BIOLOGY OF REPRODUCTION **78,** 796–806 (2008)
Published online before print 28 November 2007.
DOI 10.1095/biolreprod.107.064568

Minireview

Phylogenetic Analysis and Identification of Pseudogenes Reveal a Progressive Loss of Zona Pellucida Genes During Evolution of Vertebrates[1]

Ghylène Goudet,[2,3] Sylvie Mugnier,[3] Isabelle Callebaut,[4] and Philippe Monget[3]

Physiologie de la Reproduction et des Comportements,[3] INRA-CNRS-Université de Tours-Haras Nationaux, IFR 135, 37380 Nouzilly, France
Département de Biologie Structurale,[4] IMPMC UMR 7590, Universités Pierre et Marie Curie-Paris 6 et Denis Diderot-Paris 7, CNRS, Campus Boucicaut, 75015 Paris, France

ABSTRACT

Vertebrate eggs are surrounded by an extracellular matrix with similar functions and conserved individual components: the zona pellucida (ZP) glycoproteins. In mammals, chickens, frogs, and some fish species, we established an updated list of the *ZP* genes, studied the relationships within the *ZP* gene family using phylogenetic analysis, and identified *ZP* pseudogenes. Our study confirmed the classification of *ZP* genes in six subfamilies: *ZPA/ZP2*, *ZPB/ZP4*, *ZPC/ZP3*, *ZP1*, *ZPAX*, and *ZPD*. The identification of a *Zpb* pseudogene in the mouse genome, *Zp1* pseudogenes in the dog and bovine genomes, and *Zpax* pseudogenes in the human, chimpanzee, macaque, and bovine genomes showed that the evolution of *ZP* genes mainly occurs by death of genes. Our study revealed that the extracellular matrix surrounding vertebrate eggs contains three to at least six ZP glycoproteins. Mammals can be classified in three categories. In the mouse, the ZP is composed of three ZP proteins (ZPA/ZP2, ZPC/ZP3, and ZP1). In dog, cattle and, putatively, pig, cat, and rabbit, the zona is composed of three ZP proteins (ZPA/ZP2, ZPB/ZP4, and ZPC/ZP3). In human, chimpanzee, macaque, and rat, the ZP is composed of four ZP proteins (ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, and ZP1). Our review provides new directions to investigate the molecular basis of sperm-egg recognition, a mechanism which is not yet elucidated.

evolution, fertilization, gamete biology, oocyte development, ovum, pseudogene, zona pellucida

INTRODUCTION

Vertebrate eggs are surrounded by an extracellular matrix called the chorion in fish, the vitelline envelope in amphibians, the perivitelline envelope in reptiles and birds, and the zona pellucida (ZP) in mammals. These extracellular matrices have similar functions. They participate in taxon-specific sperm-egg binding during fertilization and protect the embryo during early development. Moreover, they have similar ultrastructures, composed of fibrous matrices with conserved individual components featuring common protein domains.

The ZP matrix that surrounds all mammalian oocytes is composed of three to four major glycoproteins. Remarkably, the different ZP glycoproteins share an apparent overall similar architecture, around a conserved C-terminal ZP domain, preceding their transmembrane segment. The ZP domain is a large domain (~260 residues) found in a wide variety of extracellular proteins with various functions [1]. Sequences downstream from the ZP domains are much variable and present marked differences between the ZP proteins (see below). The initial nomenclature of the ZP glycoproteins was based on the apparent molecular weight of the mouse ZP proteins after migration by SDS-PAGE. These proteins were named ZP1, ZP2, and ZP3 from the highest to the lowest apparent molecular weight, respectively [2]. Zona pellucida glycoproteins that have been subsequently described in other species have been named according to several criteria, including apparent molecular weight and charge following two-dimensional gel electrophoresis [3], the size of the cDNAs (with *Zpa* being the largest and *Zpc* the smallest) [4], as well as sequence identity comparison [5, 6]. This has resulted in a confused nomenclature.

The situation for the pig species was particularly confusing. Electrophoretic analysis suggested that the pig ZP family was composed of four distinct proteins [7]. Other investigators reported that the pig ZP family was composed of three proteins, referred to as PZI, PZII, and PZIII [8, 9], or ZP1, ZP2, and ZP3 [10]. Later studies suggested the presence of a fourth pig ZP protein, referred to as PZIV [11] or 25K [12]. Additional studies suggested that the ZP3 component was actually a mixture of two different proteins, referred to as ZP3alpha and ZP3beta [13].

In an attempt to clarify the relationship between the different classes of *ZP* genes, Harris et al. [4] proposed a unified system of nomenclature in which *ZP* genes were named in order of the length of their encoded protein sequences with a new letter-based system. Thus, *ZP2* became *ZPA*, *ZP1* became *ZPB*, and *ZP3* became *ZPC*. However, the number system and the letter system are used concurrently by different groups, increasing the confusion. For example, in humans, *ZPB* and *ZP1* are two distinct genes [14]. In the pig, *ZP1* and *ZP2*, described as separate genes, are two names of the same gene [15, 16]. Later, Conner et al. [17] proposed a simplified naming system in which ZP genes are termed ZP1, ZP2, ZP3, and ZP4. They

[1]Supported by grants from Les Haras Nationaux in France and La Région Centre in France to S.M.
[2]Correspondence: FAX: 33 2 47 42 77 43; e-mail: goudet@tours.inra.fr

Received: 26 July 2007.
First decision: 29 August 2007.
Accepted: 7 November 2007.

ISSN: 0006-3363. http://www.biolreprod.org

focused on the four main classes of ZP protein, omitting ZPAX and ZPD subfamilies. Spargo and Hope [15] have also created a logical system of nomenclature on the basis of which an updated classification can be proposed.

Here, we have established an updated list of the ZP genes in mammals, chickens, and frogs. In these species and some fish species we have established the relationships within the *ZP* gene family using phylogeny. Finally, we have identified *ZP* pseudogenes in several species. This will help to understand the characteristic evolution of these reproductive proteins.

ANNOTATION OF VERTEBRATE ZP GENES BY PHYLOGENETIC ANALYSIS

The nomenclature used to describe *ZP* genes and proteins from different mammalian species is very confusing. We established an updated list of the *ZP* genes in mammals, chickens, and frogs, and we identified orthologous and paralogous genes on the basis of phylogenetic analyses. Phylogenetic analyses were carried out using FIGENIX software (http://www.up.univ-mrs.fr/evol/figenix) [18], which produces a consensus tree according to the three approaches, the Neighbor Joining method [19], the Maximum Parsimony method [20], and the Maximum Likelihood method [21]. The dataset of putative homologous sequences used for phylogenetic reconstruction is built by BLAST that performs local alignment of peptide sequences. Then, sequences are submitted to the MUSCLE software (http://www.drive5.com/muscle) [22, 23], which performs a multiple alignment of complete peptide sequences. The sequences of interest used as query were the amino acid sequences of the pig ZPA/ZP2 and ZPC/ZP3.

Table 1 shows an updated list of the *ZP* genes in mammals, chickens, and frog, as well as the Swissprot and GenBank accession numbers and the current name of all of these *ZP* genes. Three to six *ZP* genes are present in these species. To avoid confusion, fish *ZP* sequences have been omitted, as there have been both genome and gene duplications giving rise to multiple gene copies [24].

An Updated Annotation of ZP *Genes*

The results of the phylogenetic analyses are shown in Figures 1 and 2 and in the supplemental data (Supplementary Fig. 1 and Supplementary Fig. 2, available online at www.biolreprod.org). When the phylogenetic analysis is built with FIGENIX using the amino acid sequence of the porcine ZPA as query, the phylogenetic tree contains the genes from the *ZPA*/*ZP2*, *ZPB*/*ZP4*, *ZP1*, *ZPD*, and *ZPAX* subfamilies, but not the *ZPC*/*ZP3* subfamily (Fig. 1 and Supplementary Fig. 1). As previously suggested [5], a gene duplication gave rise to two paralogous groups of genes within the *ZPB* subfamily, *ZP1* and *ZPB*/*ZP4*, since the phylogenetic tree shows that they share a common ancestral gene. This common ancestral gene shares a common ancestral gene with the *ZPA*/*ZP2* subfamily, and the latter shares a common ancestral gene with the *ZPAX* subfamily. These genes share a common ancestral gene with the *ZPD* subfamily. This ZPD subfamily shares an ancestral gene with the CUZD1/DMBT1 gene subfamily (CUB and zona pellucida-like domains 1/Deleted in Malignant Brain Tumors 1). When the phylogenetic analysis is built using the amino acid sequence of the porcine ZPC as query, the phylogenetic tree only contains the genes from the *ZPC*/*ZP3* subfamily (Fig. 2 and Supplementary Fig. 2). This subfamily shares an ancestral gene with the DMBT1 gene subfamily. We never could obtain any consensus tree containing the six ZP subfamilies. This raises the question of an ancestral gene with *ZPC*. One could hypothesize that the first event in ZP evolution could be a gene duplication event, which gave rise to the ancestral *ZPC* gene and to the ancestral gene of the *ZPA*/*ZP2*, *ZPB*/*ZP4*, *ZPD*, *ZP1*, and *ZPAX* subfamilies. Spargo and Hope [15] provide some evidence that the first event in ZP evolution was a gene duplication event, which gave rise to the ancestral *ZPC* gene and to the precursor of all other *ZP* gene subfamilies. In fish species, there are several *ZPC*/*ZP3* genes. For example, the phylogenetic tree exhibits four Oryzias latipes *ZPC-ZP3* genes—*ZPC2*, *ZPC3*, *ZPC4*, and *ZPC5*—and three Danio rerio *ZPC*/*ZP3* genes—*ZP3*, *ZP3a* and *ZP3b*. As stated above, due to the degree of diversity of ZP in these species, the clarification of fish ZP gene classification requires further analysis that is not presented here because it would have increased the complexity of our data.

The ZP subfamilies share an ancestral gene with the CUZD1/DMBT1 gene subfamily. The CUZD1/DMBT1 subfamily contains proteins that incorporate two domains, the CUB domain and the ZP domains [25–29]. These two domains are present in two separate families of proteins implicated in sperm-egg recognition: the CUB domain is found in the spermadhesins [30], and the ZP domain is found in the ZP proteins [31]. Moreover, the CUZD1/DMBT1 proteins are expressed in the female genital tract under estrogen regulation [25–27, 32]. This suggests that these proteins may possess a role in fertilization.

The phylogenetic analysis allowed us to classify *ZP* genes in six subfamilies: the *ZPA*/*ZP2* subfamily, the *ZPB*/*ZP4* subfamily, the *ZPC*/*ZP3* subfamily, the *ZP1* subfamily, the *ZPAX* subfamily, and the *ZPD* subfamily. It should be noted that two so-called Macaque (Mra) *ZP1* were identified, one being reclassified here in the *ZPB*/*ZP4* subfamily [33], and the other in the *ZP1* subfamily. Moreover, the so-called marmoset (Cja) *ZP1* gene [34] was reclassified here in the *ZPB*/*ZP4* subfamily, and the so-called pig *ZP1* gene [35] was reclassified here in the *ZPA*/*ZP2* subfamily, because the genes previously named *ZP1* [35], *ZP2* [36], and *ZPA* [37] in this species encode for the same protein. There are actually three distinct *ZP* genes in the pig: the *ZPA*/*ZP2* gene previously called *ZP1* [10], the *ZPB*/*ZP4* gene previously called *ZP3alpha* [13], and the *ZPC*/*ZP3* gene previously called *ZP3beta* [13].

The murine ZP is composed of three major glycoproteins, ZP1, ZP2, and ZP3 [2], and it was suggested that the human ZP was composed of three distinct proteins, ZPA, ZPB, and ZPC [4]. The complete cDNA sequence of the mouse *Zp1* gene was determined subsequently to the sequence of the human *ZPB* cDNA sequence [4, 38]. It was assumed that the mouse ZP1 was orthologous to human ZPB [38] because it shared greater amino acid sequence identity with human ZPB than with human ZPA or ZPC, until a human genomic sequence orthologous to the mouse *Zp1* gene and paralogous to the human *ZPB* gene was identified [39]. Soon after, Bausek et al. [5] suggested that a gene duplication gave rise to two paralogous groups of genes within the *ZPB* subfamily: *ZP1* and *ZPB*. Our bioinformatic analysis now clarifies the classification of the *ZP1* and *ZPB* genes, with the *ZPB*/*ZP4* subfamily containing the ortholog of human *ZPB*/*ZP4*, and the *ZP1* subfamily containing the human, chimpanzee, macaque, mouse, rat, and chicken *ZP1*.

The avian oocyte is surrounded by the perivitelline membrane, which is equivalent to the ZP. Whereas the mammalian ZP has three to four main glycoproteins, the chicken perivitelline membrane is composed of six glycoproteins: ZP1/ZPB1, ZPA/ZP2, ZPB/ZP4/ZPB2, ZPC/ZP3, ZPD/ZPX2, and ZPAX/ZPX1 [40]. The genetic and physical mapping of these genes has been performed [40]. The *Xenopus* egg envelope contains five main glycoprotein components: ZPA, ZPB, ZPC, ZPD/ZPX2, and ZPAX/ZPX1 [15]. As stated above, the fish gene family is complex, and all of the *ZP* genes

TABLE 1. Characterized genes belonging to the *ZP* gene family.

Current symbol	GenBank accession no.	SwissProt UniProt accession no.
Sus scrofa (Ssc), Pig		
ZPA/ZP2 (also named *ZP1* [10, 15])	NM_213848	P42099
	D45064	
	L22170	
	S74651	
ZPB/ZP4 (also named *ZP3alpha* [13])	NM_214045	Q07287
	L11000	
ZPC/ZP3 (also named *ZP3beta* [13])	NM_213893	P42098
	L22169	
	D45065	
Bos taurus (Bta), Cow		
ZPA/ZP2	NM_173973	Q9BH10
	AB042653	
ZPB/ZP4 (also named *ZPB2* [16])	NM_173975	Q9BH11
	AB042652	
ZPC/ZP3 (also named *ZP3B* [4])	NM_173974	P48830
	U05775	
	BT021613	
Canis familiaris (Cfa), Dog		
ZPA/ZP2	NM_001003304	P47983
	D45069	
	U05779	
ZPB	AY573930	
ZPC/ZP3	NM_001003224	P48831
	U05780	
	D45070	
Felis catus (Fca), Cat		
ZPA/ZP2	NM_001009875	P47984
	D45067	
	U05776	
ZPB/ZP4 (also named *ZPB2* [16])	NM_001009260	P48834
	U05777	
ZPC/ZP3	NM_001009330	P48832
	U05778	
	D45068	
Oryctolagus cuniculus (Ocu), Rabbit		
ZPA/ZP2 (also named *75kDa* [16])	L12167	P48829
ZPB/ZP4 (also named *ZPB2* [16])	M58160	Q00193
ZPC/ZP3	U05782	P48833
Homo sapiens (Hsa), Human		
ZPA/ZP2	NM_003460	Q05996
	AF001550	
	BC096304	
	BC096305	
	BC096306	
	BC096307	
	M90366	
ZPB/ZP4 (also named *ZPB2* [16])	NM_021186	Q12836
	AL359924	
	BC069521	
	U05781	
ZPC/ZP3	NM_007155	P21754
	M60504	
	X56777	
	A18567	
ZP1	NM_207341	P60852
	AC004126	
Macaca radiata (Mra), Bonnet monkey		
ZPA/ZP2	Y10690	O77726
ZP1 (also named *ZPB2* [16])	Y10381	O19027
	Y10383	
ZPC/ZP3	X82639	P53785
ZP1	EF530200	
	ABP88868	
Macaca fascicularis (Mfa), Crab-eating macaque		
ZPA	AY222645	Q864C1
ZPB	AY222647	Q86150
	AY222648	
ZPC	AY222644	Q864C2
Macaca mulatta (Mmul), Rhesus macaque		
ZPA/ZP2	XM_001091029	
	XM_001091147	
ZPB/ZP4	XM_001096956	

TABLE 1. Continued.

Current symbol	GenBank accession no.	SwissProt UniProt accession no.
	XM_001096846	
ZPC/ZP3	XM_001114760	
ZP1	XM_001084628	
Callithrix jacchus (Cja), White-tufted-ear marmoset		
ZP2	Y10767	P79160
ZP1 (also named *ZPB2* [16])	Y10822	P79159
Callithrix sp.(Csp), Marmoset		
ZPC/ZP3	S71825	P53786
Pan troglodytes (Ptr), Chimpanzee		
ZPA/ZP2	XM_510869	
ZPB/ZP4	XM_525105	
ZPC/ZP3	XM_519164	
ZP1	XM_522022	
Papio cynocephalus (Pcy), Yellow baboon		
ZPB	AY222646	
Mus musculus (Mmus), House mouse		
ZPA/ZP2	NM_011775	P20239
	BC071183	
	M34148	
ZPC/ZP3	NM_011776	P10761
	X14376	
	M20026	
ZP1 (also named *ZPB1* [16])	NM_009580	Q62005
	U20448	
	U24227	
	U24228	
	U24229	
	U24230	
Rattus norvegicus (Rno), Rat		
ZPA/ZP2	NM_031150	O54767
	AB000929	
ZPB/ZP4 (also named *ZPB2* [16])	NM_172330	Q8CH34
	AF456325	
ZP3	NM_053762	P97708
	Y10823	
	D78482	
ZP1 (also named *ZPB1* [16])	NM_053509	O54766
	AB000928	
Gallus gallus (Gga), Chicken		
ZPA/ZP2	BN000517	Q5CZI6
	BAE72123	Q2PGY2
ZPB/ZP4/ZPB2	AB025428	Q6WV21
	BAA76739	
	NM_204879	
ZPC/ZP3	AY628622	P79762
	NM_204389	
	AB031033	
ZP1/ZPB1	NM_204683	Q9DER4
	AJ289697	Q6WV24
ZPAX/ZPX1	NM_001045837	Q684L7
	AJ698915	Q6WV22
ZPD/ZPX2	AB114441	Q766V2
	BAD13713	
	NP_998741	
	NM_213576	
Xenopus laevis (Xla), African clawed frog		
ZPA/ZP2/69kDa	AAD12172	O73735
	BC079825	Q6AX05
	AF038151	
ZPB/ZP4/gp37	XLU44950	
	BC123370	
	AAA91465	
ZPC/ZP3	U44952	Q91675
	AAB39079	
	BC072326	
ZPAX/ZPX1	AF225906	
	AY079195	
	AAF43011	
ZPD/ZPX2	XLU44949	Q91672
	AY079194	
	AAA91467	
Xenopus tropicalis (Xtr), Western clawed frog		
ZPA	AY079191	

TABLE 1. Continued.

Current symbol	GenBank accession no.	SwissProt UniProt accession no.
	NM_203524	
	NP_988855	
ZPB/ZP4	AY079192	Q28CH4
	NM_203523	
ZPC/ZP3	AY079193	Q28G34
	NM_203522	
ZPAX	AY079195	
	NM_203520	
ZPD	AY079194	Q28FD6
	NM_203521	

were not present in the tree that we generated, likely due to the high degree of divergence with mammalian *ZP* genes.

The automatic alignments provided by the BLAST2 software between paralogous genes, which mainly focused on the much conserved ZP domains, show relatively high levels of divergence. These alignments did not take into account the N-terminal extensions which exist in ZP proteins, with the exception of ZPC (as well as the chicken ZPD). We recently showed that these extensions are made of divergent copies of ZP-N domains [41]. ZP-N domains consist in the N-terminal part of ZP domains, which have been shown to exist independently of the C-terminal regions in proteins, such as

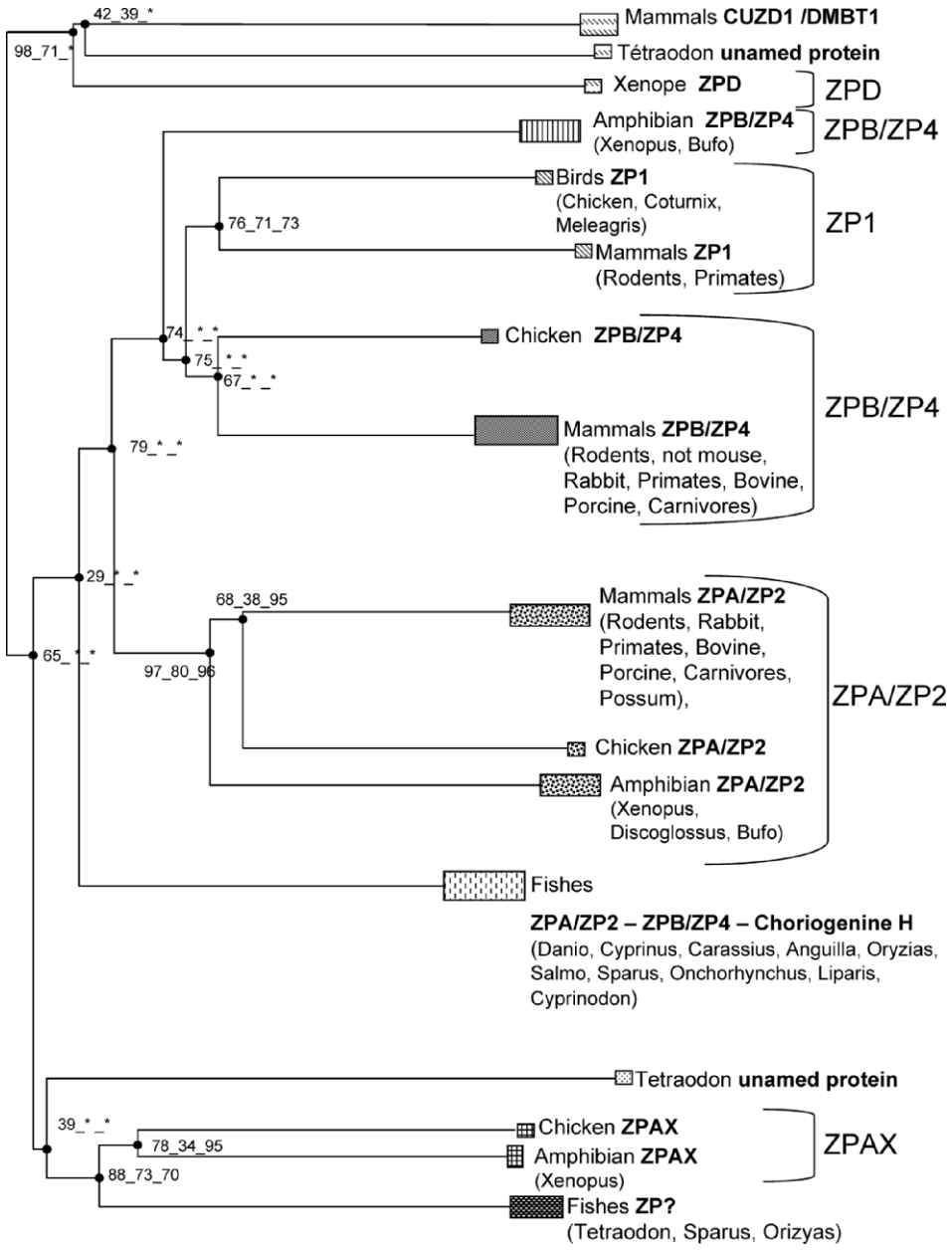

FIG. 1. Phylogenetic tree of the *ZP* gene family obtained when the FIGENIX BLAST is performed from the porcine ZPA/ZP2 sequence. Relative branch lengths indicate rates of evolution along a particular branch. The values at the tree nodes represent bootstrap values.

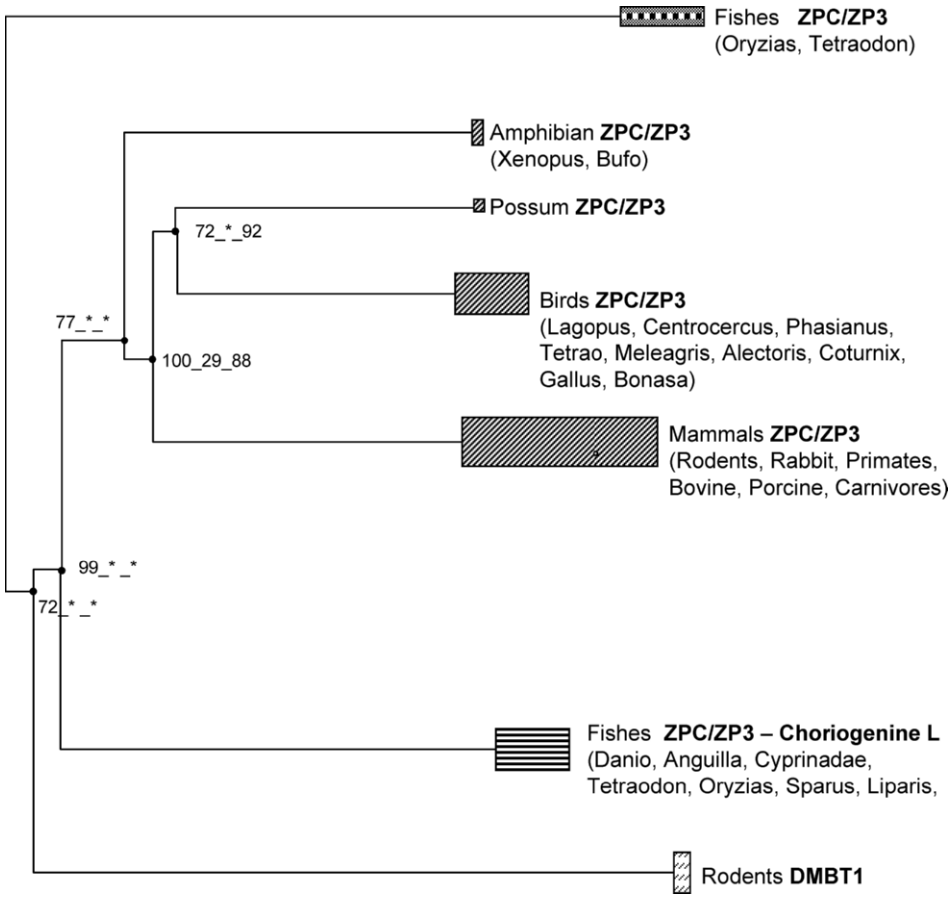

FIG. 2. Phylogenetic tree of the *ZP* gene family obtained when the FIGENIX BLAST is performed from the porcine ZPC/ZP3 sequence. Relative branch lengths indicate rates of evolution along a particular branch. The values at the tree nodes represent bootstrap values.

PLAC1 and Oosp1, and fold autonomously [42]. ZP1 and ZPB possess one divergent copy of the ZP-N domain in their N-termini, whereas ZPA and ZPAX have several (the first one being much more related to the single ZP-N copy of ZP1 and ZPB; Fig. 3).

Verification of Orthology Relationship by Localization of Genes in Syntenic Chromosomal Regions

To further ascertain the identity of *ZPA*, *ZPB*, *ZPC*, and *ZP1* orthologous genes in the rat, mouse, and human, we verified that they are all mapped in syntenic regions in these three species (Table 2). Mapping of the human, mouse, and rat *ZP* genes in the corresponding genomes was performed using both the Ensembl genome browser software (http://www.ensembl.org/; release 47, October 2007) [43] and the BLAT Search software from the UCSC Genome Bioinformatics address (http://genome.ucsc.edu) [44]. In particular, the human chromosome 16 region that contains *ZPA*/*ZP2* is syntenic to the murine chromosome 7 region that contains *Zpa*/*Zp2*. The chromosomal localization of the rat *Zpa*/*Zp2* is not established. However, the *ZPA*/*ZP2* gene on human chromosome 16 is localized close to (approximately 200 kb) the dynein axonemal heavy polypeptide 3 gene (*DNAH3*, GenBank accession no. NM_017539, gene map locus 16p12) [45, 46]. The rat gene orthologous to human *DNAH3* maps to chromosome 1 (gene map locus 1q35), which shows conserved synteny with the region of human chromosome 16, in which both *DNAH3* and *ZPA* are localized. Moreover, the human chromosome 1 region that contains *ZPB*/*ZP4* is syntenic to the rat chromosome 17 region that contains *Zpb*/*Zp4*. The human chromosome 7 region that contains *ZPC*/*ZP3* is syntenic to the murine chromosome 5 region and to the rat chromosome 12 region that contain *Zpc*/*Zp3*. The human chromosome 11 region that contains *ZP1* is syntenic to the murine chromosome 19 region and to the rat chromosome 1 region that contain *Zp1*.

Table 3 summarizes the diversity of ZP genes expressed in mammals, chicken, and *Xenopus*.

IDENTIFICATION OF ZP PSEUDOGENES

The phylogenetic analysis suggested that in several mammals the presence of *ZP1*, *ZPB*, *ZPD*, and/or *ZPAX* gene is lacking. Moreover, a search in EST databases revealed that there are no sequences with a high level of identity with human *ZP1* or chicken *ZPAX* in cattle, pig, dog, and rabbit, and no sequence with a high level of identity with rat *Zpb*/*Zp4* in the mouse. This suggested the presence in the genome of these species of pseudogenes (genes that have evolved by generating stop codon and/or insertion/deletion disrupting the reading frame and resulting in the loss of their protein-coding ability).

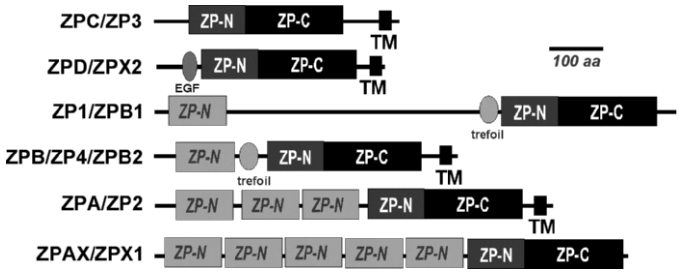

FIG. 3. Schematic drawing of chicken ZP proteins, illustrating domain organizations. TM indicates a transmembrane segment. Note that ZP1 and ZPB/ZP4 possess one divergent copy of the ZP-N domain in their N-termini, whereas ZPA/ZP2 and ZPAX have several.

TABLE 2. Chromosomal localization of the human, rat, and mouse *ZP* genes.

Gene symbol	Chromosome	Start	End
Homo sapiens			
ZPA or *ZP2*	16	21116274 bp	21130369 bp
ZPB or *ZP4*	1	234371751 bp	234379976 bp
ZPC or *ZP3* or *ZP3A*	7	75698923 bp	75716038 bp
ZP1	11	60394561 bp	60399742 bp
Mus musculus			
Zpa or *Zp2*	7	107428985 bp	107441914 bp
Zpc or *Zp3*	5	133425213 bp	133433732 bp
Zp1	19	10013396 bp	10019739 bp
Rattus norvegicus			
Zpa or *Zp2*	Not localized		
Zpb or *Zp4*	17	69587790 bp	69594343 bp
Zp3	12	21792841 bp	21799576 bp
Zp1	1	213568592 bp	213574972 bp

Identification of a Zpb *Pseudogene in the Mouse Genome*

To identify an eventual *Zpb*/*Zp4* pseudogene in the mouse, we submitted the nucleotide and the amino acid sequences of the rat *Zpb*/*Zp4* to the BLASTn and the tBLASTn softwares against the mouse genome. We observed on the mouse genome an alignment between the nucleotide sequence of rat *Zpb* and sequences from the mouse syntenic region (Supplementary Fig. 3, available online at www.biolreprod.org). Interestingly, we observed a microdeletion of 19 nucleotides beginning at 438 147 bp, leading to a frame shift in the open reading frame and the appearance of premature stop codons at 438 187 and 438 220 bp (Fig. 4). This microdeletion is likely responsible for the loss of ZPB protein in the mouse. We also verified that the newly identified mouse *Zpb* pseudogene, located in the region A2 of chromosome 13, mapped in a genome region syntenic to human and rat genome regions carrying corresponding *Zpb*/*Zp4* genes (gene map locus: 1q43 for human and 17q12.1 for rat). Overall, these in silico evidences are supported by recent data using mass spectrometry analysis, which failed to identify mouse ZPB/ZP4 [47]. Numerous peptides from ZP1, ZPA/ZP2, and ZPC/ZP3 were identified, but the authors failed to identify any peptides that could correspond to ZPB/ZP4.

TABLE 3. Classification of *ZP* genes: each column contains orthologous genes within one subfamily, each line contains paralogous genes within one species.[a]

Species	*ZPA*/*ZP2*	*ZPB*/*ZP4*	*ZPC*/*ZP3*	*ZP1*	*ZPAX*	*ZPD*
Human (*Homo sapiens*)	*ZPA = ZP2*	*ZPB = ZP4*	*ZPC = ZP3 = ZP3A*	*ZP1*	*ZPAX* pseudogene	Not found
Chimpanzee (*Pan troglodytes*)	*ZPA = ZP2*	*ZPB = ZP4*	*ZPC = ZP3*	*ZP1*	*ZPAX* pseudogene	Not found
Rhesus macaque (*Macaca mulata*)	*ZP2*	*ZP4*	*ZP3*	*ZP1*	*ZPAX* pseudogene	Not found
Bonnet monkey (*Macaca radiate*)	*ZPA = ZP2*	*ZP1 = ZPB2*	*ZPC = ZP3*	*ZP1*	Not found	Not found
Crab-eating macaque (*Macaca fascicularis*)	*ZPA*	*ZPB*	*ZPC*	Not found	Not found	Not found
White-tufted-ear marmoset (*Callithrix jacchus*)	*ZP2*	*ZP1*		Not found	Not found	Not found
Marmoset (*Callithrix sp.*)			*ZPC = ZP3*	Not found	Not found	Not found
Yellow baboon (*Papio cynocephalus*)		*ZPB*		Not found	Not found	Not found
House mouse (*Mus musculus*)	*Zpa = Zp2*	*Zpb* pseudogene	*Zpc = Zp3*	*Zp1*	Not found	Not found
Rat (*Rattus norvegicus*)	*Zpa = Zp2*	*Zpb = Zp4*	*Zp3*	*Zp1*	Not found	Not found
Pig (*Sus scrofa*)	*ZPA = ZP2 = ZP1*	*ZPB = ZP4 = ZP3alpha*	*ZPC = ZP3 = ZP3beta*	Not found	Not found	Not found
Cow (*Bos taurus*)	*ZPA = ZP2*	*ZPB = ZP4*	*ZPC = ZP3 = ZP3B*	*ZP1* pseudogene	*ZPAX* pseudogene	Not found
Dog *Canis familiaris*)	*ZPA = ZP2*	*ZPB*	*ZPC = ZP3*	*ZP1* pseudogene	Not found	Not found
Cat (*Felis catus*)	*ZPA = ZP2*	*ZPB = ZP4*	*ZPC = ZP3*	Not found	Not found	Not found
Rabbit (*Oryctolagus cuniculus*)	*ZPA = ZP2 = 75kDa*	*ZPB = ZP4 = ZPX*	*ZPC = ZP3*	Not found	Not found	Not found
Chicken (*Gallus gallus*)	*Zpa = Zp2*	*Zpb = Zp4 = Zpb2*	*Zpc = Zp3*	*Zp1 = Zpb1*	*Zpax = Zpx1*	*Zpd = Zpx2*
African clawed frog (*Xenopus laevis*)	*ZPA = ZP2*	*ZPB*	*ZPC = ZP3*	Not found	*ZPAX = ZPX1*	*ZPD = ZPX2*
Western clawed frog (*Xenopus tropicalis*)	*ZPA*	*ZPB*	*ZPC*	Not found	*ZPAX*	*ZPD*

[a] Not found: no pseudogene was found, either because genomic fragment was completely lost, or because the genome was not completely sequenced.

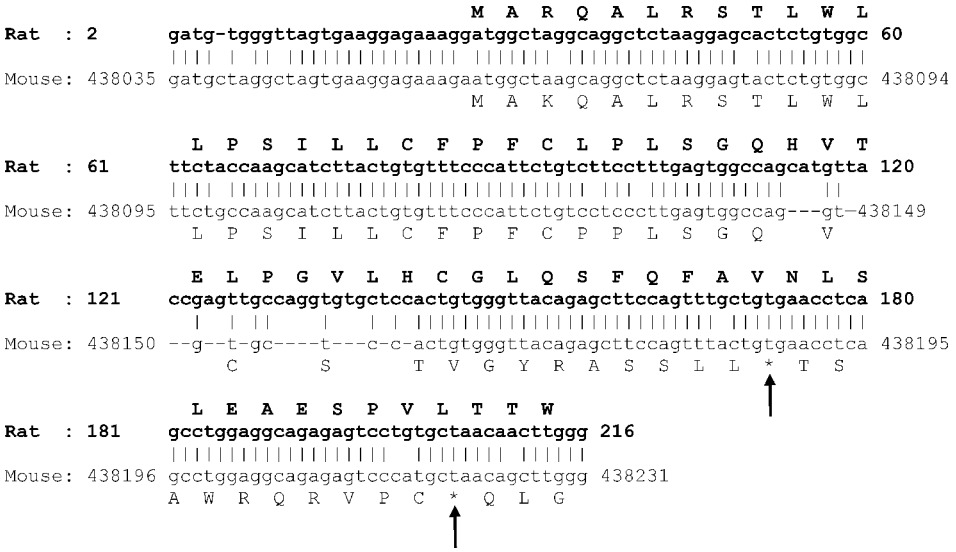

FIG. 4. Alignment of the cDNA sequence and translation of rat *Zpb* (bold) with the mouse genome. We observed on the mouse genome an alignment between the nucleotide sequence of rat *Zpb* and sequences from the mouse syntenic region: nucleotides 438 035 to 438 231 (85% identity), as presented in the figure, 438 474 to 438 565 (92% identity), 440 449 to 440 634 (85% identity), 442 409 to 442 585 (88% identity), 443 070 to 443 152 (88% identity), and 444 748 to 444 891 (86% identity). We observed a microdeletion of 19 nucleotides beginning at 438 147 bp, leading to a frame shift in the open reading frame and the appearance of premature stop codons at 438 187 and 438 220 bp.
*The tga stop codon.

Identification of ZP1 *Pseudogenes in the Dog and Bovine Genome*

By using the same strategy, we searched for a pseudogene of *ZP1* in the dog genome. A tBLASTn analysis with the human ZP1 protein sequence revealed a significant alignment with the exons 2, 3, 5, 7, 8, and 9 (from 42% to 82% similarity). This analysis also revealed the existence of at least three stop codons at positions 151, 279, and 421, replacing, respectively, a proline in exon 3, a glutamine in exon 5, and a tyrosine in exon 8 of the human ZP1 protein sequence (Fig. 5). The alignment also revealed the existence of microinsertions/deletions in the same exons, but it was not possible with a BLASTn analysis to describe them precisely. We also verified that the *ZP1* dog

FIG. 5. Alignment of human ZP1 protein sequence (upper sequence) with the dog genome by using a tBLASTn analysis. Conservative substitutions (substitutions of one amino acid with another with generally similar properties: size, hydrophobicity, etc.) are indicated with +. Arrows indicate the presence of stop codons. Three stop codons are present in the dog genomic sequence at positions 151, 279, and 421, replacing, respectively, a proline in exon 3, a glutamine in exon 5, and a tyrosine in exon 8 of the human ZP1 protein sequence. Lines and dotted lines represent microinsertions and microdeletions, respectively. Two alternate stop codons are represented in exon 3, likely due to the frame shift caused by insertions/deletions.

Exon 5

```
Chicken  190   QGITFSAVRSSTFYKQQWMILMVDTAVACPV
               QGITF+AVRSSTFYK  WMILMVDTA+ACPV
Human          QGITFTAVRSSTFYK*CWMILMVDTAMACPV
```

FIG. 6. Alignment of chicken ZPAX protein sequence (upper sequence) with the human genome using a tBLASTn analysis. We observed an alignment on nucleotides 1 661 981 to 1 662 175 (43% identity), 1 663 855 to 1 664 187 (35% identity), 1 665 125 to 1 665 235 (56% identity), 1 665 755 to 1 665 979 (50% identity), 1 667 308 to 1 667 628 (48% identity, presented in the figure), and a stop codon in the exon 5, as presented in the figure. *Presence of the stop codon.

pseudogene is located on chromosome 18 in the dog genome, near *GPR44* and *PRPF19* canine genes, in a region corresponding to a syntenic region of chromosome 11 in human genome, near the GPR44 and PRPF19 human counterpart genes.

A similar work allowed the identification of a *ZP1* pseudogene on chromosome 29 in the bovine genome, with microinsertions and a stop codon (Supplementary Fig. 3).

Identification of ZPAX *Pseudogenes in the Human, Chimpanzee, Rhesus Macaque, and Bovine Genome*

A tBLASTn analysis with the chicken ZPAX protein sequence on the human genome showed an alignment between the nucleotide sequence of chicken *ZPAX* and sequences from the human syntenic region (Supplementary Fig. 3). We observed a stop codon in the exon 5 (Fig. 6) and an insertion of six nucleotides and a stop codon in exon 3. We also verified that the *ZPAX* human pseudogene is located on chromosome 2p24.2 in the human genome, between the *MSGN1* and *KCNS3* human genes and in a region corresponding to a syntenic region of chromosome 3 in the chicken genome, near the *MSGN1* and *KCNS3* chicken counterpart genes.

A similar work allowed the identification of a *ZPAX* pseudogene located on chromosome 2A in the chimpanzee genome, on chromosome 13 in the Rhesus macaque (*Macaca mulata*) genome, and on chromosome 11 in the bovine genome, between the *MSGN1* and *KCNS3* genes (Supplementary Fig. 3). An alignment of the human *ZPAX* pseudogene and the chimpanzee *ZPAX* pseudogene showed the great homology between these two pseudogenes (88% identity). Moreover, the mutations observed in these two pseudogenes have similar localizations.

We did not find any traces of *ZPD* pseudogenes in any mammalian genomes. In the chicken genome, the ZPD gene is located on chromosome 11. There is no predicted coding sequence in the syntenic region of all mammalian genomes examined, suggesting that the *ZPD* gene has completely disappeared in these species.

A summary of the pseudogenes that have been identified is presented in Table 3. Note that for those that have not been found, either they are present but not yet revealed because the genome is not completely sequenced, or the genomic fragment has been completely lost. The complete sequencing of other vertebrate genomes will definitively elucidate which species have lost *ZP1*, *ZPAX*, and *ZPD* genes during evolution.

ZP GENE EVOLUTION: DEATH OF GENES FROM LOWER VERTEBRATES TO MAMMALS

The phylogenetic analysis allowed us to classify *ZP* genes in six subfamilies: the *ZPA*/*ZP2* subfamily, the *ZPB*/*ZP4* subfamily, the *ZPC*/*ZP3* subfamily, the *ZP1* subfamily, the *ZPAX* subfamily, and the *ZPD* subfamily. The mammalian genome contains three to four *ZP* genes. The *Xenopus* genome contains five *ZP* genes. The chicken genome contains six *ZP* genes. In fish, there are at least seven genes encoding ZP proteins, ZPC being in particular highly duplicated in Medaka and Zebrafish. The identification of pseudogenes showed that *ZP* gene evolution mainly occurs by death of genes.

Since the chicken genome contains a *ZP1* gene, the duplication of the ancestor of this gene can be dated before the divergence between birds and mammals (300Mya). After this divergence, *ZPAX* and *ZPD* genes seem to have disappeared in mammals, not in birds. In particular, the death of

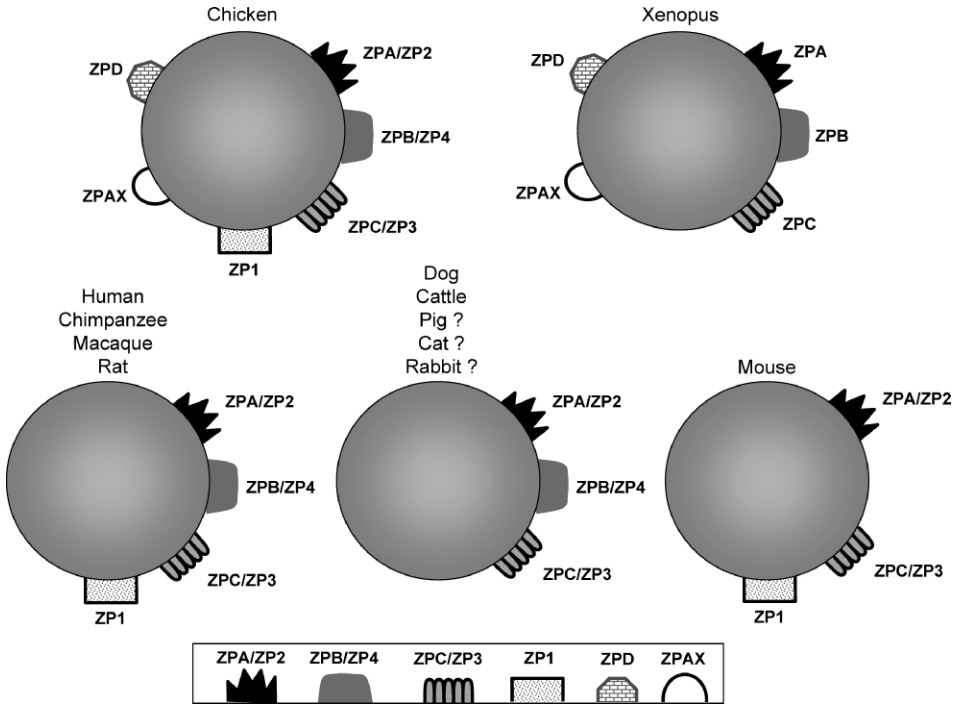

FIG. 7. Composition of the ZP in vertebrates. In chicken, the zona is composed of six ZP proteins (ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, ZP1, ZPAX, ZPD). In *Xenopus*, the zona is composed of five ZP proteins (ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, ZPAX, ZPD). In human, chimpanzee, macaque, and rat, the zona is composed of four ZP proteins (ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, ZP1). In dog, cattle and, putatively, pig, cat, and rabbit, the zona is composed of three ZP proteins (ZPA/ZP2, ZPB/ZP4, ZPC/ZP3). In the mouse, the zona is composed of three ZP proteins (ZPA/ZP2, ZPC/ZP3, ZP1).

the *ZPAX* gene occurred before the divergence between humans and monkeys, since similar mutations were observed in human and chimpanzee *ZPAX* pseudogenes. In mammals, only primates and rodents contain a *ZP1* gene, whereas we observed a *ZP1* pseudogene in cows and dogs, which suggests that the death of the *ZP1* gene in those species occurred after the divergence between primate and rodent groups, and other mammals. Finally, we observed a *ZPB* pseudogene in the mouse, which shows that the death of this *ZPB* gene occurred after the divergence between the mouse and the rat. The persistence of both *ZPA*/*ZP2* and *ZPC*/*ZP3* genes across the vertebrates indicates that both genes have functional importance.

Overall, it seems that there is a sort of discontinuity in the loss of the ZP genes along the phylogeny, suggesting that this loss of ZP genes along the evolution of mammals was independent between species. The significance of this loss of *ZP* genes in mammals compared with other vertebrates remains obscure. It can be argued that a higher number of *ZP* genes are necessary for species for which fertilization is external to the female genital tract, whereas a lower number are necessary for species for which fertilization is internal. However, since fertilization is internal in chickens, this would not be the good or the sole explanation.

SPERM-ZP INTERACTION

The amino acid sequences of the ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, and ZP1 families present a high degree of identity between species. This implies a structural similarity in the ZP structures of these species and may suggest some similarities in the mechanisms of sperm-zona interaction. However, this review raises new insights into the mechanisms of sperm-ZP interaction.

Our analysis revealed that mammals can be classified in three categories (Fig. 7). In the first category, composed of human, chimpanzee, macaque, and rat, the ZP is composed of four ZP proteins: ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, and ZP1. In the second category, composed of dog, cattle and, putatively, pig, cat, and rabbit, the ZP does not contain ZP1 protein. In the third category, composed of only the mouse, ZP does contain ZP1 but not ZPB/ZP4. So all ZPs in mammals share ZPA/ZP2 and ZPC/ZP3 proteins, and one or both of the ZP1 and ZPB/ZP4 proteins. Interestingly, these two latter genes share an ancestral gene, suggesting a closest similarity between the two corresponding proteins than with ZPA/ZP2 and ZPC/ZP3. Overall, this suggests that the mechanism of sperm-zona interaction requires the presence of both ZPA/ZP2 and ZPC/ZP3 proteins, and one or both of the ZP1 and ZPB/ZP4 proteins.

Current models for the structure of mammalian ZP are based upon the existence of three ZP proteins. In particular, data obtained in the mouse raise the hypothesis that the mammalian ZP has three proteins, of which ZPC/ZP3 is the primary sperm receptor [48]. Since four *ZP* genes are expressed in the human, chimpanzee, macaque, and rat oocyte, a re-evaluation would be required, both on the structure of the ZP and, potentially, on the mechanisms of sperm-ZP interaction. Moreover, in the pig, sperm membranes present a high affinity for ZPB-ZPC heterocomplexes but not free ZPB or ZPC glycoprotein subunits [49], whereas in the mouse, ZPC/ZP3 is the primary sperm receptor. This result was puzzling, since porcine ZPB/ZP4 and ZPC/ZP3 were thought to be orthologous to mouse ZP1 and ZPC/ZP3. Here, we show that the ZPB/ZP4 protein has been lost in the mouse species. This would explain that spermatozoa only interact with ZPC/ZP3 in this species. This also suggests that different mechanisms of sperm-egg interaction have evolved in mammals. Knockout experiments in which murine ZP proteins were replaced by human equivalents [50, 51] showed that although mouse sperm is able to bind mouse oocytes engineered to express human ZPA/ZP2 and ZPC/ZP3 proteins, human sperm did not bind. These findings suggest either that the species-specific recognition of the ZP is mediated by the mouse-specific glycosylation patterns, regardless of the amino acid sequence, or that human sperm has evolved to interact with a ZP composed of four, not three, proteins. The role of carbohydrate moieties remains to be clarified. In fact, the amino acid sequence could be sufficient for sperm recognition in some mammals, in which ZP proteins expressed in *Escherichia coli*, a species incapable of synthesizing glycoproteins, bind to homologous spermatozoa and stimulate acrosomal exocytosis. This is the case in the macaque [52, 53] and the bovine [54]. In humans, contradictory results have been observed [55, 56]. Moreover, if human sperm interacts with a ZP composed of four proteins, it may interact with rat ZP. However, sperm binding assays to ovulated eggs indicate that rat ZPA/ZP2, ZPB/ZP4, ZPC/ZP3, and ZP1 are not sufficient to support human sperm binding [57], whereas mouse and rat sperm bind to mouse ZP (composed of three proteins) and rat ZP (composed of four proteins) [57]. Moreover, mouse and rat sperm bind to zonae composed solely of mouse ZPA/ZP2 and ZPC/ZP3 [58, 59]. Thus, whether the number of glycoproteins in the ZP affects taxon-specific sperm-egg recognition remains to be determined.

In conclusion, the molecular basis of sperm binding to the ZP, an essential first step in egg fertilization, remains an enigma. Our review clarifies the classification and evolution of the ZP gene family and provides new directions to investigate sperm-egg recognition in mammals.

ACKNOWLEDGMENTS

We wish to thank Pierre Pontarotti and Philippe Gouret for their help in the phylogenetic study and for helpful discussion, and Lauranne Paillard-Laurence for technical assistance.

REFERENCES

1. Jovine L, Darie CC, Litscher ES, Wassarman PM. Zona pellucida domain proteins. Annu Rev Biochem 2005; 74:83–114.
2. Bleil JD, Wassarman PM. Structure and function of the zona pellucida: identification and characterization of the proteins of the mouse oocyte's zona pellucida. Dev Biol 1980; 76:185–202.
3. Sacco AG, Yurewicz EC, Subramanian MG, DeMayo FJ. Zona pellucida composition: species cross reactivity and contraceptive potential of antiserum to a purified pig zona antigen (PPZA). Biol Reprod 1981; 25:997–1008.
4. Harris JD, Hibler DW, Fontenot GK, Hsu KT, Yurewicz EC, Sacco AG. Cloning and characterization of zona pellucida genes and cDNAs from a variety of mammalian species: the ZPA, ZPB, and ZPC gene families. DNA Seq 1994; 4:361–393.
5. Bausek N, Waclawek M, Schneider WJ, Wohlrab F. The major chicken egg envelope protein ZP1 is different from ZPB and is synthesized in the liver. J Biol Chem 2000; 275:28866–28872.
6. Wang H, Gong Z. Characterization of two zebrafish cDNA clones encoding egg envelope proteins ZP2 and ZP3. Biochim Biophys Acta 1999; 1446:156–160.
7. Menino AR Jr, Wright RW Jr. Characterization of porcine oocyte zonae pellucidae by polyacrylamide gel electrophoresis. Proc Soc Exp Biol Med 1979; 160:449–452.
8. Dunbar BS, Raynor BD. Characterization of porcine zona pellucida antigens. Biol Reprod 1980; 22:941–954.
9. Dunbar BS, Liu C, Sammons DW. Identification of the three major proteins of porcine and rabbit zonae pellucidae by high resolution two-dimensional gel electrophoresis: comparison with serum, follicular fluid, and ovarian cell proteins. Biol Reprod 1981; 24:1111–1124.
10. Subramanian MG, Yurewicz EC, Sacco AG. Specific radioimmunoassay for the detection of a purified porcine zona pellucida antigen (PPZA). Biol Reprod 1981; 24:933–943.

11. Dunbar BS, Dudkiewicz AB, Bundman DS. Proteolysis of specific porcine zona pellucida glycoproteins by boar acrosin. Biol Reprod 1985; 32:619–630.
12. Wardrip NJ, Hedrick JL. Pig zona pellucida 25K and 65K glycoproteins are derived from hydrolysis and reduction of the 90K family. J Cell Biol 1985; 101:378a.
13. Yurewicz EC, Sacco AG, Subramanian MG. Structural characterization of the Mr=55,000 antigen (ZP3) of porcine oocyte zona pellucida. J Biol Chem 1987; 262:564–571.
14. Hughes DC, Barratt CL. Identification of the true human orthologue of the mouse Zp1 gene. Evidence for greater complexity in the mammalian zona pellucida? Biochim Biophys Acta 1999; 1447:303–306.
15. Spargo SC, Hope RM. Evolution and nomenclature of the zona pellucida gene family. Biol Reprod 2003; 68:358–362.
16. Smith J, Paton IR, Hughes DC, Burt DW. Isolation and mapping the chicken zona pellucida genes: an insight into the evolution of orthologous genes in different species. Mol Reprod Dev 2005; 70:133–145.
17. Conner SJ, Lefièvre L, Hughes DC, Barratt CLR. Cracking the egg: increased complexity in the zona pellucida. Hum Reprod 2005; 20: 1148–1152.
18. Gouret P, Vitiello V, Balandraud N, Gilles A, Pontarotti P, Danchin EGJ. FIGENIX: intelligent automation of genomic annotation: expertise integration in a new software platform. BMC Bioinformatics 2005; 6:198.
19. Saitou N, Nei M. The neighbour-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4:406–425.
20. Fitch WM. Toward defining the course of evolution: minimum change for a specific tree topology. Sys Zool 1971; 20:406–416.
21. Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 1981; 17:368–376.
22. Edgar RC, Robert C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004; 32:1792–1797.
23. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004; 5:113.
24. Conner SJ, Hughes DC. Analysis of fish ZP1/ZPB homologous genes—evidence for both genome duplication and species-specific amplification models of evolution. Reproduction 2003; 126:347–352.
25. Leong CTC, Ng CY, Ng CP, Ma ZS, Nguyen TH, Tay SK, Huynh H. Molecular cloning, characterization and isolation of novel spliced variants of the human ortholog of a rat estrogen-regulated membrane-associated protein, UO-44. Oncogene 2004; 23:5707–5718.
26. Kasik JW. A cDNA cloned from pregnant mouse uterus exhibits temporo-spatial expression and predicts a novel protein. Biochem J 1998; 330:947–950.
27. Chen D, Xu X, Zhu LJ, Angervo M, Li Q, Bagchi MK, Bagchi IC. Cloning and uterus/oviduct-specific expression of a novel estrogen-regulated gene (ERG1). J Biol Chem 1999; 274:32215–32224.
28. Bisgaard HC, Holmskov U, Santoni-Rugiu E, Nagy P, Nielsen O, Ott P, Hage E, Dalhoff K, Rasmussen LJ, Tygstrup N. Heterogeneity of ductural reactions in adult rat and human liver revealed by novel expression of Deleted in Malignant Brain Tumor 1. Am J Pathol 2002; 161:1187–1198.
29. Matsushita F, Miyawaki A, Mikoshiba K. Vomeroglandin/CRP-Ductin is strongly expressed in the glands associated with the mouse vomeronasal organ: identification and characterization of mouse vomeroglandin. Biochem Biophys Res Commun 2000; 268:275–281.
30. Töpfer-Petersen E, Calvete JJ. Sperm-associated protein candidates for primary zona pellucida-binding molecules: structure-function correlations of boar spermadhesins. J Reprod Fertil Suppl 1996; 50:55–61.
31. Monné M, Han L, Jovine L. Tracking down the ZP domain: from the mammalian zona pellucida to the molluscan vitelline envelope. Semin Reprod Med 2006; 24:204–216.
32. Tynan S, Pacia E, Haynes-Johnson D, Lawrence D, D'Andrea MR, Guo JZ, Lundeen S, Allan G. The putative tumor suppressor Deleted in Malignant Brain Tumors 1 is an estrogen-regulated gene in rodent and primate endometrial epithelium. Endocrinology 2005; 146:1066–1073.
33. Gupta SK, Sharma M, Behera AK, Bisht R, Kaul R. Sequence of complementary deoxyribonucleic acid encoding bonnet monkey (Macaca radiata) zona pellucida glycoprotein-ZP1 and its high-level expression in Escherichia coli. Biol Reprod 1997; 57:532–538.
34. Kerr LE, Wilson MR, Aitken RJ. Molecular characterization of zona pellucida 1 (ZP1) in the marmoset monkey, *Callithrix jacchus*. J Reprod Fertil 1996; Abstr Ser 18:29–30.
35. Taya T, Yamasaki N, Tsubamoto H, Hasegawa A, Koyama K. Cloning of a cDNA coding for porcine zona pellucida glycoprotein ZP1 and its genomic organization. Biochem Biophys Res Commun 1995; 207:790–799.
36. Hatanaka Y, Nagai T, Tobita T, Nakano M. Changes in the properties and composition of zona pellucida of pigs during fertilization in vitro. J Reprod Fertil 1992; 95:431–440.
37. Kudo K, Yonezawa N, Katsumata T, Aoki H, Nakano M. Localization of carbohydrate chains of pig sperm ligand in the glycoprotein ZPB of egg zona pellucida. Eur J Biochem 1998; 252:492–499.
38. Epifano O, Liang LF, Dean J. Mouse *Zp1* encodes a zona pellucida protein homologous to egg envelope proteins in mammals and fish. J Biol Chem 1995; 270:27254–27258.
39. Hughes DC, Barratt CLR. Identification of the true human orthologue of the mouse Zp1 gene: evidence for greater complexity in the mammalian zona pellucida? Biochim Biophys Acta 1999; 1447:303–306.
40. Smith J, Paton IR, Hughes DC, Burt DW. Isolation and mapping the chicken zona pellucida genes: an insight into the evolution of orthologous genes in different species. Mol Reprod Dev 2005; 70:133–145.
41. Callebaut I, Mornon JP, Monget P. Isolated ZP-N domains constitute the N-terminal extensions of zona pellucida proteins. Bioinformatics 2007; 23: 1871–1874.
42. Jovine L, Janssen WG, Litscher ES, Wassarman PM. The PLAC1-homology region of the ZP domain is sufficient for protein polymerisation. BMC Biochem 2006; 7:11.
43. Hubbard TJP, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T, Down T, Dyer SC, et al.Nucleic Acids Res 2007; 35(Database issue):D610–D617.
44. Kent WJ. BLAT—The BLAST-like alignment tool. Genome Res 2002; 4: 656–664.
45. Maiti AK, Mattei MG, Jorissen M, Volz A, Zeigler A, Bouvagnet P. Identification, tissue specific expression, and chromosomal localisation of several human dynein heavy chain genes. Eur J Hum Genet 2000; 8: 923–932.
46. Chapelin C, Duriez B, Magnino F, Goossens M, Escudier E, Amselem S. Isolation of several human axonemal dynein heavy chain genes: genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment. FEBS Lett 1997; 412:325–330.
47. Boja ES, Hoodbhoy T, Fales HM, Dean J. Structural characterization of native mouse zona pellucida proteins using mass spectrometry. J Biol Chem 2003; 278:34189–34202.
48. Wassarman PM. Mammalian fertilization: egg and sperm (glycol)proteins that support gamete adhesion. Am J Reprod Immunol 1995; 33:253–258.
49. Yurewicz EC, Sacco AG, Gupta SK, Xu N, Gage DA. Hetero-oligomerization dependent binding of pig oocyte zona pellucida glycoproteins ZPB and ZPC to boar sperm membrane vesicles. J Biol Chem 1998; 273:7488–7494.
50. Rankin TL, Tong ZB, Castle PE, Lee E, Gore-Langton R, Nelson LM, Dean J. Human ZP3 restores fertility in Zp3 null mice without affecting order-specific sperm binding. Development 1998; 125:2415–2424.
51. Rankin TL, Coleman JS, Epifano O, Hoodbhoy T, Turner SG, Castle PE, Lee E, Gore-Langton R, Dean J. Fertility and taxon-specific sperm binding persist after replacement of mouse sperm receptors with human homologs. Dev Cell 2003; 5:33–43.
52. Govind CK, Gahlay GK, Choudhury S, Gupta SK. Purified and refolded recombinant Bonnet Monkey (*Macaca radiata*) zona pellucida glycoprotein-B expressed in *Escherichia coli* binds to spermatozoa. Biol Reprod 2001; 64:1147–1152.
53. Gahlay GK, Srivastava N, Govind CK, Gupta SK. Primate recombinant zona pellucida proteins expressed in *Escherichia coli* bind to spermatozoa. J Reprod Immunol 2002; 53:67–77.
54. Hinsch E, Aires VA, Hedrich F, Oehninger S, Hinsch KD. A synthetic decapeptide from a conserved ZP3 protein domain induces the G protein-regulated acrosome reaction in bovine spermatozoa. Theriogenology 2005; 63:1682–1694.
55. Chapman NR, Kessopoulou E, Andrews PD, Hornby DP, Barratt CLR. The polypeptide backbone of recombinant human zona pellucida glycoprotein-3 initiates acrosomal exocytosis in human spermatozoa in vitro. Biochem J 1998; 330:839–845.
56. Chakravarty S, Suraj K, Gupta SK. Baculovirus-expressed recombinant human zona pellucida glycoprotein-B induces acrosomal exocytosis in capacitated spermatozoa in addition to zona pellucida glycoprotein-C. Mol Hum Reprod 2005; 11:365–372.
57. Hoodbhoy T, Joshi S, Boja ES, Williams SA, Stanley P, Dean J. Human sperm do not bind to rat zonae pellucidae despite the presence of four homologous glycoproteins. J Biol Chem 2005; 280:12721–12731.
58. Rankin T, Talbot P, Lee E, Dean J. Abnormal zonae pellucidae in mice lacking ZP1 result in early embryonic loss. Development 1999; 126: 3847–3855.
59. Boja ES, Hoodbhoy T, Garfield M, Fales HM. Structural conservation of mouse and rat zona pellucida glycoproteins. Probing the native rat zona pellucida proteome by mass spectrometry. Biochemistry 2005; 44: 16445–16460.