Convergent Evolution of Endometrial Prolactin Expression in Primates, Mice, and Elephants Through the Independent Recruitment of Transposable Elements

Deena Emera,*[1] Claudio Casola,[2] Vincent J. Lynch,[1] Derek E. Wildman,[3] Dalen Agnew,[4] and Günter P. Wagner[1]

[1]Department of Ecology and Evolutionary Biology and Yale Systems Biology Institute, Yale University

[2]Department of Biology, Indiana University

[3]Center for Molecular Medicine and Genetics and Department of Obstetrics and Gynecology, Wayne State University School of Medicine

[4]Diagnostic Center for Population and Animal Health, Michigan State University

***Corresponding author:** E-mail: deena.emera@yale.edu.

Associate editor: David Irwin

Research article

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

Abstract

Prolactin (PRL) is a multifunctional signaling molecule best known for its role in regulating lactation in mammals. Systemic PRL is produced by the anterior pituitary, but extrapituitary PRL has also been detected in many tissues including the human endometrium. Prolactin is essential for pregnancy in rodents and one of the most dramatically induced genes in the endometrium during human pregnancy. The promoter for human endometrial *Prl* is located about 5.8 kb upstream of the pituitary promoter and is derived from a transposable element called MER39. Although it has been shown that prolactin is expressed in the pregnant endometrium of a few mammals other than humans, MER39 has been described as primate specific. Thus, in an effort to understand mechanisms of prolactin regulatory evolution, we sought to determine how uterine prolactin is transcribed in species that lack MER39. Using a variety of complementary strategies, including reverse transcriptase–polymerase chain reaction, 5′ rapid amplification of cDNA ends, and whole-transcriptome sequencing, we show that endometrial *Prl* expression is not a shared character of all placental mammals, as it is not expressed in rabbits, pigs, dogs, or armadillos. We show that in primates, mice, and elephants, prolactin mRNA is transcribed in the pregnant endometrium from alternative promoters, different from the pituitary promoter and different from each other. Moreover, we demonstrate that the spider monkey promoter derives from the long terminal repeat (LTR) element MER39 as in humans, the mouse promoter derives from the LTR element MER77, and the elephant promoter derives from the lineage-specific LINE retrotransposon L1-2_LA. We also find surprising variation of transcriptional start sites within these transposable elements and of *Prl* splice variants, suggesting a high degree of flexibility in the promoter architecture even among closely related species. Finally, the three groups shown here to express endometrial prolactin—the higher primates, the rodents, and the elephant—represent three of the four lineages that showed adaptive evolution of the *Prl* gene in an earlier study (Wallis M. 2000. Episodic evolution of protein hormones: molecular evolution of pituitary prolactin. J Mol Evol. 50:465–473), which supports our findings and suggests that the selective forces responsible for accelerated *Prl* evolution were in the endometrium. This is the first reported case of convergent evolution of gene expression through the independent recruitment of different transposable elements, highlighting the importance of transposable elements in gene regulatory, and potentially adaptive, evolution.

Key words: transposable elements, alternative promoters, convergent evolution, endometrial prolactin.

Introduction

Prolactin (PRL), a polypeptide hormone primarily known for its role in regulating lactation in humans and other mammals, is produced by the anterior pituitary and released into the blood stream. Extrapituitary PRL production has also been detected in a variety of tissues, including the decidua, lymphoid cells, myometrium, mammary glands, and brain cells, other than pituitary cells to name a few (for a review, see BenJonathan et al. 1996). In humans, *Prl* is one of the most dramatically induced genes in decidualized endometrial stromal cells, which contribute the maternal component of the placenta, and is one of the most abundant secretory products in the amniotic fluid. Rodents also express the *Prl* gene in decidualized stromal cells during pregnancy (Prigent-Tessier et al. 1999; Bao et al. 2007). The prolactin hormone in rodents is required to sustain progesterone production in the ovary during pregnancy, but it has also been shown that prolactin action at the level of the decidua is necessary to sustain rodent pregnancies (Bao et al. 2007). *Prl*/*Prl-R* knockout mice treated with progesterone have a high rate of abortion starting from midpregnancy onward, which is caused by elevated endometrial levels of IL-6 and 20αHSD, two proteins detrimental to the maintenance of pregnancy (Bao et al. 2007). Expression of IL-6 and 20αHSD, which are absent from the decidua of wild-type mice, is repressed by prolactin (Bao et al. 2007).

In humans, the PRL protein expressed in the pituitary and in the endometrium are identical, but the mRNAs

© The Author 2011. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

differ in their 5′-untranslated regions (UTRs) and the transcripts originate from alternative promoters. The transcriptional start site (TSS) of the human *Prl* transcript expressed in the endometrium is about 5.8 kb upstream of the first coding exon of *Prl* and is located within a long terminal repeat (LTR) transposable element previously described as primate specific (Jurka et al. 2005; Gerlo et al. 2006). This LTR element, named MER39, is a member of the ERV1 class of endogenous retroviruses.

Alternative promoters are a widespread feature of mammalian genes; ~50% of human genes have alternative promoters (Davuluri et al. 2008), some of which are derived from LTR elements (Cohen et al. 2009). The contribution of transposable elements to gene regulatory innovation is an emerging theme from many genomic investigations (Feschotte 2008; Oliver and Greene 2009) and may be specifically important in the origin of evolutionary novelties (Wagner and Lynch 2010). Thus, a detailed understanding of how transposable elements contribute to the origin of novel promoters is essential for understanding how gene regulation evolves.

Prl expression at the maternal–fetal interface (MFI) has been reported from human, rhesus macaque, rat, and mouse (Riddick et al. 1978; Brown and Bethea 1994; Prigent-Tessier et al. 1999; Bao et al. 2007). In an attempt to determine whether *Prl* expression at the MFI is a derived feature of placental mammals, we previously tested MFI tissue from opossum and elephant and the oviduct of chickens for the presence of *Prl* transcripts (Lynch et al. 2008). We found no expression in chicken or opossum, but strong expression in elephant. Since the most recent common ancestor of humans and elephants is also the most recent common ancestor of all placental mammals, the most parsimonious interpretation of our data was that *Prl* expression at the MFI is a shared derived (synapomorphic) character of all placental mammals (Lynch et al. 2008). However, since the TSS of the human and macaque *Prl* transcript at the MFI is derived from a lineage-specific transposable element, the question arose: what is the mechanism of *Prl* expression in species that lack MER39? In this paper, we address this question by presenting 5′ rapid amplification of cDNA ends (RACE), transcriptomic and genomic evidence that *Prl* expression at the MFI is not a shared character of all placental mammals. We demonstrate that endometrial *Prl* expression in primates, mice, and elephants is not synapomorphic but based on the independent derivation of alternative promoters in these lineages. Moreover, we show that in all three groups, the promoters are derived from different transposable elements. This is the first example of convergent evolution of gene expression where orthologous genes utilize different transposable elements to obtain a similar pattern of tissue-specific expression.

Materials and Methods

Alignment and Sequence Comparisons

MER39 sequences were retrieved from Ensembl (www.ensembl.org), and the L1-2_LA sequence including exon 1LA was obtained from the University of California–Santa Cruz (UCSC) Genome Browser (Kent et al. 2002). Sequences were aligned using ClustalW2 (Larkin et al. 2007) and MAFFT (Katoh et al. 2005) and adjusted by eye. Pairwise sequence identities were determined from these adjusted alignments in ClustalW2 and BLAST (bl2seq; Altschul et al. 1997). Consensus sequences used for BLAST analysis were obtained from Repbase (Jurka et al. 2005). Phylogenetic trees were built with PhyML 3.0 (Guindon and Gascuel 2003) and MEGA4 (Tamura et al. 2007) and edited with MEGA4.

Samples

Although our goal was to isolate decidual tissue from pregnant females of all the species we tested, it was not technically possible to completely separate maternal and fetal tissues at the MFI of pregnant animals. Thus, we refer to all of our samples as originating from the MFI. We obtained *Sus scrofa* (pig) tissue from a cesarian section near term. Two samples of *Oryctolagus cuniculus* (rabbit) tissue were obtained: one from a female at term and one from a female sacrificed after 8 days of pregnancy. *Canis familiaris* (dog) tissue was obtained from a female near term. Mouse RNA was collected from antimesometrial tissue at stage E7.5 of C57Bl6 mice. Tissue was obtained from midstage pregnant *Dasypus novemcinctus* (armadillo). *Loxodonta africana* (elephant) and *Ateles fusciceps* (black-headed spider monkey) tissues were obtained from females near term. Tissues were stored in the RNA stabilization agent RNAlater (Qiagen) until RNA extractions were performed. RNA was extracted and purified from tissues using Trizol and/or the Qiagen RNA-Easy Midi RNA-extraction kit in conjunction with the RNase-Free DNase Set following the manufacturer's protocol to remove any contaminating DNA from the sample (Valencia, CA). 5′ RACE ready cDNA was made using the SMART RACE cDNA Amplification Kit (Clontech, Mountain View, CA).

5′ RACE, Polymerase Chain Reaction, and mRNAseq

For each species, a 5′ RACE was attempted with at least two different 5′ RACE reverse primers and the reagents in the SMART RACE cDNA Amplification Kit (supplementary table S1, Supplementary Material online). In addition, the 5′ RACE primers were tested on genomic DNA of each species (with additional forward primers designed within the same exon) to test whether the 5′ RACE primers are effective. A nested 5′ RACE was also attempted. In species in which there was no 5′ RACE product, amplification of other genes expressed in the decidua of humans (IGFBP1, HOXA11, GAPDH) was attempted to confirm the quality and identity of the cDNA in question. Also, a polymerase chain reaction (PCR) of prolactin with primers designed to span introns (also checked on genomic DNA) was attempted in case prolactin transcripts were present but the 5′ RACE was failing.

For the elephant, spider monkey, and mouse, we obtained 5′ RACE products. For the elephant, we used the

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

universal primer mix (UPM) from the SMART RACE cDNA Amplification Kit and a reverse primer specific to elephant *Prl* exon 1 (*L. africana* PRLX1R; supplementary table S1, Supplementary Material online) for the 5′ RACE PCR. For the spider monkey, we did a nested PCR. The first PCR used the UPM from the SMART RACE cDNA Amplification Kit and the species-specific *A. fusciceps* PRLX3R (supplementary table S1, Supplementary Material online). The nested PCR used the product from the first PCR as template, the nested universal primer from the SMART RACE cDNA Amplification Kit, and *A. fusciceps* PRLX2R (supplementary table S1, Supplementary Material online). For the mouse, we did two nested PCRs. The first set of PCRs used the UPM from the SMART RACE cDNA Amplification Kit and two different species-specific primers, *Mus musculus* PRLX4R and PRLX3R (supplementary table S1, Supplementary Material online). The nested PCR used the products from the first set of PCRs as template, the nested universal primer from the SMART RACE cDNA Amplification Kit, and *M. musculus* PRLX3R (with PRLX4R product) and PRLX2R (with PRLX3R product) (supplementary table S1, Supplementary Material online). RACE products were visualized on a 1% agarose gel, gel extracted, and were cloned using the TOPO TA Cloning Kit for Sequencing. At least three clones from each gel band were sequenced.

Of the species in which we failed to find evidence of prolactin expression at the MFI by PCR, we chose two on which to perform whole-transcriptome sequencing (mRNAseq) to confirm that prolactin is absent. We extracted total RNA using the methods above. Total RNA was assayed with a Bioanalyzer 2100 (Agilent) and found to be of excellent quality. Aliquots were sequenced using the Illumina Genome Analyzer II platform, following the protocol suggested by Illumina for sequencing of cDNA samples. Sequence analysis was performed with Bowtie and reads mapped to the dog (canFam2.0) and armadillo (dasNov2) cDNA builds at Ensembl; two mismatches were allowed, and reads aligning to more than one cDNA were discarded. Sequencing was performed at the W.M. Keck Microarray at the Yale University Medical School.

Transient Transfection Experiment

To make the elephant *Prl* promoter reporter construct (elephant.dPRL), a PCR on genomic *L. africana* DNA was performed which amplified a region from −9771 to −8912 relative to the start codon of elephant *Prl*. This 856 bp product was cloned into the pGL4.70 reporter vector (Promega). Human (human.dPRL) and spider monkey (*Ateles paniscus*) (spidermonkey.dPRL) *Prl* promoter reporter constructs were made in a similar way: a roughly 720 bp region including MER39 was cloned into pGL4.70. For transfection, human endometrial stromal cells (ATCC Number CRL-4003) were plated at a density of about 6.0×10^4 cells/well in 24-well plates in media containing 10% charcoal-stripped Fetal bovine serum (FBS). Four hundred and fifty nanograms of elephant.dPRL, human.dPRL, spidermonkey.dPRL, or empty PGL4.70 was cotransfected with 50 ng of the internal control vector PGL4.23 (Promega) using the Lipofectamine LTX reagent (Invitrogen). Six hours after transfection, the media was changed to one containing 2% charcoal-stripped FBS. Twenty-four hours after transfection, the media was changed to one containing 10^{-6} M medroxyprogesterone acetate (MPA, a stable synthetic progestin) and 0.5 mM 8-bromoadenosine cyclic monophosphate (8-Br-cAMP, a stable cAMP analog) or a control medium containing ethanol as a vehicle control. Forty-eight hours later, cells were harvested, and a luciferase assay was performed using the Dual-Luciferase Reporter Assay System (Promega). *Renilla* luciferase activity was normalized to firefly luciferase activity, and fold change values were obtained by comparing normalized luciferase activity from MPA/cAMP-treated cells and control cells. Transfections were performed with six replicates, and the experiments were performed twice; values are represented as means ± standard error of the mean.

Results

First, we investigated whether MER39 at the *Prl* locus is indeed limited to primates by conducting BLASTn searches of the consensus MER39 sequence against nonprimate genomes and carefully examining the prolactin locus in other mammals. We found that MER39 is present within 8 kb upstream of the first coding exon of *Prl* in all primates with genome sequence available (supplementary table S2, Supplementary Material online). In the bush baby (*Otolemur garnettii*), a basal primate, it appears that there was a recombination event between MER39 and a downstream transposable element (MER57C1), resulting in the deletion of most of MER39 and about 2.5 kb of genomic sequence. Unexpectedly, we also found MER39 upstream of the orthologous *Prl* locus in the mouse and rat, though it differs greatly from that in humans (56% and 55% identity, respectively); this extreme divergence is likely why it was not identified in an earlier study (Gerlo et al. 2006). Although there is a large family of prolactin paralogs in rodents clustered together on the same chromosome (13 in mice and 17 in rats; Wiemers et al. 2003), MER39 is found associated with only one copy (*Prl*) in the region, about 9 kb upstream of the *Prl* locus orthologous to that in humans. In addition to mouse and rat, we identified the orthologous MER39 in squirrel (66% identity) and partial sequences in guinea pig, rabbit, and tree shrew. Extensive searches of other mammalian genomes failed to identify additional species with MER39 elements (supplementary table S2, Supplementary Material online). Thus, the MER39 family of LTR transposable elements has a deeper phylogenetic presence than was previously thought and is specific to the clade that includes primates and rodents, that is, Euarchontoglires (fig. 1).

Given that prolactin is essential during rodent pregnancy and so dramatically upregulated in the endometrium during human pregnancy, we next explored if and from what TSS *Prl* is transcribed at the MFI of other placental mammals using a variety of complementary strategies, including 5′ RACE, nested 5′ RACE, and reverse transcriptase–PCR. We identified *Prl* transcripts in a New World monkey (*A. fusciceps*) and African elephant

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

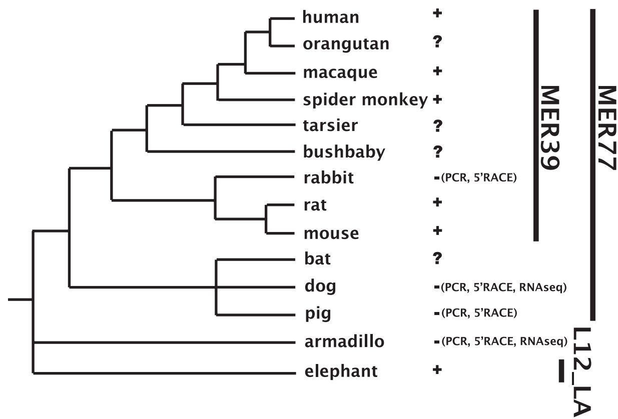

Fig. 1. Phylogeny of various eutherian mammals showing distribution of transposable element families (MER39, MER77, and L1-2_LA) and *Prl* expression at the MFI. Species with a plus are those in which there is evidence of *Prl* expression from prior studies (human: Riddick et al. 1978; macaque: Brown and Bethea 1994; mouse: Bao et al. 2007; and rat: Prigent-Tessier et al. 1999) and this study (spider monkey, mouse, and elephant). Species with a minus are those in which *Prl* is not expressed at the MFI (strategies used in this study to detect *Prl* are in parentheses). Species with a question mark are those that have not been tested. (Tree modified from Murphy et al. 2007; Churakov et al. 2009.)

(*L. africana*) and confirmed *Prl* expression in mouse (*M. musculus*); however, after exhaustive searches, we were unable to identify *Prl* transcripts in rabbit, dog, pig, and armadillo samples (fig. 1). To confirm that *Prl* is not expressed at the MFI of these species, we sequenced the transcriptome of dog and armadillo MFI tissue and found no *Prl* transcripts in over 13.6 and 20 million reads, respectively.

Primate Endometrial *Prl* Promoter

We were successful at obtaining 5′-UTR sequences from all three species in which we identified transcripts. For *A. fusciceps*, the black-headed spider monkey, we performed a nested 5′ RACE and identified three distinct transcripts (fig. 2*a* and *e*; supplementary fig. S1, Supplementary Material online). Mapping these transcripts against the marmoset (*Callithrix jacchus*) genome, the most closely related New World monkey with a genome available, indicates that transcription of *Prl* at the MFI initiates at two locations within the MER39-derived promoter, 25 bp downstream and 52 bp upstream of the major human start site (fig. 2*e*; Dimattia et al. 1990). In addition, the spider monkey transcript has gained a novel non–protein-coding exon (fig. 2*e*). In human, the 80 bp non–protein-coding exon (termed exon 1a) is spliced to a position 98 bp upstream of the start codon (Dimattia et al. 1990). In spider monkey, however, exon 1a is spliced to a position 1 kb downstream, where an additional 54 bp exon (termed here exon 1b) is spliced to a position 98 bp upstream of the start codon. In the shortest spider monkey transcript, transcription is initiated at the more downstream TSS, and exon 1b is spliced to the beginning of exon 2 (fig. 2*e*). As this shortest transcript lacks exon 1, and thus the canonical start codon, it is unclear if/how this transcript is translated. Although we do not know its function, if any, it is worth noting that there are genes known to produce functional non–protein-coding and protein-coding transcripts, as is the case with the *SRA1* gene. Alternative splicing of *SRA1* produces a non-coding RNA and protein, both of which are thought to regulate steroid receptor signaling pathways (see Leygue 2007).

Mouse Endometrial *Prl* Promoter

In *M. musculus*, we used a nested 5′ RACE strategy to obtain the 5′-UTR of endometrial *Prl*, which yielded a number of products between 100 and 500 bp (fig. 2*b* and *f*). Sequences of these products indicate that transcription of the *Prl* gene at the MFI in mouse does not initiate at the same position as the pituitary *Prl* transcript, nor from MER39 as it does in primates. Rather, it initiates in a region about 53 kb upstream of its coding region, in three locations within 60 bp of each other (fig. 2*f*; supplementary fig. S1, Supplementary Material online). In some transcripts, this first non–protein-coding exon (termed here exon 1MMa) is spliced to a position about 2 kb downstream of it (to another non–protein-coding exon termed here exon 1MMb); in other transcripts, exon 1MMa is spliced to a position about 2.4 kb upstream of coding exon 1 to a third non–protein-coding exon (termed here exon 1MMc); and in a third set of transcripts, exon 1MMa is spliced directly to coding exon 2. The only transcript that includes coding exon 1 initiates at the more upstream initiation site, is spliced to exon 1MMc, and is spliced to a position 90 bp upstream of the start codon (fig. 2*f*). In the transcripts without coding exon 1, it is unclear if and where translation initiates (see above for possible role of transcripts missing the canonical start codon).

The TSS of all sequenced mouse *Prl* transcripts is located in a region of the mouse genome annotated as containing an LTR element called MER77. MER77 belongs to the ERV3 class of endogenous retroviruses. The MER77 element upstream of the *Prl* locus in mouse shows orthologous MER77 elements in a variety of placental mammals in the clades Euarchontoglires and Laurasiatheria, including human, dog, and horse, but not in members of the other two major clades of placentals, Xenarthra and Afrotheria (supplementary table S2, Supplementary Material online). According to annotation at the UCSC Genome Browser, exon 1MMa from the mouse transcripts maps to a position about 40 bp downstream of the 5′-end of MER77 (which is in the reverse orientation relative to *Prl*). However, annotation of the orthologous MER77 in the human genome is more complete at the 5′-end and includes the region orthologous to exon1 MMa in mouse. In fact, a careful examination of the MER77 region in mouse suggests that MER77 is incompletely annotated at its 5′-end in mouse: the extended mouse MER77 (which includes exon 1MMa) shares 59% identity with the consensus MER77 sequence from humans (http://www.girinst.org/). Thus, it appears that the promoter for *Prl* at the MFI in mouse is derived from the transposable element MER77, an LTR element specific to members of the Boreoeutheria clade of placental mammals. A scan of the genomic region in rodents containing the prolactin family of genes

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

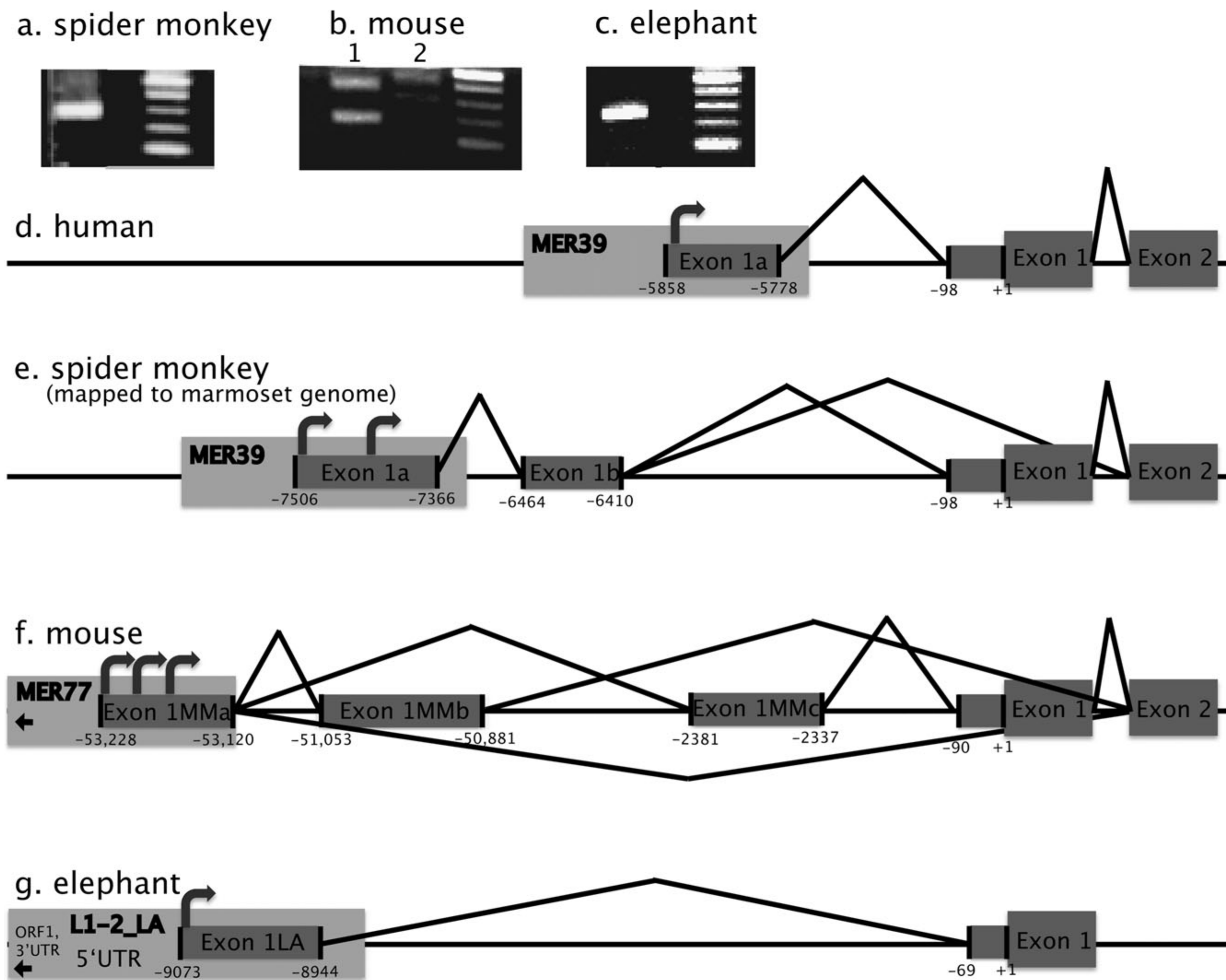

Fig. 2. Expression (*a*, *b*, and *c*) and splicing scheme (*d*, *e*, *f*, and *g*) of prolactin transcripts from the MFI in spider monkey (*a*, *e*), mouse (*b*, *f*), elephant (*c*, *g*), and human (*d*; Dimattia et al. 1990). (*a*, *b*, and *c*) Left lane(s), RACE product (cDNA template; forward primer from SMART RACE cDNA Amplification Kit; *Prl* species-specific reverse primer); right lane, 100-bp ladder. In (*b*), lanes 1 and 2 use different sets of *Prl* mouse-specific reverse primers—see Materials and Methods. (*d*, *e*, *f*, and *g*) Positions are relative to human *Prl* translational start site (+1) or orthologous position in mouse, elephant, and marmoset genomes (the marmoset was used to map the spider monkey transcript as it is the most closely related New World monkey with a genome available). Bent arrows indicate TSSs at the MFI. For human, only the major TSS is indicated (see Dimattia et al. 1990 for minor start sites, all located within 25 bp of major start site). Black carrots indicate splice junctions. Transposable elements are shaded in light gray. Exons are shaded in dark gray; UTRs are shown as narrow boxes, and coding regions are shown as thicker boxes.

reveals that MER77 is only associated with one paralogue, *Prl* (as described for MER39).

Elephant Endometrial *Prl* Promoter

The elephant *Prl* 5′-UTR sequence was obtained by 5′ RACE, which generated one 229 bp sequence (fig. 2*c* and *g*; supplementary fig. S1, Supplementary Material online). Mapping this sequence against the elephant genome indicates that transcription of *Prl* at the MFI in elephant initiates about 9 kb upstream of the translational start site (fig. 2*g*). A 130 bp noncoding exon located at this position (termed here exon 1LA) is spliced to a position 70 bp upstream of coding exon 1. We have shown above that both MER39 and MER77 elements are absent from Afrotheria; thus, the TSS of this *Prl* transcript from the MFI in elephant is clearly different from that in primates and rodents.

As annotation of the elephant genome is less complete than that of the human and mouse and as *Prl* expression in the elephant endometrium has not been described in detail, we performed a deeper examination of the promoter sequence and function in this species. An initial investigation of exon 1LA revealed several copies scattered through the elephant genome, suggesting that it could also be derived from a transposable element. To confirm this possibility, we retrieved the sequence surrounding exon 1LA and delimited the longest possible repetitive unit of the putative transposable element by sequence similarity searches on the elephant genome (supplementary information, Supplementary Material online). This allowed us to identify a 4,208 bp sequence that includes exon 1LA at position 370–498 and has multiple copies in the elephant genome. A BLASTn search on the Repbase library of transposable

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

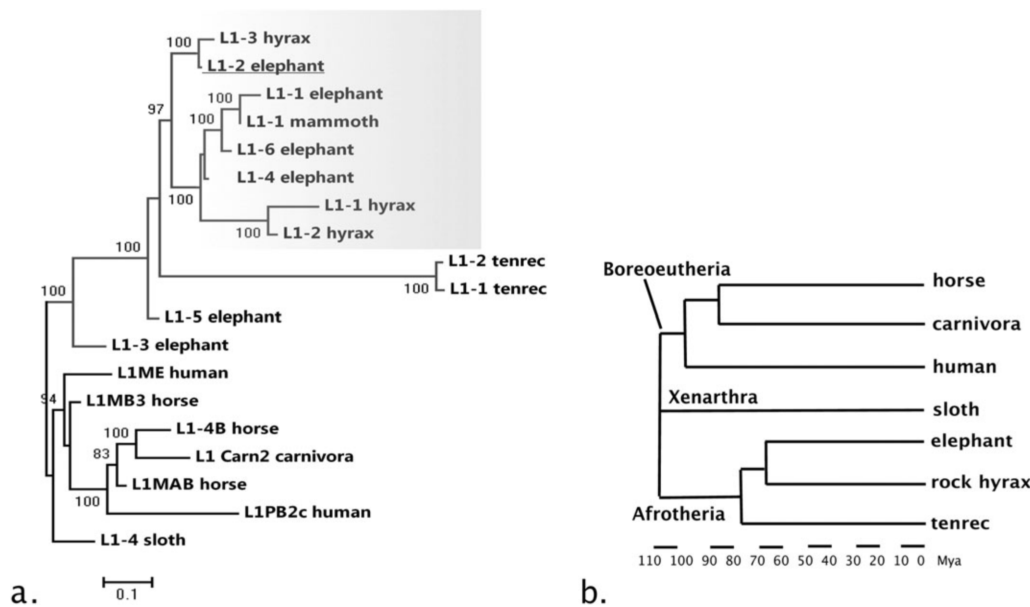

Fig. 3. Evolutionary relationships of L1 families in placental mammals. (*a*) Phylogenetic tree of L1-2_LA and 18 related mammalian L1 families. The tree was built using the PhyML 3.0 algorithm on the multialignment of the 3,307 bp long ORF2. The L1-2_LA family name is underlined. Branches of the Afrotheria clade are highlighted. The shaded box delimits L1 families found in Proboscidea and Hyracoidea. Species names adopted by Repbase are removed (i.e., L1-2_LA is reported as L1-2 elephant). Bootstrap values lower than 70 are not shown. (*b*) Evolutionary relationships and timescale of mammals with analyzed L1 families (modified from Murphy et al. 2007; Churakov et al. 2009).

elements (http://www.girinst.org/) revealed that the 4,208 bp sequence shares ~85% identity over 3705 bp with the LINE retrotransposon family L1-2_LA recently identified in the elephant genome (supplementary fig. S2, Supplementary Material online). This is similar to the ~90% identity that the L1-2_LA consensus sequence shares with the most closely related copies of this family (Jurka et al. 2005), suggesting that exon 1LA originated from an L1-2_LA element that inserted upstream of the prolactin gene (supplementary information, Supplementary Material online). The L1-2_LA consensus sequence itself does not include exon 1LA, likely because of incomplete annotation of the L1-2_LA 5′-UTR; indeed, 531 of the 572 exon 1LA-related sequences we identified in the elephant genome lie within a kilobase upstream of an L1-2_LA element.

To determine the lineage specificity of the L1-2_LA family of transposable elements, we searched for evolutionarily related L1 families in other mammalian genomes (supplementary information, Supplementary Material online). We found closely related L1-2_LA families in elephant and other Afrotherians, but not in other species of placental mammals. We used the consensus sequences of these families to reconstruct their evolutionary relationships and found that L1-2_LA and other L1 families from African elephant, woolly mammoth, lesser hedgehog tenrec, and rock hyrax form a distinct clade of elements that is afrotherian specific (fig. 3). Several of these afrotherian-specific L1 families, including L1-2_LA, diversified further in two sister afrotherian orders, the Proboscidea and Hyracoidea, which diverged around 64 Ma (Murphy et al. 2007). L1 elements from the only other available afrotherian genome, the tenrec, branch outside of the main elephant–rock hyrax L1 group (fig. 3). Since the elephant and rock hyrax lineage diverged from the tenrec about 78 Ma (Murphy et al. 2007), the approximate origin of the L1-2_LA family is 64–78 Ma. However, while the L1-2_LA family predates the divergence of the elephant and hyrax lineages, the *Prl* locus in hyrax does not have an L1-2_LA insertion (data not shown). Thus, exon 1LA originated from an afrotherian-specific transposable element that inserted upstream of *Prl* sometime after the divergence of elephants and hyraxes. Given these constraints, we can date the earliest possible use of L1-2_LA as a promoter, and by extension the origin of *Prl* expression at the MFI in elephant, at ~64 Ma.

To better characterize the evolutionary pathway of L1-2_LA from a transposable element into a promoter, we reannotated the 5′-UTR of the L1-2_LA family to include the first ~570 bp that are missing from the consensus sequence. This reconstructed 5′-UTR contains two direct repeats of 86 bp, one of which is partly contained in exon 1LA (supplementary fig. S3, Supplementary Material online). Mapping this reconstructed 5′-UTR onto the elephant *Prl* locus indicates that the L1-2_LA element inserted in the opposite direction compared with the *Prl* gene (fig. 2g). Interestingly, the 5′-UTR of primate L1 elements has an antisense promoter that is involved in the transcription of human genes (Speek 2001), suggesting that a mechanism for domesticating transposable elements for host functions may be through co-option of antisense promoters into alternative promoters for nearby host genes.

To directly test if L1-2_LA can function as a promoter, we cloned approximately 900 bp of the region including and surrounding exon 1LA into a promoterless luciferase reporter vector (elephant.dPRL; fig. 4) and tested its ability to activate luciferase expression in human endometrial

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

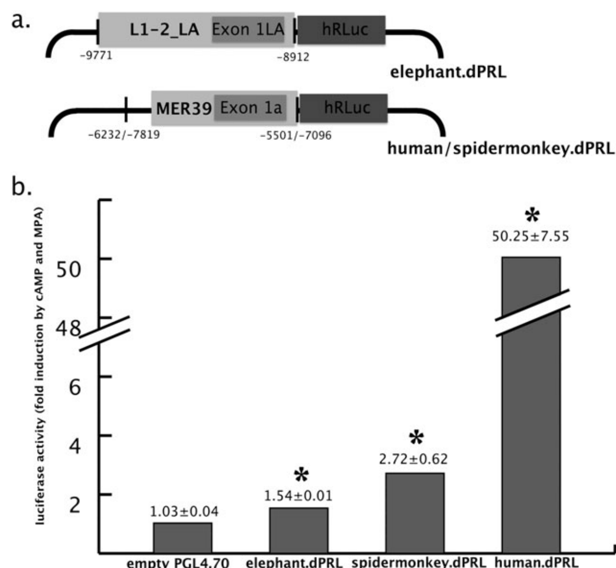

FIG. 4. Activity of elephant (*Loxodonta africana*), human (*Homo sapiens*), and spider monkey (*Ateles paniscus*) decidual *Prl* promoters in human endometrial stromal cells. (*a*) Structure of L1-2_LA and MER39 cloned into the PGL4.70 vector. hRLuc, renilla luciferase gene (dark gray box). Positions relative to translational start site of *Prl* (spider monkey positions are based on marmoset genome). (*b*) Fold induction over unstimulated controls of empty PGL4.70, elephant.dPRL, spidermonkey.dPRL, and human.dPRL in human endometrial stromal cells in response to 0.5 mM 8-bromoadenosine cyclic monophosphate (8-Br-cAMP) and 10^{-6} M MPA. Values are expressed as the mean ± standard error of the mean ($n = 12$). *, *t*-test, $P = 0.003$ (elephant.dPRL), $P = 0.020$ (spidermonkey.dPRL), $P = 0.001$ (human.dPRL) versus fold change empty PGL4.70.

stromal cells. We also tested the human and spider monkey (*A. paniscus*) MER39-derived promoters for comparative purposes. We found that exon1 LA and the surrounding sequence from elephant induced a 1.5-fold increase in luciferase activity compared with untreated controls in human endometrial stromal cells transiently transfected with elephant.dPRL and treated with 8-Br-cAMP/MPA to induce differentiation into decidual cells (fig. 4). The spider monkey promoter induced a 2.7-fold increase, and the human, a 50-fold increase of luciferase in response to the differentiation agents (primate decidual *Prl* promoter activities will be further explored in a future study). The finding that the elephant promoter induced a response in human cells is remarkable given that this sequence is only found in afrotherians, and the *trans*-regulatory landscape of human and elephant decidual cells is likely to vary after more than 200 My of independent evolution. Thus, the modest but highly significant activity of the elephant promoter (*t*-test, $P = 0.003$ vs. fold change of empty vector) indicates that the L1-2_LA situated upstream of the elephant prolactin gene contains the promoter for prolactin expression at the MFI in elephants.

Discussion

Our data show that *Prl* expression at the MFI evolved at least three times, twice in Euarchontoglires and once in Afrotheria. We found no evidence for *Prl* expression at the MFI in the two other major clades of mammals, the Xenarthra with armadillo as a representative species and the Laurasiatheria represented by the pig and the dog. In the higher primates, mice, and elephants, *Prl* mRNA at the MFI is transcribed from alternative promoters, different from the pituitary promoter and different from each other. In each case, the alternative promoter is derived from a different transposable element. In the higher primates, *Prl* transcription initiates from the LTR element MER39, which is specific to Euarchontoglires. In the mouse, transcription initiates from the LTR element MER77, which is present in Euarchontoglires and Laurasiatheria. In the elephant, the alternative promoter is derived from a member of the LINE L1-2_LA family, which is unique to Afrotherians.

While there are known cases of transposable elements being co-opted for promoter and enhancer functions, this is the first reported case of the convergent evolution of gene expression through the independent recruitment of different transposable elements. Since endometrial prolactin expression is not a shared character of all placental mammals as shown in this paper, co-option of MER39 in the higher primates, L1-2_LA in elephants, and MER77 in mice was involved in the evolution of its expression in these groups. However, it is also clear that the insertion of these elements alone (at least in the case of Euarchontoglires) was not sufficient to recruit prolactin into uterine expression during pregnancy. Primates and rodents have both MER39 and MER77, but only the higher primates use MER39 and rodents MER77 for *Prl* expression. In addition, rabbits have both elements but do not express endometrial prolactin. Also, a recent report on *Prl* expression in guinea pigs suggests that, like the rabbit, they have both elements but do not express endometrial *Prl* during pregnancy (Alam et al. 2010). Thus, it appears that the elements provided the substrate necessary for *Prl* expression, but additional mutations were necessary to initiate its expression in the endometrium.

While it is unclear what precise role(s) prolactin has in the decidua of groups in which endometrial expression evolved, it is notable that these groups—the higher primates, the rodents, and the elephant—represent three of the four groups that show accelerated, and arguably adaptive, evolution of the *Prl* gene (Wallis 2000). Wallis (2000) demonstrated that the rate of *Prl* sequence evolution was slow during much of tetrapod evolution but that this rate increased four times in the history of eutherian mammals: in primates, rodents, elephants, and ruminant artiodactyls (ruminants were not tested in our study for endometrial prolactin expression for lack of available tissue). Moreover, Li et al. (2005) showed that the *Prl-R* gene has coevolved with *Prl* in these groups. It has been suggested that the basis of this accelerated evolution is maternal–fetal conflict (Haig 2008), or "function switching," whereby prolactin acquired a second function whose importance fluctuated over time (Forsyth and Wallis 2002). While this question has not been answered, the remarkable correlation between groups that express endometrial

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

prolactin and those that experienced adaptive evolution of the ligand and its receptor suggests that the selective force(s) responsible for the adaptive changes were in fact in the endometrium. Future work should test whether the fourth group that showed accelerated evolution of prolactin, the ruminants, also expresses endometrial prolactin. Since a possible consequence of the accelerated changes in prolactin was a change in ligand specificity (see Li et al. 2005), future work should also investigate whether prolactin ligand–receptor interactions are different in decidual tissue in these groups. It is known, for example, that humans, rodents, and ruminants express multiple *Prl-R* isoforms and that humans and rodents produce a full-length and cleaved form of prolactin that act through different receptors and have opposing functions (Freeman et al. 2000; Jabbour and Critchley 2001). Whether this variety is also seen in groups that have not experienced adaptive evolution of the ligand/receptor and whether there are any functional consequences in the uterus remains to be determined.

If, indeed, adaptive evolution of *Prl* occurred because of selective forces in the pregnant endometrium, this would suggest that the TEs co-opted for endometrial *Prl* expression in the primate, rodent, and elephant lineages were involved in these adaptive events since *Prl* expression in the endometrium likely preceded evolution of the protein-coding sequence. Thus, our work not only demonstrates for the first time the convergent evolution of gene expression through the recruitment of different transposable elements, but it suggests that adaptive evolution can be mediated by TEs. This highlights the importance of transposable elements as drivers in gene regulatory, and potentially adaptive, evolution.

Supplementary Material

Supplementary figures S1–S4, tables S1 and S2, and supplementary information are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org).

Acknowledgments

We thank the Broad Institute and the Baylor College of Medicine for providing the elephant and hyrax genome assemblies, respectively, Dr Birgit Gellersen for advice on the transient transfection experiment, Drs Jimmy Zhou and Martín Garcia-Castro for rabbit endometrial tissue, and Dr Michael Soares and Naveen K. Neradugomma for mouse endometrial tissue. This work was supported by the John Templeton Foundation (grant #12793). The opinions expressed in this article are not that of the John Templeton Foundation.

References

Alam SMK, Konno T, Rumi MAK, Dong YF, Weiner CP, Soares MJ. 2010. Prolactin family of the guinea pig, Cavia porcellus. *Endocrinology* 151:3918–3928.

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25:3389–3402.

Bao L, Tessier C, Prigent-Tessier A, Li FX, Buzzio OL, Callegari EA, Horseman ND, Gibori G. 2007. Decidual prolactin silences the expression of genes detrimental to pregnancy. *Endocrinology* 148:2326–2334.

BenJonathan N, Mershon JL, Allen DL, Steinmetz RW. 1996. Extrapituitary prolactin: distribution, regulation, functions, and clinical aspects. *Endocr Rev*. 17:639–669.

Brown NA, Bethea CL. 1994. Cloning of decidual prolactin from rhesus macaque. *Biol Reprod*. 50:543–552.

Churakov G, Kriegs JO, Baertsch R, Zemann A, Brosius J, Schmitz J. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res*. 19:868–875.

Cohen CJ, Lock WM, Mager DL. 2009. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* 448:105–114.

Davuluri RV, Suzuki Y, Sugano S, Plass C, Huang THM. 2008. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet*. 24:167–177.

Dimattia GE, Gellersen B, Duckworth ML, Friesen HG. 1990. Human prolactin gene-expression—the use of an alternative noncoding exon in decidua and the IM-9-P3 lymphoblast cell-line. *J Biol Chem*. 265:16412–16421.

Feschotte C. 2008. Opinion—transposable elements and the evolution of regulatory networks. *Nat Rev Genet*. 9:397–405.

Forsyth IA, Wallis M. 2002. Growth hormone and prolactin-molecular and functional evolution. *J Mammary Gland Biol Neoplasia*. 7:291–312.

Freeman ME, Kanyicska S, Lerant A, Nagy G. 2000. Prolactin: structure, function, and regulation of secretion. *Physiol Rev*. 80:1523–1631.

Gerlo S, Davis JRE, Mager DL, Kooijman R. 2006. Prolactin in man: a tale of two promoters. *Bioessays* 28:1051–1055.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.

Haig D. 2008. Placental growth hormone-related proteins and prolactin-related proteins. *Placenta*. 29:S36–S41.

Jabbour HN, Critchley HOD. 2001. Potential roles of decidual prolactin in early pregnancy. *Reproduction* 121:197–205.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 110:462–467.

Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*. 33:511–518.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res*. 12:996–1006.

Larkin MA, Blackshields G, Brown NP, et al. (13 co-authors) 2007. Clustal W and clustal X version 2.0. *Bioinformatics* 23:2947–2948.

Leygue E. 2007. Steroid receptor RNA activator (*SRA1*): unusual bifaceted gene products with suspected relevance to breast cancer. *Nucl Recept Signal*. 5:e006.

Li Y, Wallis M, Zhang YP. 2005. Episodic evolution of prolactin receptor gene in mammals: coevolution with its ligand. *J Mol Endocrinol*. 35:411–419.

Lynch VJ, Tanzer A, Wang YJ, Leung FC, Gellersen B, Emera D, Wagner GP. 2008. Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proc Natl Acad Sci U S A*. 105:14928–14933.

Murphy WJ, Pringle TH, Crider TA, Springer MS, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res*. 17:413–421.

Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. *Bioessays* 31:703–714.

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022

Prigent-Tessier A, Tessier C, Hirosawa-Takamori M, Boyer C, Ferguson-Gottschall S, Gibori G. 1999. Rat decidual prolactin—identification, molecular cloning, and characterization. *J Biol Chem*. 274:37982–37989.

Riddick DH, Luciano AA, Kusmik WF, Maslar IA. 1978. Denovo synthesis of prolactin by human decidua. *Life Sci*. 23:1913–1921.

Speek M. 2001. Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol*. 21:1973–1985.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24:1596–1599.

Wagner GP, Lynch VJ. 2010. Evolutionary novelties. *Curr Biol*. 20:R48–R52.

Wallis M. 2000. Episodic evolution of protein hormones: molecular evolution of pituitary prolactin. *J Mol Evol*. 50:465–473.

Wiemers DO, Shao LJ, Ain R, Dai GL, Soares MJ. 2003. The mouse prolactin gene family locus. *Endocrinology* 144:313–325.

Downloaded from https://academic.oup.com/mbe/article/29/1/239/1749212 by guest on 21 August 2022