



HAL
open science

Gene birth, death, and divergence: the different scenarios of reproduction-related gene evolution.

Xin Tian, Géraldine Pascal, Sophie Fouchécourt, Pierre Pontarotti, Philippe Monget

► **To cite this version:**

Xin Tian, Géraldine Pascal, Sophie Fouchécourt, Pierre Pontarotti, Philippe Monget. Gene birth, death, and divergence: the different scenarios of reproduction-related gene evolution.. *Biology of Reproduction*, 2009, 80 (4), pp.616-21. 10.1095/biolreprod.108.073684 . hal-00831221

HAL Id: hal-00831221

<https://hal.science/hal-00831221v1>

Submitted on 30 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Minireview

Gene Birth, Death, and Divergence: The Different Scenarios of Reproduction-Related Gene Evolution¹

Xin Tian,^{3,4} Géraldine Pascal,^{3,4} Sophie Fouchécourt,⁴ Pierre Pontarotti,⁵ and Philippe Monget^{2,4}

Physiologie de la Reproduction et des Comportements,⁴ Unité Mixte de Recherche 6175, Institut National de la Recherche Agronomique (INRA)-Centre National de la Recherche Scientifique-Université François Rabelais de Tours-Haras Nationaux, Nouzilly, France
Laboratoire de PhyloGénomique,⁵ EA 3781 Evolution Biologique, Université de Provence, Marseille, France

ABSTRACT

Reproductive genes are known to evolve more rapidly than genes expressed in other organs. In this paper we present an overview and bring some new data on the evolutionary study of reproduction-related genes by integrating phylogeny with gene genomic localization. We focus on the gene evolutionary processes of gene birth, death, and divergence. We show that phylogenetic gene birth is confirmed by gene location in genomes, which definitively localized the “place of birth” of new genes (such as *Obox* and *KHDC1/DPPA5/ECAT1/OOEP* gene families). By finding their “place of death” in genomes, it also demonstrates that *ZP* genes *TGM4* and *OVGP1* have been lost in certain species during vertebrate evolution. Moreover, in the case of gene divergence, comparison of gene locations across different genomes establishes orthologous relationships that are weakly supported by the phylogenetic tree. Specifically, genomic localization demonstrates that the fish and bird *mtnr1c* (*Mel1C*) receptor is orthologous to mammalian *GPR50*, and that ungulate genomes contain new seminal vesicle-specific *BSP* genes that are not present in other species. Overall, the phylogenomic approach to gene evolution presented in this paper offers more insight into gene function, such as species-specific duplications for speciation, changes in gene expression due to gene divergence, and functional loss by gene death.

BSP, evolution of reproduction-related genes, genomic location, Obox, OVGP1, ovum, phylogeny, prostate, seminal vesicles, TGM4

INTRODUCTION

For more than a decade, increasing evidence has shown that reproduction-related genes are the most highly divergent and evolve more rapidly than genes expressed in other organs [1–5]; therefore, phylogenetic study of these genes is particularly

challenging. Until now, molecular phylogeny (based on sequence alignment) has remained the most relevant method to reconstruct gene evolutionary history. However, the divergent evolutionary rate of reproductive genes frequently generates problems in the phylogenetic study, such as the artificial structures caused by long-branch attraction [6] in reconstructed trees. Although the putative “birth” of a new gene in a phylogenetic tree might be predicated by the identification of paralogs (genes in the same species that originate from a single gene via duplication) and/or orthologs (genes in a different species that originate from a single gene in the last common ancestor via speciation) [7], the event might also be a false scenario caused by divergent evolutionary rates among gene family members [8]. This uncertainty can be resolved by confirming the gene’s “place of birth” in the genome. Similarly, gene “death” can be proven definitively when the pseudogene is found in the corresponding place in the genome. Now, with the explosion of sequences in databases and the availability of more and more sequenced genomes, the comparison of gene location in genomes has shown to be an advantage in gene evolution studies. By investigating gene location within blocks of conserved genomic regions in different organisms, we can establish their orthologous relationships, as well as localize the place of “gene birth” and “gene death” in genomes. In this paper we make an overview of the evolution of several reproduction-related genes that have been analyzed by phylogeny within the context of fully sequenced genomes. Hopefully, this will help us to better understand the main evolutionary processes of these genes, such as gene birth, gene divergence, and gene death.

GENE BIRTH

Gene duplication is thought to be the essential source of new genes with new or altered functions, as shown by the widespread existence of gene families. When the duplicate is fixed in the genome and is functionally preserved by natural selection, the newborn gene may diverge under two primary scenarios. The first classic scenario is neofunctionalization [9], in which the duplicate evolves toward a new function that results in a sequence that is significantly divergent from its ancestral gene. The alternative scenario is subfunctionalization [10], in which the ancestral functions are partitioned between new duplicates in spatial or temporal expression patterns [11]. In all cases, the origin of a new gene can only be demonstrated when its place of gene birth is localized in the genome.

¹Supported by an INRA postdoctoral fellowship to X.T.

²Correspondence: Philippe Monget, Physiologie de la Reproduction et des Comportements, UMR 6175 INRA-CNRS-Université François Rabelais de Tours-Haras Nationaux, 37380 Nouzilly, France.

FAX: 33 0 2 47 42 77 43; e-mail: pmonget@tours.inra.fr

³These authors have contributed equally to this work.

Received: 24 September 2008.

First decision: 10 October 2008.

Accepted: 10 December 2008.

© 2009 by the Society for the Study of Reproduction, Inc.

eISSN: 1259-7268 <http://www.biolreprod.org>

ISSN: 0006-3363

There are several examples of reproductive genes that are subjected to intense duplication processes. For example, *Obox* (oocyte-specific homeobox) genes have been referred to as the first homeobox gene family preferentially expressed in mouse adult germ cells [12, 13]. Phylogeny shows that *Obox* genes are related to some other homeobox genes, such as *MEOX2* and *CRX* (Supplemental Fig. S1a available at www.biolreprod.org). Although available data indicate rodent-specific expression of *Obox*, the origin of this family is not clear. Genomic analysis shows that *Obox* genes are clustered next to a retina-expressed homeobox gene, *Crx*, in the mouse genome. Both sides of this *Obox* gene cluster are conserved in mammalian genomes (Supplemental Fig. S1b), and no *Obox* homolog can be identified in nonrodents (such as human, cow, and dog) by a tBlastn [14] search. Thus, we may hypothesize that the *Obox* family has emerged specifically in rodents after the divergence of primates and rodents. Moreover, it offers evidence that this gene family might have originated from a tandem duplication of a *Crx* ancestor gene, with further multiple duplications resulting in the tandem organization of *Obox* in murine genomes. Notably, a similar tandem multigene cluster is also found in another rodent-specific reproductive homeobox gene family (*Rhox*), which is thought to be responsible for species differences in embryonic stem cell derivation and maintenance in the mouse, rat, and human [15]. The emergence of several gene families exclusively expressed in mouse oocytes (*Oog*, *Nlrp*, *Tcll*, etc.) has been described recently [16–18]. The biological function of these expanding families in rodents is unknown, but it might be a biological novelty of germ cell differentiation, which is important in rodent speciation.

If the gene duplication event is too ancient and/or genes have evolved rapidly, one may not be able to find the ancestral gene that gave birth to the present genes. For example, we have recently identified a new eutheria-specific *KHDC1/DPPA5/ECAT1/OOEP* gene family. These genes are specifically expressed in oocyte and/or embryonic stem cells, but their biological roles are not clear. Genomic analysis showed that in eutherian genomes, these genes had been inserted into a conserved gene order of vertebrates [19]. With the genomic localization of the gene's place of birth, we can be certain that these genes are absent in fish, bird, and marsupial genomes and appeared (and evolved rapidly) in eutherian mammals. However, in the present case we have not found the ancestral gene whose duplication gave rise to the *KHDC1/DPPA5/ECAT1/OOEP* family. The corresponding protein family is characterized by a degenerated KH domain, which confers the ability to bind RNA when canonical. Thus, it is likely that this family arose from a duplication of a KH domain-containing gene, potentially followed by translocations and a particularly fast evolution.

GENE DIVERGENCE

As mentioned above, the divergent evolution of duplicates (paralogs) is suggested to be a preservation mechanism through which some daughter copies keep ancient functions, whereas others evolve toward new biological functions. Divergent evolution may also occur for a single gene among different species lineages, leading to an unequal evolutionary rate of genes (among orthologs). Such divergent evolution could explain specific reproductive strategies in certain lineages due to genes that have evolved particularly quickly. In phylogenetic studies, a divergence of gene evolutionary rates could result in weakly resolved phylogenetic trees and artifacts (such as long-branch attraction) [6, 20]. In these cases, the comparison of

gene location in genomes is also able to help phylogeny identify orthologous relationships.

For example, melatonin, a hormone secreted by the pineal gland in vertebrates, is responsible for the synchronization of seasonal biological functions, such as reproduction in vertebrates [21]. Until recently, it was believed that there were four genes encoding melatonin receptors: *MTNR1A* and *MTNR1B*, shared by vertebrates, *mtnr1c* (*mellc*), shared by fishes, frogs, and chicken but not mammals, and *GPR50*, which has lost the ability to bind melatonin, shared by only mammals [22–24]. We note that the phylogenetic tree of these genes exhibited low bootstrap values at the branch section of *GPR50*. By comparing the genomic location of these genes among different genomes, we have recently shown that mammalian *GPR50* does correspond to nonmammalian *mtnr1c*, but that these two genes have evolved so quickly that they were not considered orthologs [25]. Subsequent studies aimed to understand how *mtnr1c* evolved to give rise to the *GPR50* gene after divergence of mammals, and to identify amino acid sites that were subjected to positive selection. It was demonstrated that substitutions of amino acids responsible for the loss of melatonin-binding ability were not made by positive selection, but rather under neutrality [25]. This suggests that the loss of melatonin-binding ability by *GPR50* in mammals was made by relaxation, because this binding was no longer required in these species. We presume that the divergence of nonmammalian *mtnr1c* and mammalian *GPR50* might be associated with the difference in seasonal sensitivity between mammals and other vertebrates.

Another example concerns genes expressed in the seminal vesicles of the mammalian male genital tract. It is well known that the size of seminal vesicles differs greatly in mammals, being very large in the mouse and absent in the dog. Moreover, the composition of seminal fluid is characteristic of each mammalian species. Thus, the evolution of genes specifically expressed in the seminal vesicle is particularly interesting. The *BSP* (Binder of Sperm Protein, also known as Bovine Seminal Plasma) gene family is believed to be involved in sperm capacitation, that is, final sperm maturation before fertilization in the female genital tract [26, 27]. A published phylogenetic study [28] separated *BSP* into three subfamilies: *BSP*, *BSPH1*, and *BSPH2* (Fig. 1a). (The nomenclature for *BSP* genes and subfamilies follows the new published nomenclature for mammalian *BSP* genes [27].) Surprisingly, the *BSP* subfamily seemed to be common to all mammals except murines and primates. When we localized the members of this gene family in genomes, two conserved regions were identified. One concerns the *BSP* subfamily, expressed in the seminal vesicle, and the other encompasses both *BSPH1* and *BSPH2* subfamilies, specifically expressed in the epididymis (Fig. 1b). We note that in the phylogenetic tree (Fig. 1a), dog *CDK105* is orthologous to the other members of *BSP* subfamily but with a low bootstrap value and a long branch. However, in the dog genome, the sequence of *CDK105* is not found in the corresponding region of the *BSP* subfamily, but is found in the region encompassing the *BSPH1/BSPH2* subfamilies. Thus, *CDK105* is more likely to be a member of the *BSPH1/BSPH2* subfamilies than of the *BSP* subfamily. Note that the dog does not have seminal vesicles. Similarly, the rabbit *BSP* homolog (*EP52C1*) is located in the phylogenetic tree as ortholog to the other *BSP* members, but also with a low bootstrap value (Fig. 1a). The *EP52C1* gene is expressed in the epididymis, like other members of *BSPH1/BSPH2* subfamilies. Unfortunately, it is impossible to localize this gene because the rabbit genome is not fully sequenced, but from previous observations we can hypothesize that the rabbit *BSP* homolog, *EP52C1*, might

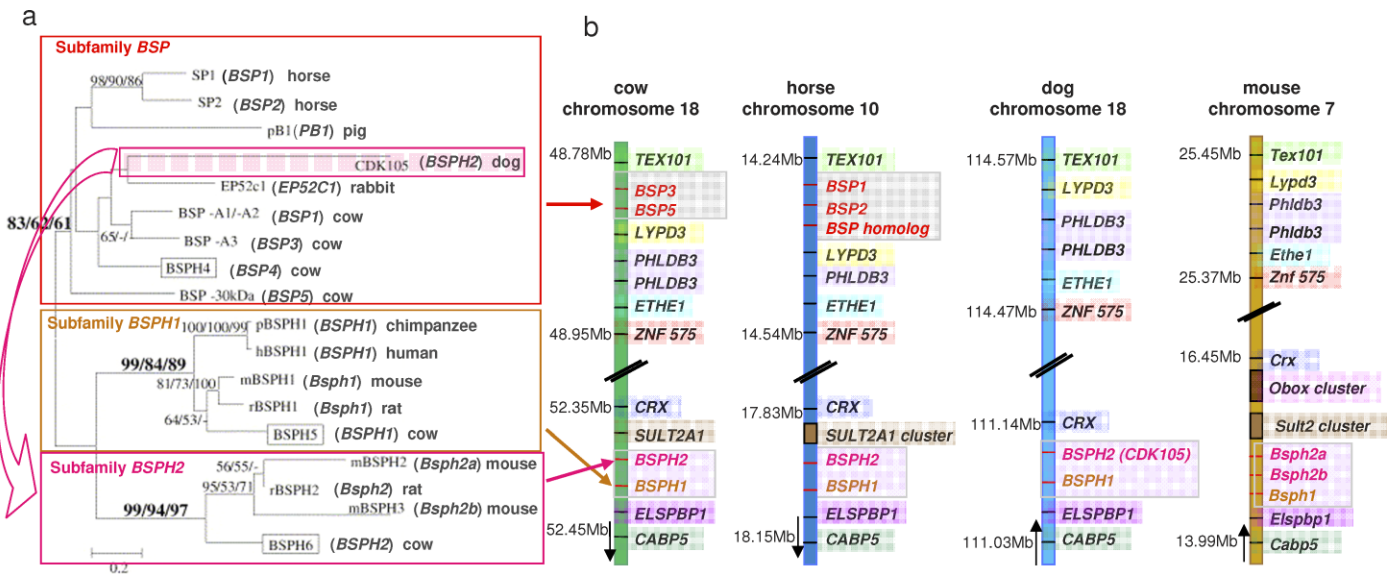
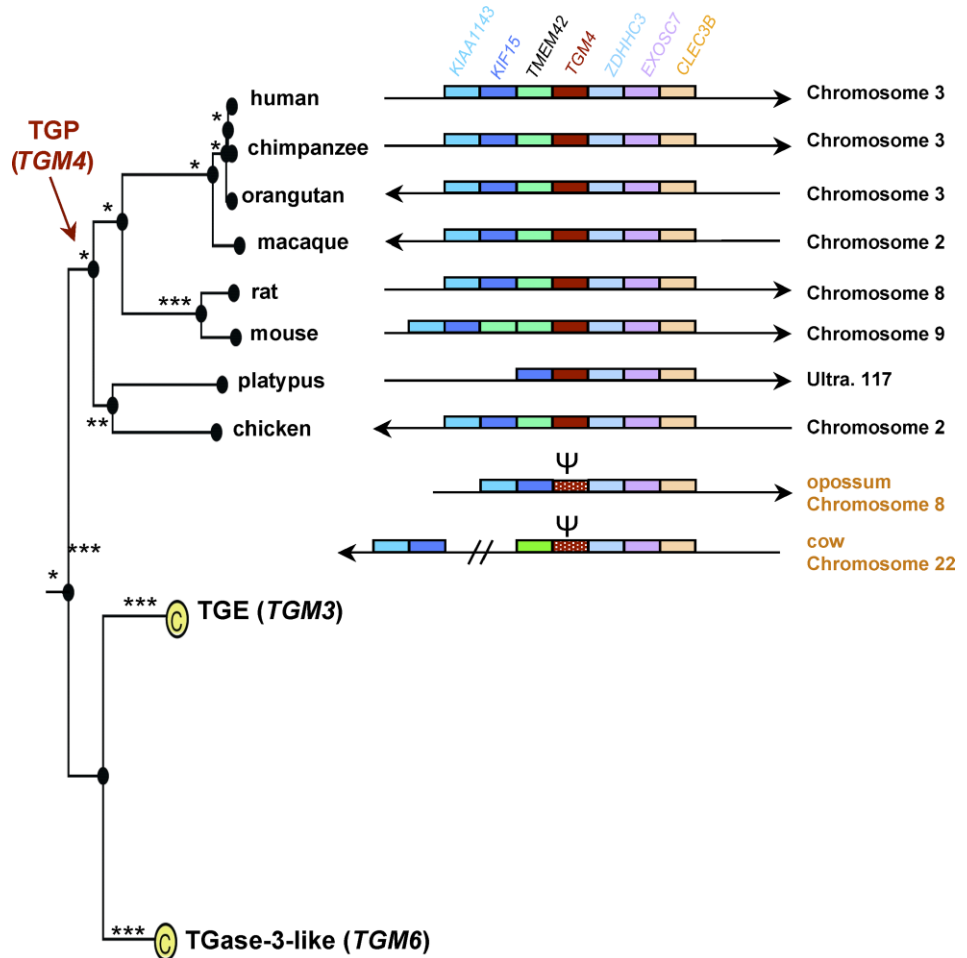


FIG. 1. Gene divergence of mammalian *BSP* gene family. **a**) Phylogenetic tree from Fan et al. [28] (reprinted with permission from Elsevier). *BSP* gene family is composed of three subfamilies: *BSP*, *BSPH1*, and *BSPH2*. (The nomenclature for these genes has been corrected in Manjunath et al. [27].) Based on syntenic comparison, the dog *BSPH2* (*CDK105*) should be a member of *BSPH1/2* subfamilies. **b**) Genomic localization of *BSP* genes in cow, horse, dog, and mouse. Dog *BSP-homolog* (*CDK105*) and its newly identified paralog (*BSPH1*) are not located in the *BSP*-related conserved region, but in the *BSPH1/BSPH2*-related one. No *BSP* ortholog is present in dog and mouse genomes. Bovine *BSP1* and *BSP4* have not been localized because of the incomplete genomic sequencing, but no counterpart of these two genes was identified in other species. The orthologous genes are shaded in the same color.

FIG. 2. Gene death of *TGM4* in opossum and cow. Phylogenetic tree of TGP protein (encoded by *TGM4*) and its relative based on FIGENIX analysis [37]. Bootstrap values were calculated by neighbor-joining, maximum parsimony, and maximum likelihood methods for each clade (*, **, and *** represent the values higher than 80% supported by one, two, and all three methods, respectively). Genomic region encompassing *TGM4* gene is shown for each species. In opossum and cow genomes, the pseudogenized *TGM4* (shaded and indicated by Ψ) is localized in the same conserved region that is predicted by the tBlastn search. The accession numbers of the sequences used for phylogeny are shown in Supplemental Table S1. The detailed information on the identified pseudogene is listed in Supplemental Figures S3a and S3b. The orthologous genes are depicted in the same color. Ultra., UltraContig.



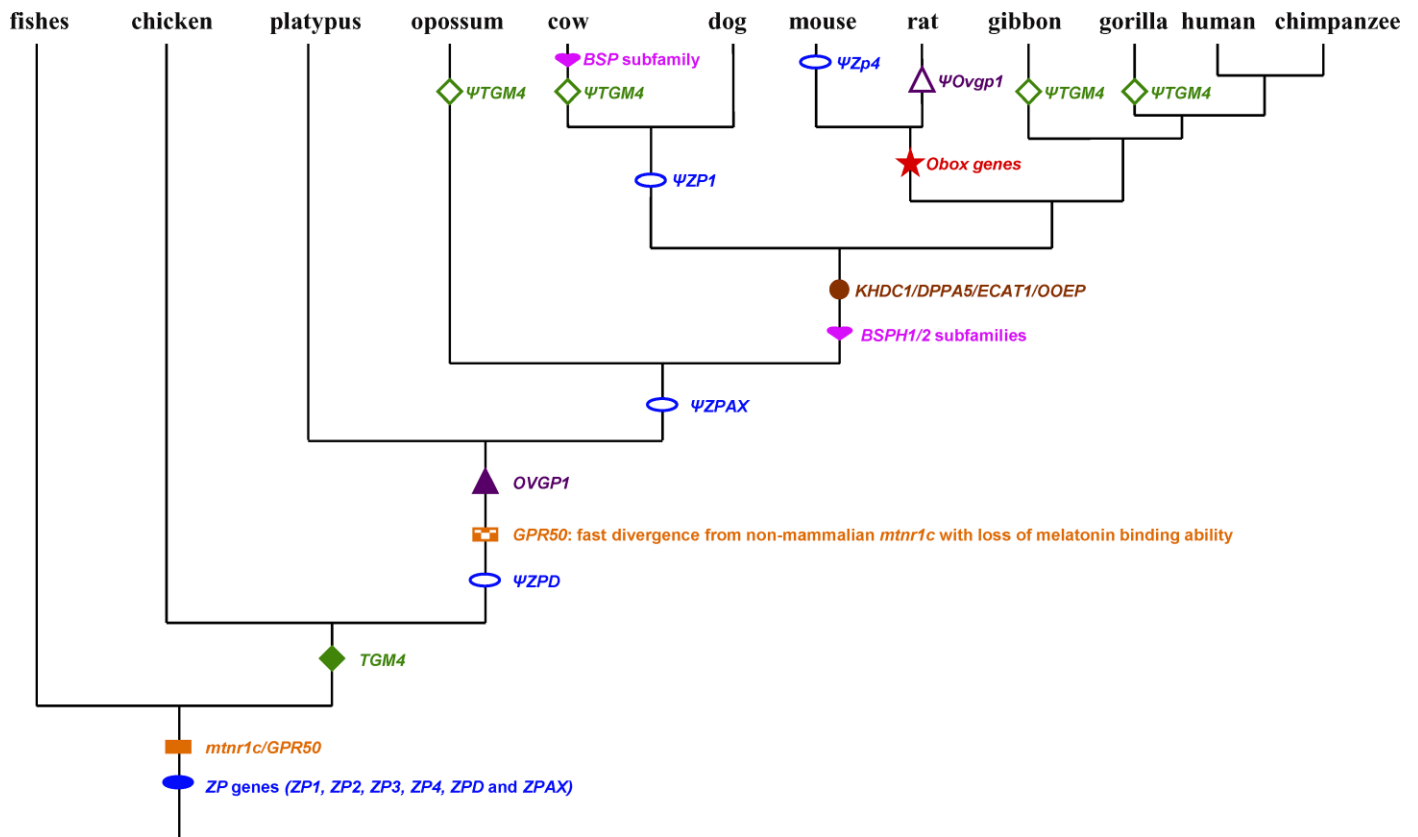


FIG. 3. Evolutionary scenarios of some reproduction-related genes. The birth, death, and divergence of seven genes or gene families mentioned in this paper are presented in the branches of a life tree. In detail, 1) *Obox* gene family was specifically born in rodents after their divergence from primates; 2) *KHDC1/DPPA5/ECAT1/OOEP* gene family was born before the divergence of eutherian mammals; 3) mammalian *GPR50* is orthologous to nonmammalian *mtnr1c* gene, which was born before the divergence of vertebrates, and in mammals *GPR50* rapidly diverged with the loss of melatonin-binding ability; 4) *BSPH1/2* subfamilies were born, at the latest, before the divergence of eutherian mammals, and *BSP* subfamily was born specifically in cow, maybe ungulate clade; 5) *ZP* gene family was born before the divergence of marsupials (opossum) and eutherian mammals, loss of *ZPD* gene before the divergence of cow and dog, and loss of *Zp4* gene specifically in mouse; 6) *TGM4* gene was born, at the latest, in land vertebrates, the pseudogene being found in opossum, cow, gibbon, and gorilla genomes; and 7) *OVGP1* gene was born before the divergence of mammals and has been lost specifically in rat. Different genes are represented by different symbols in different colors. The birth of a gene is depicted by a solid symbol, and the death of a gene is depicted by a hollow symbol and Ψ .

belong to the *BSPH1/BSPH2* subfamilies instead of the *BSP* subfamily. Consequently, the *BSP* subfamily seems to be shared only by ungulates, like cow, horse, and perhaps pig (which is not fully demonstrated because the genomic sequence is unavailable). Moreover, our tBlastn search found neither orthologs nor pseudogenes of the *BSP* gene subfamily in nonungulate genomes, such as dog, human, and mouse, whereas the orthologs of bovine *BSPH1/BSPH2* were identified in the corresponding genomic regions of horse, dog, and mouse genomes (Fig. 1b). Thus, the most plausible scenario would be an early duplication event of a *BSP* ancestor giving rise to a mammalian epididymis-expressed *BSPH1/BSPH2* ancestor before the divergence of mammals, and further independent duplications in ungulates that resulted in the seminal vesicle-expressed *BSP* gene subfamily. The role of these genes in the biological and biochemical properties of ungulate seminal fluid remains to be investigated.

GENE DEATH

Gene death is usually caused by introducing any changes that lead to a premature stop codon. Now, with the availability of complete animal genome sequences, it is possible to test hypotheses concerning gene death by searching for the trace of a pseudogene in certain species. To demonstrate pseudogene-

zation suspected by phylogeny, one can implement a tBlastn search to identify the trace of a pseudogene. If the pseudogene can be localized in the corresponding position of a conserved genomic region, we can conclude that gene pseudogenization has occurred.

Using this approach, we have recently classified genes encoding the proteins that constitute the zone pellucida (zp) surrounding oocytes, and we have shown that there are at least eight *ZP* genes in several fish species, six in the chicken (*ZPAX*, *ZPD*, *ZP1*, *ZP2*, *ZP3*, and *ZP4*), four in primates and the rat (*ZP1*, *ZP2*, *ZP3*, and *ZP4*), three in the cow and dog (*ZP2*, *ZP3*, and *ZP4*), and three in the mouse (*Zp1*, *Zp2*, and *Zp3*) [29]. A systematic tBlastn analysis against different genomes revealed the presence of the *ZPAX* pseudogene in cow, opossum (data not shown), and certain primates, the *ZP1* pseudogene in dog and cow genomes, as well as the *Zp4* pseudogene in the mouse genome. Overall, this study revealed that the evolution of vertebrates was accompanied by the progressive loss of *ZP* genes. The significance of such loss among vertebrates remains to be further investigated, but it may be associated with the fact that in the latter case, unlike with fishes, the fertilization of oocytes occurs within the female genital tract, where the specificity of gamete interactions is better facilitated than in seawater.

The second example concerns prostate-specific transglutaminase (TGP), which is involved in the formation of semen coagulum and copulatory plugs in primates [30]. Phylogeny shows that TGP-encoding gene *TGM4* is present in chicken, platypus, some primate, and rodent genomes, but no homolog seems to exist in other mammalian lineages, such as dog, cow, and opossum (Fig. 2). Genomic comparison shows that *TGM4* is located in a conserved genomic region in vertebrates (Fig. 2). Furthermore, by tBlastn analyses we identified traces of *TGM4* pseudogenization in the corresponding positions in opossum and cow genomes (Fig. 2 and Supplemental Fig. S3), which indicates that *TGM4* has been lost in these species. The expression of *TGM4* in rodents and certain primates is consistent with the formation of copulatory plugs in these species (not observed in dog or ungulates). Interestingly, it was suggested that the formation of this copulatory plug would have been selected for during the evolution of species exhibiting male competition for female mating [30]. Moreover, *TGM4* pseudogenes have been identified in gorilla and the lar gibbon genomes [30]. Both species are considered monoandrous; that is, males do not compete to mate with females. The functional loss of gorilla *TGM4* is consistent with the lack of semen coagulation in this species [31] and with the mutations of other genes responsible for copulatory plugs in the gorilla, such as semenogelins I and II (*SEMG1* and *SEMG2*) [32] and kallikrein 2 [30]. Because a *TGM4* homolog is also present in the chicken genome, it is reasonable to conclude that the ancestral *TGM4* gene appeared at least in land vertebrates, likely acquired a plug formation function in mammals, and was lost in species for which a copulatory plug was no longer necessary, such as for marsupials, cow, and some primates. Until now, the existence of copulatory plugs has not been reported in any other species besides primates and rodents, so the function of *TGM4* in the platypus and chicken remains to be investigated.

Another example concerns oviduct-specific glycoprotein 1 (OVGP1), also known as oviductin, which has been shown to play a biological role in the fertilization process and/or early embryonic development [33]. A genomic search has enabled us to find the trace of a rat *Ovgpl* pseudogene that explains the absence of rat *Ovgpl* in the phylogenetic tree (see Supplemental Figs. S2 and S3c). As with many other reproduction-related genes, *OVGP1* has recently been considered to have evolved under positive selection in primates [34]. Interestingly, targeted inactivation of *Ovgpl* in the mouse is without any consequence in female reproduction [35], implying that at least in both murine species, *Ovgpl* is not essential for fertility. Overall, these data suggest that the *OVGP1* gene under positive selection in primates seems to be relaxed in rodents.

Additionally, a recent phylogenomic study [36] has shown that the emergence of lactation and placenta genes in the common mammalian ancestor and the development of placentation in marsupial and eutherian mammals allowed for the gradual loss of yolk-dependent nourishment genes, such as vitellogenin genes, during mammalian evolution.

Overall, the systematic study of gene birth and gene death enables us to construct original scenarios of gene evolution in vertebrates. A schematic phylogenetic tree of life illustrates the evolutionary steps of reproduction-related genes noted in this paper (Fig. 3).

CONCLUSION

Among genes involved in reproductive functions, some elements are well conserved among vertebrates and are subjected to purifying selection (such as *FSH* or activin genes). These genes are likely to be involved in well-conserved

biological processes. However, many genes involved in reproduction are highly divergent and/or have evolved rapidly in vertebrates. They are thought to contribute to species-specific reproductive processes and particularly to speciation. In this paper we have summarized the evolutionary scenarios of some reproduction-related genes in three main processes, including gene birth, gene death, and gene divergence, by combining phylogeny with gene genomic localization. Because of continually increasing, comprehensive databases of sequenced genomes, this phylogenomic method is expected to be a source of new information on both the evolution and predicated functions of reproductive genes.

ACKNOWLEDGMENTS

We are grateful to John Williams, Kelly Everhart, and Eben Clattenburg for their critical reading of this manuscript.

REFERENCES

1. Singh RS. Patterns of species divergence and genetic theories of speciation. In: Wohrmann K, Jain SK (eds.), *Topics in Population Biology and Evolution*. New York: Springer-Verlag; 1990:231–265.
2. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. *Nat Rev Genet* 2002; 3:137–144.
3. Torgerson DG, Kulathinal RJ, Singh RS. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol Biol Evol* 2002; 19:1973–1980.
4. Torgerson DG, Singh RS. Rapid evolution through gene duplication and subfunctionalization of the testes-specific alpha4 proteasome subunits in *Drosophila*. *Genetics* 2004; 168:1421–1432.
5. Wyckoff GJ, Wang W, Wu CI. Rapid evolution of male reproductive genes in the descent of man. *Nature* 2000; 403:304–309.
6. Anderson FE, Swofford DL. Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA. *Mol Phylogenet Evol* 2004; 33:440–451.
7. Fitch WM. Distinguishing homologous from analogous proteins. *Syst Zool* 1970; 19:99–113.
8. Zhang P, Gu Z, Li WH. Different evolutionary patterns between young duplicate genes in the human genome. *Genome Biol* 2003; 4:R56.
9. Ohno S. *Evolution by Gene Duplication*. Berlin: Springer Verlag; 1970.
10. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999; 151:1531–1545.
11. Hurles M. Gene duplication: the genomic trade in spare parts. *PLoS Biol* 2004; 2:E206.
12. Cheng WC, Hsieh-Li HM, Yeh YJ, Li H. Mice lacking the Obox6 homeobox gene undergo normal early embryonic development and are fertile. *Dev Dyn* 2007; 236:2636–2642.
13. Rajkovic A, Yan C, Yan W, Klysiak M, Matzuk MM. Obox, a family of homeobox genes preferentially expressed in germ cells. *Genomics* 2002; 79:711–717.
14. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25:3389–3402.
15. Jackson M, Watt AJ, Gautier P, Gilchrist D, Driehaus J, Graham GJ, Keebler J, Prugnolle F, Awadalla P, Forrester LM. A murine specific expansion of the RhoX cluster involved in embryonic stem cell biology is under natural selection. *BMC Genomics* 2006; 7:212.
16. Dade S, Callebaut I, Mermillod P, Monget P. Identification of a new expanding family of genes characterized by atypical LRR domains. Localization of a cluster preferentially expressed in oocyte. *FEBS Lett* 2003; 555:533–538.
17. Dade S, Callebaut I, Paillisson A, Bontoux M, Dalbies-Tran R, Monget P. In silico identification and structural features of six new genes similar to MATER specifically expressed in the oocyte. *Biochem Biophys Res Commun* 2004; 324:547–553.
18. Paillisson A, Dade S, Callebaut I, Bontoux M, Dalbies-Tran R, Vaiman D, Monget P. Identification, characterization and metagenome analysis of oocyte-specific genes organized in clusters in the mouse genome. *BMC Genomics* 2005; 6:76.
19. Pierre A, Gautier M, Callebaut I, Bontoux M, Jeanpierre E, Pontarotti P, Monget P. Atypical structure and phylogenomic evolution of the new eutherian oocyte- and embryo-expressed KHDC1/DPPA5/ECAT1/OOEP gene family. *Genomics* 2007; 90:583–594.

20. Hendy MD, Penny D. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 1989; 38:297–309.
21. Malpoux B. Seasonal regulation of reproduction in mammals. In: Neill JD ed, Knobil and Neill's *Physiology of Reproduction*, vol. 3, 3rd ed. New York: Elsevier; 2006:2231–2281.
22. Ebisawa T, Kame S, Lerner MR, Reppert SM. Expression cloning of a high-affinity melatonin receptor from *Xenopus* dermal melanophores. *Proc Natl Acad Sci U S A* 1994; 91:6133–6137.
23. Reppert SM, Weaver DR, Cassone VM, Godson C, Kolakowski LF Jr. Melatonin receptors are for the birds: molecular analysis of two receptor subtypes differentially expressed in chick brain. *Neuron* 1995; 15:1003–1015.
24. Roca AL, Godson C, Weaver DR, Reppert SM. Structure, characterization, and expression of the gene encoding the mouse Mel1a melatonin receptor. *Endocrinology* 1996; 137:3469–3477.
25. Dufourny L, Levasseur A, Migaud M, Callebaut I, Pontarotti P, Malpoux B, Monget P. GPR50 is the mammalian ortholog of Mel1c: evidence of rapid evolution in mammals. *BMC Evol Biol* 2008; 8:105.
26. Manjunath P, Therien I. Role of seminal plasma phospholipid-binding proteins in sperm membrane lipid modification that occurs during capacitation. *J Reprod Immunol* 2002; 53:109–119.
27. Manjunath P, Lefebvre J, Jois PS, Fan J, Wright MW. New Nomenclature for Mammalian BSP Genes. *Biol Reprod* 2009; 80:394–397.
28. Fan J, Lefebvre J, Manjunath P. Bovine seminal plasma proteins and their relatives: a new expanding superfamily in mammals. *Gene* 2006; 375:63–74.
29. Goudet G, Mugnier S, Callebaut I, Monget P. Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. *Biol Reprod* 2008; 78:796–806.
30. Clark NL, Swanson WJ. Pervasive adaptive evolution in primate seminal proteins. *PLoS Genet* 2005; 1:e35.
31. Dixon AL, Anderson MJ. Sexual selection, seminal coagulation and copulatory plug formation in primates. *Folia Primatol (Basel)* 2002; 73:63–69.
32. Jensen-Seaman MI, Li WH. Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. *J Mol Evol* 2003; 57:261–270.
33. Killian GJ. Evidence for the role of oviduct secretions in sperm function, fertilization and embryo development. *Anim Reprod Sci* 2004; 82–83:141–153.
34. Swanson WJ, Nielsen R, Yang Q. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 2003; 20:18–20.
35. Araki Y, Nohara M, Yoshida-Komiya H, Kuramochi T, Ito M, Hoshi H, Shinkai Y, Sendai Y. Effect of a null mutation of the oviduct-specific glycoprotein gene on mouse fertilization. *Biochem J* 2003; 374:551–557.
36. Brawand D, Wahli W, Kaessmann H. Loss of egg yolk genes in mammals and the origin of lactation and placentation. *PLoS Biol* 2008; 6:e63.
37. Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EG. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. *BMC Bioinformatics* 2005; 6:198.