for GPCR142. *J. Biol. Chem.* **278**: 50765–50770.
- Liu, C., Eriste, E., Sutton, S., Chen, J., Roland, B., Kuei, C., Farmer, N., Jornvall, H., Sillard, R., and Lovenberg, T.W. 2003b. Identification of relaxin-3/*INSL7* as an endogenous ligand for the orphan G-protein-coupled receptor GPCR135. *J. Biol. Chem.* **278**: 50754–50764.
- Long, M., Betran, E., Thornton, K., and Wang, W. 2003. The origin of new genes: Glimpses from the young and old. *Nat. Rev. Genet.* **4**: 865–875.
- Lynch, M. and Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151–1155.
- Ma, S., Roozendaal, B., Burazin, T.C., Tregear, G.W., McLaugh, J.L., and Gundlach, A.L. 2005. Relaxin receptor activation in the basolateral amygdala impairs memory consolidation. *Eur. J. Neurosci.* **22**: 2117–2122.
- Mackay, S., Xie, Q., Ullmann, S.L., Gilmore, D.P., and Payne, A.P. 2004. Postnatal development of the reproductive system in the grey short-tailed opossum, *Monodelphis domestica*. *Anat. Embryol. (Berl.)* **208**: 121–133.
- Marriott, D., Gillette-Castro, B., and Gorman, C.M. 1992. Prohormone convertase-1 will process prorelaxin, a member of the insulin family of hormones. *Mol. Endocrinol.* **6**: 1441–1450.
- Nef, S. and Parada, L.F. 1999. Cryptorchidism in mice mutant for *Ins13*. *Nat. Genet.* **22**: 295–299.
- Postlethwait, J., Amores, A., Cresko, W., Singer, A., and Yan, Y.L. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. *Trends Genet.* **20**: 481–490.
- Rokas, A., Kruger, D., and Carroll, S.B. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310**: 1933–1938.
- Rosengren, K.J., Lin, F., Bathgate, R.A., Tregear, G.W., Daly, N.L., Wade, J.D., and Craik, D.J. 2006a. Solution structure and novel insights into the determinants of the receptor specificity of human relaxin-3. *J. Biol. Chem.* **281**: 5845–5851.
- Rosengren, K.J., Zhang, S., Lin, F., Daly, N.L., Scott, D.J., Hughes, R.A., Bathgate, R.A., Craik, D.J., and Wade, J.D. 2006b. Solution structure and characterization of the receptor binding surface of insulin-like peptide 3. *J. Biol. Chem.* **281**: 28287–28295.
- Semyonov, J., Park, J.I., Chang, C.L., and Hsu, S.Y.T. 2008. GPCR genes are preferentially retained after whole genome duplication. *PLoS ONE* **3**: e1903. doi: 10.1371/journal.pone.0001903.
- Sudo, S., Kumagai, J., Nishi, S., Layfield, S., Ferraro, T., Bathgate, R., and Hsueh, A.J. 2002. H3 relaxin is a specific ligand for LGR7 and activates the receptor by interacting with both the ectodomain and the exoloop 2. *J. Biol. Chem.* **278**: 7855–7862.
- Tanaka, M., Iijima, N., Miyamoto, Y., Fukusumi, S., Itoh, Y., Ozawa, H., and Ibata, Y. 2005. Neurons expressing relaxin 3/*INSL7* in the nucleus incertus respond to stress. *Eur. J. Neurosci.* **21**: 1659–1670.
- Taylor, J.S. and Raes, J. 2004. Duplication and divergence: The evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**: 615–643.
- Taylor, J.S., Van de Peer, Y., and Meyer, A. 2001. Genome duplication, divergent resolution and speciation. *Trends Genet.* **17**: 299–301.
- Thornton, J.W. 2004. Resurrecting ancient genes: Experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5**: 366–375.
- Thornton, J.W. and Kolaczowski, B. 2005. No magic pill for phylogenetic error. *Trends Genet.* **21**: 310–311.
- van der Schoot, P. 1996. Foetal genital development in *Hyrax capensis*, a species with primary testicondia: Proposal for the evolution of Hunter's gubernaculum. *Anat. Rec.* **244**: 386–401.
- Werdelin, L. and Nilsson, A. 1999. The evolution of the scrotum and testicular descent in mammals: A phylogenetic view. *J. Theor. Biol.* **196**: 61–72.
- Wilhelm, D. and Koopman, P. 2006. The makings of maleness: Towards an integrated view of male sexual development. *Nat. Rev. Genet.* **7**: 620–631.
- Wilkinson, T.N., Speed, T.P., Tregear, G.W., and Bathgate, R.A. 2005a. Coevolution of the relaxin-like peptides and their receptors. *Ann. N.Y. Acad. Sci.* **1041**: 534–539.
- Wilkinson, T.N., Speed, T.P., Tregear, G.W., and Bathgate, R.A. 2005b. Evolution of the relaxin-like peptide family. *BMC Evol. Biol.* **5**: 14. doi: 10.1186/1471-2148-5-14.
- Wilkinson, T.N., Speed, T.P., Tregear, G.W., and Bathgate, R.A. 2005c. Evolution of the relaxin-like peptide family: From neuropeptide to reproduction. *Ann. N.Y. Acad. Sci.* **1041**: 530–533.
- Woodburne, M.O., Rich, T.H., and Springer, M.S. 2003. The evolution of

- tribospheny and the antiquity of mammalian clades. *Mol. Phylogenet. Evol.* **28**: 360–385.
- Yang, Z. 1997. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- Yang, Z., Kumar, S., and Nei, M. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**: 1641–1650.
- Yokoyama, S. 2002. Evaluating adaptive evolution. *Nat. Genet.* **30**: 350–351.
- Zhang, J., Zhang, Y.P., and Rosenberg, H.F. 2002. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat. Genet.* **30**: 411–415.
- Zhao, L., Roche, P.J., Gunnarsen, J.M., Hammond, V.E., Tregear, G.W., Wintour, E.M., and Beck, F. 1999. Mice without a functional relaxin gene are unable to deliver milk to their pups. *Endocrinology* **140**: 445–453.
- Zimmermann, S., Steding, G., Emmen, J.M., Brinkmann, A.O., Nayernia, K., Holstein, A.F., Engel, W., and Adham, I.M. 1999. Targeted disruption of the *Insl3* gene causes bilateral cryptorchidism. *Mol. Endocrinol.* **13**: 681–691.

Received September 7, 2007; accepted in revised form February 12, 2008.