Gamete interaction protein evolution evinces Pantherines hybridizing history

Francisco Pisciottano[1], Clara Campos[1†], Clementina Penna[1†], Carlos David Bruque[2], Toni Gabaldón[3], Patricia Saragüeta[1*]

[1]Laboratorio de Regulación Hormonal del Tracto Reproductivo Femenino, Instituto de Biología y Medicina Experimental (IByME-CONICET), Buenos Aires, Argentina

[2]Unidad de Conocimiento Traslacional Hospitalaria Patagónica, Hospital de Alta Complejidad El Calafate SAMIC, Santa Cruz, Argentina

[3] Life Sciences Department, Barcelona Supercomputing Center (BSC), Mechanisms of Disease Department, Institute for Research in Biomedicine (IRB), and Universitat Pompeu Fabra (UPF), 08003, Barcelona, Spain

*** Correspondence:** patriciasaragueta2@gmail.com

Keywords: Jaguar, ZP, evolution, Felidae, inbreeding, pantherines.

Abstract

The jaguar (*Panthera onca*) is the most relevant predator in South America and an endangered species in Argentina. Therefore genomic, evolutionary and functional studies of this species are strategic to gain biological insights and support conservation efforts. With the advent of sequencing advances the study of relevant character species evolution has to be revisited. Recent comparative genomic analysis revealed a long history of hybridization among extant big cats, which indicates an absence of inter-specific isolation barriers. We aimed to uncover the molecular basis of this inter-species crossing and the evolution of the functional capacities of pantherine gamete interaction Zona Pellucida proteins (ZP) and JUNO. Hi-C long-range sequencing and chromosome-resolved jaguar genome assembly was used to resolve ZP genes loci and reconstruct the evolutionary history of these proteins. Exploring sequence variability in more that 30 Carnivora species, Feliforme and Caniforme sister groups show significant differences in average sequence identity for all studied proteins except ZP4, always showing Caniformes as a more diverse group than Feliformes. Molecular evolution of ZP isoforms involved in sperm-recognition, ZP3 and ZP2, showed similar phylogenetic histories, displaying signatures of positive selection among Carnivores and Caniformes, but not among Pantherines or Feliformes. On the other hand, structural cross-linking ZP subunits, ZP1 and ZP4, displayed a different evolutionary history. The fusion protein IZUMO1R was found to be highly conserved in all studied phylogenetic groups. These results strengthen the idea that ZPs are proteins more related to specific-specificity isolation than IZUMO1R, which is specialized in fusion events. Altogether, our findings indicate that sperm-oocyte interaction and fusion proteins lack the degree of diversification necessary to establish a prezygotic reproductive isolation barrier between pantherines, which would explain observed reticulated patterns of evolution in this clade.

1. Introduction

The zona pellucida is the extracellular matrix synthesized in the ovaries of mammals during follicular development. It is found surrounding the oocyte and early embryo plasma membrane and constitutes a three-dimensional structure composed of ZP glycoprotein filaments. Oocyte formation, fertilization and early embryo development are biological processes involving ZP proteins (Yanagimachi, 1994; Dean, 2007; Wassarman & Litscher, 2009; Gupta and Bhandari, 2011; Gupta et al., 2012; Tanihara et al., 2013; Shu et al., 2015; Moros-Nicolás et al., 2018a; Moros-Nicolás et al., 2021).

The zona pellucida of most, but not all, eutherians is composed of four ZP protein subunits: ZP1-4.

ZP3 was the first glycoprotein of the zona pellucida to be considered indispensable for fertilization in mammals when two groups independently discovered that mouse oocytes with a disrupted *Zp3* gene lack zona pellucida and are infertile (Liu et al., 1996; Rankin et al., 1996). Therefore, ZP3 was, for some time, regarded as the key and main sperm-binding protein.

Further research revealed that ZP2 is also a protein essential for mouse fertility as it is required for sperm binding (Rankin et al., 2001). Capacitated sperm attach to an N-terminal domain of ZP2 in the extracellular zona pellucida surrounding eggs (Baibakov et al., 2012; Avella et al., 2014). After fertilization, ovastacin, an oocyte-specific metalloendoproteases, is exocytosed from egg cortical granules and cleaves extracellular ZP2 (Burkart et al., 2012). The cleavage site for this event is located in the ZP-N2 domain of ZP2. Postfertilization proteolytic cleavage at the N terminus of ZP2 protein prevents sperm binding to the zona pellucida and thus provides a block to polyspermy (Baibakov et al., 2012).

ZP1 is a ZP protein not directly related to gamete interaction, but its function is no less relevant for the formation of the zona pellucida. ZP1 is involved in ZP protein filaments cross-linking (Monné & Jovine, 2011; Nishimura et al., 2019; Stsiapanava et al., 2020), thus giving and maintaining the structure of such extracellular matrix.

ZP4 was the last of the eutherian ZP proteins to be identified (Lefièvre et al., 2004), as it is not present in the house mouse (*Mus musculus*), the traditional mammalian model in which most studies of the zona pellucida had been performed. ZP4 is considered to be a paralog of ZP1 and the origin of these two modern genes can be traced back to a duplication of a common ancestral gene in the early evolution of amniotes (Goudet et al., 2008). Similar to ZP1, ZP4 plays a structural and mechanical role that is not essential during fertilization *in vivo* but fundamental for the protection of the developing embryo (Lamas-Toranzo et al., 2019).

The study of ZP proteins family evolutionary history has uncovered many events of gene gains and losses. The duplication that gave rise to *Zp1* and *Zp4* occurred during the early evolution of the amniotes (Hughes and Barratt, 1999; Bausek et al., 2000; Goudet et al., 2008) and multiple events of pseudogenization resulted in modern placental mammals whose zona pellucida structure is only composed of three subunits instead of four. *Zp1* was lost in some species of the Caniformia suborder, such as dog (*Canis familiaris*), dingo (*Canis lupus dingo*) and fox (*Vulpes vulpes*) (Goudet et al.,

2008; Moros-Nicolás et al., 2018a). *Zp1* gene was also lost among some Pinnipeds (Moros-Nicolás et al., 2018a) and at least one more pseudogenization of *Zp1* took place in the evolution of Cetartiodactyla as many even-toed ungulates and cetaceans lack a functional version of this gene (Goudet et al., 2008; Stetson et al., 2012). Other events of *Zp1* gene pseudogenization are reported in small primates lineages, such as marmosets and tarsiers (Stetson et al., 2012). On the other hand, pseudogenization of *Zp4* gene was described in the house mouse (*Mus musculus*) (Lefièvre et al., 2004) and other members of the *Mus* genus (Izquierdo-Rico et al., 2021). However, ZP3 and ZP2 proteins have not been reported as lacking in any mammalian species. This indicates that their function is essential and they both have been reported as necessary for gamete interaction and fertilization (Liu et al., 1996; Rankin et al., 1996, 2001).

The structure of ZP proteins is characterized by different domains crucial to the secretion and assembly of ZP subunits (Monné et al., 2006). These are an N-terminal secretory signal peptide (SP) that identifies secreted proteins; the ZP domain, a conserved sequence of ~260 amino acids including 8 or 10 invariant cysteine residues forming two alternative disulfide bonds; a recognition site for members of the proprotein convertase family of proteolytic enzymes, a consensus furin cleavage site (CFCS); and a C-terminal propeptide that includes a single-spanning transmembrane (TM) domain (Jovine et al., 2005; Monné et al., 2006; Wilburn & Swanson, 2017).

The mammalian fertilization process has been extensively described by both physiology and molecular biology. ZP and Juno oocyte proteins play a key role in this process, facilitating the interaction and fusion between the oocyte and the sperm respectively. Their interactions with sperm counterparts may give rise to a prezygotic reproductive isolation mechanism. Owing to protein specificity of these interactions, they are involved in the species-specific barrier (Florman & Ducibella, 2006).

These proteins are representative components involved in reproduction and are usually found evolving under positive selection and this favored rapid evolution plays an important role in reproductive isolation of diverging taxa (Swanson et al., 2001; Swanson & Vacquier, 2002; Turner & Hoekstra, 2008). Also, traits of coevolution between sperm and egg protein counterparts were reported to conduct reproductive isolation (Swanson et al., 2003; Aagaard et al., 2006). The counterparts of sperm interactive couples are not well established in mammals; therefore, in this work we will specifically focus on egg coat protein positive selection marks in order to look into their role in speciation.

Our non-model species of study -the jaguar (*Panthera onca*)- is the most important predator in the South American bioma and in danger of extinction in Argentina (Paviolo et al., 2016). Hence, it is crucial to develop genomic, evolutionary and functional studies. Genomic introgression analysis carried out alongside with the first release of the jaguar complete genome exposed a complex pattern of historical inter-specific hybridization in the Pantherinae lineage (Figueiró et al., 2017). These findings revealed high rates of fertilization among pantherines, particularly between *Panthera leo* and *Panthera onca*, a fact that would indicate an absence of inter-specific isolation barriers. We aimed to

uncover the underlying molecular basis of this inter-species crossing and evolution of the functional capacities of pantherine ZP and JUNO proteins.

We hypothesize that sperm-oocyte interaction and fusion proteins are particularly conserved or accumulated few diversifying changes along the Pantherinae lineage, preventing prezygotic reproductive isolation.

Previous results that support our hypothesis are a), pervasive genealogical discordance across Panthera genomes, caused by both incomplete lineage sorting and complex patterns of historical interspecific hybridization (Li et al., 2016; Figueiró et al., 2017); b), natural hybrids between all panther species (Yadav et al., 2019); c), the high efficiency to obtain hybrids by IVF of oocytes with whole Zona Pellucida between different felines (Duque Rodriguez et al., 2021); and d), ZP genes with remarkably fast evolution that could make them good candidates to contribute to prezygotic species isolation (Turner & Hoekstra, 2008., 2008, Meslin et al., 2012).

This work analyzes the genetic basis of the lack of prezygotic isolation among pantherine species through an evolutionary analysis of ZP1, ZP2, ZP3, ZP4 and IZUMO1R proteins in Carnivores, Feliformes, Caniformes and Pantherines.

2. Materials and Methods

2.1 Sequence retrieval and alignment

Jaguar sequences were obtained from the first version of the jaguar reference genome (Figueiró et al, 2017) and/or the HiC assembled jaguar genome (DNA zoo) by performing a local BLASTn using leopard (*Panthera pardus*) ortholog sequence as query, obtained from GenBank (NCBI Resource Coordinators, 2016). Other mammalian species sequences were obtained from Ensembl (Howe et al., 2021) and GenBank Nucleotide database (Supplementary tables 9, 10, 11, 12, 13). Visualization and multiple sequence alignments were performed with MEGA 5.2 (Kumar et al., 2018). Clustal and/or Muscle, as implemented in MEGA 5.2, were used for sequence alignments. A careful manual curation of the alignments was performed in order to maximize quality of aligned sequences. Low quality sequences that could not be replaced or significantly improved were discarded. Low quality or non-aligning portions of coding sequences in one species were deleted and, whenever possible, replaced using additional sequence sources such as Nscan and GeneScan, accessed through UCSC Genome Browser in order to guarantee homology of the aligned sequences.

2.2 Positive selection analysis

In order to detect signatures of molecular adaptation we used the modified Model A branch-site test 2 of positive selection (Zhang et al., 2005) applied in codeml program from the PAML4 package (Yang, 2007). This model has 2 variants: the focal branch variant and the subtree variant. In each case this model is able to detect signatures of positive selection in preassigned specific branches or

clades respectively, and in both cases these changes can be assigned to individual codon positions. We applied both models for each of the tested groups in this work: Carnivora, Feliformia, Caniformia, Pantherinae, *Panthera onca*. Test 2 of the Model A of codeml consists of two nested hypotheses that must be compared: in the alternative hypothesis, positive selection is allowed only in the selected branch or clade, but not in the background branches or clades; and in the null hypothesis no positive selection is allowed in any branch or clade of the tree. The heterogeneity among protein sites was handled by site classes with different omega values (ω=dN/dS). In this model three omegas were used: $\omega_0<1$, $\omega_1=1$ (fixed) and $\omega_2>1$. In the alternative hypotheses site classes with ω_2 were enabled in the foreground branch or clade allowing the presence of positive selection, but in the null hypotheses only site classes with $\omega 0$ and $\omega 1$ were enabled, restricting sites to purifying selection or neutral evolution only. The fit of the gene data (sequence alignment and tree) to these hypotheses was obtained by maximum likelihood (ML) and evaluated by a likelihood ratio test (LRT) in which a statistic calculated from twice the difference of the ML obtained was compared to a chi-square distribution (1 degree of freedom) to infer the significance between hypotheses.

Given that this test for positive selection detection is a very parameter-rich model, we implemented it by following an ad-hoc methodology designed to prevent local minimum solutions and maximize rerun results consistency. Instead of running all parameters optimization while evaluating the selection model, we pre-computed all branch lengths under the Goldman & Yang codon-based evolutionary model (Goldman & Yang, 1994) using M0 model, also known as one-ratio model, from PAML package. Thus the total number of parameters was split and only model-specific parameters were optimized during the positive selection analysis.

2.3 Tree selection

Two phylogenetic trees were reconstructed for each of the studied genes: a species tree that follows the canonical mammalian phylogeny (as depicted in https://www.ensembl.org/info/about/speciestree.html) and a gene tree reconstructed by ML from our sequence alignments data using the Jones-Taylor-Thornton (JTT) amino acid substitution model with partial deletion treatment for missing data (85% site coverage cutoff). Although tree topologies were defined in these ways, branch lengths of the trees used in the branch-site positive selection analyses were estimated using M0 (one ratio) model from PAML package (see above). All positive selection analyses for each gene were carried out using the species tree and the gene tree. The fitting of each of these phylogenies to the sequence data in the corresponding alignment was assessed based on the results of KH (Kishino & Hasegawa, 1989) and AU (unbiased) tests, performed with CONSEL v0.20 (Shimodaira & Hasegawa., 2001). When the species tree was rejected ($p<0.05$), only the results obtained from the gene tree were considered. When none of the phylogenies were rejected, both result sets were considered.

2.4 Positive selected sites

Positive selected sites were obtained by the Bayesian posterior analysis (BEB) implemented in codeml's branch-site specific positive selection test a posteri of the maximum likelihood estimates (Yang et al, 2005). As site-specific models are sensitive to codon parameters, we repeated each analysis 6 times using 2 different codon equilibrium frequencies estimation parameter variants

(CodonFreq parameter in PAML). Sites reported correspond to those consistent with at least 3 replicates in each of the 2 CodonFreq variants applied ('codon table' and '3x4') with a p-value of at least 0.7 in all reruns. For cases where neither the species tree nor the gene tree were rejected, all these criteria had to be met in both species tree and gene tree based analysis.

2.5 Sequence Identities

Mean sequence identities for our taxonomic groups of interest were obtained from pairwise sequence identities comparisons from each of our alignments using Sequence Identity And Similarity online tool (SIAS – http://imed.med.ucm.es/Tools/sias.html). Caniformes vs Feliformes mean identity comparison for each protein were performed through a two sample T-Test (Welch's T-test). Boxen plot graphs were drawn with Python 2.7 using the Seaborn library (version 0.9.1).

2.6 Predicted structure model of ZP proteins

The predicted structural models for jaguar ZP proteins were performed with the DeepMind AlphaFold2 (Jumper et al., 2021) as implemented in ColabFold (Mirdita *et al*., 2021) using 20 templates (phenix versión). The University of California in San Francisco (UCSF) Chimera v1.16 program (Pettersen et al, 2004) was used for visualization and the backbone-dependent rotamer library was used for structural interpretation (Shapovalov et al., 2011). N-glycosylation and O-glycosylation sites were obtained from the UniProt database (The UniProt Consortium, 2021) and converted to jaguar positions with our multiple sequence alignments. The Coulombic Surface Coloring calculates electrostatic potential according to Coulomb's law (coulombic surface coloring, −10 to +10 kcal/(mol*e), red to blue).

2.7 Protein domain identification

Protein domains and their corresponding positions in jaguar were defined using specific associated bibliography (Nishimura et al., 2019; Raj et al., 2017; Han et al., 2010; Moros-Nicolás et al., 2018a; Aydin et al., 2016) and our highly curated sequence alignments for position conversion along ortholog sequences. Start and end positions of ZP protein domains were later corrected based on the three-dimensional models and the observed and expected secondary structure of each domain: secondary structure prediction for each ZP protein was obtained using PDBsum web server (Laskowski et al., 1997), PSIPRED 4.0 (Jones, 1999) and DISOPRED3 (Jones et al., 2014) and contrasted with the canonical ZP hemi-domains topology (Bokhove et al., 2018) in order to check if obtained domains fulfilled the structural characteristic features in number and type of secondary structures and also if they displayed all disulfide-bonds described as necessary for each domain correct folding.

Lineal schematic representation showing ZP protein domains were based on Bokhove et al., 2018.

3. Results

The study of protein evolution has greatly changed thanks to the growing availability of complete genome sequences. For the jaguar, our no-model system of study, global protein evolution analysis is now possible owing to the availability of the jaguar genome draft (Figueiró et al., 2017). Furthermore, we have complemented and improved our evolutionary analyses of gamete interaction

proteins using a second version of jaguar chromosome-resolved genome assembled by Hi-C long-range sequencing (Dudchenko et al., 2017). Our analysis starts with ZP isoforms directly involved in sperm-recognition, ZP3 and ZP2, and continues with the analysis of structural cross-linker ZP proteins, ZP1 and ZP4; finally, we extended our results to IZUMO1R and added two control proteins: one under non-selection and another under previously reported positive selection in jaguar.

3.1 Jaguar ZP3

Jaguar *Zp3* gene sequence (reported in online supplementary material) was obtained performing a local BLASTn over the chromosome-resolution Hi-C genome assembly of *Panthera onca* (from DNA Zoo) using leopard (*Panthera pardus*) sequence (GeneID:109250568) as query. Our previous attempts to obtain this sequence from the first version of the jaguar genome failed, as only a partial and fragmented sequence could be retrieved. With the complete sequence, we could confirm that *Zp3* gene in the jaguar consists of 1272 nucleotides organized in 8 exons (Figure 1A, top) which matches with all panterine species orthologs and also in most sequenced felid species, with the only annotated exception of the puma (*Puma concolor*) ortholog, which displays an extra exon, 9 in total.

ZP3 protein presents the simplest domain composition among all members of the ZP protein family as it displays the basic structural elements common to all ZP proteins: a signal peptide, one complete zona pellucida domain, divided into its N-terminal and C-terminal moieties (Figure 1A bottom), a furin cleavage site and an hydrophobic C-terminal trans-membrane domain (TMD), all necessary for their secretion and incorporation in the zona pellucida matrix.

The *Zp3* positive selection analysis was performed with the ortholog sequences from a total of 34 species presented in the tree shown in Figure 1B. KH and AU tests did not reject either species or gene ZP3 trees. Therefore, only results consistently obtained in both analyses were considered. *Zp3* evolutionary analysis displayed no differential adaptations for the pantherines when compared to its sister groups. In fact, there is no evidence of molecular adaptation in *Zp3* within the whole Felidae family. However, Felidae do display a confirmed signal of episodic adaptation at its basal branch (red line) (Figure 1B and Table S1).

In contrast to the absence of signatures of positive selection among felids, subtree analysis revealed events of adaptive change for ZP3 in its sister group, Caniformia (brown highlighted area), and also for the whole Carnivora clade (green highlighted area) (Figure 1B and Table S1).

ZP3 positive selected amino acids were found at positions 26, 27, 28, 46 and 344 (Figures 1A,1C, and table S1). Three sites were detected congruent with the episodic adaptation event identified in the basal branch of Feliformia: amino acid positions 26 W>A/T, 46 E>H and 344 S/L>M. Amino acid positions related to the evidence of molecular adaptation in ZP3 within the subtree of the whole Carnivora order were: 26 W>T/A/P/S, 27 L/D/Q>E/G/W/P/S and 28 L/D>R/G/T, with the last two sites (27 L/D/Q>W/P/S and 28 L/D>R/G/T) also involved in events of molecular adaptation within Carnivora subtree (Figures 1A and 1B and table S1). It is interesting to note that, with the exception

of position 344, which stands in the most extreme region of the active ZP3 protein, just before the furin cleavage site, all remaining positive selected sites detected for ZP3 are located in the linking peptide located between signal peptide and ZP-N domain (Figures 1A and 1E and table S1).

3.2 Jaguar ZP2

Jaguar *Zp2* gene complete sequence (reported in online supplementary material) was obtained using leopard orthologous sequence (GeneID:109255525) as query for a BLASTn over the jaguar chromosome-resolution Hi-C genome assembly (from DNA Zoo). Jaguar *Zp2* gene consists of 2151 nucleotides organized in 18 coding exons (Figure 2A top) and their number, grouping and distribution along the locus is conserved among all sequenced felid species, with the exceptions of jaguarundi (*Puma yagouaroundi*) ortholog (GeneID:121019416) that displays 19 exons, and 2 alternative transcripts (XM_025916776.1 and XM_025916778.1) of the puma (*Puma concolor*) ortholog, that also display an extra exon.

Contrary to ZP3, ZP2 protein exhibits the most intricate composition in terms of protein domains among all members of ZP protein family. Besides the characteristic ZP module (subdivided in its 2 N-terminal and C-terminal moieties) present in all ZP proteins, ZP2 contains 3 ZP-N domains located N-terminal from the complete ZP domain, which are named ZP-N1 to ZP-N3 (Figure 2A bottom).

The positions of all jaguar ZP2 protein domains are indicated in Figure 2A and were assigned in accordance with the corresponding positions in ZP2 mouse sequence (Raj et al., 2017).

In order to search for positive selection signatures in the evolutionary history of ZP2, orthologs from 35 mammal species presenting a complete ZP2 public gene sequence were aligned and used to reconstruct the tree shown in Figure 2B. KH and AU tests from CONSEL failed to reject both the species and the gene trees reconstructed for ZP2. Therefore, both evolutionary contexts were considered and only consistent results in both scenarios were kept.

Similar to the results obtained for ZP3, ZP2 evolutionary analyses showed no signatures of selection among pantherines or even among felids. Once more, these analyses displayed signatures of positive selection among the sister group Caniformia (brown highlighted area) and for the whole Carnivora subtree (green highlighted area) (Figure 2B and table S2). Nevertheless, and in contrast to the findings for ZP3, the felids basal branch did not present signatures of episodic adaptive change and also jaguar ZP2 branch did present a positive selection signal associated to residues located at positions 231 Q/E>V and 713 M/T>I in the jaguar sequence (Figure 2A bottom, 2C and 2D). Considering that the residue 713 is located beyond the furin cleavage site and is, therefore, not a part of the functional protein, it is not taken into account in our report. No consistent positive selected site could be robustly identified in correlation with the evolutionary signal established for Caniformia subtree. However, we could effectively reveal 3 changes related to the positive selection events in ZP2 among species of the Carnivora subtree: positions 180 Q/K/E>L/G/P and 181 M/T>R/K/P in

ZP-N2 domain, and position 346 P/L/G>Q/R/M/V in ZP-N3 domain (Figure 2A bottom, 2D and supplementary table S2).

3.3 Jaguar ZP1

Jaguar *Zp1* complete sequence (reported in online supplementary material) was obtained through local BLASTn over the Illumina genome assembly (Figueiró et al., 2017) using leopard *Zp1* sequence as a query (GeneID:109246518). Jaguar *Zp1* gene comprises 1818 nucleotides organized in 12 exons and follows the same gene structure in quantity, grouping and distribution of exons along the locus as all other sequenced felids.

Beyond the ZP protein common structural organization of domains, ZP1 displays a single ZP-N domain and a trefoil domain (Figure 3A), considered important for the cross-linking between ZP filaments (Killingbeck & Swanson, 2018).

Pseudogenization is much more common for *Zp1* than for any other *Zp* gene along placental mammal species. *Zp1* appears as a pseudogene in all Caninae species (dog, dingo and red fox), cattle (cow, pig, goat, sheep) and even in some cetacean (dolphin) and Pinniped species (Antarctic fur seal and Weddell Seal) (Goudet et al., 2008; Moros-Nicolás et al., 2018a). As these species are included in the species set we used for the analyses of the other members of the ZP protein family, our ZP1 species collection was strongly reduced. In order to keep a similar number of species for each evolutionary analysis, in the case of ZP1 we included other outgroup species, such as bats, Western European hedgehog, European shrew and star-nosed mole (Figure 3B). The resulting species tree was rejected by KH and AU CONSEL tests ($p<0.001$), but not the reconstructed gene tree ($p=0.999$). Therefore only results obtained from the evolutionary analysis based in the gene tree were considered.

Among the proteins studied in this work, ZP1 is the one that shows the largest number of positive selected sites: 12 sites (additional sites detected beyond the furin cleavage site were not considered or depicted in the figures) (Figure 3A and 3C). As with the proteins studied so far, there are no adaptive changes in Pantherines for ZP1. However, unlike the cases of ZP3 and ZP2, ZP1 does display signatures of positive selection in Feliformia subtrees, as well as Carnivora and Caniformia subtrees (Figure 3B and table S1). Positive selected sites 116 P>R/H, 356 F>V/I and 422 Q>D detected in the Feliformia subtree correspond to the ZP-N1, ZP-N and ZP-C protein domains, respectively (Figure 3A yellow sites). Most sites detected in the analysis of the Caniformia subtree (Figure 3A brown sites: 35 S>G/H, 147 T/A>M/V, 308 L>F, 371 P>W, 384 R>Q/H, 398 Y/F>H) and in the whole Carnivora subtree analysis (Figure 3A light green sites: 146 P>R/H, 207 T>A/V, 285 L>V/M, 371 P>W, 384 R>Q/H, 398 Y/F>H and 422 Q>D/H) also correspond to those same protein domains, specially the ZP module that gathers an important cluster of positive selected sites.

3.4 Jaguar ZP4

Jaguar *Zp4* gene has 11 exons with a total length of 1707 nucleotides and encodes a 569 aa long polypeptide. Its complete coding sequence (reported in online supplementary material) was obtained via local BLASTn over the Illumina jaguar genome assembly (Figueiró et al., 2017) using leopard *Zp4* sequence as query (GeneID:109248980).

Given that ZP4 and ZP1 are considered to be paralogous genes formed as a consequence of the duplication of a common ancestral gene, both have a of 45 aa long trefoil domain and a ZP-N1 subdomain in their amino-terminal region (Figures 3A and 4A).

ZP4 alignment and tree were reconstructed with ortholog sequences from 34 species (Figre 4B). KH and AU tests from CONSEL rejected the species tree ($p=1x10^{-5}$). Therefore only results obtained from the evolutionary analyses based in the gene tree were considered.

Unlike the other three ZP proteins, ZP4 does not show adaptation signals in the subtrees of carnivores and caniformes. Instead, we found positive selection marks in the subtrees of feliformia and pantherines. ZP4 is the only protein studied in this work that does not follow our proposed hypothesis.

Regarding ZP4 positive selected sites, jaguar position 505 S/F>L is the only positive selected change detected for the subtree of Feliformes. This site is located beyond the furin cleavage site (positions 496-501); therefore, it is not present in the mature functional protein.

The other two detected sites: 309 T>S and 396 G>V/A correspond changes that took place among pantherine species and are located in the central region of ZP domain, more specifically in ZP-N moiety and ZP-C moiety, respectively.

3.5 Jaguar IZUMO1R

JUNO protein, also known as IZUMO1R, is present at the cell surface of the oocyte and is the receptor of the sperm cell surface protein IZUMO1. The interaction between JUNO and IZUMO1 is necessary for the adhesion and fusion of sperm and egg (Bianchi et al., 2014).

Jaguar *Izumo1r* complete sequence (reported in online supplementary material) was obtained through local BLASTn over the Illumina genome assembly (Figueiró et al., 2017) and the chromosome-resolution Hi-C genome assembly (from DNA Zoo) using leopard ortholog coding sequence as query (GeneID:109267358). Jaguar *Izumo1r* gene has only 729 nucleotides in length and is composed of 4 coding exons and its locus structure, number and disposition of exons is well conserved along all available felid species.

Jaguar IZUMO1R/JUNO protein is 243aa long peptide that starts with a 19aa long signal peptide and contains a folate receptor domain between its positions 26 and 201.

No signatures of positive selection were detected along the IZUMO1R Carnivora clades either using the species or their gene trees. All tested branches and subtrees (Carnivora, Caniformia, Feliformia, Pantherines and *Panthera onca*) rendered non-significant results for positive selection (supplementary table S5).

3.6 Sequence Identities

Looking more broadly at the whole range of amino acid changes accumulated for these proteins along Carnivora evolutionary history, we can appreciate that IZUMO1R acquired a similar amount of changes to those reported for ZP proteins. However, while all ZP proteins retained some amount of adaptive mutations, none of the conserved changes in IZUMO1R were positively selected. The average identity values of IZUMO1R for the studied groups is not lower than those of ZP proteins (supplementary table S8).

It is worth noting that Feliformes and Caniformes sister groups show significant differences in average sequence identity for all studied proteins except ZP4, always evincing Caniformes as a more diverse group and Feliformes as the one that features less variability (Figure 5).

Even though Pantherines display the highest average identity in each case, the higher percentage of identity observed in felids when compared to caniformes is not due to a pantherines effect exclusively, as non-pantherine felids always produce an average identity similar to those of the whole Feliformes group (Figure 5).

3.7 Control proteins: Tpm3, a jaguar positively selected protein and Actn1, a neutrally evolved protein.

To demonstrate the reliability of the method used to detect positive selection and inferring sites potentially subjected to selection, we performed likelihood ratio tests on control proteins, using *Tpm3* one of the positively jaguar specific selected genes previously reported (Figueiró el al., 2017). Results shown in supplementary table S6 shows a p-value=0.0006.

To demonstrate that the likelihood methods do not spuriously detect selection, we performed likelihood ratio tests for the conserved housekeeping *Actn1* gene (Holterhoff et al., 2019) (a negative control protein). Supplementary table S7 shows the negative expected results.

4. Discussion

There are two possible mechanisms of selection in gamete interaction protein evolution that could explain how key genes of reproduction could evolve rapidly to contribute to reproductive isolation.

In one of them, positive selection may result from interactions between individuals of the same species. Differences between sexual strategies that involve gamete interaction possibilities in finding optimal fertilization rates can lead to sexual conflict. This conflict may maintain either polymorphism or extend a co-evolutionary competition between egg and sperm proteins. This mechanism will not be approached in this study because the counterpart proteins of gamete interaction in sperm have not been defined in mammals. In the second case, selection may occur if there are hetero-specific interactions. When gametes from different species meet, resulting in hybrids that have reduced fitness, selection changes protein interactions, reducing hetero-specific fertilization. Consequently, this implies an economy of reproductive efforts (Turner & Hoekstra, 2008). Here, we tested the latter scenario by investigating the evolution of hetero-specific interactions along a wide curated universe of mammalian sequences. To do so, we measured and compared amino acid diversity for gamete interaction proteins among different Carnivora taxa and delved into their evolutionary histories assessing selective forces influence and uncovering events of rapid evolution and positive selection.

Although our evolutionary analyses revealed positive selection events in different groups along the history of Carnivora for every ZP protein but not for the fusion protein IZUMO1R, sequence identity results show that IZUMO1R was not kept immutable along mammalian evolution, but rather showed a similar number of changes as the ZP proteins. Selection analyses of the gene *Izumo1r* indicate that the driving evolutionary forces that ruled the history of this protein were purifying selection and neutral evolution. Sequence identity analyses of IZUMO1R and ZP proteins also revealed that Feliformia comprises low amino acid diversity when compared to its sister group, Caniformia, with the Pantherines exhibiting the lowest diversity. This difference between Caniformia and Feliformia was statistically significant in all but one of our studies proteins, ZP4, due to the fact that Caniformia has a comparatively higher degree of conservation in this protein.

Each jaguar *Zp* and *Izumo1r* gene shows a well-conserved gene structure among feline species. In terms of numbers of coding exons and their distribution and grouping along the locus, each of our studied genes presents a high consistency for all feline species, with the exception of the species of the genus Puma (*P. concolor* and *P. yagouaroundi*), which present a higher variability in this regard. When compared to all the other felid species gene structures, puma displays an extra exon in *Zp3* gene and in its *Zp2* alternative annotated transcript, while jaguarundi displays this extra exon only for *Zp2* gene.

The search for signatures of episodic adaptation at the basal branches of our taxonomic groups of interest (Carnivora, Feliformia, Caniformia and Pantherinae) revealed no punctual event of selection at any of these specific lineages for any of our studied proteins but one. The only exception was the basal branch of feliformes in ZP3, which did display a signature of episodic positive selection. The results obtained for ZP3 in the basal branch of Feliformia mean that the detected adaptive changes (26 W>A/T, 46 E>H and 344 S/L>M) occurred early in the history of this suborder, after their split from Caniformia, and were fixed by selection. Taking into account that the first two sites (26 W>A/T & 46 E>H) are located in the amino terminal region of ZP3 and the fact that the changes took place in the early history of the Feliformes and were later fixed, these changes could be involved in the surge of a prezygotic isolating mechanism between this group and its sister group Caniformia.

However, we can conclude that no subsequent adaptations took place later among felid species in *Zp3* gene, as no other molecular adaptation signals were detected in its subtree, evincing no contribution of ZP3 to isolation among felid species, including pantherines.

The absence of this kind of episodic selection in any other of the basal branches tested indicate that, with the exception of Feliformes in ZP3, none of the other groups of interest acquired changes during their early evolutionary history in these proteins that were later fixed by means of adaptive evolution.

A different general scenario was revealed when searching for positive selection signals in these proteins within the subtrees of the same taxonomic groups. Signatures of molecular adaptation were repeatedly identified within the subtrees of some of these phylogenetic groups, mainly in the Carnivora order and Caniformia suborder in ZP proteins. ZP3 and ZP2 displayed the same pattern of subtree evolution showing signatures of selection in Carnivora and Caniformia subtrees, but not in Feliformia or Pantherinae subtrees. ZP1 evolution deviated from this pattern as it also presented signatures of positive selection in the subtree of Feliformes, although the result for the subtree of pantherines was again negative. ZP4 is the subunit whose pattern of evolution differs more from the ones previously detailed, as it does not present signatures of selection among the subtrees of Carnivora or Caniformia. Moreover, Zp4 is the only gamete interaction protein studied here that shows evidence of diversifying positive selection in the subtree of Pantherinae as well as for the broader Feliformia suborder.

Positive selected sites detected for ZP3 among Carnivora and Caniformia subtrees overlap and cluster with Feliformia basal branch related sites gathering in the amino terminal region, in particular the inter-domain peptide that links the signal peptide and ZP domain (Figure 1A bottom). As to positions with selective changes clustered in this portion of ZP3 protein, position 26 of the jaguar sequence is found as a tryptophan in most mammals but it varies among different Caniformia subgroups, particulary in Pinnipeds: 26 W>P/S. This is also one of the sites detected for the basal branch of Feliformes, along with position 46, and due to these selective changes in both clades it emerges as a positive selected site in the subtree of Caniformia. Regarding positions 26 and 27, they both show an increased selected diversity among Caniformia subgroups and thus appear as positive selected sites both in Caniformia and in Carnivora. Altogether, these results indicate that positive selected changes for ZP3 may have played an important role in the specialization events that occurred inside Carniformia suborder or in the divergence between this group and Feliformia, but none of these changes would contribute to a prezygotic isolation of Panthera species or even between felid species. A similar reasoning can be made for ZP2 as positive selection for this protein was detected inside Carnivora and Caniformia subtrees but not inside Feliformia or Pantherinae subtrees. However, ZP2 evolutionary analysis also displayed a signature of positive selection in the jaguar terminal branch. This species-specific adaptation of jaguar ZP2 could be traced to a single amino acid position: site 231 in the jaguar sequence, where a valine (V) replaces the highly conserved glutamine (Q) in nearly all mammals. The other 3 detected sites (180, 181 and 346) correspond to positions that were selected within the Carnivora subtree. Although a clear signature of adaptation was detected for the subtree of caniformia, no sites could be individually identified by BEB analysis. This could be explained by the combination of multiple restrictive criteria that we requested for a site to be unmistakably identified

as positively selected in the final results.
Residues 180 and 181, detected in Carnivora subtree analysis and residue 231, detected as positive selected exclusively in the jaguar branch, are all located in ZP-N2, the domain that contains ZP2 cleavage site from which gamete recognition in mice depends (Galhay et al., 2010; Avella et al., 2014; 2016). In the jaguar sequence the cleft appears between positions 171-172. All these sites are located in the C-terminal from the mentioned cleavage site and outside ZP2 sperm-binding region, which in the jaguar protein extends from position 55 to position 154. Therefore, the 3 positive selected sites detected in ZP-N2 domain of ZP2 would not compromise gamete recognition and sperm penetration through the zona pellucida.

Among all studied proteins, ZP1 was the one that showed a higher number of positive selected sites, 12 sites, compared to the rest of the ZP proteins that showed between 3 and 5. Although ZP1 and Zp4 remain by far the less studied components of the mammalian ZP, Jovine's lab structurally characterized hZP1 cross-linking function (Monné & Jovine, 2011; Nishimura et al., 2019). Studies of hZP1 I390fs404X mutant protein suggest that infertile patients carrying the corresponding mutation lack the ZP that could result in hZP1 secretion. As this step is prior to the cross linking with ZP3 it cannot discriminate how this region of hZP1 would interact with hZP3 if at all (Huang et al., 2014). On the other hand, the trefoil/ZP module region of hZP1 does not form non-covalent homodimers, which leads us to think that hZP1-N1 necessarily participates in cross-linking to maintain human ZP with its consequences over fertility.

Two other mutations were reported to negatively affect the cross-linking ability of hZP1. One is the replacement of a residue of hZP1-N1 β- strand C with an additional Cys (R61C) (Yuan et al., 2019) or the truncation of the protein at the level of the hZP1-N1 bc loop (P54fs64X) (Sun et al., 2019). Then, there are two different regions to take into account in the crosslinking function of ZP1: one is the ZP1-N1 domain, in which we found two positive selective aa (35 inside Canifornia subtree and 116 in Feliformia) but both are outside the influence of the critical C66 region. The other important functional region is the ZP-C domain in which three aa were positively selected (384, 398 in Carnivora and Canifornia, and 422 in Carnivora and Felifornia). These three residues could alter ZP properties. Although molecular basis of egg coat filament cross-linking by ZP1 is conserved between birds and mammals, it is less plausible that ZP1 could be involved in the species-specific isolation.

ZP4 was the last of the mammalian ZP proteins to be identified (Lefièvre et al., 2004) as it not present in the mouse (*Mus musculus*) the most extended and used mammalian model. Its coding gene is considered, along with *Zp1*, a consequence of a recent duplication event (Goudet et al., 2008) and both proteins conserve the same arrangement of domains, being the only two members of the ZP proteins family that display a trefoil domain. As a consequence, their cross-linking functions can be considered to be similar, overlapping or even redundant and this would be supported by the fact that the absence of any of these proteins in some species does not cause sterility (Goudet et al., 2008).

All in all, we have found that sperm-oocyte interaction and fusion proteins lack the degree of diversification necessary to fix a prezygotic reproductive isolation barrier between pantherines. Regarding the evolutionary histories revealed for ZP proteins, it is interesting to note that *Zp3* and

Zp2, genes that encode proteins directly involved in sperm recognition, display a similar pattern of adaptive evolution showing signatures of positive selection for Carnivora and Caniformia subtrees but not in Feliformia or Pantherinae subtrees. On the other hand, the evolution of *Zp1* and *Zp4*, the subunits involved in filaments cross-linking, does not concur with that pattern. Instead, they show signatures of positive selection for Feliformia in the case of *Zp1,* and Feliformia and Panthera genus subtrees in the case of *Zp4*. Furthermore, this also extends to the location of positive selected sites along protein domains as both ZP1 and ZP4 display adaptive changes in their ZP domain while in ZP3 and ZP2 this domain is noticeably devoid of such changes.

These results are consistent with the previously reported pseudogenization events of *Zp1* in some carnivores, even-toed ungulates and cetacean species (Moros-Nicolás et al., 2018a; Goudet et al., 2008), the pseudogenization of *Zp4* in the house mouse (*Mus musculus*) (Lefièvre et al., 2004) and in two South American marsupials (common opossum and gray short-tailed opossum) (Moros-Nicolás et al., 2018b), the fact that *Zp4* appears to have diverged among vertebrates (Nishimura et al., 2019) and, finally, that ZP4 is not responsible for gamete interaction (Izquierdo-Rico et al, 2021). As *Zp4* and *Zp1* are recent paralogs that keep the same structure and domains, their functions would at least partially overlap and in those species with just one of these cross-linking proteins, the other would fulfill its function. *Izumo1r* evolutionary analysis showed high levels of conservation in all studied groups (Carnivora, Feliformia, Caniformia and Pantherines) and no signature of positive selection was detected for these groups either in their basal branches or inside their subtrees. These results strengthen the idea that ZP3 and ZP2 are proteins that give more specific-specificity isolation than IZUMO1R, which is specialized in fusion events.

All these results are consistent with previous evidence that shows ZP3 and ZP2 are essential for gamete recognition and, therefore conservation in these proteins would prevent the origin of a reproductive barrier between closely related species, allowing inter-species hybridization. This becomes more significant in the Panthera genus, characterized by species of big hypercarnivore predators that occupy large home ranges, breed small litters and display low population densities. All these characteristics make these species more vulnerable to adverse environmental changes and especially to anthropogenic effects over their habitat that can cause an even further reduction in population size, fragmentation of the habitat and isolation of these small populations. All these effects will eventually bring to a decrease in genetic diversity of a population, a diminishing viability and a genomic erosion of the species, which could bring it to the edge of extinction (Abascal et al., 2016). One of the escape mechanisms from this fate is inter-species hybridization as a way to quickly increase genome diversity. As it is proposed, in species characterized by small population sizes, as is the case with pantherines, hybridization would pose an advantage to escape Allee effect and avoid extinction (Gómez-Sánchez et al., 2018). However, if key gamete interaction proteins have accumulated significant diversifying changes in these proteins between related species, cross-fertilization would be impossible. Conservation of these key proteins between such species would represent an advantage in challenging times and even save them from extinction. Moreover, the use of hybridization as a way to rapidly increase genetic diversity in a waning population or in reintroduction programs can serve as an effective tool in conservation efforts (Zecherle et al., 2021)

and would, in some cases, even be useful to save a highly endangered species from extinction. This, of course, brings up an interesting debate in the field of species conservation as it transgresses a purist conception of a species identity and evolves the concept of hybridization-induced-extinction (Todesco et al., 2016).

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial o*r financial relationships that could be construed as a potential conflict of interest*.

Author Contributions

CC, CP and FP performed the bioinformatic analyses. FP designed the evolution study. CDB, FP, CC and CP performed jaguar ZP proteins structural models. TG contributes with phylogenic approaches and all concerning Vagalume jaguar genome data. PS conceived the project. FP and PS wrote the paper. All authors discussed the results and commented on the manuscript.

Funding

FP received the Wood-Wheelan Research Fellowship from International Union of Biochemistry and Molecular Biology (IUBMB) and travel fellowship from CONICET partial funding program for short stays abroad for postdoctoral fellows. PS and FP are CONICET researchers. PS is a senior researcher of PIP 2015 and PICT 2015 grants.

Acknowledgments

We thank DNAZoo, particularly Olga Dudchenko and Erez Lieberman-Aiden for discussion about HiC jaguar genome data, Luca Jovine and our colleagues for fruitful paper and idea discussions. We also thank Fundación Williams, FIByME for the administration of grants and aided the author's work.

Supplementary Material

Supplementary tables for this article can be found online attached to this manuscript.
Additional supplementary information will be reported online upon article publication or will be made available by the authors upon request.

References

Aagaard J.E., Yi X., MacCoss M.J., Swanson W.J. (2006). Rapidly evolving zona pellucida domain proteins are a major component of the vitelline envelope of abalone eggs. Proc Natl Acad Sci U S A. 103(46):17302-7. doi: 10.1073/pnas.0603125103.

Avella M.A., Baibakov B., Dean J. (2014). A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. J Cell Biol. 205(6):801-9. doi: 10.1083/jcb.201404025.

Avella M. A., Baibakov B. A., Jimenez-Movilla M., Sadusky A. B., Dean J. (2016). ZP2 peptide beads select human sperm in vitro, decoy mouse sperm in vivo, and provide reversible contraception. Science translational medicine, 8(336), 336ra60-336ra60.

Aydin H., Sultana A., Li S., Thavalingam A., Lee J.E. (2016). Molecular architecture of the human sperm IZUMO1 and egg JUNO fertilization complex. Nature. 534(7608):562-5. doi: 10.1038/nature18595.

Bausek N., Waclawek M., Schneider W. J., Wohlrab F. (2000). The major chicken egg envelope protein ZP1 is different from ZPB and is synthesized in the liver. J. Biol. Chem. 275, 28866–28872. doi: 10.1074/jbc.275.37.28866

Bianchi E., Doe B., Goulding D., Wright G.J. (2014). Juno is the egg Izumo receptor and is essential for mammalian fertilization. Nature. 508(7497):483-7. doi: 10.1038/nature13203.

Bokhove M., Jovine L. (2018). Structure of Zona Pellucida Module Proteins. In Current Topics in Developmental Biology (1st ed., Vol. 130). Elsevier Inc. https://doi.org/10.1016/bs.ctdb.2018.02.007

Chen S., Costa V., Beja-Pereira A. (2011). Evolutionary patterns of two major reproduction candidate genes (Zp2 and Zp3) reveal no contribution to reproductive isolation between bovine species. BMC Evol Biol. 11:24. doi: 10.1186/1471-2148-11-24.

Dean J. (2007). The enigma of sperm-egg recognition in mice. Society of Reproduction and Fertility supplement, 63, 359-365.

Duque Rodriguez M., Gambini A., Ratner L.D., Sestelo A.J., Briski O., Gutnisky C., Rulli S.B., Fernández Martin R.,Cetica P., Salamone D.F. (2021). Aggregation of Leopardus geoffroyi hybrid embryos with domestic cat tetraploid blastomeres. Reproduction 161(5):539-548. doi: 10.1530/REP-20-0589.

Dudchenko O., Batra S.S., Omer A.D., Nyquist S.K.,Hoeger M.,Durand N.C.,Shamim M.S.,Machol I.,Lander E.S., Aiden A.P., Aiden E.L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 356(6333):92-95. doi: 10.1126/science.aal3327.

Figueiró H.V., Li G., Trindade F.J., Assis J., Pais F., Fernandes G., Santos S.H.D., Hughes G.M., Komissarov A., Antunes A., Trinca C.S., Rodrigues M.R., Linderoth T., Bi K., Silveira L., Azevedo F.C.C, Kantek D., Ramalho E., Brassaloti R.A., Villela P.M.S., Nunes A.L.V., Teixeira R.H.F., Morato R.G., Loska D.,Saragüeta P.,Gabaldón T., Teeling E.C., O'Brien S.J., Nielsen R., Coutinho L.L., Oliveira G., Murphy W.J., Eizirik E. (2017). Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci Adv. 3(7):e1700299. doi: 10.1126/sciadv.1700299.

Florman H.M., Ducibella T. (2006). Fertilization in Mammals. In Knobil and Neill's Physiology of Reproduction (Third Edition). St Louis: Academic Press; p.p. 55-112.

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(5988):216-9. doi: 10.1126/science.1188178.

Goldman N., Yang Z. (1994). A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biology and Evolution, 11(5), 725–736.

Gómez-Sánchez D., Olalde I., Sastre N., Enseñat C., Carrasco R., Marques-Bonet T., Lalueza-Fox C., Leonard J.A., Vilà C., Ramírez O. (2018). On the path to extinction: Inbreeding and admixture in a declining grey wolf population. Mol Ecol. Sep;27(18):3599-3612. doi: 10.1111/mec.14824.

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(5):796-806. doi: 10.1095/biolreprod.107.064568.

Gupta S. K., Bhandari B. (2011). Acrosome reaction: relevance of zona pellucida glycoproteins. Asian journal of andrology, 13(1), 97.

Gupta S. K., Bhandari B., Shrestha A., Biswal B. K., Palaniappan C., Malhotra S. S., Gupta N. (2012). Mammalian zona pellucida glycoproteins: structure and function during fertilization. Cell and tissue research, 349(3), 665-678.

Han L., Monné M., Okumura H., Schwend T., Cherry A.L., lot D., Matsuda T., Jovine L. (2010). Insights into Egg Coat Assembly and Egg-Sperm Interaction from the X-Ray Structure of Full-Length ZP3, Cell, 143(3):404-415. doi.org/10.1016/j.cell.2010.09.041.

Holterhoff C.K., Saunders R.H., Brito E.E. Wagner D.S. (2009), Sequence and expression of the zebrafish alpha-actinin gene family reveals conservation and diversification among vertebrates. Dev. Dyn., 238: 2936-2947.

Howe K.L., Achuthan P., Allen J., et al. (2021). Ensembl 2021. *Nucleic Acids Res.* vol. 49(1):884–891

Huang H.L., Lv C., Zhao Y.C., Li W., He X.M., Li P., Sha A.G., Tian X., Papasian C.J., Deng H.W., Lu G.X., Xiao H.M. (2014). Mutant ZP1 in familial infertility. N Engl J Med. 370(13):1220-6. doi: 10.1056/NEJMoa1308851.

Hughes D. C., Barratt C. L. R. (1999). Identification of the true human orthologue of the mouse Zp1 gene: evidence for greater complexity in the mammalian zona pellucida? Biochim. Biophys. Acta Gene Struct. Expr. 1447, 303–306. doi: 10.1016/S0167-4781(99)00181-5

Izquierdo-Rico M.J., Moros-Nicolás C., Pérez-Crespo M., Laguna-Barraza R., Gutiérrez-Adán A., Veyrunes F., Ballesta J., Laudet V., Chevret P., Avilés M. (2021). ZP4 Is Present in Murine Zona Pellucida and Is Not Responsible for the Specific Gamete Interaction. Front Cell Dev Biol. 8:626679. doi: 10.3389/fcell.2020.626679.

Jones D.T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292:195–202

Jones D.T., Cozzetto D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics. Mar 15; 31(6): 857–863.

Jovine L., Darie C.C., Litscher E.S., Wassarman P.M. (2005). Zona pellucida domain proteins. Annu Rev Biochem.;74:83-114. doi:10.1146/annurev.biochem.74.082803.133039.

Jumper J., Evans R., Pritzel A. et al. (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589

Killingbeck E.E., Swanson W.J. (2018). Egg Coat Proteins Across Metazoan Evolution. Curr Top Dev Biol.130:443-488. doi: 10.1016/bs.ctdb.2018.03.005.

Kishino H., Hasegawa M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. Journal of Molecular Evolution. 29(2):170–179.

Kumar S., Stecher G., Li M., Knyaz C., and Tamura K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549.

Lamas-Toranzo I., Balvís, N. F., Querejeta-Fernández A., Izquierdo-Rico, M. J. González-Brusi L., Lorenzo P. L., García-Rebollar P., Avilés M., Bermejo-álvarez P. (2019). Zp4 confers structural properties to the zona pellucida essential for embryo development. ELife, 8, 1–18.

Laskowski R. A., Hutchinson E. G., Michie A. D., Wallace A. C., Jones M. L., Thornton J. M. (1997). PDBsum: A Web-based database of summaries and analyses of all PDB structures. Trends in Biochemical Sciences, 22(12), 488–490.

Lefièvre L., Conner S. J., Salpekar A., Olufowobi O., Ashton P., Pavlovic B., Lenton W., Afnan M., Brewis I. A., Monk M., Hughes D. C., Barratt, C. L. R. (2004). Four zona pellucida glycoproteins are expressed in the human. Human Reproduction, *19*(7).

Li G., Davis B. W., Eizirik E., Murphy W. J. (2016). Phylogenomic evidence for ancient hybridization in the genomes of living cats (Felidae). Genome Res. 26, 1–11.

Liu C., Litscher E. S., Mortillo S., Sakai Y., Kinloch R. A., Stewart C. L., et al. (1996). Targeted disruption of the mZP3 gene results in production of eggs lacking a zona pellucida and infertility in female mice. Proc. Natl. Acad. Sci. U.S.A. 93, 5431–5436. doi: 10.1073/pnas.93.11.5431

Meslin C., Mugnier S., Callebaut I., Laurin M., Pascal G., Poupon A., Goudet G., Monget P. (2012). Evolution of genes involved in gamete interaction: evidence for positive selection, duplications and losses in vertebrates. PLoS One. (9):e44548. doi: 10.1371/journal.pone.0044548.

Mirdita M., Ovchinnikov S., Steinegger M. (2021). ColabFold - Making protein folding accessible to all. bioRxiv, 2021.08.15.456425

Monné M., Han L., Jovine L. (2006). Tracking down the ZP domain: From the mammalian zona pellucida to the molluscan vitelline envelope. Semin Reprod Med. (4):204-16. doi: 10.1055/s-2006-948550.

Monné M., Jovine L. (2011). A structural view of egg coat architecture and function in fertilization. Biology of Reproduction, 85(4), 661-669.

Moros-Nicolás C., Chevret P., Jiménez-Movilla M., Algarra B., Cots-Rodríguez P., González-Brusi L., Avilés M., Izquierdo-Rico M.J. (2021). New Insights into the Mammalian Egg Zona Pellucida. Int J Mol Sci. 22(6):3276. doi: 10.3390/ijms22063276.

Moros-Nicolás C., Leza A., Chevret P., Guillén-Martínez A., González-Brusi L., Boué F., Lopez-Bejar M. D , Ballesta J.A, Avilés M. A., Izquierdo-Rico M. J. (2018a). Analysis of ZP1 gene reveals differences in zona pellucida composition in carnivores. Reprod. Fertil. Dev. 30, 272–285. doi: 10.1071/RD17022.

Moros-Nicolás C., Chevret P., Izquierdo-Rico M. J., Holt W. V., Esteban-Díaz D., López-Béjar M., et al. (2018b). Composition of marsupial zona pellucida: a molecular and phylogenetic approach. Reprod. Fertil. Dev. 30, 721–733. doi: 10.1071/RD16519

NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. Nucleic acids research, *44*(D1), D7–D19. https://doi.org/10.1093/nar/gkv1290

Nishimura K., Dioguardi E., Nishio S., Villa A., Han L., Matsuda T., Jovine, L. (2019). Molecular basis of egg coat cross-linking sheds light on ZP1-associated female infertility. Nat Commun 10, 3086 https://doi.org/10.1038/s41467-019-10931-5.

Paviolo A., De Angelo C., Ferraz K.M., Morato R.G., Martinez Pardo J., Srbek-Araujo A.C., Beisiegel B.M., Lima F., Sana D., Xavier da Silva M., Velázquez M.C., Cullen L., Crawshaw P. Jr., Jorge M.L., Galetti P.M., Di Bitetti M.S., de Paula R.C., Eizirik E., Aide T.M., Cruz P., Perilli M.L., Souza A.S., Quiroga V., Nakano E., Ramírez Pinto F., Fernández S., Costa S., Moraes E.A. Jr., Azevedo F. (2016). A biodiversity hotspot losing its top predator: The challenge of jaguar conservation in the Atlantic Forest of South America. Sci Rep. Nov 16;6:37147. doi: 10.1038/srep37147.

Pettersen EF, Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. 25:1605–1612

Raj I., Sadat Al Hosseini H., Dioguardi E., Nishimura K., Han L., Villa A., de Sanctis D., Jovine L. (2017). Structural Basis of Egg Coat-Sperm Recognition at Fertilization. Cell. 169(7):1315-1326.e17. doi: 10.1016/j.cell.2017.05.033.

Rankin T., Familari M., Lee E., Ginsberg A., Dwyer N., Blanchette-Mackie J., Drago J., Westphal H., Dean J. (1996). Mice homozygous for an insertional mutation in the Zp3 gene lack a zona pellucida and are infertile. Development, 122(9), 2903-2910.

Rankin T. L., O'Brien M., Lee E., Wigglesworth K., Eppig J., Dean J. (2001). Defective zonae pellucidae in Zp2-null mice disrupt folliculogenesis, fertility and development. Development 128, 1119–1126.

Shapovalov MV, Dunbrack RL. Jr (2011) A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. Structure. 2011;19:844–858

Shimodaira H., Hasegawa M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics, 17(12), 1246-1247.

Shu L., Suter M. J. F., Räsänen K. (2015). Evolution of egg coats: linking molecular biology and ecology. Molecular Ecology, 24(16), 4052-4073.

Stetson I., Izquierdo-Rico M. J., Moros C., Chevret P., Lorenzo P. L., Ballesta J., et al. (2012). Rabbit zona pellucida composition: a molecular, proteomic and phylogenetic approach. J. Proteomics 75, 5920–5935. doi: 10.1016/j.jprot.2012.07.027

Stsiapanava A., Xu C., Brunati M., Zamora-Caballero, S., Schaeffer C., Bokhove M., Han L., Hebert H., Carroni M., Yasumasu S., Rampoldi L., Wu B., Jovine L. (2020). Cryo-EM structure of native human uromodulin, a zona pellucida module polymer. The EMBO journal, 39(24), e106807.

Sun L., Fang X., Chen Z., Zhang H., Zhang Z., Zhou P., Xue T., Peng X., Zhu Q., Yin M., Liu C., Deng Y., Hu H., Li N. (2019). Compound heterozygous ZP1 mutations cause empty follicle syndrome in infertile sisters. Hum. Mutat. 40, 2001–2006. doi: 10.1002/humu.23864.

Swanson W. J., Yang Z., Wolfner M. F., Aquadro C. F. (2001). Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. Proceedings of the National Academy of Sciences, 98(5), 2509-2514.

Swanson W. J., Vacquier V. D. (2002). The rapid evolution of reproductive proteins. Nature reviews genetics, 3(2), 137-144.

Swanson W.J., Nielsen R., Yang Q. (2003). Pervasive Adaptive Evolution in Mammalian Fertilization Proteins. Molecular Biology and Evolution, Volume 20, Issue 1, January 2003, Pages 18–20, https://doi.org/10.1093/oxfordjournals.molbev.a004233

Tanihara F., Nakai M., Kaneko H., Noguchi J., Otoi T., Kikuchi K. (2013). Evaluation of zona pellucida function for sperm penetration during in vitro fertilization in pigs. Journal of Reproduction and Development.

The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49:D1

Tian J., Gong H., Lennarz W.J. (1999). Xenopus laevis sperm receptor gp69/64 glycoprotein is a homolog of the mammalian sperm receptor ZP2. Proc Natl Acad Sci U S A. Feb 2;96(3):829-34. doi: 10.1073/pnas.96.3.829.

Todesco M., Pascual M. A., Owens G. L., Ostevik K. L., Moyers B. T., Hübner S., Heredia S. M., Hahn M. A., Caseys C., Bock D. G., Rieseberg L. H. (2016). Hybridization and extinction. *Evolutionary applications*, *9*(7), 892–908.

Turner L.M., Hoekstra H.E. (2008). Reproductive protein evolution within and between species: maintenance of divergent ZP3 alleles in Peromyscus. Mol Ecol. (11):2616-28. doi: 10.1111/j.1365-294X.2008.03780.x.

Wassarman P. M., Litscher E. S. (2009). The multifunctional zona pellucida and mammalian fertilization. Journal of reproductive immunology, 83(1-2), 45-49.

Wilburn D. B., Swanson W. J. (2017). The "ZP domain" is not one, but likely two independent domains. Molecular reproduction and development, 84(4), 284–285.

Yadav A., Jain A., Sahu J., Dubey A., Gadpayle R., Barwa D.K., Verma U. and Kashyap K. (2019). An overview on species hybridization in animals. International Journal of Fauna and Biological Studies 6(5): 36-42.

Yanagimachi R. (1994). Fertility of mammalian spermatozoa: its development and relativity. *Zygote*, *2*(4), 371-372.

Yang Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution, 24(8), 1586–1591.

Yang Z., Wong W. S., Nielsen R. (2005). Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular biology and evolution*, *22*(4), 1107–1118.

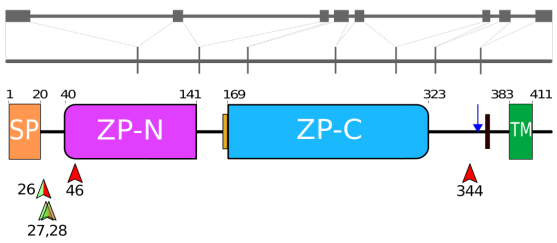
Yuan P., Li R., Li D., Zheng L., Ou S., Zhao H., Zhang Q., Wenjun W. (2019). Novel mutation in the ZP1 gene and clinical implications. J. Assist. Reprod. Genet. 36, 741–747. doi: 10.1007/s10815-019-01404-1.

Zecherle LJ, Nichols HL, Bar-David S, Brown PB, Hipperson H, Hipperson H, Horsburgh GJ, Templeton AR. (2021) Subspecies hybridization as a potential conservation tool in species reintroductions. *Evol Appl*. 14: 1216– 1224

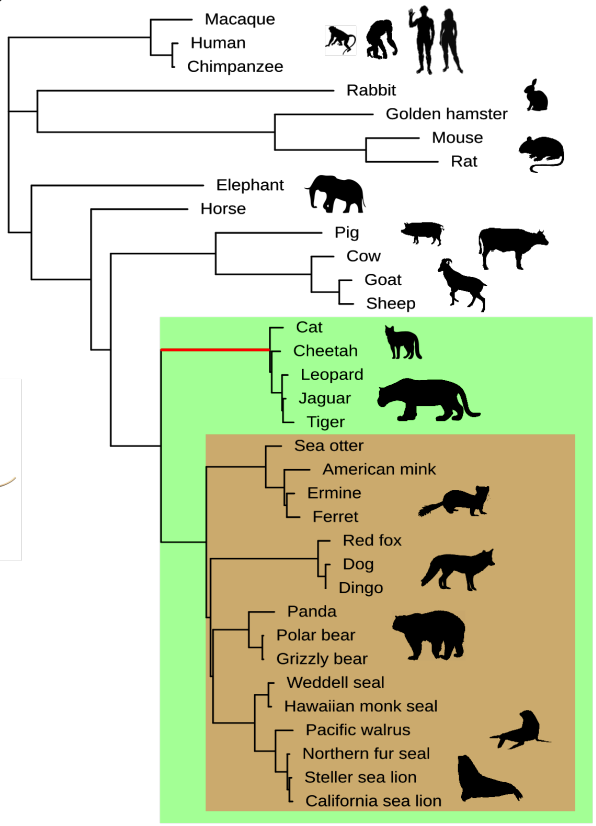
Zhang J., Nielsen R., Yang Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution, 22(12), 2472–2479.

Figure 1

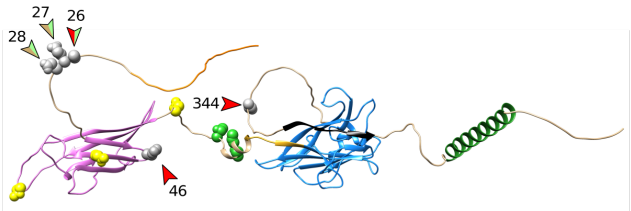
A



B



C



D

Position	Outgroups	Carnivora	
		Feliformia	Caniformia
26	W	T,A	P,S,W
27	L,D,Q	E,G	W,P,S
28	L,D	D	R,G,T
46	E	H	E
344	S,L	M	S

E

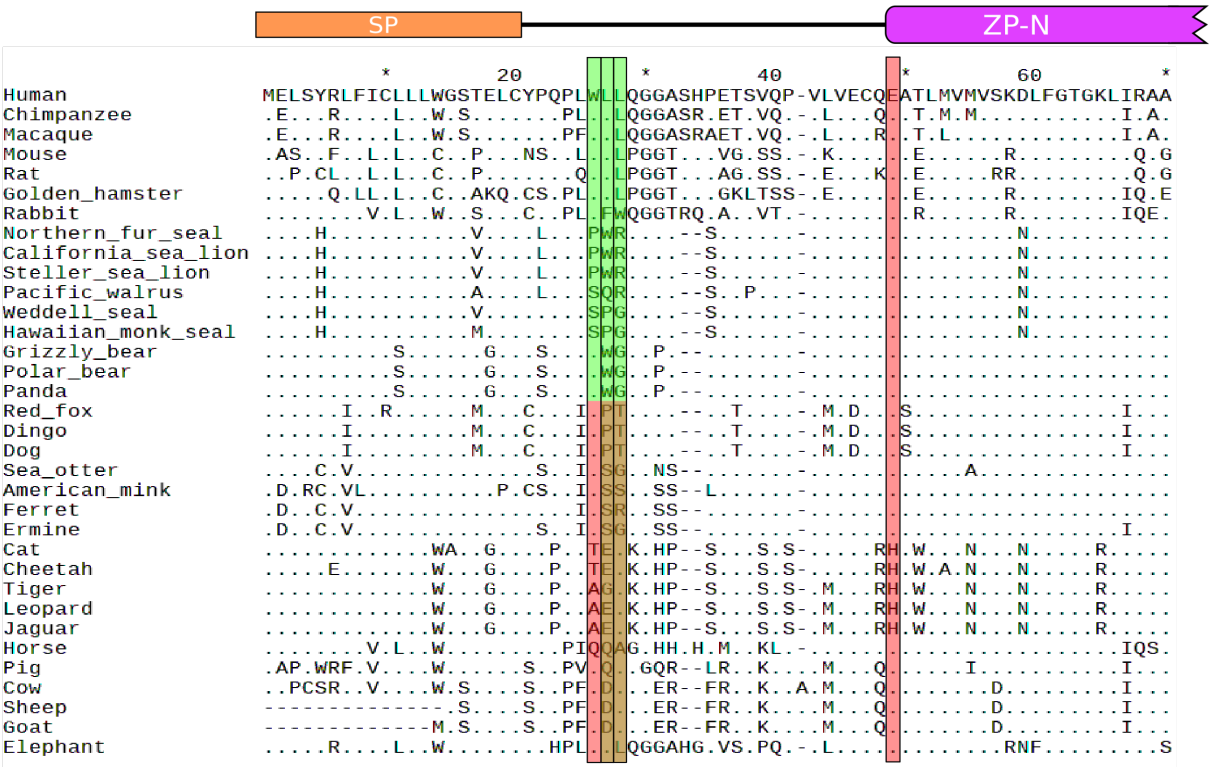
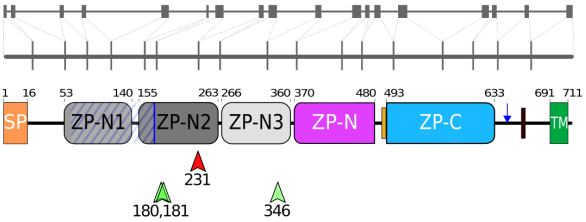
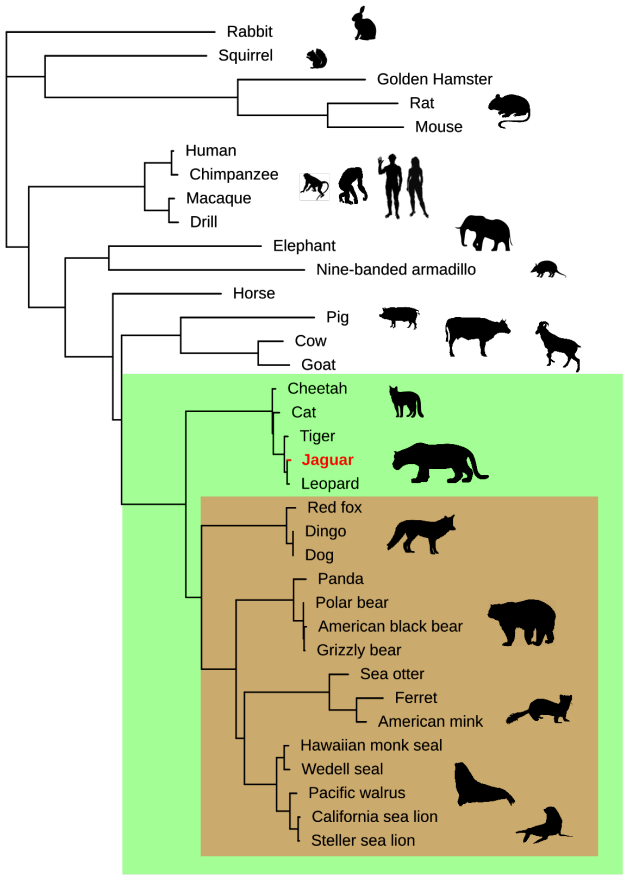


Figure 2

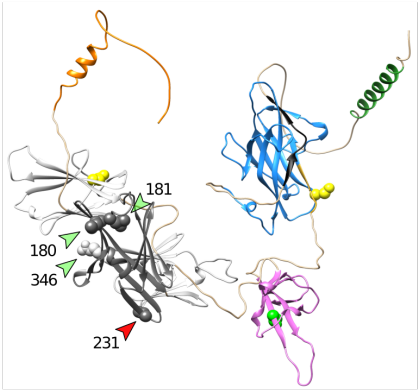
A



B



C



D

Position	Outgroups	Carnivora	
		Feliformia	Caniformia
180	Q,K,E	P,L	G,L
181	M,T	M	K,M,R,P
231	Q,E	Q,V	Q
346	P,L,G	Q	R,M,L,V

E

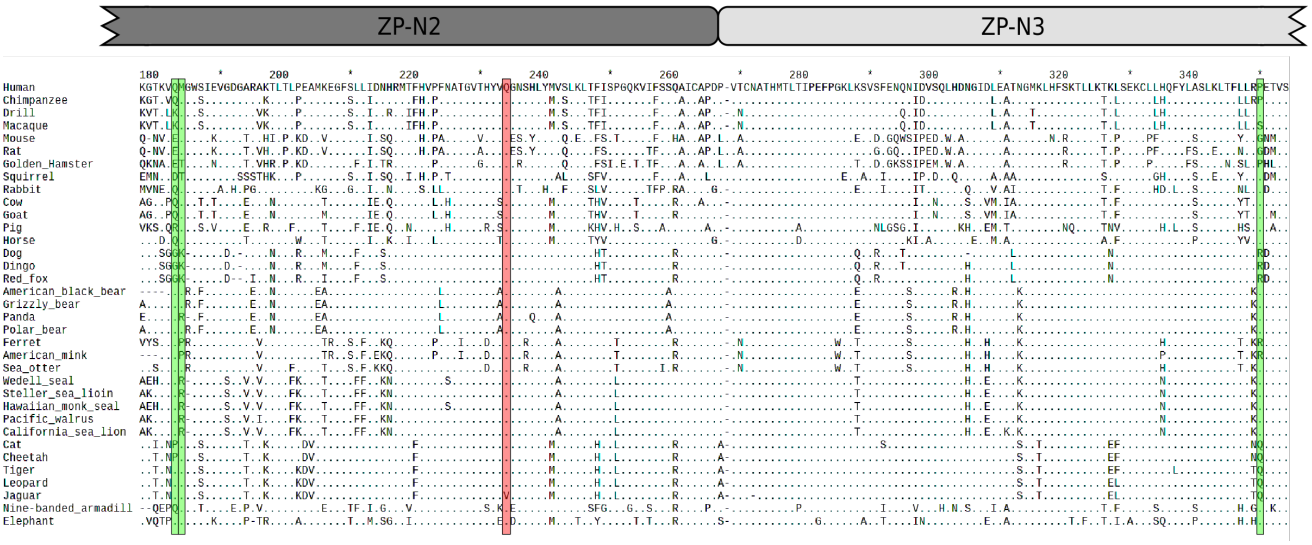
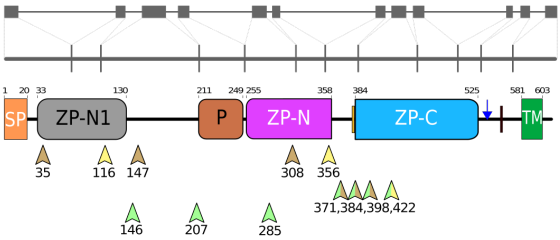
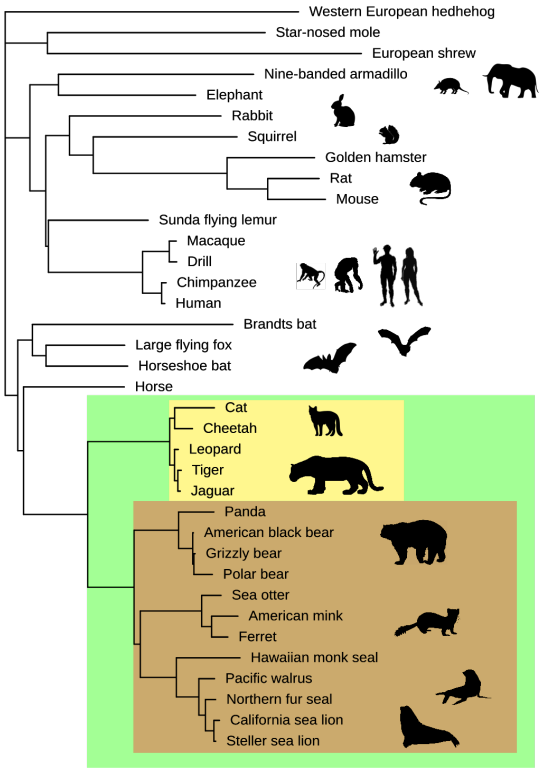


Figure 3

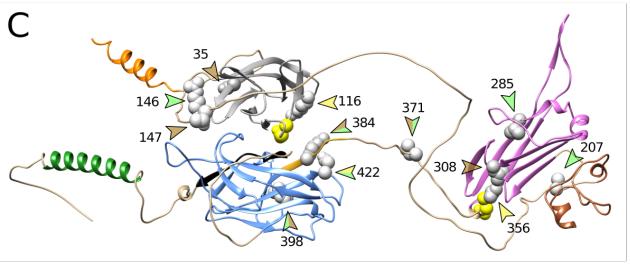
A



B



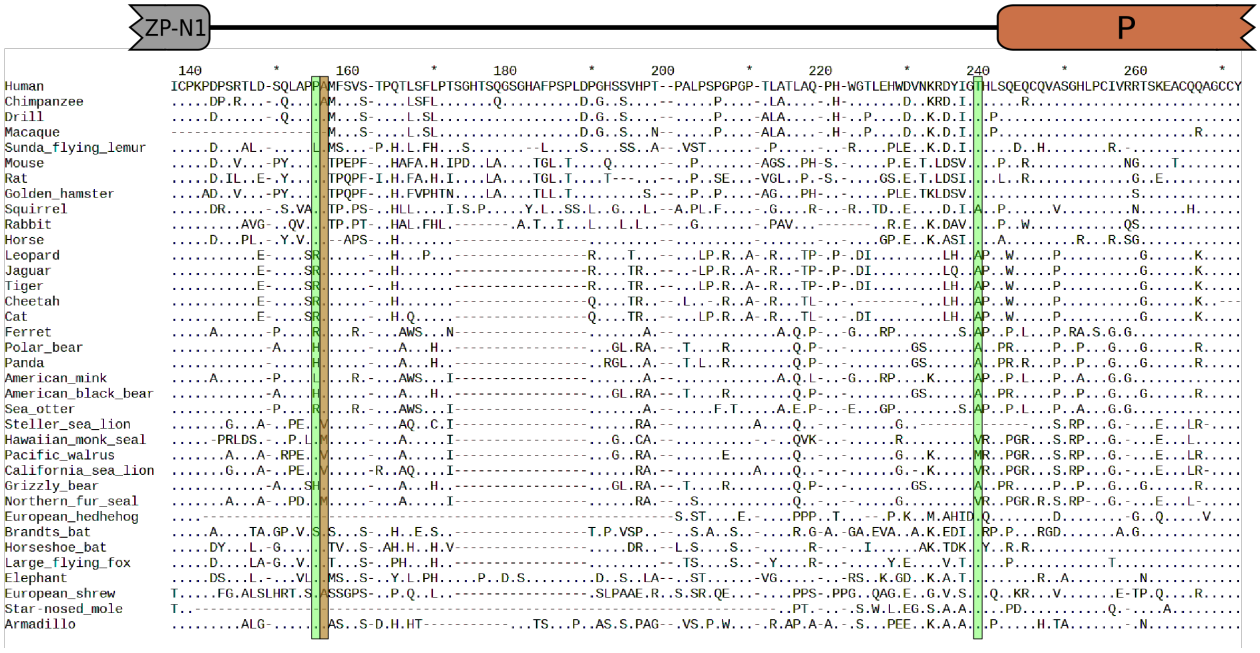
C



D

Position	Outgroups	Carnivora	
		Feliformia	Caniformia
35	S	S	G,H,S
116	P	R,H	P
146	P	R	R,H,P
147	T,A	T	M,V,T
207	T	A	A,V
285	L	V	M,L
308	L	L	F,L
356	F	V,I	F
371	P	P	P,W
384	R	R	Q,R,H
398	Y,F	F	Y,F,H
422	Q	D	Q,H

E



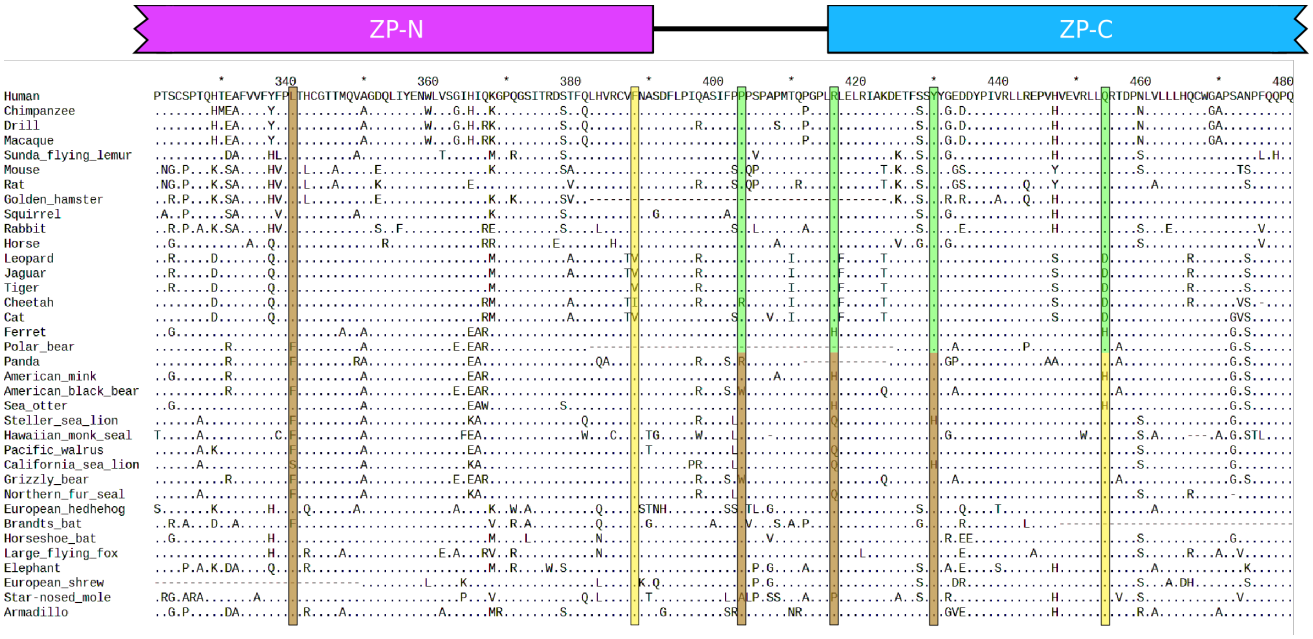
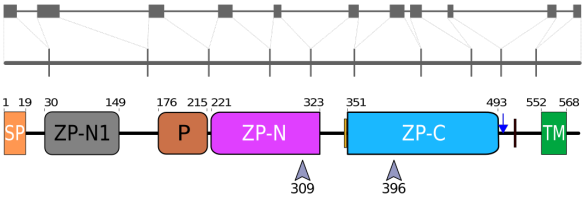
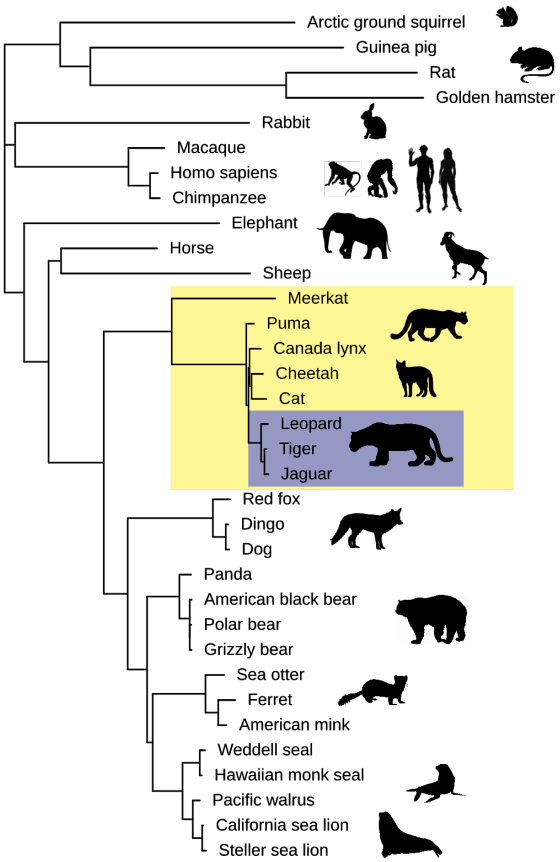

ZP-N
ZP-C
Human
Chimpanzee
Drill
Macaque
Sunda_flying_lemur
Mouse
Rat
Golden_hamster
Squirrel
Rabbit
Horse
Leopard
Jaguar
Tiger
Cheetah
Cat
Ferret
Polar_bear
Panda
American_mink
American_black_bear
Sea_otter
Steller_sea_lion
Hawaiian_monk_seal
Pacific_walrus
California_sea_lion
Grizzly_bear
Northern_fur_seal
European_hedhehog
Brandts_bat
Horseshoe_bat
Large_flying_fox
Elephant
European_shrew
Star-nosed_mole
Armadillo

Figure 4

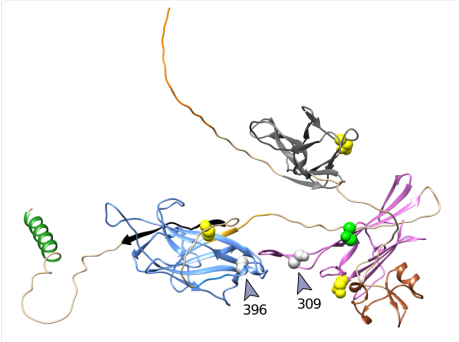
A



B



C



D

Position	Outgroups	Feliformia	Panthera
309	T	T	T,S
396	G,E	G	V,A

E

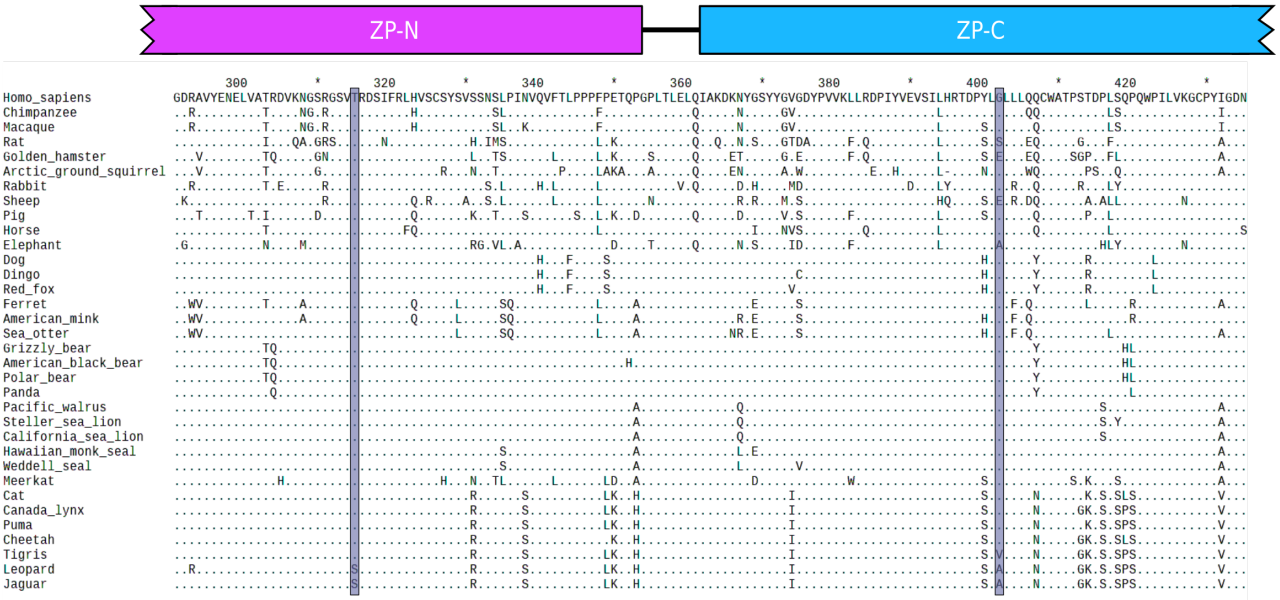


Figure 5

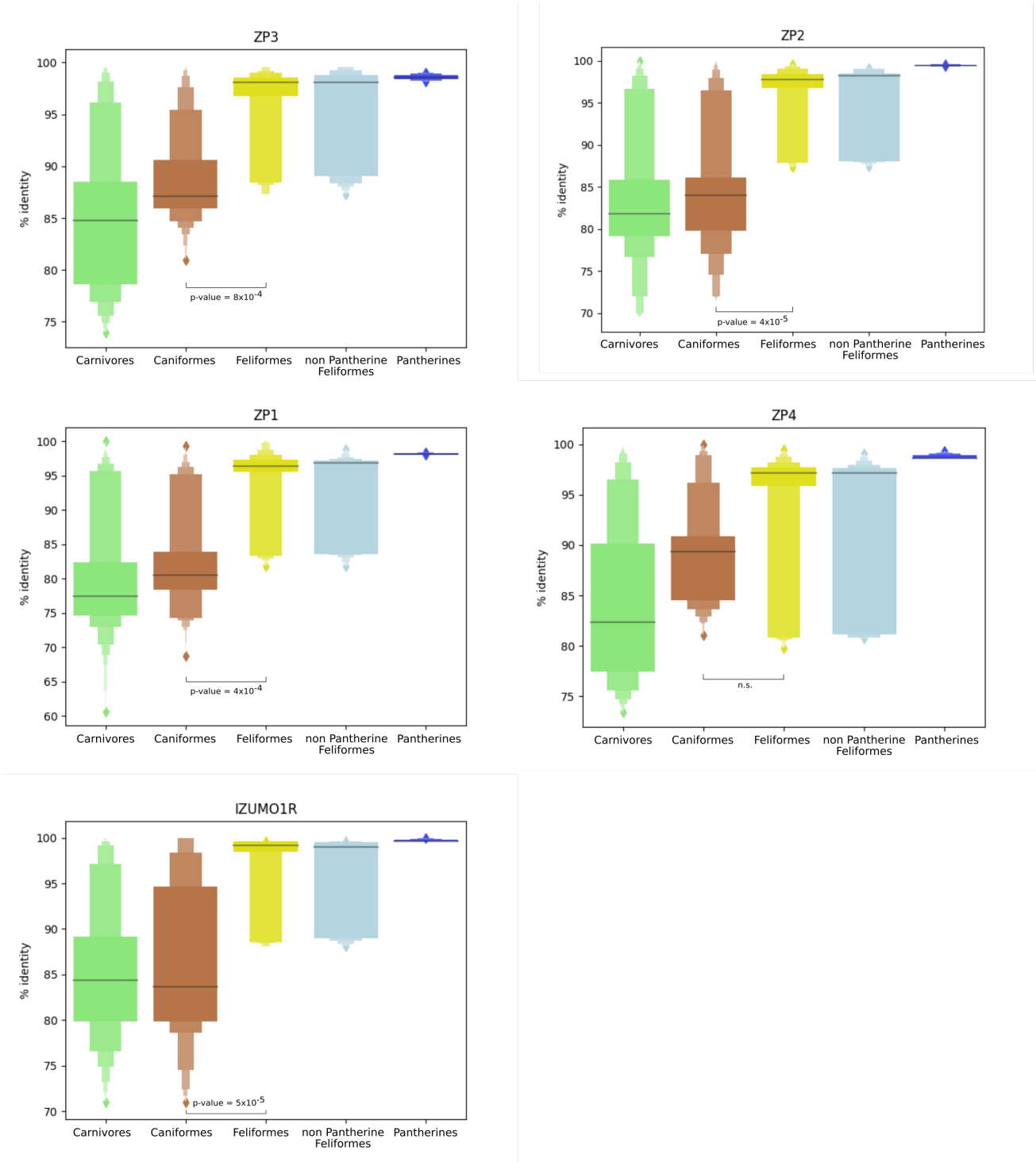

ZP3
% identity
p-value = 8x10-4
Carnivores
Caniformes
Feliformes
non Pantherine Feliformes
Pantherines
ZP2
% identity
p-value = 4x10-5
Carnivores
Caniformes
Feliformes
non Pantherine Feliformes
Pantherines
ZP1
% identity
p-value = 4x10-4
Carnivores
Caniformes
Feliformes
non Pantherine Feliformes
Pantherines
ZP4
% identity
n.s.
Carnivores
Caniformes
Feliformes
non Pantherine Feliformes
Pantherines
IZUMO1R
% identity
p-value = 5x10-5
Carnivores
Caniformes
Feliformes
non Pantherine Feliformes
Pantherines

Figure legends

1. Figure 1: ZP3 evolution. A. Jaguar *Zp3* gene schematic representation showing exons and introns (first line). Mature transcript of *Zp3* gene (second line). Start and end positions for each domain in the jaguar protein (third line) are displayed above and positive selected sites are indicated with arrowheads (fourth line) below the schematic representation of ZP3 protein domains. Red solid arrowheads indicate Feliformia basal branch positively selected sites. Split arrowheads with two colors denote sites that were detected in more that one analysis and their filling colors correspond to the highlighted subtrees or branches. SP: Signal Peptide; ZP-N: Zona Pellucida domain N-terminal moiety; ZP-C: Zona Pellucida domain C-terminal moiety; dark gray area: ZP3-specific subdomain; blue arrow pointing down: furin cleavage site; TM: transmembrane domain. B. Maximum likelihood ZP3 phylogenetic reconstruction for 34 species used in this analysis. Highlighted zones show positively selected Carnivora (light green) and Caniformia subtrees (brown). The basal branch of Feliformia suborder is colored in red to show the detection of positive episodic evolution along that branch. C. Three-dimensional model of the protein ZP3 shown in ribbons. N-glycosylated and O-glycosylated sites are shown as yellow and bright green atom spheres, respectively. Positive selected sites are indicated and shown as gray atom spheres. D. Positive selected sites table showing identity for each phylogenetic group. The identities in feline species of positive selected sites for Feliformes are shown in red. Autapomorphies in the background are not included. All residues located beyond the furin cleavage site were excluded. E. Detailed view of the first portion of ZP3 alignment showing the first 4 positive selected sites. The corresponding portion of the domain scheme is displayed above as a reference.

2. Figure 2: ZP2 evolution. A. Jaguar *Zp2* gene exons and introns model (first line). Mature transcript of *Zp2* gene (second line). Start and end positions for each domain in the jaguar protein (third line) are displayed above and positive selected sites are indicated with arrowheads (fourth line) below the schematic representation of ZP2 protein domains. The solid red arrowhead indicates the only positively selected sites detected for the jaguar lineage. Light green filled arrowheads indicate sites under selection in the Carnivora subtree. SP: Signal Peptide; ZP-N1/2/3: ZP-N like domains; diagonal striped blue pattern: ZP2 sperm-binding region; blue solid vertical line: ZP2 ovastatin cleavage site; ZP-N: Zona Pellucida domain N-terminal moiety; ZP-C: Zona Pellucida domain C-terminal moiety; blue arrow pointing down: furin cleavage site; green rectangle: transmembrane domain. B. ZP2 phylogeny reconstructed by maximum likelihood for 35 mammal species. Carnivora and Caniformia subtrees are featured in light green and brown respectively to indicate that signatures of selection were detected among them. Jaguar terminal branch is colored in red to show the detection of positive selection in that particular terminal branch. C. Three-dimensional model of the protein ZP2 shown in ribbons. N-glycosylated and O-glycosylated sites are shown as yellow and bright green atom spheres, respectively. Positive selected sites are indicated and shown as gray atom spheres. D. Table of positive selected sites detected for *Zp2* showing their position in the jaguar protein sequence and identity for each phylogenetic group. The identity in jaguar for the positive selected site in this species is shown in bold red. Autapomorphies in the background species are not included. E. Detailed view of the ZP-N2 and ZP-N3 portion of ZP2 alignment containing all 4

positive selected sites detected.

3. Figure 3: ZP1 evolution. A. Jaguar *Zp1* gene model showing 12 exons and 11 introns (first line). Mature transcript of *Zp1* gene (second line) aligned with the schematic representation of ZP1 protein. Start and end positions for each domain in the jaguar protein (third line) are displayed above and positive selected sites are indicated with arrowheads (fourth line) below the schematic representation of ZP1 protein domains. Below it, positive selected sites are indicated with colored arrowheads. Brown arrowheads correspond to positive selected sites inside Caniformia subtree. Yellow arrowheads correspond to positive selected sites for the Feliformia subtree. Light green arrows indicate sites under selection in the Carnivora subtree. Split arrowheads filled with two colors denote sites that were detected in more than one analysis and their filling colors correspond to the highlighted subtrees. SP: Signal Peptide; ZP-N1: ZP-N like domain; P: trefoil domain; ZP-N and ZP-C: Zona Pellucida domain N-terminal and C-terminal moieties; blue arrow pointing down: furin cleavage site; TM: transmembrane domain. B. ZP1 phylogeny reconstructed by maximum likelihood for 36 mammal species. Carnivora, Feliformia and Caniformia subtrees are featured in light green, yellow and brown respectively to indicate that signatures of selection were detected among them. C. Three-dimensional model of the protein ZP1 shown in ribbons. N-glycosylated sites are shown as yellow atom spheres. Positive selected sites are indicated and shown as gray atom spheres. D. Table of positive selected sites detected for *Zp1* showing their position in the jaguar protein sequence and their identity for each phylogenetic group of interest. Autapomorphies in the background species are not included. All residues located beyond the furin cleavage site were excluded. E1. Detailed view showing the interdomain region between ZP-N1 and trefoil domains of ZP1 alignment. E2. Detailed view for most of ZP1 ZP domain alignment. Positive selected sites in those regions are highlighted following the color code mentioned above. The Domain scheme is represented above each alignment.

4. Figure 4: ZP4 evolution. A. Jaguar *Zp4* gene model displaying its 11 exons and 10 introns (first line). Mature transcript of *Zp4* gene (second line) aligned with the schematic representation of ZP4 protein. Start and end positions for each domain in the jaguar protein (third line) are displayed above and positive selected sites among the Panthera genus are indicated with blue arrowheads (fourth line) below the schematic representation of ZP3 protein domains. SP: Signal Peptide; ZP-N1: ZP-N like domain; P: trefoil domain; ZP-N and ZP-C: Zona Pellucida domain N-terminal and C-terminal moieties; blue arrow pointing down: furin cleavage site; TM: transmembrane domain. B. Maximum likelihood phylogenetic reconstruction for ZP4 using 34 mammal species. Feliformia suborder and Panthera genus are highlighted in yellow and blue respectively to indicate that positive selection was detected in these subtrees. C. Three-dimensional model of the protein ZP4 shown in ribbons. N-glycosylated and O-glycosylated sites are shown as yellow and bright green atom spheres, respectively. Positive selected sites are indicated and shown as gray atom spheres. D. Table of positive selected sites detected for *Zp4* showing their position in the jaguar protein sequence and their identity for each phylogenetic group of interest. Autapomorphies in the background species are not included. All residues located beyond the furin cleavage site were excluded. E. ZP4 ZP-N and ZP-C domains alignments showing the 2 Panthera positive selected positions highlighted in blue. Domains corresponding to that area are aligned above.

5. Figure 5: ZPs and IZUMO1R sequence identities. Boxenplots graphs for ZP3, ZP2, ZP1, ZP4 and IZUMO1R showing mean sequence identity values (black horizontal line) for Carnivores (light green), Caniformes (brown), Feliformes (yellow), Pantherines (blue) and Feliformes species without Pantherines (light blue). Caniformes vs Feliformes mean identity comparisons are significant for ZP3, ZP2, ZPp1 and IZUMO1R, but not for ZP4.