



**HAL**  
open science

## Evolution of genes involved in gamete interaction: Evidence for positive selection, duplications and losses in vertebrates

Camille Meslin, Sylvie Mugnier, Isabelle Callebaut, Michel Laurin, Géraldine Pascal, Anne Poupon, Ghylène Goudet-Guitton, Philippe Monget

### ► To cite this version:

Camille Meslin, Sylvie Mugnier, Isabelle Callebaut, Michel Laurin, Géraldine Pascal, et al.. Evolution of genes involved in gamete interaction: Evidence for positive selection, duplications and losses in vertebrates. PLoS ONE, 2012, 7 (9), pp.1-8. 10.1371/journal.pone.0044548 . hal-01136282

**HAL Id: hal-01136282**

**<https://hal.science/hal-01136282v1>**

Submitted on 28 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.

# Evolution of Genes Involved in Gamete Interaction: Evidence for Positive Selection, Duplications and Losses in Vertebrates

Camille Meslin<sup>1,2,3,4,9</sup>, Sylvie Mugnier<sup>5,9</sup>, Isabelle Callebaut<sup>6</sup>, Michel Laurin<sup>7</sup>, Géraldine Pascal<sup>1,2,3,4</sup>, Anne Poupon<sup>1,2,3,4</sup>, Ghylène Goudet<sup>1,2,3,4</sup>, Philippe Monget<sup>1,2,3,4\*</sup>

**1** UMR85 Physiologie de la Reproduction et des Comportements, INRA, Nouzilly, France, **2** UMR6175, CNRS, Nouzilly, France, **3** Université François Rabelais de Tours, Tours, France, **4** IFCE, Nouzilly, France, **5** Département Agronomie Agro-équipement Élevage Environnement, AgroSup Dijon, Dijon, France, **6** IMPMC UMR 7590, Université Pierre et Marie Curie-Paris 6, Paris, France, **7** UMR 7207, CNRS/MNHN/UPMC, Muséum National d'Histoire Naturelle, Paris, France

## Abstract

Genes encoding proteins involved in sperm-egg interaction and fertilization exhibit a particularly fast evolution and may participate in prezygotic species isolation [1,2]. Some of them (ZP3, ADAM1, ADAM2, ACR and CD9) have individually been shown to evolve under positive selection [3,4], suggesting a role of positive Darwinian selection on sperm-egg interaction. However, the genes involved in this biological function have not been systematically and exhaustively studied with an evolutionary perspective, in particular across vertebrates with internal and external fertilization. Here we show that 33 genes among the 69 that have been experimentally shown to be involved in fertilization in at least one taxon in vertebrates are under positive selection. Moreover, we identified 17 pseudogenes and 39 genes that have at least one duplicate in one species. For 15 genes, we found neither positive selection, nor gene copies or pseudogenes. Genes of teleosts, especially genes involved in sperm-oolemma fusion, appear to be more frequently under positive selection than genes of birds and eutherians. In contrast, pseudogenization, gene loss and gene gain are more frequent in eutherians. Thus, each of the 19 studied vertebrate species exhibits a unique signature characterized by gene gain and loss, as well as position of amino acids under positive selection. Reflecting these clade-specific signatures, teleosts and eutherian mammals are recovered as clades in a parsimony analysis. Interestingly the same analysis places *Xenopus* apart from teleosts, with which it shares the primitive external fertilization, and locates it along with amniotes (which share internal fertilization), suggesting that external or internal environmental conditions of germ cell interaction may not be the unique factors that drive the evolution of fertilization genes. Our work should improve our understanding of the fertilization process and on the establishment of reproductive barriers, for example by offering new leads for experiments on genes identified as positively selected.

**Citation:** Meslin C, Mugnier S, Callebaut I, Laurin M, Pascal G, et al. (2012) Evolution of Genes Involved in Gamete Interaction: Evidence for Positive Selection, Duplications and Losses in Vertebrates. PLoS ONE 7(9): e44548. doi:10.1371/journal.pone.0044548

**Editor:** Olivier Lespinet, Université Paris-Sud, France

**Received:** February 27, 2012; **Accepted:** August 7, 2012; **Published:** September 5, 2012

**Copyright:** © 2012 Meslin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** CM is funded by a MENRT PhD fellowship. This work was supported by INRA. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: philippe.monget@tours.inra.fr

<sup>9</sup> These authors contributed equally to this work.

## Introduction

A lot of barriers are able to prevent a successful reproduction. Dobzhansky proposed a classification of species isolation from prezygotic to postzygotic isolation [5]. In postzygotic isolation, a zygote is formed but the hybrid offspring is not viable or sterile. In prezygotic isolation, species could not mate because of a difference in sexual behavior or for mechanical issues. Another step of prezygotic isolation is the lack of interaction between male and female gametes, the sperm-egg interaction being a several steps process that leads to a successful fertilization. Although details vary between species, a successful fertilization requires three main steps in sperm-egg interaction [6]. The first step is the recognition and the binding between sperm and the extracellular matrix surrounding eggs (ECM), called the chorion in teleosts, the vitelline envelope in birds, and the zona pellucida (ZP) in mammals. The second step, required for the sperm to penetrate

through the egg coat, is the acrosome reaction. The third step is the binding and the subsequent fusion between sperm and the oolemma, the plasma membrane of the oocyte. Many experiments have been performed in order to determine the importance of candidate proteins during fertilization. Surprisingly, knock-out experiment of B4GT1 or targeted mutation that abolished the protease activity of ACR in mouse show that these proteins are important but not essential [7,8]. Several studies showed that genes involved in sperm-egg interaction exhibit a particular evolution. For example, a recent study shows that several ZP genes have been lost during evolution of vertebrates [9]. ZP1 gene, one of the glycoproteins that compose the ZP, has been lost during evolution in cattle, pig, cat and dog, whereas ZP4 gene has been lost in mouse. Moreover, ZPAX and ZPD genes, which are present in chicken, have been lost in mammals. In addition, several genes involved in fertilization such as ZP3, ADAM1, ADAM2, ACR and CD9 have been shown to evolve under

positive selection [3,4]. The authors hypothesized that these positively selected amino acids might play a role in the specificity of sperm-egg interaction between species because they are located in the putative sperm-egg binding domain. Despite the fact that several studies suggested the involvement of positive Darwinian selection on sperm-egg interaction genes, none has been done exhaustively. In our study, we have listed all genes which have been experimentally shown to be involved in one vertebrate and in at least one step of the sperm-egg interaction, on both sperm and/or oocyte, and on 19 species with fully-sequenced genomes. We have studied all these genes with an evolutionary perspective by searching for gene gains and losses, as well as positive selection, a divergent evolution of these genes among vertebrate species potentially leading to reproductive isolation.

## Results and Discussion

### Rates of Gene Gain, Loss and Pseudogenization are Clade-dependent Variables

We identified 69 genes from the literature (Figure 1 and Table S1) that encode proteins experimentally known to be involved in at least one of the three steps described above. These genes are well studied in mouse and cattle that are common models for studies on physiology and reproduction. We classified all genes into three main groups according to which step they are involved in, excepted CRISP1 that is involved in all three steps of fertilization [10,11,12]. Thanks to phylogenetic analyses, we found several gene duplications, as well as pseudogenizations, among genes involved in the three main steps of fertilization. In particular, we found evidence for (i) gain of FUT5 and Zp3r genes in primates and rodents, respectively, (ii) pseudogenization of SPA17 and ADAM15 genes in chicken, ZAN, ADAM1a and ADAM3 in human, ADAM3 in chimpanzee, spermadhesins in dog, IGSF8 and SPESP1 in horse, as well as ZPAX in dog, and the known ZP pseudogenes published elsewhere [9] (ZPB/ZP4 in mouse, ZP1 in cow and dog, ZPAX in primates and cow); and (iii) specific duplication events of ADAM genes in therian mammals, spermadhesins in ungulates, as well as the ACR gene in *Xenopus* (14 copies) and HSPE1 in macaque (11 copies). For each gene, we have systematically assessed, by optimizing gene presence, number, and type (functional or pseudogene) on a time-calibrated reference phylogeny, the absolute rate of the following four types of events: gene gain, duplication, inactivation (resulting in a pseudogene) and complete loss (presumably through deletion, although incomplete genome annotation may have led to overestimation of loss rate). The overall rate of gene appearance (genes that appear only in some monophyletic groups), pseudogenization and deletion events differs strongly between clades and type of event (Figure 2a). By taking into consideration the number and the geological age of sampled lineages, we found that these three types of events are far more frequent in eutherians than in birds and teleosts. A phylogeny-informed binomial test shows that these differences between eutherians and teleosts are statistically significant, differences between eutherians and birds are also statistically significant for deletions and gene appearance. However, a possible bias stems from the fact that our list of gene is composed almost exclusively of genes identified in eutherians and thus, the rates of the different events may be overestimated in this clade. No pseudogene has been identified in teleosts. The duplications occur at more comparable rates across these three taxa but tend to occur slightly more frequently in eutherians than in teleosts and aves. The consensual conclusion is that duplication rates for most of the genes are higher in teleosts than in other taxa due to the third whole genome duplication, specific to teleosts

[13,14]. Consequently, our results are surprising. They reflect the greater geological age of lineages of teleosts, resulting in a phylogenetic diversity twice as great as eutherians, despite a lower number of included species (Figure 2a). Thus, a greater number of events per lineage is required in teleosts to yield comparable rates.

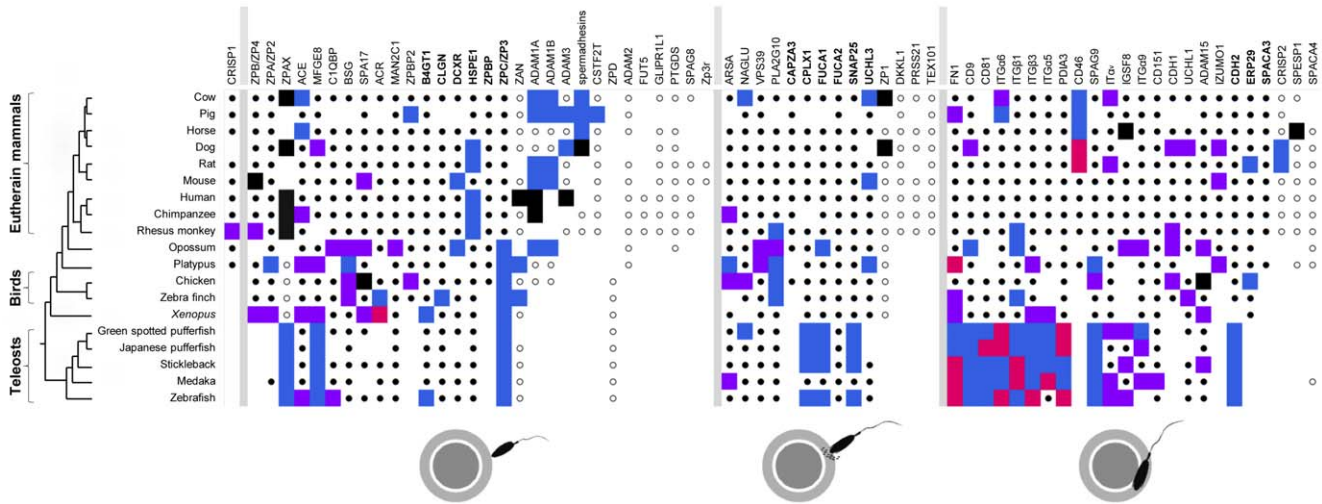
### Genes of Teleosts are More Subjected to Positive Selection

We have also submitted the protein sequence to the PhyleasProg web server [15] (<http://phyleasprog.inra.fr/>) to determine selective pressures that shape the evolution of these genes and used a new phylogenetically-informed test to compare the amount of positive selection in various clades. The calculation of positive selection is globally based on rate of non-synonymous ( $dN$ ) and synonymous ( $dS$ ) substitutions [16]. The accumulation of synonymous substitutions (saturation of  $dS$ ) may create a problem to detect positive selection with sequences from highly divergent species such as human and zebrafish with some methods. However, a previous study showed that the method we use to detect positive selection, i.e. maximum likelihood estimate, is robust to  $dS$  saturation [17]. The calculation was possible for 48 genes and we found amino acids under positive selection for 33 of them (Figure 1). Among teleosts, birds and eutherians taxa, we found various patterns of positive selection. The FN1 gene has sites under positive selection in 7 species for which data are available. The number of genes under positive selection in each species ranges from 0 (horse and human) to 10 (opossum). Positive selection occurs in the three groups of genes, suggesting the presence of an evolutionary pressure in the three steps of fertilization.

Our comparison of the intensity of positive selection shows that genes involved in sperm-oolemma fusion, including genes of the integrin family (IGSF8, ITG $\alpha$ 9, ITG $\alpha$ , ITG $\alpha$ 6), are mostly subjected to positive selection in teleosts, but much less so in mammals (Figure 2b). The difference between teleosts and birds ( $p < 0.05$ ) and between teleosts and eutherian mammals ( $p < 0.001$ ) are statistically significant whereas the difference between birds and eutherian mammals is not. This result is surprising for teleosts because we expected to find more amino acids under positive selection for genes involved in the sperm-ECM binding, which would contribute to the first barrier in free-spawning conditions. In particular, positive Darwinian selection has been shown to play a role in the rapid evolution of genes involved in sperm-egg binding in three other free-spawning taxa, sea urchin, mussel and abalone [18,19,20,21]. The comparisons of the intensity of positive selection between taxa for sperm-ECM binding and between taxa for acrosome reaction are not significant. When analyzed globally, genes involved in the three steps of gamete interaction show a similar, but less marked, pattern in which the only statistically significant difference is between teleosts and eutherian mammals ( $p < 0.05$ ) (Figure 2c). Besides, in contrast to previous studies [4], we did not find positive selection for ZPC/ZP3, in agreement with more recent data obtained on a larger number of species [22]. We also show that several genes are positively selected in more than one species, such as ACE that exhibits positive selection in chimpanzee, opossum, *Xenopus*, and zebrafish.

### Amino Acids under Positive Selection are not Located in known Functional Domains of Proteins

In order to project our previous results of positive selection on 3D structures, we have built 23 tridimensional models of 6 proteins (ACE, ARSA, C1QBP, ITG $\alpha$ , PDIA3 and UCHL3) from species



**Figure 1. Signature of evolution of genes involved in sperm-oocyte interaction in Vertebrates.** Genes (names are indicated on the top) are classified into three main functional groups (from left to right): sperm-zona pellucida binding, acrosome reaction, and sperm-oolemma fusion. Species are indicated in the tree (left). Blue square: duplication; purple square: positive selection; pink square: duplication + positive selection. Black square: pseudogene; Black dot: no event. Black circle: no calculation of positive selection. White square: no gene found. doi:10.1371/journal.pone.0044548.g001

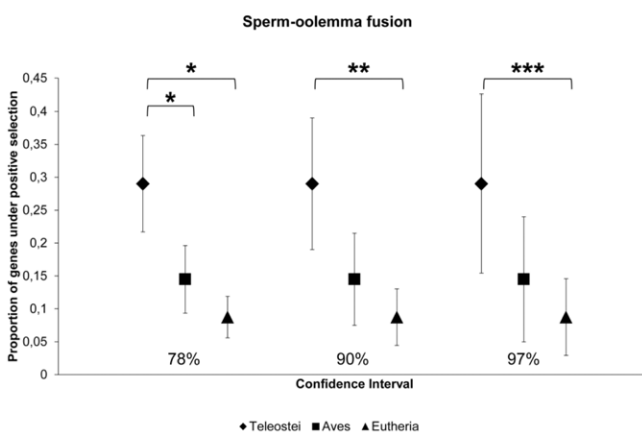
where positive selection has been detected (Figure 3 and Figure S1 to S5). These 6 proteins were chosen because PDB template structures were available [23]. Firstly, we show that amino acids under positive selection are never located in the domain involved

in known biological role (for ACE, ARSA, PDIA3 and UCHL3). This suggests that these residues are involved in other secondary functions such as interaction with partners or in the folding of the protein. Secondly, we show that amino acids under positive

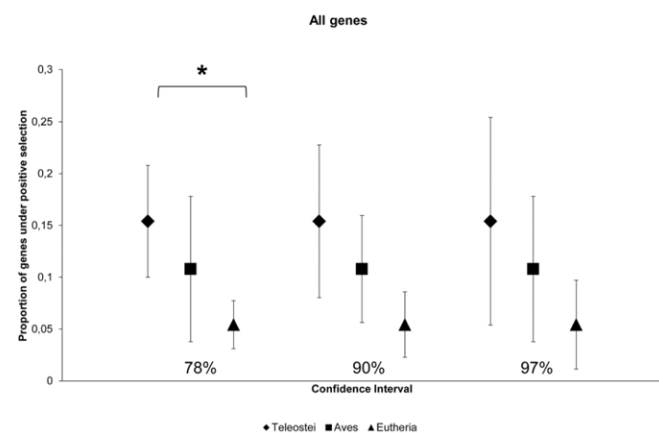
**a**

Taxon	Phylogenetic diversity (Myr)	Rate (event per Myr and lineage)			
		Gene appearance	Gene duplication	Pseudogenization	Deletion
Teleostei	1228	0,00321	0,04147	0	0,01739
Aves	415	0,00723	0,03325	0,00665	0,01995
Eutheria	645	0,07551	0,05034	0,01573	0,05663

**b**



**c**



**Figure 2. Rate of gene appearance, gene duplication, pseudogenization, loss and intensity of positive selection in Teleostei, Aves and Eutheria.** **a.** Rate of events per million years (Myr) per lineage for the set of genes studied. This includes up to 69 genes, but in most cases, the number of relevant genes is lower because some types of events cannot happen in some taxa (losses or duplications cannot happen if a gene is primitively absent, and gains cannot occur in a clade if the gene appeared before the base of a given clade), and we have incorporated that factor into our calculations to obtain comparable rates across taxa. The geological age of lineages is used in the rate calculation through the phylogenetic diversity of the clades. **b** and **c.** Proportion of genes under positive selection in Teleostei, Aves and Eutheria for genes belonging to the third step, the sperm-oolemma fusion (B), and for all genes (C). Three types of confidence intervals are shown: 78%, 90% and 97% representing testing at 0.05, 0.01 and 0.001 probability thresholds, respectively. Statistically significant comparisons, for which confidence intervals do not overlap, are indicated. doi:10.1371/journal.pone.0044548.g002

selection, for the same protein, have different location between species. Thus, we suppose that the function of these proteins in the process of fertilization may differ between species. This hypothesis is supported by a recent study that has emphasized the importance of structural architecture of the egg coat and its importance in fertilization [24]. Thus, we postulate that positive Darwinian selection could contribute to slight modifications of the architecture of egg or sperm coat by changing key amino acids in some taxa. These architectural modifications may also create or reinforce reproductive barriers. Experimental investigations must be done (i) to identify amino acids involved in sperm-egg interaction in these proteins, (ii) to study the expression and the role of these genes in gametes and (iii) to identify their function in sperm-egg interaction in a greater number of species.

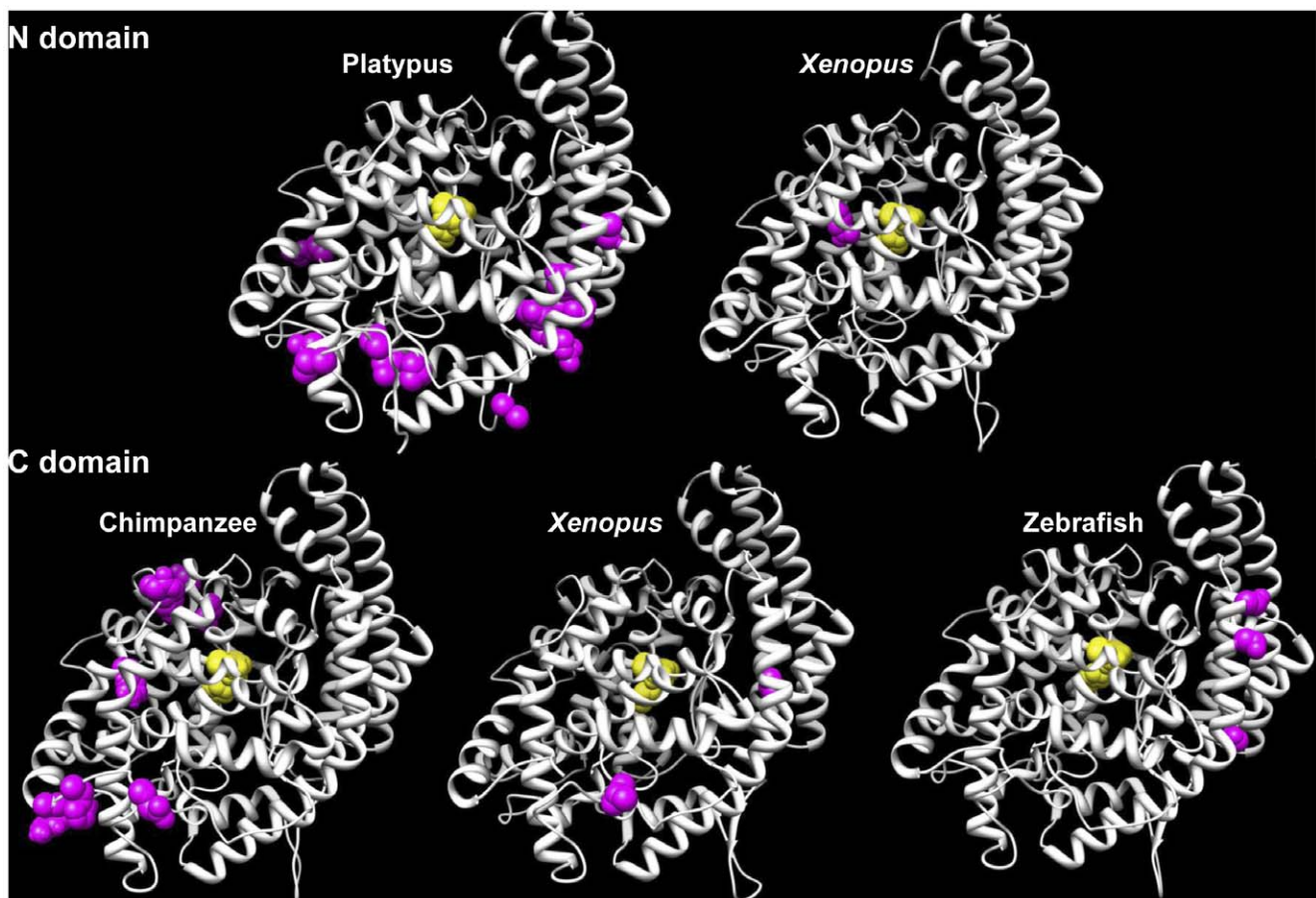
In conclusion, we study the evolution of 69 genes involved in the sperm-egg interaction. This work should help to improve our understanding of the evolution and the mechanism of fertilization in vertebrates, which still remains puzzling. In particular, the evolution of proteins present on male and female gametes can lead to reproductive barriers at the level of sperm-egg interaction in vertebrates. The variation between species regarding positive selection patterns as well as gain or loss of certain genes involved in fertilization could actually lead to the appearance or the

reinforcement of reproductive barriers. The relationship between positive selection and molecular adaptation has been experimentally shown previously [25,26]. Our work provides a list of potential targets, i.e. amino acids under positive selection, for targeted mutagenesis experiments to see if these amino acids are indeed involved in reproductive barriers between species. However, this list of genes should be completed after further experiments on other species. All this should improve significantly our understanding of the appearance of prezygotic isolation mechanisms, and hence, of the speciation process, which is of paramount importance in generating biodiversity.

## Materials and Methods

### Selection of Data

This study has sampled the genome of 19 bony vertebrate species that have been fully sequenced (*Bos taurus*, *Sus scrofa*, *Equus caballus*, *Canis familiaris*, *Rattus norvegicus*, *Mus musculus*, *Homo sapiens*, *Pan troglodytes*, *Macaca mulatta*, *Monodelphis domestica*, *Ornithorhynchus anatinus*, *Gallus gallus*, *Taeniopygia guttata*, *Xenopus tropicalis*, *Tetraodon nigroviridis*, *Takifugu rubripes*, *Gasterosteus aculeatus*, *Oryzias latipes* and *Danio rerio*). For all identified genes, the corresponding Ensembl protein ID was retrieved from Ensembl database and submitted to



**Figure 3. 3D structures of N and C domain of ACE in platypus, *Xenopus*, chimpanzee and zebrafish.** ACE structures have been modeled for four species based on human PDB structure (PDB: 1O8A) [36]. This enzyme is composed of two homolog N and C domains, each containing a catalytic site composed by two histidines in a conserved -HEXXH- zinc-binding motif [37] and a glutamate as the third ligand 24 residues downstream this motif [38]. These three binding sites are indicated in yellow, whereas amino acids under positive selection are in pink. ACE protein is expressed by sperm. Gene knockout experiments of ACE in mouse results in an impaired uterotubular sperm migration and a reduced ability to bind the zona pellucida of the egg [39,40,41]. doi:10.1371/journal.pone.0044548.g003

the PhyleasProg web server [15]. For each submitted Ensembl protein ID, the PhyleasProg web server interrogates the Ensembl database to retrieve orthologs and paralogs, the orthology/paralogy relationships assignment being based on a phylogenetic approach in Ensembl. All phylogenetic trees were reconstructed by PhyleasProg using the TreeBeST program (<http://treesoft.sourceforge.net/treebest.shtml>) and carefully examined before interpreting selective pressure results, also calculated by PhyleasProg. Selective pressure results are systematically verified and eventually corrected by synteny analysis as previously described by Tian, et al. [27]. Thus, calculations were performed with supposed correct orthologs. To score the presence, number and type of genes in the data matrix, a search for pseudogenes was systematically performed in the concerned genomes for genes for which no ortholog has been identified in at least one of the species of interest. The pseudogene status was inferred for a gene in a genome only if we have found a stop codon or an indel in the sequence identified by the similarity search in the syntenic locus in comparison with the other species of interest. Moreover, a search for duplicates was realized by using phylogenetic trees available in Ensembl and accessible via the Ensembl ID used for PhyleasProg and summarized in the Table S1.

### Inference of Positive Selection

The inference of selection was performed by PhyleasProg with branch-site models of codeml of the PAML package, with each ortholog branch tested for positive selection. Multiple alignments were systematically and carefully examined to avoid all false positive results. In particular, amino acids predicted to be under positive selection that were at the boundary of the alignments were not considered because doubtful. We also eliminated genes for which positive selection was due to sequence errors in Ensembl according to a comparison with other available sequences from other database such as RefSeq in NCBI (Figure S6 to S8). Each branch of each phylogenetic tree was tested for positive selection. So we performed multiple test corrections by controlling for the false discovery rate (FDR) using the R package QVALUE [28]. Results are considered significant with a threshold of  $q = 10\%$  of false positives. Sites with posterior probabilities of Bayes Empirical Bayes analyses superior to 95 or 99% were considered as positively selected. Datasets with less than 10 sequences, the minimum threshold required to obtain significant results, with excessively divergent sequences, or with sequences of genes for which annotations are not reliable were not retained for subsequent analysis (Figure 1). MODELLER [29] was used to build homology models of the 3D structure of proteins when possible and amino acids under positive selection and essential functional residues were localized on them.

### Comparison of Evolutionary Rates and the Intensity of Positive Selection Across Clades

Character history was inferred using parsimony optimization onto our rooted tree. In most cases, this yielded an unambiguous polarity. In a few cases, such as for the character ZPB/ZP4, the ancestral condition could not be unambiguously resolved using parsimony; it was then resolved using additional criteria. For instance, in that case, one of the two main clades, Teleostei, had the state "absent", whereas the other main clade, Tetrapoda, had the state "single copy". In such cases, we assumed that absence was primitive, and scored a gain for Tetrapoda. All cases of losses located on other parts of the tree (not on one of the two daughter-branches of the basalmost node) were inferred from parsimony optimizations whose polarity, in that case, is unproblematic. Similarly, when one of the two basalmost branches can be inferred

to have had a single copy whereas the other could most parsimoniously have two or more, we inferred that the presence of a single copy was primitive, based on the principle that the most common condition is likely to be primitive. This situation was encountered in the character ZPAX, for which Tetrapoda have a single copy, whereas Teleostei can most parsimoniously be inferred to have had two or three. To assess evolutionary rates, we optimized the 69 discrete characters representing gene presence, number and type (functional gene or pseudogene) onto the reference tree (Figure S9) and scored the occurrence of events of four types (gene appearance, duplication, transformation into a pseudogene, or loss) in three clades (Teleostei, Aves, Eutheria). We considered the basal branch subtending each of these three taxa as being part of it, and this is consistent with the fact that 19 sampled taxa in this study do not include representatives of both basalmost branches of these clades (e.g. paleognaths are not among the two birds sampled). We then divided the number of events of each type in each clade by that clade's sampled phylogenetic diversity index [30] (the sum of branch lengths, including of the basal branch). We then divided the resulting number by the proportion of the 69 genes in which such changes could occur in each clade. The last operation compensates for the fact that several genes present in eutherian mammals are absent in teleosts, and this avoids underestimating the rate of gene duplication in teleosts. Thus, origin of these genes could occur in teleosts, but not duplication, inactivation, or deletion. To statistically compare the rates between teleosts and eutherians, we took into consideration the sampled phylogenetic diversity of each clade and the number of genes for which each event was possible to assess the relative probability of occurrence of each event under the null hypothesis that these events are clade-independent. We then assessed the probability that a distribution at least as extreme as the observed one arises under the null hypothesis using a binomial test implemented in GraphPad (<http://www.graphpad.com/quickcalcs/binomial2.cfm>).

To compare the intensity of positive selection in the three clades of interest (Teleostei, Aves, Eutheria), we computed the inferred ancestral value of each clade using squared-change parsimony [31] in Mesquite [32]. Confidence intervals of these values were computed using phylogenetic independent contrasts [33] using the PDAP:PD TREE module [34] of Mesquite. These were used to assess statistical significance of the differences in ancestral values of each clade. Our method is derived from the continuous analysis developed for heterochrony analysis of continuous characters [35], a method whose statistical properties (especially power and type I error rate) have already been assessed using simulations. However, we improved a bit the method to more fully incorporate uncertainty about nodal estimates; instead of comparing a nodal value with the confidence interval (CI) of another clade, we compare the confidence intervals of both clades. This allows using narrower CIs because the probability that both nodal values actually lie outside the computed CI at a given error threshold is equal to the square of that threshold. As a result, instead of comparing a nodal value with the 95% CI (0.05 probability threshold), we compared the CIs at 78% of both clades because the square of 0.22 is just under 0.05 (we rounded off to the nearest higher percentage, which should make the test slightly conservative). Thus, if the two compared CIs do not overlap, we conclude that the difference is statistically significant. This method thus incorporates uncertainty in the estimates of all compared nodal values. The fact that data from all taxa are used to compute ancestral values should also make the test conservative because this should tend to make ancestral estimates of sister-taxa more similar

to each other than if the ancestral values were estimated for each clade separately.

## Supporting Information

**Figure S1 3D structure of ARSA in chicken, chimpanzee and medaka.** The three structures were modeled with human ARSA structure as template (PDB: 1N2K) [105]. The active site of the enzyme is represented in yellow, amino acids under positive selection in pink. ARSA is expressed by the spermatozoa; its involvement in the fertilization process was demonstrated with antibodies directed against ARSA. When the sperm is pretreated with these antibodies, their ability to bind the zona pellucida is reduced [71].

(TIF)

**Figure S2 3D structures of C1QBP in opossum and zebrafish.** Both structures were modeled with human P32 as template (PDB: 1P32) [106]. Amino acids under positive selection are in pink. The C1QBP glycoprotein is localized on sperm; its participation in the fertilization process was demonstrated using anti-C1QBP antibodies. The interaction between sperm and zona pellucida is suppressed when the sperm is pretreated with these antibodies [107].

(TIF)

**Figure S3 3D structures of IT $\alpha_v$  in cow, rat, green spotted pufferfish, medaka and zebrafish.** Four structures were modeled with human  $\alpha_v\beta_3$  integrin as template (PDB: 1L5G) [108]. The structure presented for the cow is the template itself. Amino acids under positive selection are in pink. This integrin is expressed by both oocyte and sperm. The pretreatment of sperm with anti IT $\alpha_v$  antibodies significantly decreases the sperm-zona pellucida binding and the fertilization [109].

(TIF)

**Figure S4 3D structures of PDIA3 in green spotted pufferfish, japanese pufferfish and zebrafish.** The three structures were modeled with human ERp57 as template (PDB: 3F8U) [110]. Catalytic motifs of the protein are in yellow, amino acids under positive selection are in pink. PDIA3 is expressed on sperm in the acrosomal region. Its role in the fusion process of sperm and oolemma was demonstrated with the use of anti-PDIA3 antibodies [89].

(TIF)

**Figure S5 Models of the 3D structures of UCHL3 in opossum and stickleback.** The two structures were modeled with human UCHL3 as template (PDB: 1UCH) [111]. Amino acids corresponding to the catalytic site of the enzyme are represented in yellow, amino acids under positive selection in pink. UCHL3 protein is present on both sperm acrosome and oocyte cortex. A study showed that UCHL3 is an ubiquitin C-terminal hydrolase and is involved in antipolyspermy defense during porcine fertilization [78].

## References

- Swanson WJ, Vacquier VD (2002) The rapid evolution of reproductive proteins. *Nat Rev Genet* 3: 137–144.
- Wyckoff GJ, Wang W, Wu CI (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature* 403: 304–309.
- Swanson WJ, Nielsen R, Yang Q (2003) Pervasive adaptive evolution in mammalian fertilization proteins. *Mol Biol Evol* 20: 18–20.
- Swanson WJ, Yang Z, Wolfner MF, Aquadro CF (2001) Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc Natl Acad Sci U S A* 98: 2509–2514.
- Dobzhansky T (1964) Columbia University Press.
- Yanagimachi R (1994) Mammalian fertilization. In: Knobil E, Neill JD, editors. *The physiology of reproduction*. New York: Raven Press. 189–317.
- Baba T, Azuma S, Kashiwabara S, Toyoda Y (1994) Sperm from mice carrying a targeted mutation of the acrosin gene can penetrate the oocyte zona pellucida and effect fertilization. *J Biol Chem* 269: 31845–31849.
- Lu Q, Shur BD (1997) Sperm from beta 1,4-galactosyltransferase-null mice are refractory to ZP3-induced acrosome reactions and penetrate the zona pellucida poorly. *Development* 124: 4121–4131.
- Goudet G, Mugnier S, Callebaut I, Monget P (2008) Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. *Biol Reprod* 78: 796–806.
- Busso D, Cohen DJ, Maldera JA, Dematteis A, Cuasnicu PS, et al. (2007) A novel function for CRISP1 in rodent fertilization: involvement in sperm-zona

(TIF)

**Figure S6 An example of a sequence error in Ensembl.** The sequence presented in (a) is the protein sequence of the gene PDIA3 from *Xenopus*. Amino acids in red and orange are those which have been detected as positively selected by PhyleasProg. A tBLASTn has been performed with this sequence against EST database of NCBI. The result of the BLAST is presented in (b), the domain which contains amino acids under positive selection in (a) is not retrieved. A new sequence for *Xenopus* was searched in NCBI RefSeq database. The sequence of the best BLAST hit is submitted to a tBLASTn against the EST database of NCBI for verification (c). The new identified sequence from NCBI replaced in this case the sequence from Ensembl.

(TIF)

**Figure S7 An example of a doubtful amino acid under positive selection.** The amino acid in red in the sequence of the rat (*Rattus norvegicus*) was found as positively selected in the alignment of protein sequences of the gene CD46. Because it is located at the boundary of the alignment, this amino acid is not considered, because doubtful.

(TIF)

**Figure S8 An example of reliable amino acids under positive selection.** Amino acids indicated in red in the sequence of the platypus (*Ornithorhynchus anatinus*) are considered as reliable in this multiple sequence alignment of ACE protein sequences.

(TIF)

**Figure S9 Time-calibrated reference phylogeny used for all phylogeny-informed analyses.** This tree follows classical time-calibrated phylogenies of teleosts [112] and mammals [113]. The divergence time between the chicken and zebra finch follows a recent review paper [114]. The other divergence times derive from a review of paleontological and molecular dates [115]. The names of several higher taxa are placed, with those most discussed in the text in bold type. The scale to the right gives approximate divergence times (in millions of years).

(TIF)

**Table S1 Genes included in the analysis.**

(DOCX)

## Acknowledgments

We acknowledge Pierre Pontarotti, Hugues Roest Crolius and Marc Robinson-Rechavi for helpful discussion and useful comments on the manuscript.

## Author Contributions

Conceived and designed the experiments: GG PM. Performed the experiments: CM SM ML IC. Analyzed the data: CM SM IC ML GP AP GG PM. Wrote the paper: CM SM PM.

- pellucida interaction-Evidence for the involvement of testicular protein CRISP2 in mouse sperm-egg fusion. *Biol Reprod* 77: 848–854.
11. Cohen DJ, Busso D, Da Ros V, Ellerman DA, Maldera JA, et al. (2008) Participation of cysteine-rich secretory proteins (CRISP) in mammalian sperm-egg interaction. *Int J Dev Biol* 52: 737–742.
  12. Da Ros VG, Maldera JA, Willis WD, Cohen DJ, Goulding EH, et al. (2008) Impaired sperm fertilizing ability in mice lacking Cysteine-Rich Secretory Protein 1 (CRISP1). *Dev Biol* 320: 12–18.
  13. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, et al. (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431: 946–957.
  14. Sato Y, Nishida M (2010) Teleost fish with specific genome duplication as unique models of vertebrate evolution. *Environmental Biology of Fishes* 88: 169–188.
  15. Busset J, Cabau C, Meslin C, Pascal G (2011) PhyleasProg: a user-oriented web server for wide evolutionary analyses. *Nucleic Acids Res* 39: W479–485.
  16. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586–1591.
  17. Studer RA, Penel S, Duret L, Robinson-Rechavi M (2008) Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. *Genome Res* 18: 1393–1402.
  18. Aagaard JE, Vacquier VD, MacCoss MJ, Swanson WJ (2010) ZP domain proteins in the abalone egg coat include a paralog of VERL under positive selection that binds lysin and 18-kDa sperm proteins. *Mol Biol Evol* 27: 193–203.
  19. McCartney MA, Lessios HA (2004) Adaptive Evolution of Sperm Bindin Tracks Egg Incompatibility in Neotropical Sea Urchins of the Genus *Echinometra*. *Molecular Biology and Evolution* 21: 732–745.
  20. Riginos C, McDonald JH (2003) Positive selection on an acrosomal sperm protein, M7 lysin, in three species of the mussel genus *Mytilus*. *Mol Biol Evol* 20: 200–207.
  21. Vacquier VD, Swanson WJ, Lee YH (1997) Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence? *J Mol Evol* 44 Suppl 1: S15–22.
  22. Berlin S, Smith NG (2005) Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. *BMC Evol Biol* 5: 65.
  23. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28: 235–242.
  24. Monne M, Jovine L (2011) A Structural View of Egg-Coat Architecture and Function in Fertilization. *Biol Reprod*.
  25. Sawyer SL, Wu LJ, Emerman M, Malik HS (2005) Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A* 102: 2832–2837.
  26. Weinberger H, Moran Y, Gordon D, Turkov M, Kahn R, et al. (2009) Positions under positive selection - key for selectivity and potency of scorpion {alpha}-toxins. *Mol Biol Evol* 17: 17.
  27. Tian X, Pascal G, Fouchecourt S, Pontarotti P, Monget P (2009) Gene birth, death, and divergence: the different scenarios of reproduction-related gene evolution. *Biol Reprod* 80: 616–621.
  28. Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 100: 9440–9445.
  29. Eswar N, Eramian D, Webb B, Shen MY, Sali A (2008) Protein structure modeling with MODELLER. *Methods Mol Biol* 426: 145–159.
  30. Faith DP (1992) Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61: 1–10.
  31. Maddison WP (1991) Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. *Systematic Zoology* 40: 304–314.
  32. Maddison WP, Maddison DR (2010) Mesquite: a modular system for evolutionary analysis.
  33. Felsenstein J (1985) Phylogenies and the comparative method. *American Naturalist* 125: 1–15.
  34. Midford P, Garland TJ, Maddison WP (2003) PDAP Package for Mesquite.
  35. Germain D, Laurin M (2009) Evolution of ossification sequences in salamanders and urodele origins assessed through event-pairing and new methods. *Evol Dev* 11: 170–190.
  36. Natesh R, Schwager SL, Sturrock ED, Acharya KR (2003) Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature* 421: 551–554.
  37. Soubrier F, Alhenc-Gelas F, Hubert C, Allegrini J, John M, et al. (1988) Two putative active centers in human angiotensin I-converting enzyme revealed by molecular cloning. *Proc Natl Acad Sci U S A* 85: 9386–9390.
  38. Williams TA, Corvol P, Soubrier F (1994) Identification of two active site residues in human angiotensin I-converting enzyme. *J Biol Chem* 269: 29430–29434.
  39. Esther CR, Jr., Howard TE, Marino EM, Goddard JM, Capecchi MR, et al. (1996) Mice lacking angiotensin-converting enzyme have low blood pressure, renal pathology, and reduced male fertility. *Lab Invest* 74: 953–965.
  40. Hagaman JR, Moyer JS, Bachman ES, Sibony M, Magyar PL, et al. (1998) Angiotensin-converting enzyme and male fertility. *Proc Natl Acad Sci U S A* 95: 2552–2557.
  41. Krege JH, John SW, Langenbach LL, Hodgin JB, Hagaman JR, et al. (1995) Male-female differences in fertility and blood pressure in ACE-deficient mice. *Nature* 375: 146–148.
  42. Cohen DJ, Ellerman DA, Busso D, Morgenfeld MM, Piazza AD, et al. (2001) Evidence that human epididymal protein ARP plays a role in gamete fusion through complementary sites on the surface of the human egg. *Biol Reprod* 65: 1000–1005.
  43. Chiu PC, Wong BS, Chung MK, Lam KK, Pang RT, et al. (2008) Effects of native human zona pellucida glycoproteins 3 and 4 on acrosome reaction and zona pellucida binding of human spermatozoa. *Biol Reprod* 79: 869–877.
  44. Gahlay G, Gauthier L, Baibakov B, Epifano O, Dean J (2010) Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. *Science* 329: 216–219.
  45. Smith J, Paton IR, Hughes DC, Burt DW (2005) Isolation and mapping the chicken zona pellucida genes: an insight into the evolution of orthologous genes in different species. *Mol Reprod Dev* 70: 133–145.
  46. Deguchi E, Tani T, Watanabe H, Yamada S, Kondoh G (2007) Dipeptidase-inactivated ACE action in vivo: selective inhibition of sperm-zona pellucida binding in the mouse. *Biol Reprod* 77: 794–802.
  47. Ensslin MA, Shur BD (2003) Identification of mouse sperm SED1, a bimotif EGF repeat and discoidin-domain protein involved in sperm-egg binding. *Cell* 114: 405–417.
  48. Ghosh I, Datta K (2003) Sperm surface hyaluronan binding protein (HABP1) interacts with zona pellucida of water buffalo (*Bubalus bubalis*) through its clustered mannose residues. *Mol Reprod Dev* 64: 235–244.
  49. Saxena DK, Oh-Oka T, Kadomatsu K, Muramatsu T, Toshimori K (2002) Behaviour of a sperm surface transmembrane glycoprotein basigin during epididymal maturation and its role in fertilization in mice. *Reproduction* 123: 435–444.
  50. Yamasaki N, Richardson RT, O'Rand MG (1995) Expression of the rabbit sperm protein Sp17 in COS cells and interaction of recombinant Sp17 with the rabbit zona pellucida. *Mol Reprod Dev* 40: 48–55.
  51. Furlong LI, Harris JD, Vazquez-Levin MH (2005) Binding of recombinant human proacrosin/acrosin to zona pellucida (ZP) glycoproteins. I. Studies with recombinant human ZPA, ZPB, and ZPC. *Fertil Steril* 83: 1780–1790.
  52. Cornwall GA, Tulsiani DR, Orgebin-Crist MC (1991) Inhibition of the mouse sperm surface alpha-D-mannosidase inhibits sperm-egg binding in vitro. *Biol Reprod* 44: 913–921.
  53. Lin YN, Roy A, Yan W, Burns KH, Matzuk MM (2007) Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Mol Cell Biol* 27: 6794–6805.
  54. Tengowski MW, Wassler MJ, Shur BD, Schatten G (2001) Subcellular localization of beta1,4-galactosyltransferase on bull sperm and its function during sperm-egg interactions. *Mol Reprod Dev* 58: 236–244.
  55. Ikawa M, Nakanishi T, Yamada S, Wada I, Kominami K, et al. (2001) Calmegin is required for fertilin alpha/beta heterodimerization and sperm fertility. *Dev Biol* 240: 254–261.
  56. Boue F, Berube B, De Lamirande E, Gagnon C, Sullivan R (1994) Human sperm-zona pellucida interaction is inhibited by an antiserum against a hamster sperm protein. *Biol Reprod* 51: 577–587.
  57. Walsh A, Whelan D, Bielanowicz A, Skinner B, Aitken RJ, et al. (2008) Identification of the molecular chaperone, heat shock protein 1 (chaperonin 10), in the reproductive tract and in capacitating spermatozoa in the male mouse. *Biol Reprod* 78: 983–993.
  58. Yu Y, Xu W, Yi YJ, Sutovsky P, Oko R (2006) The extracellular protein coat of the inner acrosomal membrane is involved in zona pellucida binding and penetration during fertilization: characterization of its most prominent polypeptide (IAM38). *Dev Biol* 290: 32–43.
  59. Bi M, Hickox JR, Winfrey VP, Olson GE, Hardy DM (2003) Processing, localization and binding activity of zonadhesin suggest a function in sperm adhesion to the zona pellucida during exocytosis of the acrosome. *Biochem J* 375: 477–488.
  60. Nishimura H, Kim E, Nakanishi T, Baba T (2004) Possible function of the ADAM1a/ADAM2 Fertilin complex in the appearance of ADAM3 on the sperm surface. *J Biol Chem* 279: 34957–34962.
  61. Yamaguchi R, Yamagata K, Ikawa M, Moss SB, Okabe M (2006) Aberrant distribution of ADAM3 in sperm from both angiotensin-converting enzyme (Ace)- and calmegin (Clgn)-deficient mice. *Biol Reprod* 75: 760–766.
  62. Veselsky L, Peknicova J, Cechova D, Kraus M, Geussova G, et al. (1999) Characterization of boar spermadhesin by monoclonal and polyclonal antibodies and their role in binding to oocytes. *Am J Reprod Immunol* 42: 187–197.
  63. Tardif S, Akrofi AS, Dass B, Hardy DM, MacDonald CC (2010) Infertility with impaired zona pellucida adhesion of spermatozoa from mice lacking TauCstF-64. *Biol Reprod* 83: 464–472.
  64. Okumura H, Kohno Y, Iwata Y, Mori H, Aoki N, et al. (2004) A newly identified zona pellucida glycoprotein, ZPD, and dimeric ZP1 of chicken egg envelope are involved in sperm activation on sperm-egg interaction. *Biochem J* 384: 191–199.
  65. Nishimura H, Myles DG, Primakoff P (2007) Identification of an ADAM2-ADAM3 complex on the surface of mouse testicular germ cells and cauda epididymal sperm. *J Biol Chem* 282: 17900–17907.
  66. Chiu PC, Chung MK, Koistinen R, Koistinen H, Seppala M, et al. (2007) Glycodelin-A interacts with fucosyltransferase on human sperm plasma



- membrane to inhibit spermatozoa-zona pellucida binding. *J Cell Sci* 120: 33–44.
67. Gibbs GM, Lo JC, Nixon B, Jamsai D, O'Connor AE, et al. (2010) Glioma pathogenesis-related 1-like 1 is testis enriched, dynamically modified, and redistributed during male germ cell maturation and has a potential role in sperm-oocyte binding. *Endocrinology* 151: 2331–2342.
  68. Goncalves RF, Barnabe VH, Killian GJ (2008) Pre-treatment of cattle sperm and/or oocyte with antibody to lipocalin type prostaglandin D synthase inhibits in vitro fertilization and increases sperm-oocyte binding. *Anim Reprod Sci* 106: 188–193.
  69. Cheng GY, Shi JL, Wang M, Hu YQ, Liu CM, et al. (2007) Inhibition of mouse acrosome reaction and sperm-zona pellucida binding by anti-human sperm membrane protein 1 antibody. *Asian J Androl* 9: 23–29.
  70. Buffone MG, Zhuang T, Ord TS, Hui L, Moss SB, et al. (2008) Recombinant mouse sperm ZP3-binding protein (ZP3R/sp56) forms a high order oligomer that binds eggs and inhibits mouse fertilization in vitro. *J Biol Chem* 283: 12438–12445.
  71. Rattanachaiyanont M, Weerachayanukul W, Leveille MC, Taylor T, D'Amours D, et al. (2001) Anti-SLIP1-reactive proteins exist on human spermatozoa and are involved in zona pellucida binding. *Mol Hum Reprod* 7: 633–640.
  72. Zitta K, Wertheimer EV, Miranda PV (2006) Sperm N-acetylglucosaminidase is involved in primary binding to the zona pellucida. *Mol Hum Reprod* 12: 557–563.
  73. Brahmaraju M, Shoeb M, Laloraya M, Kumar PG (2004) Spatio-temporal organization of Vam6P and SNAP on mouse spermatozoa and their involvement in sperm-zona pellucida interactions. *Biochem Biophys Res Commun* 318: 148–155.
  74. Escoffier J, Jemel I, Tanemoto A, Taketomi Y, Payre C, et al. (2010) Group X phospholipase A2 is released during sperm acrosome reaction and controls fertility outcome in mice. *J Clin Invest* 120: 1415–1428.
  75. Sosnik J, Buffone MG, Visconti PE (2010) Analysis of CAPZA3 localization reveals temporally discrete events during the acrosome reaction. *J Cell Physiol* 224: 575–580.
  76. Zhao L, Reim K, Miller DJ (2008) Complexin-I-deficient sperm are subfertile due to a defect in zona pellucida penetration. *Reproduction* 136: 323–334.
  77. Song XX, Park CK, Piao YJ, Niwa K (2007) Effect of monosaccharide L-fucose and polysaccharide fucoidan on sperm alpha-L-fucosidase activity and relation to sperm-oocyte interaction in pig. *Asian-Australasian Journal of Animal Sciences* 20: 351–358.
  78. Yi YJ, Manandhar G, Sutovsky M, Li R, Jonakova V, et al. (2007) Ubiquitin C-terminal hydrolase-activity is involved in sperm acrosomal function and anti-polyspermy defense during porcine fertilization. *Biol Reprod* 77: 780–793.
  79. Ganguly A, Bukovsky A, Sharma RK, Bansal P, Bhandari B, et al. (2010) In humans, zona pellucida glycoprotein-1 binds to spermatozoa and induces acrosomal exocytosis. *Hum Reprod* 25: 1643–1656.
  80. Kohn MJ, Szein J, Yagi R, DePamphilis ML, Kaneko KJ (2010) The acrosomal protein Dickkopf-like 1 (DKKL1) facilitates sperm penetration of the zona pellucida. *Fertil Steril* 93: 1533–1537.
  81. Yamashita M, Honda A, Ogura A, Kashiwabara S, Fukami K, et al. (2008) Reduced fertility of mouse epididymal sperm lacking Prss21/Tesp5 is rescued by sperm exposure to uterine microenvironment. *Genes Cells* 13: 1001–1013.
  82. Yin L, Chung CM, Huo R, Liu H, Zhou C, et al. (2009) A sperm GPI-anchored protein elicits sperm-cumulus cross-talk leading to the acrosome reaction. *Cell Mol Life Sci* 66: 900–908.
  83. Thys M, Nauwynck H, Maes D, Hoogewijs M, Vercauteren D, et al. (2009) Expression and putative function of fibronectin and its receptor (integrin alpha(5)beta(1)) in male and female gametes during bovine fertilization in vitro. *Reproduction* 138: 471–482.
  84. Ziyat A, Rubinstein E, Monier-Gavelle F, Barraud V, Kulski O, et al. (2006) CD9 controls the formation of clusters that contain tetraspanins and the integrin alpha 6 beta 1, which are involved in human and mouse gamete fusion. *J Cell Sci* 119: 416–424.
  85. Takahashi Y, Bigler D, Ito Y, White JM (2001) Sequence-specific interaction between the disintegrin domain of mouse ADAM 3 and murine eggs: role of beta1 integrin-associated proteins CD9, CD81, and CD98. *Mol Biol Cell* 12: 809–820.
  86. Barraud-Lange V, Naud-Barriant N, Saffar L, Gattegno L, Ducot B, et al. (2007) Alpha6beta1 integrin expressed by sperm is determinant in mouse fertilization. *BMC Dev Biol* 7: 102.
  87. Ziyat A, Naud-Barriant N, Barraud-Lange V, Chevalier F, Kulski O, et al. (2005) Cyclic FEE peptide increases human gamete fusion and potentiates its RGD-induced inhibition. *Hum Reprod* 20: 3452–3458.
  88. Boissonnas CC, Montjean D, Lesaffre C, Auer J, Vaiman D, et al. (2010) Role of sperm alphavbeta3 integrin in mouse fertilization. *Dev Dyn* 239: 773–783.
  89. Ellerman DA, Myles DG, Primakoff P (2006) A role for sperm surface protein disulfide isomerase activity in gamete fusion: evidence for the participation of ERp57. *Dev Cell* 10: 831–837.
  90. Taylor CT, Biljan MM, Kingsland CR, Johnson PM (1994) Inhibition of human spermatozoan-oocyte interaction in vitro by monoclonal antibodies to CD46 (membrane cofactor protein). *Hum Reprod* 9: 907–911.
  91. Jagadish N, Rana R, Mishra D, Garg M, Chaurasiya D, et al. (2005) Immunogenicity and contraceptive potential of recombinant human sperm associated antigen (SPAG9). *J Reprod Immunol* 67: 69–76.
  92. Linfor J, Berger T (2000) Potential role of alpha and beta1 integrins as oocyte adhesion molecules during fertilization in pigs. *J Reprod Fertil* 120: 65–72.
  93. Glazar AI, Evans JP (2009) Immunoglobulin superfamily member IgSF8 (EWI-2) and CD9 in fertilisation: evidence of distinct functions for CD9 and a CD9-associated protein in mammalian sperm-egg interaction. *Reprod Fertil Dev* 21: 293–303.
  94. Vjugina U, Zhu X, Oh E, Bracero NJ, Evans JP (2009) Reduction of mouse egg surface integrin alpha9 subunit (ITGA9) reduces the egg's ability to support sperm-egg binding and fusion. *Biol Reprod* 80: 833–841.
  95. Marin-Briggiler CI, Veiga MF, Matos ML, Echeverria MF, Furlong LI, et al. (2008) Expression of epithelial cadherin in the human male reproductive tract and gametes and evidence of its participation in fertilization. *Mol Hum Reprod* 14: 561–571.
  96. Sekiguchi S, Kwon J, Yoshida E, Hamasaki H, Ichinose S, et al. (2006) Localization of ubiquitin C-terminal hydrolase L1 in mouse ova and its function in the plasma membrane to block polyspermy. *Am J Pathol* 169: 1722–1729.
  97. Pasten-Hidalgo K, Hernandez-Rivas R, Roa-Espitia AL, Sanchez-Gutierrez M, Martinez-Perez F, et al. (2008) Presence, processing, and localization of mouse ADAM15 during sperm maturation and the role of its disintegrin domain during sperm-egg binding. *Reproduction* 136: 41–51.
  98. Inoue N, Ikawa M, Isotani A, Okabe M (2005) The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature* 434: 234–238.
  99. Ziv S, Rufas O, Shalgi R (2002) Cadherins expression during gamete maturation and fertilization in the rat. *Mol Reprod Dev* 62: 547–556.
  100. Ying X, Liu Y, Guo Q, Qu F, Guo W, et al. (2010) Endoplasmic reticulum protein 29 (ERp29), a protein related to sperm maturation is involved in sperm-oocyte fusion in mouse. *Reprod Biol Endocrinol* 8: 10.
  101. Herrero MB, Mandal A, Digilio LC, Coonrod SA, Maier B, et al. (2005) Mouse SLLP1, a sperm lysozyme-like protein involved in sperm-egg binding and fertilization. *Dev Biol* 284: 126–142.
  102. Busso D, Goldweic NM, Hayashi M, Kasahara M, Cuanicu PS (2007) Evidence for the involvement of testicular protein CRISP2 in mouse sperm-egg fusion. *Biol Reprod* 76: 701–708.
  103. Fujihara Y, Murakami M, Inoue N, Satouh Y, Kaseda K, et al. (2010) Sperm equatorial segment protein 1, SPESP1, is required for fully fertile sperm in mouse. *J Cell Sci* 2010: 7.
  104. Shetty J, Wolkowicz MJ, Digilio LC, Klotz KL, Jayes FL, et al. (2003) SAMP14, a novel, acrosomal membrane-associated, glycosylphosphatidylinositol-anchored member of the Ly-6/urokinase-type plasminogen activator receptor superfamily with a role in sperm-egg interaction. *J Biol Chem* 278: 30506–30515.
  105. Chruszcz M, Laidler P, Monkiewicz M, Ortlund E, Lebiada L, et al. (2003) Crystal structure of a covalent intermediate of endogenous human arylsulfatase A. *J Inorg Biochem* 96: 386–392.
  106. Jiang J, Zhang Y, Kraimer AR, Xu RM (1999) Crystal structure of human p32, a doughnut-shaped acidic mitochondrial matrix protein. *Proc Natl Acad Sci U S A* 96: 3572–3577.
  107. Ranganathan S, Ganguly AK, Datta K (1994) Evidence for presence of hyaluronan binding protein on spermatozoa and its possible involvement in sperm function. *Mol Reprod Dev* 38: 69–76.
  108. Xiong JP, Stehle T, Zhang R, Joachimiak A, Frech M, et al. (2002) Crystal structure of the extracellular segment of integrin alpha Vbeta3 in complex with an Arg-Gly-Asp ligand. *Science* 296: 151–155.
  109. Goncalves RF, Wolmetz CD, Killian GJ (2007) Influence of arginine-glycine-aspartic acid (RGD), integrins (alphaV and alpha5) and osteopontin on bovine sperm-egg binding, and fertilization in vitro. *Theriogenology* 67: 468–474.
  110. Dong G, Wearsch PA, Peaper DR, Cresswell P, Reinisch KM (2009) Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. *Immunity* 30: 21–32.
  111. Johnston SC, Larsen CN, Cook WJ, Wilkinson KD, Hill CP (1997) Crystal structure of a deubiquitinating enzyme (human UCH-L3) at 1.8 Å resolution. *Embo J* 16: 3787–3796.
  112. Yamanoue Y, Miya M, Inoue JG, Matsuura K, Nishida M (2006) The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. *Genes Genet Syst* 81: 29–39.
  113. Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, et al. (2007) The delayed rise of present-day mammals. *Nature* 446: 507–512.
  114. van Tuinen M (2009) *The Timetree of Life*; S. BH, S. K, editors. New York: Oxford University Press.
  115. Marjanovic D, Laurin M (2007) Fossils, molecules, divergence times, and the origin of lissamphibians. *Syst Biol* 56: 369–388.