Positive selection in gamete interaction proteins in Carnivora

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

[1] 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] Barcelona Supercomputing Center (BSC), Institute for Research in Biomedicine (IRB), and Universitat Pompeu Fabra (UPF), 08003, Barcelona, Spain

[†] These authors contributed equally to this work

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

Keywords: ZP, evolution, Felidae, IZUMO1-JUNO, pantherines, jaguar

Abstract

Positive selection in rapidly evolving gamete-interaction proteins has been studied in many mammalian groups, but no such analysis has yet focused on Carnivora species. Here we uncover evolutionary forces acting on the whole family of Zona Pellucida (ZP) proteins and the sperm-oocyte fusion protein pair IZUMO1-JUNO along the *Carnivora* order, including *Caniformia* and *Feliformia* suborers and *Pantherinae* subfamily. A chromosome-resolved jaguar genome assembly was used to retrieve ZP, IZUMO1 and JUNO coding sequences and to reconstruct the evolutionary history of these proteins. Exploring sequence variability in more that 30 *Carnivora* species, studied proteins *Feliformia* displays a significantly lower diversity than its sister taxa, *Caniformia*. Molecular evolution of ZP2 and ZP3, subunits directly involved in sperm-recognition, uncovered the presence of diversifying positive selection in *Feliformia*, *Caniformia* and *Pantherinae*, but no relevant change could be linked to sperm binding. On the other hand, structural cross-linking ZP subunits, ZP4 and ZP1, displayed lower levels or even complete absence of positive selection. The ZP4 only relevant positive selection change (163 P>Y) would affect the flexibility of a proline-rich interdomain region. The fusion proteins IZUMO1 showed the highest amount of positive selection signals and sites while JUNO showed to lack any signatures of positive selection in all tested lineages and clades. Eight positive selected sites were detected in 2 clusters of IZUMO1 that are also known to be involved in the interaction with JUNO. Altogether, our results shed light on the evolution of ZP proteins and the IZUMO1-JUNO gamete interaction pair within *Carnivora* order.

1. Introduction

The zona pellucida (ZP) 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). ZP matrix of most, but not all, eutherians is composed of four ZP protein subunits: ZP1-4 (Figure 1A).

ZP3 was the first glycoprotein of the ZP to be considered indispensable for fertilization in mammals (Bleil & Wassarman, 1980). Independent research findings revealed that mouse oocytes lacking a functional Zp3 gene exhibit infertility and the absence of the ZP (Liu et al., 1996; Rankin et al., 1996). As a result, ZP3 garnered broad consensus in the scientific community as the predominant and indispensable protein responsible for sperm binding.

However, further research revealed that ZP2 is also a protein essential for mouse fertility as it is required for sperm binding (Rankin et al., 2001, Avela et al., 2014). Capacitated sperm attaches to an N-terminal domain of ZP2 in the extracellular ZP 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. Post-fertilization proteolytic cleavage at the N terminus of ZP2 protein prevents sperm binding to the ZP and 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 ZP. ZP1 protein is involved in ZP protein filaments cross-linking and interacts with other ZP proteins, particularly ZP2 and ZP3, to form the ZP matrix, giving and maintaining its structure (Monné & Jovine, 2011; Wassarman & Litscher, 2018; Nishimura et al., 2019; Stsiapanava et al., 2020). Therefore, the main function of ZP1 protein is to contribute to the structure and organization of the ZP.

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 most extended mammalian model in which most studies of the ZP had been performed. ZP4 is considered 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). Like 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).

ZP2-ZP3 filaments are formed through heterodimerization of the ZP2 and ZP3 proteins after ther secretion to the extracellular space by the oocyte. These heterodimers then assemble end-to-end, forming long filaments that extend throughout the ZP matrix (Jovine et al., 2002; Wassarman &

Litscher, 2018). ZP1 forms a scaffold for the assembly of the ZP2-ZP3 filaments, aiding in their arrangement and providing structural integrity to the ZP matrix. ZP4 is a minor component of the ZP matrix but is thought to have a role in mediating intermolecular interactions between ZP2-ZP3 filaments. ZP4 may act as a link between adjacent ZP2-ZP3 filaments, contributing to the overall stability and architecture of the ZP matrix (Wassarman & Litscher, 2018).

The structure of ZP proteins is characterized by different regions and domains crucial for the secretion and assembly of ZP filaments. These are an N-terminal secretory signal peptide (SP) that identifies secreted proteins; the ZP module, a conserved sequence of ~260 amino acids composed of a ZP-N and a ZP-C domain, and include 8 or 10 invariant cysteine residues that form four, or alternatively five, 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; Bokhove et al., 2016; Wilburn & Swanson, 2017).

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 ZP module, divided into its N-terminal and C-terminal moieties, a furin cleavage site and a hydrophobic C-terminal trans-membrane domain (TMD), all of them necessary for their secretion and incorporation in the ZP matrix.

In contrast to the minimalistic domain composition of ZP3, the ZP2 protein stands out for having the most extensive domain count among all members of the ZP protein family. Alongside the distinctive ZP module found in all ZP proteins, which consists of N-terminal and C-terminal domains, ZP2 possesses three extra ZP-N domains situated at the N-terminal region preceding the complete ZP module. These additional domains go from ZP-N1 to ZP-N3.

Beyond the ZP protein common structural organization of domains, ZP1 displays a single ZP-N domain and a trefoil domain, which is important for the cross-linking between ZP filaments (Greve and Wassarman, 1985; Nishimura et al., 2019).

Given that ZP4 and ZP1 are paralogous genes resulting from the duplication of a common ancestral gene, both keep the same domain structure and have a 45 aa long trefoil domain and a ZP-N1 subdomain in their amino-terminal region (Goudet et al., 2008).

The evolutionary history of the ZP protein family reveals numerous gene gains and losses. During the early evolution of amniotes, the duplication event giving rise to *Zp1* and *Zp4* occurred (Hughes and Barratt, 1999; Bausek et al., 2000; Goudet et al., 2008). Subsequent pseudogenization events led to modern placental mammals, which possess a ZP structure composed of three subunits instead of four. Pseudogenization events are more prevalent in *Zp1* gene compared to other ZP genes across placental mammal species. *Zp1* loss has been observed in certain species of the *Caniformia* suborder, including the dog (*Canis familiaris*), dingo (*Canis lupus dingo*), and fox (*Vulpes vulpes*) (Goudet et al., 2008; Moros-Nicolás et al., 2018a). Pseudogenization of *Zp1* gene also occurred in various Pinniped species such as the Antarctic fur seal (*Arctocephalus gazella*) and the Weddell seal (*Leptonychotes*

weddellii) (Moros-Nicolás et al., 2018a). Additional instances of *Zp1* pseudogenization are found in the evolution of Cetartiodactyla, where many even-toed ungulates and cetaceans lack a functional version of this gene (Goudet et al., 2008; Stetson et al., 2012). Furthermore, pseudogenization events affecting *Zp1* have been reported in certain lineages of small primates, such as marmosets and tarsiers (Stetson et al., 2012). In contrast, the pseudogenization of *Zp4* gene has been documented in the house mouse (*Mus musculus*) and other members of the Mus genus (Lefièvre et al., 2004; Izquierdo-Rico et al., 2021). However, it is important to note that ZP3 and ZP2 proteins have not been reported as lacking in any mammalian species, indicating their essential function in gamete interaction and fertilization (Liu et al., 1996; Rankin et al., 1996, 2001).

Considerable efforts have been made to elucidate the sperm interacting partners of egg ZP proteins. However, no sperm surface proteins have been confirmed to interact with egg ZP proteins and be indispensable for fertilization. One of the most well known cases is the ZP3 receptor (ZP3R) protein. ZP3R was initially proposed to be essential for mouse fertilization. However, knockout studies in mice have shown that ZP3R is not indispensable for the fertilization process, suggesting either its non-involvement in sperm-ZP binding or the presence of functionally redundant proteins sperm-egg interaction (Muro et al., 2012). To date, only one essential mammalian sperm-egg interaction protein pair has been identified: IZUMO1 on sperm interacts with egg membrane-anchored JUNO to facilitate binding (Inoue et al., 2005; Bianchi et al., 2014) (Figure 1A).

IZUMO1 was first identified through a systematic proteomic analysis of proteins enriched in the sperm membrane (Inoue et al., 2005). Knockout studies in mice have highlighted the indispensability of IZUMO1 in mediating sperm-egg fusion (Inoue et al., 2005). Structural analyses have identified the IZUMO domain within IZUMO1 as a key component for binding to the oocyte-specific glycoprotein JUNO and initiating the fusion process (Aydin et al., 2016; Ohto et al., 2016). Given its significant role in fertilization, the protein was aptly named "IZUMO1" after a Japanese shrine associated with marriage, emphasizing its critical function in facilitating sperm-egg interaction and highlighting its symbolic relevance to the deity Izumo, who is traditionally believed to preside over marriages and relationships (Inoue et al., 2005).

JUNO protein, also known as folate receptor 4 (FOLR4), is encoded by the gene *Izumo1r*. Originally discovered in 2014 and renamed JUNO after the Roman queen goddess of marriage and childbirth, and it is a transmembrane glycoprotein that serves as the counterpart in the oocyte to the sperm protein IZUMO1. The interaction between JUNO and IZUMO1 is essential for sperm to dock to the surface of an egg (Bianchi et al., 2014), a crucial recognition step prior to membrane fusion (Inoue et al., 2015). Although neither protein has been conclusively shown to be a fusogen, there is recent evidence that IZUMO1 may have fusogenic activity (Brukman et al., 2023).

These proteins are representative components involved in reproduction and are usually found evolving under positive selection. It is proposed that 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). Up till

now, only IZUMO1-JUNO gamete interaction proteins have been identified in mammals; therefore, in this work we will specifically focus on ZP proteins and IZUMO1-JUNO protein positive selection marks in order to look into their role in speciation in the context of the *Carnivora* order focusing *Caniformia-Feliformia* suborders differences and also evaluating selection forces acting on the *Pantherinae* subfamily.

2. Materials and Methods

2.1 Sequence retrieval and alignment

All *Carnivora* species coding sequences of the studied gamete-interacting proteins were downloaded from Ensembl (Howe et al., 2021) and GenBank Nucleotide database (Supplementary tables S2-S7). The number of species from caniforms and feliforms was balanced for all analyses. Therefore, HiC chromosome-length DNA Zoo genomes of Black footed cat (*Felis* nigripes), Jaguarundi (*Herpailurus yagouaroundi*), Clouded leopard (*Neofelis nebulosa*), Snow leopard (*Panthera* uncia), Fishing cat (*Prionailurus viverrinus*) and Leopard cat (*Prionailurus bengalensis*) were downloaded and sequences were obtained via local BLAST searches and manually assembled. Jaguar (*Panthera onca*) sequences were obtained from the first version of the jaguar reference genome (Figueiró et al, 2017) and 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). After a manual inspection and identification of sequencing or assembly errors of both genomes, the sequences that exhibited less fragmentation and missing data were selected. Visual inspection and multiple sequence alignments were performed with MEGA 5.2 (Kumar et al., 2018) using Clustal and/or Muscle (Li, 2003; Edgar, 2004). 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 and the quality of the aligned sequences.

2.2 Positive selection analysis

In order to detect signatures of molecular adaptation the modified Model A branch-site test 2 of positive selection (Zhang et al., 2005) was used. This model is available in codeml program from the PAML4 package (Yang, 2007). This model has 2 variants: the focal branch variant, used to detect episodic positive selection, and the subtree variant, used to detect divergent evolution. 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. Both models were applied for each of the tested groups in this work: *Feliformia*, *Caniformia* and *Pantherinae*. 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 omega parameters 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 ω_0 and ω_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.

All positive selection analyses were performed using multiPAML 2.5. An in-house script developed to efficiently and easily run a large number of PAML analyses testing for multiple hypothesis in one or many genes in a single script run. The multiPAML 2.5 software code is publicly available at https://github.com/fpisciottano.

2.3 Gene trees reconstruction

Gene trees were reconstructed by ML from the sequence alignment data using the Jones-Taylor-Thornton (JTT) amino acid substitution model with partial deletion treatment for missing data. 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. Given that this test for positive selection detection is a very parameter-rich model, it was implemented 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.

2.4 Positive selected sites

Positive selected sites were obtained by the Bayesian posterior analysis (BEB) implemented in codeml's branch-site specific positive selection test *aposteri* 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 in all 3 replicates in each of the 2 CodonFreq variants applied ('codon table' and '3x4') with a p-value of at least 0.8 in all reruns.

2.5 Sequence Identities

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

Statical relevance of mean sequence identities differences between proteins was assessed by ANOVA and Scheffé post-hoc multiple comparison were conducted.

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 through the 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; Kato 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).
The analyses start with an extensive compilation of all accessible coding sequences for the orthologs of gamete interaction proteins ZPs and IZUMO1-JUNO in all *Carnivora* species for which sequencing data is available in biological databases (Supplementary tables S2-S7). Jaguar sequences for ZP proteins and the IZUMO1-JUNO pair were manually assembled and reported for the first time in this paper (see online supplementary data). The complete sequences of the jaguar *Zp2, Zp3, Zp4 and Zp1* genes were obtained using BLASTn searches over the jaguar chromosome-resolution Hi-C

genome assembly and/or the Illumina sequenced first draft of the jaguar genome assembly (Figueiró et al., 2017) (Supplementary table S1).

3.1 Orthologous sequences retrieval and multiple sequence alignment construction

Panthera pardus (leopard) orthologous coding sequences were used as queries in each case (*Zp2* GeneID: 109255525, *Zp3* GeneID:109250568, *Zp4* GeneID:109248980, *Zp1* GeneID:109246518). Jaguar *Zp2* coding sequence comprises 2,151 nucleotides distributed across 18 coding exons (Figure 1B), exhibiting a conserved exon arrangement among all sequenced felid species, except for the jaguarundi (*Puma yagouaroundi*) ortholog, which displays 19 exons, and two puma (*Puma concolor*) alternative transcripts (XM_025916776.1 and XM_025916778.1), both containing an additional exon. For the jaguar *Zp3* gene, attempts to obtain the sequence from the first version of the jaguar genome were unsuccessful, yielding only a partial and fragmented sequence. However, using the complete sequence derived from the *Panthera onca* Hi-C genome assembly, it was confirmed that the jaguar *Zp3* gene is composed of 1,272 nucleotides distributed across eight exons (Figure 1C), consistent with all pantherine species orthologs and also with most sequenced felid species, except for the puma (*Puma concolor*) ortholog, which contains an additional exon, totaling nine exons. As for the jaguar *Zp4* gene coding sequence, it consists of 1,707 nucleotides organized into 11 exons (Figure 1D), encoding a 569 amino acid-long polypeptide. The complete coding sequence for Jaguar *Zp4* was obtained by conducting a BLASTn search using the leopard *Zp4* sequence as the query. Finally, the jaguar *Zp1* gene encompasses 1,818 nucleotides across 12 exons (Figure 1E), mirroring the exon quantity, grouping, and distribution pattern observed in other sequenced felids. Regarding JUNO protein coding gene, *Izumo1r*, complete sequence was obtained through local BLASTn over the Illumina genome assembly (Figueiró et al., 2017) and the chromosome-resolution Hi-C genome assembly using leopard ortholog coding sequence as query (GeneID:109267358). Jaguar *Izumo1r* gene is only 729 nucleotides in length and is composed of 4 coding exons (Figure 1F) and its locus structure, number and disposition of exons is well conserved along all available felid species. Jaguar 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. On the other hand, the jaguar coding sequence of *Izumo1*, it was obtained through a local BLASTn of the lion *Izumo1* sequence (Ensembl gene ID: ENSPLOG00000009340) over the Hi-C genome assembly of *Panthera onca*. Jaguar *Izumo1* coding sequence is composed of 1,037 nucleotides distributed in 9 exons (Figure 1G) and encodes a 346 aa long protein.

In addition to individual jaguar sequences of the target genes, we carried out a comprehensive search, collection, assembly, and curation of all ortholog sequences available for these genes in *Carnivora* species (Supplementary tables S2-S7). It is worth noting that none of the sequences used in this study have been reported to meet high-quality standards or undergone manual annotation to date. Consequently, all sequences required rigorous review and curation during the multiple sequence alignment building process. As a consequence of the approach chosen to compensate caniforms and

feliforms number of species (see Materials and Methods section), the multiple sequence alignments (MSA) achieved displayed a more balanced representation between caniforms and feliforms (see online supplementary data). The majority of the compiled final alignments comprise 31 species, with a roughly equal distribution of caniform and feliform sequences. Among the feliform sequences, these alignments include six pantherine species, encompassing all five species within the *Panthera* genus as well as the clouded leopard of the *Neofelis* genus. However, in the ZP1 alignment, only 28 species are present due to pseudogenization of sequences from species within the *Canidae* family, allowing the use of only 12 caniform species. The IZUMO1 alignment is the other exception, resulting in a total of 33 sequences. These sequences were later used for a sequence diversity characterization of the studied proteins through identity analysis. We contrasted the levels of molecular variability of these proteins within the order *Carnivora*. Next, we further examined the two suborders -*Caniformia* and *Feliformia*- that constitute the *Carnivora* order. These MSAs were also the basis for phylogeny reconstructions and the positive selection analyses reported further below.

3.2 Gamete Interaction Protein Identity

To begin the evolutionary study of ZP proteins and the IZUMO1-JUNO pair, a general characterization of their evolution rates was determined, exploring the variability within the *Caniformia* order. The sequence identity revealed that the ZP proteins exhibited comparable levels of variability, except for ZP1, which displayed significant differences from ZP2 and ZP3, but not from ZP4. Similarly, the JUNO protein showed variability levels comparable to those of the ZP proteins. In contrast, IZUMO1 displayed a notably higher variability when compared to all other studied proteins, presenting significant differences from all of them (Figure 2A and supplementary table S8). Additionally, the levels of variability among groups within the order *Carnivora* for each protein showed that feliform sequences were similar to one another, while caniform sequences exhibited a more pronounced variability. The highest degree of similarity was observed in the sequences of the *Pantherinae* subfamily. Welch's T-test over the mean identity values of the amino acid variability indicated that all the studied proteins exhibited significant differences between caniforms and feliforms (Figure 2B and supplementary table S9). The statistical difference in ZP4 identity values between caniforms and feliforms (Welch's T-test p-value = 0.03649) proved to be smaller in comparison to the differences observed in the other proteins (Welch's T-test p-values: ZP2: $1.26e^{-4}$; ZP3: $2.57e^{-4}$; ZP1: $4.25e^{-4}$; IZUMO1: $1.28e^{-4}$; JUNO: $8.1e^{-5}$) (Supplementary table S9). No significant differences were detected between the mean identities of caniforms and feliforms for the control ACTN1 protein shown at the bottom of Figure 2B (Supplementary table S9) .

The positive selection analyses began with ZP isoforms directly involved in sperm-recognition, ZP2 and ZP3, and continued with the analysis of structural cross-linker ZP proteins, ZP4 and ZP1. Finally, we extended the analyses to the sperm–egg fusion pair IZUMO1-JUNO. Two control proteins: one under non-selection and another under previously reported positive selection in jaguar were added.

3.3 ZP proteins Positive Selection Analyses

The positive selection analyses revealed diversifying positive selection effects on the primary receptors ZP2 and ZP3 within the suborders of *Carnivora*. Specifically, ZP2 exhibits positive selection in the *Feliformia* suborder, while ZP3 shows positive selection in the *Caniformia* suborder (Figure 3A, B and and supplementary tables S10, S11). However, no signs of positive selection were found in the basal lineages of these suborders. As for the *Pantherine* family, only ZP2 exhibits diversifying selection effects (Figure 3A). Conversely, the subunits involved in cross-linking between ZP protein filaments ZP1 and ZP4 presented a notably lower prevalence of positive selection. No diversifying positive selection was detected in any of the analyzed clades, and a positive selection signal was detected solely in the basal lineage of caniforms for the ZP4 protein (Figure 3C and supplementary table S12). ZP1 did not display any selection signals in any of the analyzed clades or branches (Figure 2D and supplementary table S13).

3.4 ZP proteins Positive Selected Sites

ZP2 evolutionary analyses consistently revealed 2 candidate sites subjected to positive selection: jaguar positions 231 Q>V and 324 S>V/F/L, located in ZP-N2 and ZP-N3 respectively (Figure 4A, 4B and supplementary figurse S15, S16). Position 231 displays a glutamine (Q) for all caniform species and for some, but not all, feliform species. The remaining feliform species changed to a valine (V). While some feliform species conserve the glutamine, the feliform species that change to a valine involve different feliform families, including the pantherines. This indicates that this change occurred repeatedly and independently many times along the *Feliformia* suborder. On the other hand, position 324 is a serine (S) in all caniform species but changes to valine (V) and phenylalanine (F) in feliforms and to leucine (L) in some pantherine species (Figure 4C). It is important to mention that in jaguar ZP2 protein the sperm-binding region extends from position 55 to position 154 and ovastacine cleavage site appears between positions 171-172 (Figure 4A).

The positive selection analysis of ZP3 revealed two sites that contribute to the diversifying selection signal within the *Caniformia* suborder. These sites are located at the N-terminal end of the protein and reside within a region that does not correspond to any known domain (Figure 4D and 4E). The adjacent sites are identified as follows: 27 E/G>S/W/Q/P and 28 D>R/T/G (Figure 4F and Supp. Figure S18, S19). Both sites exhibit acidic amino acids in feliforms, while in caniforms, they display a broader variability encompassing various amino acid types. Specifically, site 27 is a glutamic acid (E) found in most feliform species, except for the tiger and the snow leopard, which present a glycine (G). Conversely, among caniforms, site 27 exhibits a variety of amino acids, including serine (S), proline (P), and tryptophan (W). On the other hand, site 28 contains an aspartic acid (D) in all feliform species and displays variability among caniforms, where it can be either an arginine (R), glycine (G) or threonine (T) (supplementari figures S18, S19). These sites are not located within any protein domain and, in the three-dimensional protein model, they are not found in proximity to N- or O-glycosylation sites (Figure 4E).

Moving on to the ZP subunits involved in cross-linking, sites under positive selection were only detected for ZP4. The single selection signal detected for this protein was particularly localized in the basal lineage of the caniforms (Figure 3C). Specifically, the changes detected in ZP4 for the basal caniform lineage were at positions 163 P>Y and 556 F>A (Figure 4I and supplemetary figure S21, S22).
Site 163 is located in the N-terminal portion of the ZP4 protein, in an inter-domain region between the ZP-N1 domain and the trefoil domain (Figure 4G and 4H). In feliforms, this site is a proline (P), while in caniforms, it is a tyrosine (Y). Jaguar ZP4 protein structural model and disorder prediction data shows that site 163 is located in an unstructured and disordered proline rich region (Figure 4H and supplementary figure S20). This site is spatially close to the N-glycosylated site at position 237, which belongs to the ZP-N domain of the ZP module (Figure 4H).

Site 556 is a phenylalanine (F) in all feliforms and an alanine (A) in all caniforms. This site it is located at the beginning of the transmembrane region that anchors the immature protein to the oocyte membrane (Figure 4G, 4H). The entire C-terminal end portion of ZP proteins, including this transmembrane region, is cleaved to release the mature protein and allow its integration into the ZP. The C-terminal portion remains anchored to the membrane and is subsequently degraded. Moreover, both alanine and phenylalanine belong to the group of hydrophobic amino acids, characteristic of a transmembrane region. The 556 F>A change does not alter the nature of this amino acid, but it could produce a steric difference due to the size discrepancy between the lateral groups of these amino acids.

There are no positive selected sites to be reported in ZP1 the analyses showed an absence of positive selection in the basal lineages of *Caniformia*, *Feliformia*, and the *Pantherine* family, as well as an absence of diversifying positive selection within these clades (Figure 4J, 4K).

All domain arrangements for each subunit of jaguar ZP proteins were successfully predicted through AlphaFold modeling (Figure 4B, 4E, 4H, and 4K) (Supp figures S14, S17, S20, S23). These structural models were also used to spatially locate the N-glycosylated and O-glycosylated sites in an attempt to discover if they were positioned close to the detected selected sites as these sites are likely involved in ZP polymerization (Monné et al., 2008).

3.5 IZUMO1-JUNO positive selection analyses

Regarding the IZUMO1-JUNO interaction pair, an evolutionary study was carried out to find out which were the main evolutionary forces that shaped the history of each of the members of this pair involved in the fusion between gametes. The first to highlight is that, of all the proteins analysed in this work, IZUMO1 was the only one for which we were able to obtain predicted sequences with good quality for all the species of the *Carnivora* order. While JUNO also yielded a substantial number of coding sequences from *Carnivora* species, two sequences from Ermin (*Mustela erminea*) and the American black bear (*Ursus americanus*) were excluded due to sequencing quality issues. Despite these minor differences, assembling and curating both alignments proved relatively

straightforward, with ample availability of coding sequences among *Carnivora* species and minimal annotation difficulties. However, as we reconstructed the phylogenetic trees for both genes, intriguing differences in their evolutionary histories began to emerge. The phylogenies of both genes yield the same general topology between clades of the *Carnivora* order. However, within the suborder Feliforma, it was not possible to make valid topology comparisons between trees beyond the split between species in the family *Felidae* and species belonging to other families of feliforms (Figure 5A). This limitation arose due to the low levels of branch support (<30%) obtained for JUNO along the whole *Felidae* family. While a few cases of low branch support were also detected in the *Felidae* family for IZUMO1, this phenomenon was significantly more noticeable in the case of JUNO, where it was impossible to draw confident conclusions about the topology of the *Felidae* family due to this limitation.

Significant differences emerged from the positive selection analyses of IZUMO1 and JUNO. The evolutionary analysis of IZUMO1 revealed signatures of adaptive evolution in the basal lineages of both *Caniformia* (p-value = $5.12e^{-10}$) and *Feliformia* (p-value = 0.0366) suborders, but not in the basal *Pantherinae* lineage. Moreover, strong diversifying positive selection marks were detected within the *Caniformia* suborder (p-value = $2.18e^{-17}$) (supplementary table S24). It is worth noting that selection signatures found in IZUMO1, both episodic and diversifying, were significantly more intense in the *Caniformia* suborder than in the *Feliformia* basal lineage (supplementary table S24). It is also important to highlight that such selection signatures were absent in the *Feliformia* suborder or the *Pantherine* subfamily (Figure 5A).

In contrast to the abundance of positive selection results for IZUMO1, the evolutionary analysis of JUNO revealed a complete absence of selection signals across all clades and basal lineages analysed (Figure 5A, 5E, 5F, Supplementary table S29 and Supplementary figure S30).

The results of the positive selection analyses for IZUMO1 revealed a considerable number of positively selected sites, precisely 69 sites (Figure 5B and Supplementary table S27). The majority of these sites correspond to adaptations in the basal lineage of caniforms and sites that are undergoing diversifying evolution within the *Caniformia* suborder. Only three of these 69 sites were detected in the basal lineage of feliforms. These sites are: 87 S/A>V; 181 R>K; 268 L/H/Q>T (Supplementary figure S26). This difference in the number of detected sites between the *Caniformia* and *Feliformia* suborders is not surprising when we contrast the resulting selection intensity values for these clades. The p-values when testing the hypothesis of selection against a null model for the basal lineage (indicative of adaptation through episodic selection) and the subclade (indicative of diversifying selection within the clade) of the *Caniformia* suborder were $5.12e^{-10}$ and $2.18e^{-17}$, respectively. In contrast, the p-value for the hypothesis test that detected positive selection in the basal lineage of the *Feliformia* suborder was 0.0366 (Supplementary table S24).

3.6 IZUMO1 Positive Selected Sites

Positively selected sites in IZUMO1 are uniformly distributed throughout the protein, with two important exceptions—the signal peptide and the transmembrane domain—which appear devoid of

positively selected sites (Figure 5B). Several of the many IZUMO1 sites that were identified as having changed under the influence of positive selection in *Caniformia* suborder correspond to sites reported as important for the interaction with JUNO (Kato et al., 2016). Given the large number of detected sites for IZUMO1, we will focus only on those (the complete list of IZUMO1 positive selected sites can be found in supplementary table S27). These sites can be divided into two clusters of closely located sites (Figure 5B and Supplementary figure 25). The first cluster includes sites 70, 71, and 72, situated in the 4HB domain, the four alpha helices bundle of the N-terminal region of the protein. The second cluster is located in the HINGE domain and comprises sites 143, 148, 149, 155, and 156.

Site 70 is a leucine (L) in all feliforms, except for the meerkat (*Suricata suricata*), which displays a serine (S). In contrast, caniforms show much more variability at this site, changing to aspartic acid (D) in bears, glutamic acid (E) in pinnipeds, and histidine (H) in mustelids (Supplementary figure S26). These changes involve the transition from a hydrophobic amino acid to variants with charge (Supplementary figure S28).

Adjacent to the previous site is site 71 E>N. This site contains a glutamic acid (E) in almost all species of the *Carnivora* alignment, except for species in the Canidae family (Supplementary figure S26). These species exhibit an asparagine (N), leading to a change from a negatively charged amino acid to an uncharged polar one (Supplementary figure S28). The electrostatic charge variation induced by the two aforementioned changes can be visualized on the Coulombic surface in figure 5D.

Site 73 A>T/S/Q presents an alanine (A) in all species of the *Felidae* family and a threonine (T) in the striped hyena (*Hyaena hyaena*) and meerkat (*Suricata suricata*), feliform species outside the *Felidae* family. On the other hand, in caniforms it shows a glutamine (Q) for species in the Canidae family and a serine (S) in all other caniform species (Supplementary figure S26).

Regarding the second cluster of positive selected sites crucial for the interaction of IZUMO1 with its counterpart JUNO, we can start with site 143 S>N/H. This site shows a serine (S) in all feliform species and an asparagine (N) in the majority of caniform species (Supplementary figure S26). This change may not be highly significant as both amino acids have uncharged polar side chains. However, in species of the Canidae family, this position shows a histidine (H), a positively charged amino acid (Supplementary figure S28).

Site 148 Q>R is a glutamine (Q) for most of the species in the alignment, both caniforms and feliforms. While it remains almost constant in feliforms, changing only in the meerkat (*Suricata suricata*), in caniforms, it changes to an arginine (R) in the ferret (*Mustela putorius*) and in canid species (Supplementary figure S26).

Site 149 L>V/I is a leucine (L) in all feliforms, but it changes to a valine (V) in caniforms and then again to an isoleucine (I) in canid species (Supplementary figure S26).

Sites 155 and 156 are adjacent and both are under the effects of diversifying positive selection. Site 155 S>P/T/H/S is a serine (S) in almost all feliform species, except for the meerkat, while in caniforms, it shows significant variability between serine (S), histidine (H), proline (P), and threonine (T) (Supplementary figure S26). Lastly, we have site 156 T/Y>A/P/S/Y. This site shows a threonine (T) in all felid species and a tyrosine (Y) in all non-felid feliform species. In caniforms, this site also displays substantial variability between alanine (A), proline (P), serine (S), and tyrosine (Y) (Supplementary figure S26).

A general summary of the positive selection results for all the proteins studied across all analyzed clades can be observed in figure 6.

3.7 Control proteins: ACTN1, a neutrally evolving protein, and TPM3, a jaguar positively selected protein

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 S31 shows the negative expected results.

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 S32 shows a p-value=0.0006.

4. Discussion

This study sheds light the evolution of ZP proteins and the IZUMO1-JUNO gamete interaction pair within the order *Carnivora*. The rapid evolution of reproductive proteins can make a significant contribution to reproductive isolation between diverging taxa, given the possibility that changes in these proteins may lead to isolation between species (Swanson et al., 2001; Swanson & Vacquier, 2002; Turner & Hoekstra, 2008). In order to assess this in in the *Carnivora* order, a comprehensive compilation of all available coding sequences from orthologs across species within this order was required. As a result, the most complete and curated collection of aligned carnivorans sequences for these fertilization-related proteins to date was obtained. Initially, sequence identity analyses were carried out to characterize and compare the amino acid diversity between these gamete-interaction proteins, as well as to evaluate differences between clades within the *Caniformia* suborder. Subsequently, site-lineage-specific analyses of positive selection were performed to detect signals of molecular adaptation and diversifying evolution. These hypotheses were tested for both *Carnivora* suborders, Caniforms and Feliforms, as well as the *Pantherinae* subfamily. In cases where sites under positive selection were identified, these were mapped onto structural models of these proteins generated for jaguar sequences. This allowed us to gain insights into the evolutionary forces acting upon gamete interaction proteins in the different clades of the *Carnivora* order.

Multiple driving forces have been proposed to explain the occurrence of positive selection in reproductive loci involved in speciation. One prominent hypothesis is sperm competition, which refers to the competition between sperm from different males for the fertilization of a female's eggs. This theory posits that sperm competition instigates a continuous and adaptive "arms race," wherein the selective intensity is expected to correlate with the degree of polyandry (Clark et al., 2006). Under this hypothesis, male proteins involved in various aspects of fertilization, including locating,

reaching, binding, penetrating, and fusing with the egg, may undergo adaptive evolution.
On the other side, given the larger energy investment in female gametes, polyspermy can have more detrimental effects on female fitness than male fitness. Consequently, female gamete proteins are expected to evolve to lower the fertilization rate, while sperm proteins continuously strive to increase it within a competitive context. This leads to the concept of sexual conflict over adaptive optima, whereby females and males engage in counter-adaptations, resulting in a co-evolutionary chase between male and female characters (Swanson & Vacquier, 2002). This conflict may maintain either polymorphism or extend a co-evolutionary competition between egg and sperm proteins (Turner & Hoekstra, 2008). Sperm competition, which promotes rapid rates of fertilization, may be advantageous for males, whereas females may benefit from a more moderate fertilization rate to prevent polyspermic fertilization.
In addition to sexual conflict and sperm competition, sexual selection plays a crucial role in shaping mating behavior and displays. It is conceivable that sexual selection also operates at the gametic level, wherein an egg exhibits a preference for a specific sperm protein allele, resulting in assortative mating (Eberhard, 1996).
Another driving force is reinforcement, which refers to the evolution of reproductive barriers to prevent hybridization. When allopatric populations hybridize and produce offspring with reduced fitness, reinforcement explains the divergence observed in gamete recognition proteins (Swanson & Vacquier, 2002).
While the aforementioned forces primarily stem from endogenous factors within the reproductive system of the species, it is important to consider exogenous forces as well. Pathogen resistance, for instance, represents an exogenous force that may drive divergence at these reproductive loci. Microbial attacks exert constant pressure on gamete surface proteins, necessitating their continual adaptation to evade these attackers (Vacquier et al.,1997).
It is important to highlight that several of these hypotheses share overlapping predictions, rendering their differentiation challenging (Clark et al., 2006; Carlisle & Swanson, 2021).

Characterization of jaguar gamete interaction proteins

All jaguar sequences for ZP proteins and the IZUMO1-JUNO fusion pair obtained during this study show a high identity levels with sequences from other species of the *Pantherinae* subfamily. This effect was also observed to some extent in the suborder *Feliformia*, but not in the sister suborder *Caniformia*. In terms of locus structure preservation, each of the studied genes presents a high consistency in number of coding exons and their distribution and grouping along the locus for all feliform species. The only exception is the species of the puma taxonomic group (*P. concolor* and H. yagouaroundi), that present a higher locus structural variability.

ZP proteins evolution

A priori, it might be proposed that reproductive genes under positive selection would be expected to play secondary roles in reproductive mechanisms, while proteins essential for fertilization would generally remain conserved. However, positive selection has been detected numerous times in reproductive proteins, challenging this expectation (Grayson et al., 2015). The findings presented in this study support this apparent contradiction. Essential ZP proteins subunits directly involved in fertilization and oocyte interaction, ZP2 and ZP3, display positive selection signatures that do not

involve changes in sperm recognition regions demonstrated in other model organisms. On the other hand, the subunits that play a secondary role structuring the ZP matrix, ZP1 and ZP4, exhibit lower or even no levels of positive selection. Moreover, ZP1 and ZP4 are the two subunits of the ZP protein family that appear pseudogenized in some mammalian species (Moros-Nicolás et al., 2018a; Goudet et al., 2008; Lefièvre et al., 2004) consistent with the fact that fertilization can still occur even in their absence, possibly due to a functional redundancy between them.

ZP2 positive selected sites, 231 Q>V and 324 S>V/F, in the jaguar would not affect ZP2 sperm-recognition region reported as essential in mice (Gahlay et al., 2010; Avella et al., 2014; 2016). After fertilization the ZP2 N-terminal region is cleaved by the ovastacine metalloproteases in order to prevent polyspermy (Burkart et al., 2012). Positive selected amino acids in ZP2 do not affect ovastacine cleavage site in feliforms or pantherines, as this sequence appears in the jaguar sequence between positions 171-172. However, as they are located in N-terminal ZP-N domains, in ZP-N2 and ZP-N3, respectively, they may be involved in the interaction between ZP filaments (Jovine et al., 2002; 2005).

Both positive selected sites revealed for ZP3, 27 E/G>S/W/Q/P and 28 D>R/T/G, are adjacently located at the N-terminal end of the mature protein. These sites hold the strongest diversifying selection signal for caniforms in this protein. In feliforms, both sites show acidic amino acids, while in caniforms, they exhibit a wider range of variability, including various amino acid types. This change in charge and the N-terminal location of these amino acids could indicate that they vary in the suborder *Caniformia* due to an external factor of selective pressure, such as sperm competition, sexual conflict, sexual selection, or escape from pathogens. However, the location of these sites in the N-terminal extreme of ZP3 protein and outside any domain or functional region would indicate that they are not involved in ZP3 function.

The selection signature detected for ZP4 is specifically located in the basal lineage of the *Caniformia* suborder. This, unlike the selection signatures detected for ZP2 and ZP3, suggests an episodic process that promoted the appearance of variants that were subsequently conserved. This means that detected adaptive changes (163 P>Y and 556 F>A) in this lineage occurred early in the history of this suborder, after their split from *Feliformia*, and were fixed by selection. Taking into account that these changes took place in the early history of the *Caniformia* suborder and were later fixed, they may be involved in the surge of a prezygotic isolating mechanism between caniforms and its sister group *Feliformia*. The change 163 P>Y is more relevant than 556 F>A, which is not present in the mature protein and keeps the same amino acid hydrophobic nature. Site 163 is located in the unstructured and disordered region between ZP-N1 and trefoil domains. The accumulation of prolines in this inter-domain sequence provides stiffness to this linker region (Krieger et al., 2005). The change 163 P>Y would reduce this stiffness and increases the mobility of the region and between the adjacent domains.

ZP1 was the only ZP protein that showed no signatures of positive selection in any of the lineages or clades tested. Although ZP1 and ZP4 remain by far the less studied components of the mammalian ZP, Jovine's lab structurally characterized ZP1 cross-linking function in human proteins (Monné &

Jovine, 2011; Nishimura et al., 2019). Studies of human ZP1 mutant protein suggest that lack of a ZP in infertile patients carrying the corresponding mutation could be due to impaired secretion of ZP1 and the consequent absence of ZP filament cross-linking. As this step is prior to the cross linking with ZP3 it cannot discriminate how this region of ZP1 would interact with ZP3 if at all (Huang et al., 2014). On the other hand, the trefoil/ZP module region of ZP1 does not form non-covalent homodimers in humans, which shows that ZP1-N1 necessarily participates in cross-linking to maintain human ZP with its consequences over fertility (Nishimura et al., 2019).

A number of background results would indicate a lack of interespecific isolation barrier in pantherines indicating that the existence of positive selection in gamete interaction proteins is permissive for hybridization. The background results 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); natural hybrids between all panther species (Yadav et al., 2019); and the high efficiency in hybrids obtained by IVF of oocytes with whole ZP matrix between different felines (Duque Rodriguez et al., 2021). Successful fertilization relies on the molecular compatibility between gametes. As long as sperm and egg can recognize and bind via their surface proteins, gamete fusion may transpire between individuals of distinct species, and this would be permissive of hybridization events, which could impact on the speciation process (Gert et al., 2023).

IZUMO1-JUNO evolution

Sequence identity analyses revealed that the sperm protein IZUMO1 displays significantly higher variability than the oocyte proteins. It is likely that this difference between sperm proteins and oocyte proteins is due to sperm competition. This would explain the widespread diversifying selection detected for IZUMO1. On the other side, its oocyte partner, JUNO, rendered no signature of positive selection in any of the lineages or clades tested. However, it displays variability comparable to other oocyte proteins. Therefore, the absence of positive selection signals in JUNO is not due to an extremely conserved nature of this protein but rather to a strictly neutral evolution process. Similar strikingly differences in evolutionary forces had been previously identified acting on gamete interaction proteins in abalone Lysin-VERL pair (Swanson et al., 2001). However, for the mammalian IZUMO1-JUNO pair a coordinated evolution hypothesis has been proposed given that both proteins experience positive selection in mammals, although these signals appear in different clades for each partner (Grayson et al., 2015). The phylogenetic scope used in the present work is tighter, as it focuses on the *Carnivora* order, and results obtained for JUNO indicate a total absence of positive selection in this group.

A recent article reported that the tryptophan at position 148 of IZUMO1 was crucial for the binding with JUNO (Brukman et al., 2023). This position is not affected by positive selection and is completely conserved in all *Carnivora* species. All 8 IZUMO1 positive selected sites detailed here are also known to be involved in the interaction with JUNO (Kato et al., 2016). Therefore, they may be used to generate mutations in future functional studies seeking to gain insight in IZUMO1-JUNO binding in carnivorans.

Final conclusions

Here we reported 8 IZUMO1 positive selected sites that are also involved in JUNO interaction. No other positively selected site that appears to be directly involved in sperm-egg binding was found in the remaining gamete-interacting proteins. ZP2 changes observed for feliforms and pantherines may be affecting the interaction of this protein with other filaments that structure the ZP matrix, but they do not reside in the known region for sperm interaction as described in other mammalian models. ZP3 positively selected sites detected in caniforms are not located within any functional domains of this protein. Lastly, ZP4 sole detected site that is present in mature protein seems to influence the flexibility of an interdomain region. ZP1 and JUNO, on the other hand, exhibited a complete absence of positively selected sites in carnivores.

The main limitation of this work lay in the difficulty of establishing a functional role for the detected positive selected sites. Another significant limitation is reliably discerning the origin of positive selection. While this process is typically associated with molecular adaptation in a traditional way, there are various hypotheses explaining its occurrence in reproductive proteins, and distinguishing between them poses a challenging task. Although it is still widely believed that positive selection directly influences egg-sperm recognition, we do not exclude the possibility that positive selection may have in fact an indirect role. Experimental evidence comes from VERL/lysin complexes in which positive selection at the level of VERL repeat 1 (and to a lesser extent VERL repeat 2) generates a binding affinity gradient for lysin that results in species-specificity (Raj et al., 2017). However, this is not done by directly affecting residues at the interface. In fact, the divergence of repeat 1 may even be due to completely different reasons - for example pathogen avoidance (Raj et al., 2017).

Altogether, these results indicate that the presence of positive selection in gamete interaction proteins cannot be considered as evidence for gamete incompatibility, in the same way that lack of detected positive selection cannot be considered as evidence for gametic compatibility. Positive selection might be present in a primary sperm-egg receptor and still be permissive for gamete binding or fusion. Similarly, a prezygotic isolation barrier can arise due to gametic incompatibility in a scenario completely devoided of positive selection, driven solely by neutral changes.

To the best of our knowledge, no study to the date has conducted such a meticulous and detailed investigation into the evolution of these gamete-interaction proteins in the *Carnivora* order. Furthermore, no previous work had been able to present differences between feliforms and caniforms for these proteins in such a comprehensive manner.

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 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.

Online supplementary data: https://figshare.com/articles/dataset/_b_Alignments_b_/23896791

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.

Adavoudi, R., & Pilot, M. (2021). Consequences of Hybridization in Mammals: A Systematic Review. *Genes*, *13*(1), 50

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.

Baibakov, B., Boggs, N. A., Yauger, B., Baibakov, G., & Dean, J. (2012). Human sperm bind to the N-terminal domain of ZP2 in humanized zonae pellucidae in transgenic mice. *The Journal of cell biology*, *197*(7), 897–905. https://doi.org/10.1083/jcb.201203062

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.

Bleil, J. D., & Wassarman, P. M. (1980). Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm. *Cell*, *20*(3), 873–882. https://doi.org/10.1016/0092-8674(80)90334-7

Bokhove M., Nishimura K., Brunati M., Han L., de Sanctis D., Rampoldi L., & Jovine L. (2016). A structured interdomain linker directs self-polymerization of human uromodulin. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(6), 1552–1557.

Bokhove M., Jovine L. (2018). Structure of Zona Pellucida Module Proteins. In Current Topics in Developmental Biology (1st ed., Vol. 130). Elsevier Inc.

Brukman, N. G., Nakajima, K. P., Valansi, C., Flyak, K., Li, X., Higashiyama, T., & Podbilewicz, B. (2023). A novel function for the sperm adhesion protein IZUMO1 in cell-cell fusion. *The Journal of cell biology*, *222*(2), e202207147. https://doi.org/10.1083/jcb.202207147

Burkart, A. D., Xiong, B., Baibakov, B., Jiménez-Movilla, M., & Dean, J. (2012). Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy. *The Journal of cell biology*, *197*(1), 37–44. https://doi.org/10.1083/jcb.201112094

Carlisle, J. A., & Swanson, W. J. (2021). Molecular mechanisms and evolution of fertilization proteins. *Journal of experimental zoology. Part B, Molecular and developmental evolution*, *336*(8), 652–665. https://doi.org/10.1002/jez.b.23004

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.

Clark, N. L., Aagaard, J. E., & Swanson, W. J. (2006). Evolution of reproductive proteins from animals and plants. *Reproduction (Cambridge, England)*, *131*(1), 11–22. https://doi.org/10.1530/rep.1.00357

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.

Eberhard W.G. (1996). Female Control: Sexual Selection by Cryptic Female Choice. Princeton: Princeton University Press.

Edgar R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

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.

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.

Gert, K.R.B., Panser, K., Surm, J. *et al.* Divergent molecular signatures in fish Bouncer proteins define cross-fertilization boundaries. *Nat Commun* 14, 3506 (2023). https://doi.org/10.1038/s41467-023-39317-4

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.

Greve J. M., & Wassarman P. M. (1985). Mouse egg extracellular coat is a matrix of interconnected filaments possessing a structural repeat. *Journal of molecular biology*, *181*(2), 253–264.

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

Inoue, N., Ikawa, M., Isotani, A., & Okabe, M. (2005). The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. *Nature*, *434*(7030), 234–238. https://doi.org/10.1038/nature03362

Inoue N., Hagihara Y., Wright D., Suzuki T., Wada I. (2015). Oocyte-triggered dimerization of sperm IZUMO1 promotes sperm-egg fusion in mice. *Nature communications*, *6*, 8858.

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.

Jovine L., Qi H., Williams Z., Litscher E. S., Wassarman P. M. (2004). A duplicated motif controls assembly of zona pellucida domain proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 5922–5927

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

Kato, K., Satouh, Y., Nishimasu, H. *et al.* Structural and functional insights into IZUMO1 recognition by JUNO in mammalian fertilization. *Nat Commun* **7**, 12198 (2016). https://doi.org/10.1038/ncomms12198

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.

Krieger, F., Möglich, A., & Kiefhaber, T. (2005). Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide chains. *Journal of the American Chemical Society*, *127*(10), 3346–3352. https://doi.org/10.1021/ja042798i

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.

Li K. B. (2003). ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics (Oxford, England)*, *19*(12), 1585–1586. https://doi.org/10.1093/bioinformatics/btg192

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., Han, L., Schwend, T., Burendahl, S., & Jovine, L. (2008). Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature*, *456*(7222), 653–657

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

Morgan, A. A., & Rubenstein, E. (2013). Proline: the distribution, frequency, positioning, and common functional roles of proline and polyproline sequences in the human proteome. *PloS one*, *8*(1), e53785. https://doi.org/10.1371/journal.pone.0053785

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

Muro, Y., Buffone, M. G., Okabe, M., & Gerton, G. L. (2012). Function of the acrosomal matrix: zona pellucida 3 receptor (ZP3R/sp56) is not essential for mouse fertilization. *Biology of reproduction*, *86*(1), 1–6. https://doi.org/10.1095/biolreprod.111.095877

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.

Ohto, U., Ishida, H., Krayukhina, E., Uchiyama, S., Inoue, N., & Shimizu, T. (2016). Structure of IZUMO1-JUNO reveals sperm-oocyte recognition during mammalian fertilization. *Nature*, *534*(7608), 566–569. https://doi.org/10.1038/nature18596

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.

Vacquier, V. D., Swanson, W. J., & Lee, Y. H. (1997). Positive Darwinian selection on two homologous fertilization proteins: what is the selective pressure driving their divergence?. *Journal of molecular evolution*, *44 Suppl 1*, S15–S22. https://doi.org/10.1007/pl00000049

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

Wassarman, P. M., & Litscher, E. S. (2018). The Mouse Egg's Zona Pellucida. *Current topics in developmental biology*, *130*, 331–356. https://doi.org/10.1016/bs.ctdb.2018.01.003

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.

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.

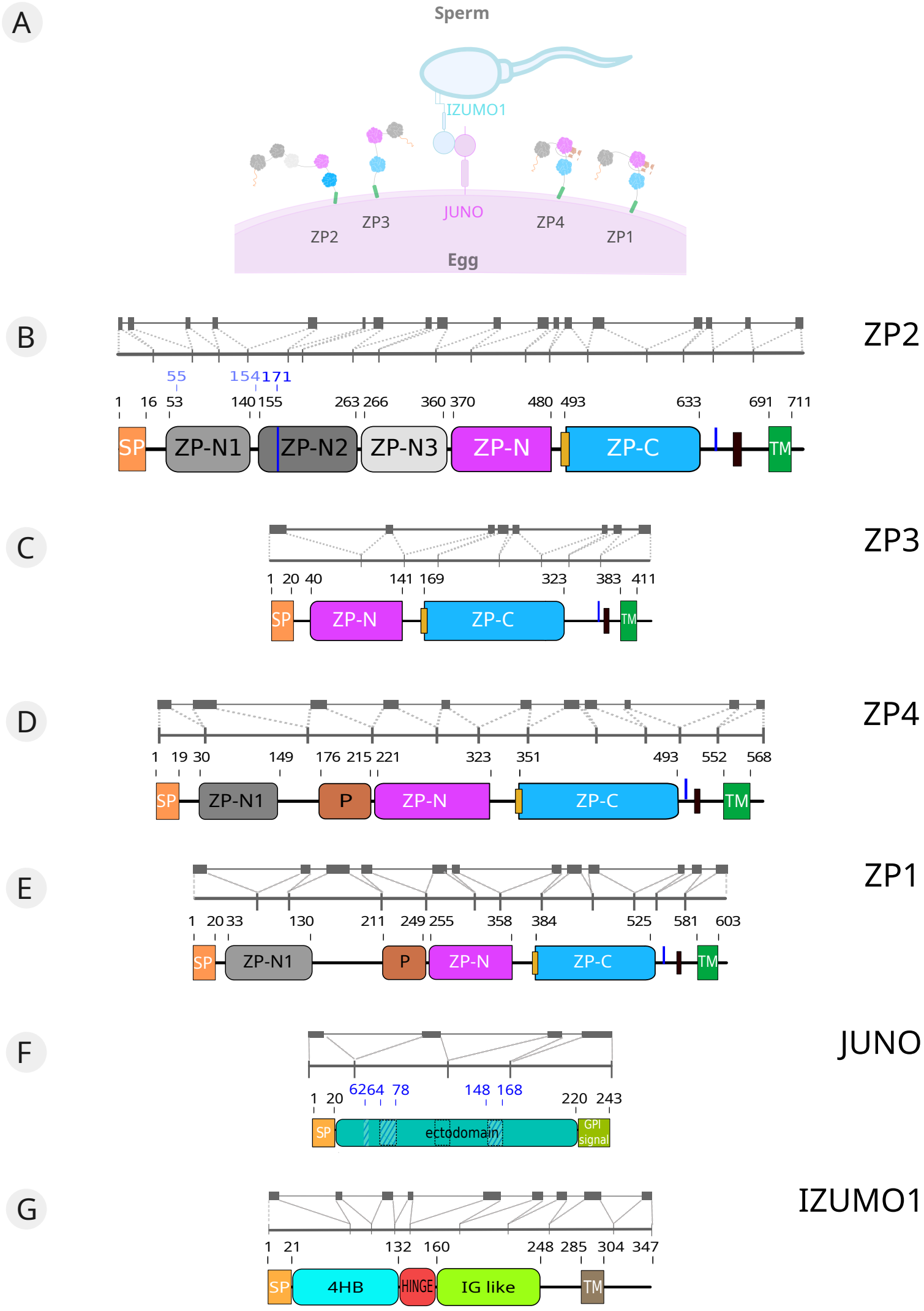
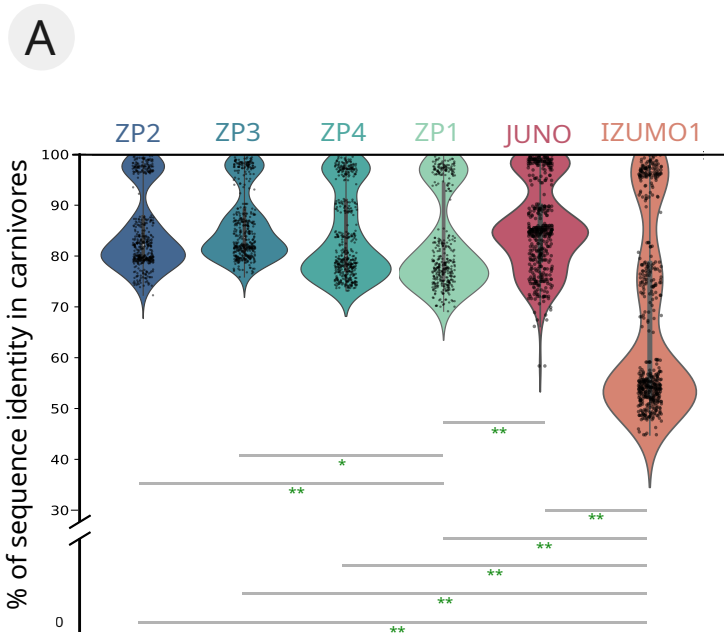

A
Sperm
IZUMO1
JUNO
ZP2
ZP3
ZP4
ZP1
Egg
B
ZP2
55
154 171
1 16 53 140 155 263 266 360 370 480 493 633 691 711
SP
ZP-N1
ZP-N2
ZP-N3
ZP-N
ZP-C
TM
C
ZP3
1 20 40 141 169 323 383 411
SP
ZP-N
ZP-C
TM
D
ZP4
1 19 30 149 176 215 221 323 351 493 552 568
SP
ZP-N1
P
ZP-N
ZP-C
TM
E
ZP1
1 20 33 130 211 249 255 358 384 525 581 603
SP
ZP-N1
P
ZP-N
ZP-C
TM
F
JUNO
1 20 62 64 78 148 168 220 243
SP
ectodomain
GPI signal
G
IZUMO1
1 21 132 160 248 285 304 347
SP
4HB
HINGE
IG like
TM

Figure 1: Schematic overview of ZP and JUNO-IZUMO1 gamete interaction proteins
A. Diagram of sperm and egg showing studied gamete interaction proteins: zona pellucida (ZP) proteins ZP1 to 4, JUNO from oocyte and IZUMO1 from sperm.
B. Schematic representation of jaguar *Zp2* gene, illustrating the arrangement of exons and introns (first line). Visualization of the mature transcript of the *Zp2* gene (second line). Schematic representation of the protein domains comprising ZP2 with jaguar start and end positions for each domain (third line). Start and end positions of the putative sperm-binding region are indicated with blue numbering above the protein scheme. SP: Signal Peptide; ZP-N1/2/3: ZP-N like domains; ZP-N and ZP-C: Zona Pellucida N-terminal and C-terminal domains; TM: Transmembrane domain. Blue vertical full line in ZP-N2: ZP2 ovastatin cleavage site. Blue vertical line: furin cleavage site. Diagonal striped blue pattern: ZP2 sperm-binding region. Yellow rectangle: internal hydrophobic patch. Black rectangle: external hydrophobic patch.
C. Schematic representation of jaguar *Zp3* gene, illustrating the arrangement of exons and introns (first line). Visualization of the mature transcript of the *Zp3* gene (second line). Schematic representation of the protein domains comprising ZP3 with jaguar start and end positions for each domain (third line). SP: Signal Peptide; ZP-N and ZP-C: Zona Pellucida N-terminal and C-terminal domains; TM: Transmembrane domain; Blue vertical line: furin cleavage site. Yellow rectangle: internal hydrophobic patch. Black rectangle: external hydrophobic patch.
D. Schematic representation of jaguar *Zp4* gene, illustrating the arrangement of exons and introns (first line). Visualization of the mature transcript of the *Zp4* gene (second line). Schematic representation of the protein domains comprising ZP4 with jaguar start and end positions for each domain (third line). SP: Signal Peptide; ZP-N1: ZP-N like domain; P: Trefoil domain; ZP-N and ZP-C: Zona Pellucida N-terminal and C-terminal domains; TM: Transmembrane domain. Blue vertical line: furin cleavage site.
E. Schematic representation of the *Zp1* gene, illustrating the arrangement of exons and introns (first line). Visualization of the mature transcript of the *Zp1* gene (second line). Schematic representation of the protein domains comprising ZP1 with jaguar start and end positions for each domain (third line). SP: Signal Peptide; ZP-N1: ZP-N like domain; P: Trefoil domain; ZP-N and ZP-C: Zona Pellucida N-terminal and C-terminal domains; TM: Transmembrane domain. Blue vertical line: furin cleavage site.
F. Schematic representation of jaguar *Juno* gene, illustrating the arrangement of exons and introns (first line). Visualization of the mature transcript of the *Juno* gene (second line). Schematic representation of the protein domains comprising JUNO with jaguar start and end positions for each domain and start and end positions of the regions involved in interaction with IZUMO1 in blue numbers (third line). Flexible regions are demarcated on the domains with a black dashed rectangles. Regions involved in interaction with the IZUMO1 are shown with blue stripes diagonal pattern. SP: Signal Peptide; GPI signal: signal for anchor addition.
G. Schematic representation of jaguar *Izumo1* gene, illustrating the arrangement of exons and introns (first line). Visualization of the mature transcript of the *Izumo1* gene (second line). Schematic representation of the protein domains comprising IZUMO1 with jaguar start and end positions for each domain (third line). Regions involved in interaction with the JUNO are shown with blue stripes diagonal pattern. SP: Signal Peptide; HB: N-terminal alpha helices bundle; HINGE: Hinge domain; IG like: C-terminal immunoglobulin-like domain; TM: Transmembrane domain.


A
ZP2
ZP3
ZP4
ZP1
JUNO
IZUMO1
% of sequence identity in carnivores
100
90
80
70
60
50
40
30
0
*
**

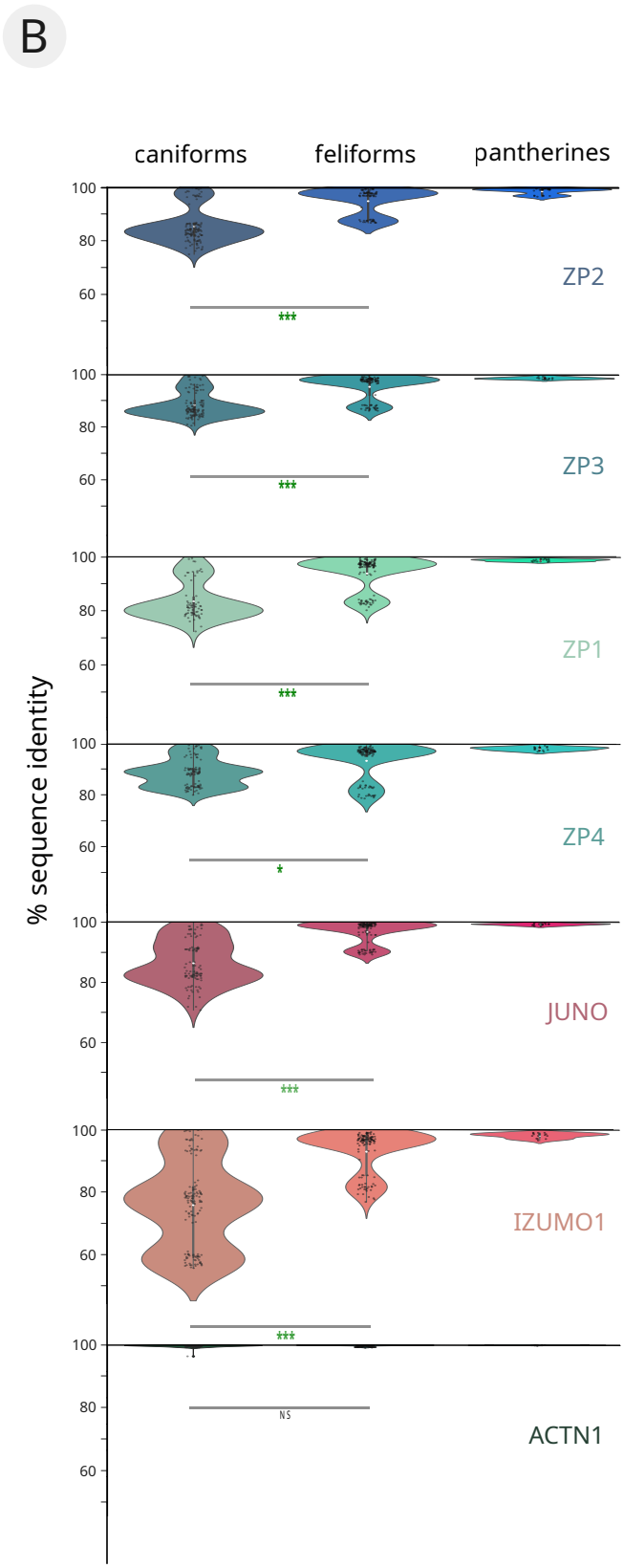

B
caniforms
feliforms
pantherines
% sequence identity
ZP2
ZP3
ZP1
ZP4
JUNO
IZUMO1
ACTN1
100
80
60
***
*
NS

Figure 2: Identity profiling of ZP and JUNO-IZUMO1 gamete interaction proteins
A. ZP proteins, JUNO and IZUMO1 carnivores sequence identities. Violin plot showing pairwise sequence identities percentages (black dots) for ZPs, JUNO, and IZUMO1 proteins. Mean percentage of sequence identities for each protein is represented in a white dot. Scheffé's T-test significance: (*) $p<0.05$ (**); $p<0.01$.
B. ZPs, JUNO and IZUMO1 sequence identities for taxonomic groups of interest (caniforms, feliforms and pantherins). Violin plot showing pairwise sequence identities percentages (black dots). Mean percentage of sequence identity for each group is represented in a white dot. Welch's T-test significance: (*) $p<0.05$ (**); $p<0.01$.

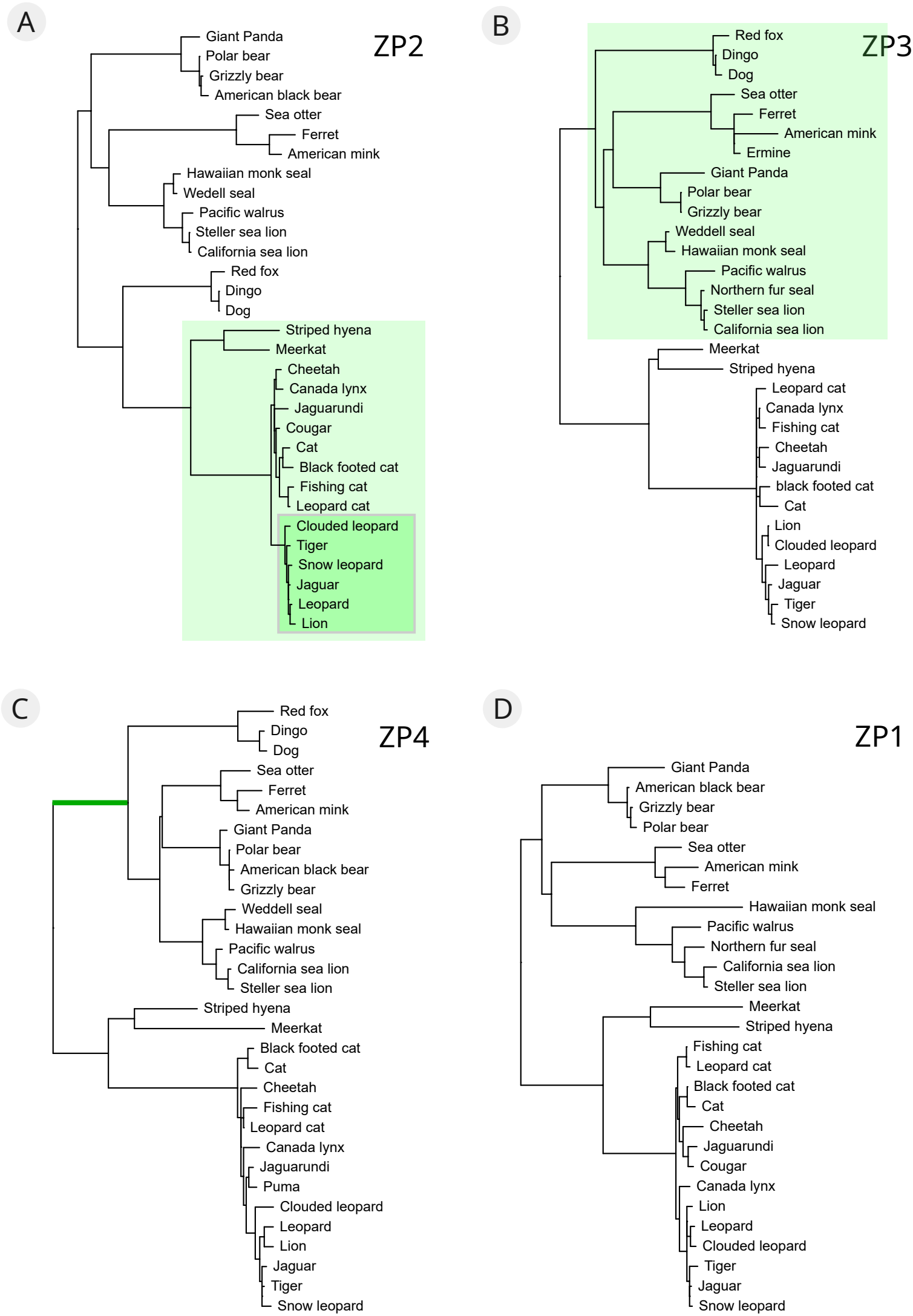

A
ZP2
Giant Panda
Polar bear
Grizzly bear
American black bear
Sea otter
Ferret
American mink
Hawaiian monk seal
Wedell seal
Pacific walrus
Steller sea lion
California sea lion
Red fox
Dingo
Dog
Striped hyena
Meerkat
Cheetah
Canada lynx
Jaguarundi
Cougar
Cat
Black footed cat
Fishing cat
Leopard cat
Clouded leopard
Tiger
Snow leopard
Jaguar
Leopard
Lion
B
ZP3
Red fox
Dingo
Dog
Sea otter
Ferret
American mink
Ermine
Giant Panda
Polar bear
Grizzly bear
Weddell seal
Hawaiian monk seal
Pacific walrus
Northern fur seal
Steller sea lion
California sea lion
Meerkat
Striped hyena
Leopard cat
Canada lynx
Fishing cat
Cheetah
Jaguarundi
black footed cat
Cat
Lion
Clouded leopard
Leopard
Jaguar
Tiger
Snow leopard
C
ZP4
Red fox
Dingo
Dog
Sea otter
Ferret
American mink
Giant Panda
Polar bear
American black bear
Grizzly bear
Weddell seal
Hawaiian monk seal
Pacific walrus
California sea lion
Steller sea lion
Striped hyena
Meerkat
Black footed cat
Cat
Cheetah
Fishing cat
Leopard cat
Canada lynx
Jaguarundi
Puma
Clouded leopard
Leopard
Lion
Jaguar
Tiger
Snow leopard
D
ZP1
Giant Panda
American black bear
Grizzly bear
Polar bear
Sea otter
American mink
Ferret
Hawaiian monk seal
Pacific walrus
Northern fur seal
California sea lion
Steller sea lion
Meerkat
Striped hyena
Fishing cat
Leopard cat
Black footed cat
Cat
Cheetah
Jaguarundi
Cougar
Canada lynx
Lion
Leopard
Clouded leopard
Tiger
Jaguar
Snow leopard

Figure 3: Phylogenetic trees and ZP proteins positive selection analyses

A. Maximum likelihood phylogenetic reconstruction for ZP2 using 31 carnivore species. Highlighted lighy green areas indicate divergent positive selection in *Feliformia* suborder and *Pantherinae* subfamily.

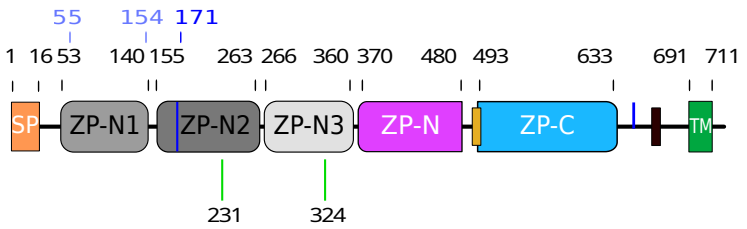
B. Maximum likelihood phylogenetic reconstruction for ZP3 using 31 carnivore species. Highlighted light green areas indicate divergent positive selection in Caniformia subtree.

C. Maximum likelihood phylogenetic reconstruction of ZP4 using 31 carnivore species. Dark green line in Caniformia basal branch indicate episodic positive selection.

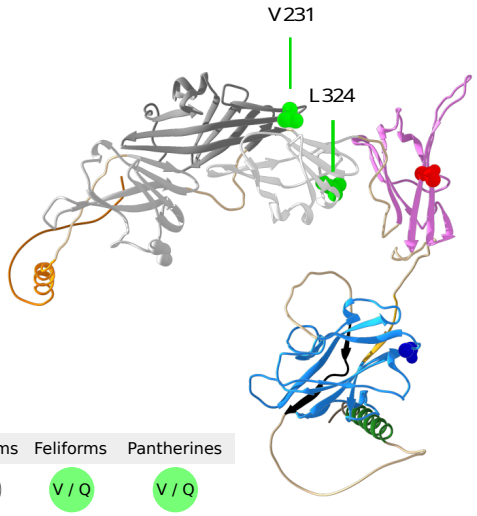
D. Maximum likelihood phylogenetic reconstruction for ZP1 using 28 carnivore species.

ZP2

A



B



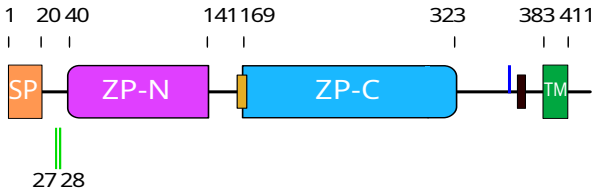
C

Position	Caniforms	Feliforms	Pantherines
231	Q	V / Q	V / Q
324	S	V / F	L / F

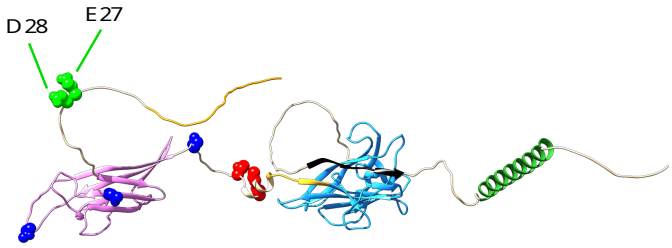
no positive selection
episodic positive selection in the basal branch of the clade
divergent evolution in the clade subtree

ZP3

D



E



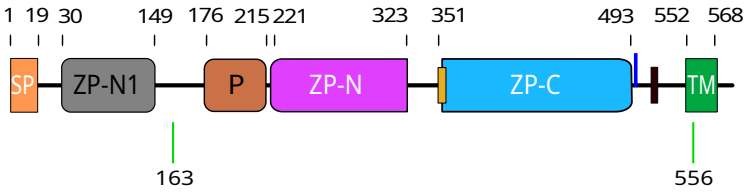
F

Position	Caniforms	Feliforms	Pantherines
27	P / W Q / S	E	E / G
28	R / S G / T	D	D

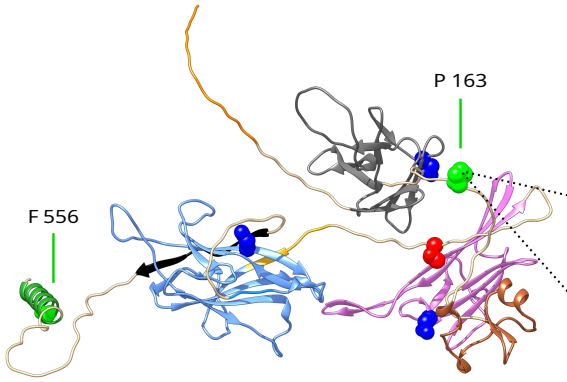
no positive selection
episodic positive selection in the basal branch of the clade
divergent evolution in the clade subtree

ZP4

G



H



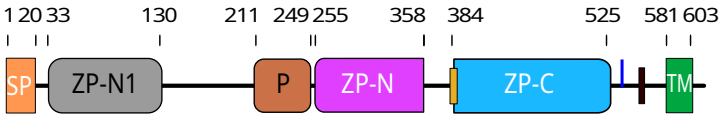
I

Position	Caniforms	Feliforms	Pantherines
163	Y	P	P
556	A	F	F

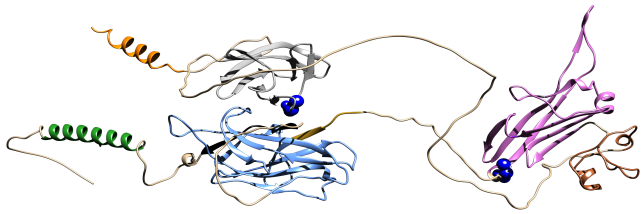
no positive selection
episodic positive selection in the basal branch of the clade
divergent evolution in the clade subtree

ZP1

J



K



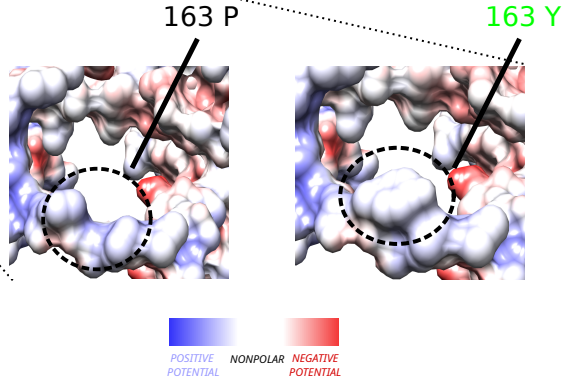


Figure 4: ZP proteins positive selected sites

A. Schematic representation of jaguar ZP2 protein domains. Idem Fig 1B. Light green lines below indicate positive selected sites.

B. Three-dimensional model of the jaguar protein ZP2 shown in ribbons. N-glycosylated and O glycosylated sites are shown as blue and red atom spheres, respectively. Position and identities of positively selected amino acids are indicated and highlighted in light green atoms spheres.

C. Table displaying sites under positive selection in ZP2 and their identities in the groups of interest.

D. Schematic representation of jaguar ZP3 protein domains. Idem Fig. 1C. Light green lines below indicate positive selected sites.

E. Three-dimensional model of the jaguar protein ZP3 shown in ribbons. N-glycosylated and O glycosylated sites are shown as blue and red atom spheres, respectively. Position and identities of positively selected amino acids are indicated and highlighted in light green atoms spheres.

F. Table displaying sites under positive selection in ZP3 and their identities in the groups of interest.

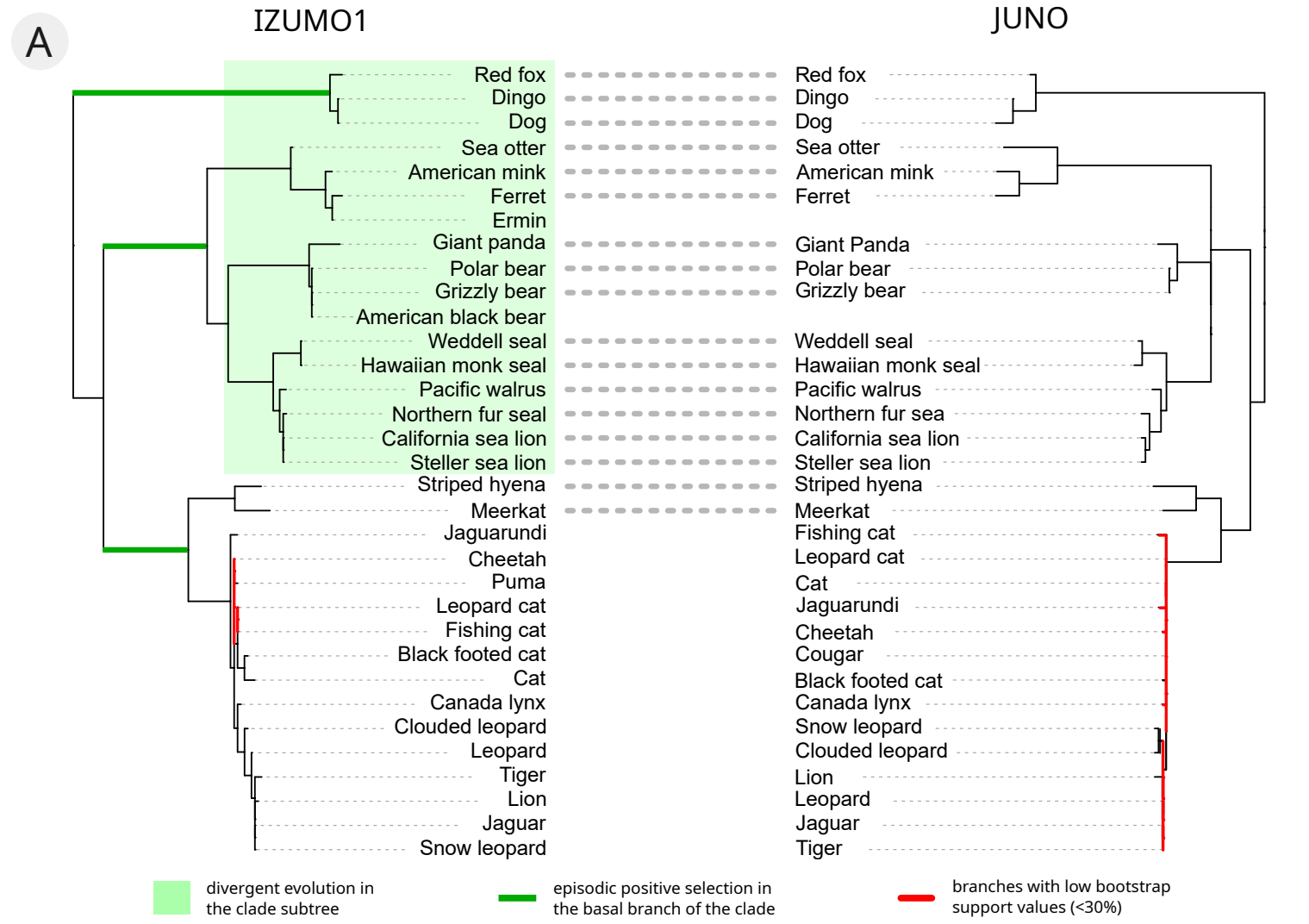
G. Schematic representation of jaguar ZP4 protein domains. Idem Fig. 1D. Light green lines below indicate positive selected sites.

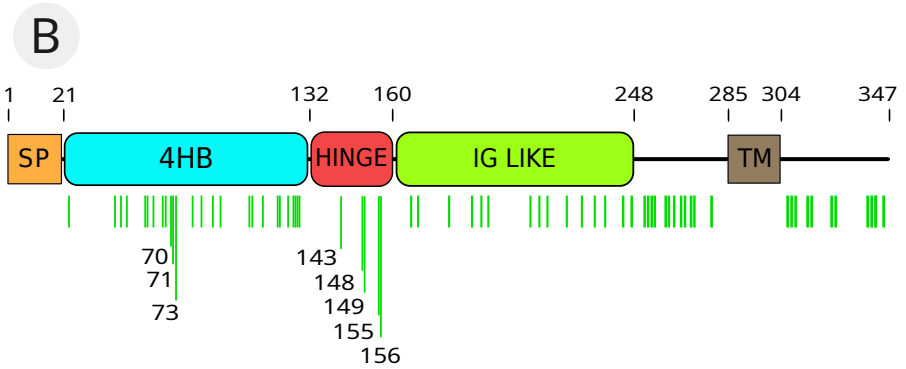
H. Three-dimensional model of the jaguar protein ZP4 shown in ribbons. N-glycosylated and O glycosylated sites are shown as blue and red atom spheres, respectively. Position and identities of positively selected amino acids are indicated and highlighted in light green atoms spheres. Coulombic surfaces depict the ancestral and derived states of positively selected amino acids, focusing solely on those sites that exhibit significant electrostatic charge changes from the ancestral to the derived state.

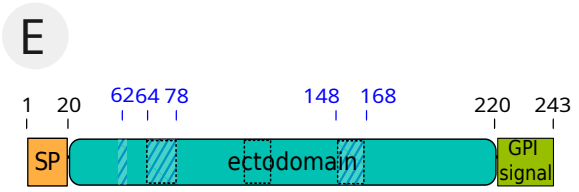
I. Table displaying sites under positive selection in ZP4 and their identities in the groups of interest.

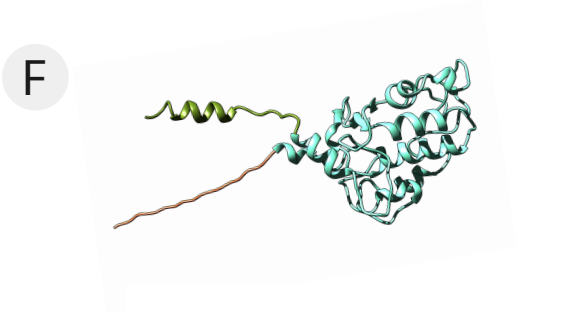
J. Schematic representation of jaguar ZP1 protein domains. Idem Fig 1 E.

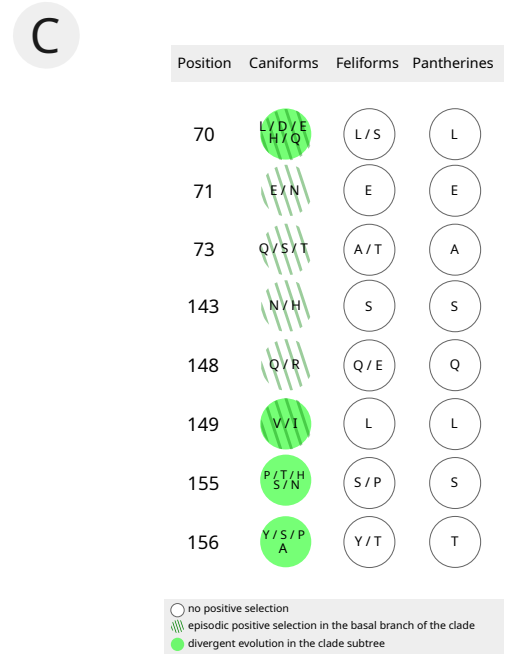
K. Three-dimensional model of the jaguar protein ZP1 shown in ribbons. N-glycosylated sites are shown as blue atom spheres.


A
IZUMO1
JUNO
Red fox
Dingo
Dog
Sea otter
American mink
Ferret
Ermin
Giant panda
Polar bear
Grizzly bear
American black bear
Weddell seal
Hawaiian monk seal
Pacific walrus
Northern fur seal
California sea lion
Steller sea lion
Striped hyena
Meerkat
Jaguarundi
Cheetah
Puma
Leopard cat
Fishing cat
Black footed cat
Cat
Canada lynx
Clouded leopard
Leopard
Tiger
Lion
Jaguar
Snow leopard
Red fox
Dingo
Dog
Sea otter
American mink
Ferret
Giant Panda
Polar bear
Grizzly bear
Weddell seal
Hawaiian monk seal
Pacific walrus
Northern fur sea
California sea lion
Steller sea lion
Striped hyena
Meerkat
Fishing cat
Leopard cat
Cat
Jaguarundi
Cheetah
Cougar
Black footed cat
Canada lynx
Snow leopard
Clouded leopard
Lion
Leopard
Jaguar
Tiger
divergent evolution in the clade subtree
episodic positive selection in the basal branch of the clade
branches with low bootstrap support values (<30%)


B
1
21
132
160
248
285
304
347
SP
4HB
HINGE
IG LIKE
TM
70
71
73
143
148
149
155
156


E
1
20
62
64
78
148
168
220
243
SP
ectodomain
GPI signal


F


C
Position
Caniforms
Feliforms
Pantherines
70
L/D/E H/Q
L/S
L
71
E/N
E
E
73
Q/S/T
A/T
A
143
N/H
S
S
148
Q/R
Q/E
Q
149
V/I
L
L
155
P/T/H S/N
S/P
S
156
Y/S/P A
Y/T
T
no positive selection
episodic positive selection in the basal branch of the clade
divergent evolution in the clade subtree

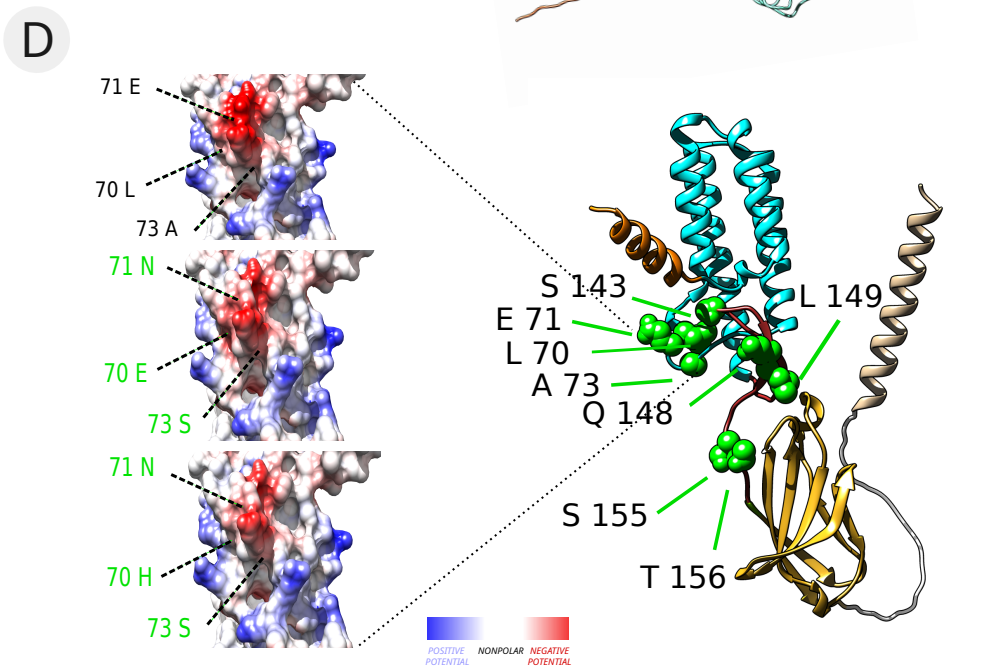

D
71 E
70 L
73 A
71 N
70 E
73 S
71 N
70 H
73 S
S 143
E 71
L 70
A 73
Q 148
L 149
S 155
T 156
POSITIVE POTENTIAL
NONPOLAR
NEGATIVE POTENTIAL

Figure 5: Evolution of IZUMO1-JUNO proteins fusion pair

A. Mirrored IZUMO1 and JUNO maximum likelihood trees showing comparable topologies (dotted lines). Episodic positive selection in the basal lineage of a group is indicated with dark green lines. Highlighted light green area indicates divergent positive selection. Branches with bootstrap values below 30% are highlighted in red.

B. Schematic representation of jaguar IZUMO1 protein domains. Idem Fig. 1F. Light green lines below indicate positive selected sites.

C. Table displaying sites under positive selection in IZUMO1 and their identities in the groups of interest. Only positive selected sites previously reported to be involved in interaction with JUNO are shown.

D. Three-dimensional model of IZUMO1 protein shown in ribbons. Position and identities of positively selected amino acids are indicated and highlighted in light green atoms spheres. Coulombic surfaces depict the ancestral and derived states of positively selected amino acids, focusing solely on those sites that exhibit significant electrostatic charge changes from the ancestral to the derived state.

E. Schematic representation of jaguar JUNO protein domains. Idem Fig. 1G

F. Three-dimensional model of the jaguar protein JUNO shown in ribbons.

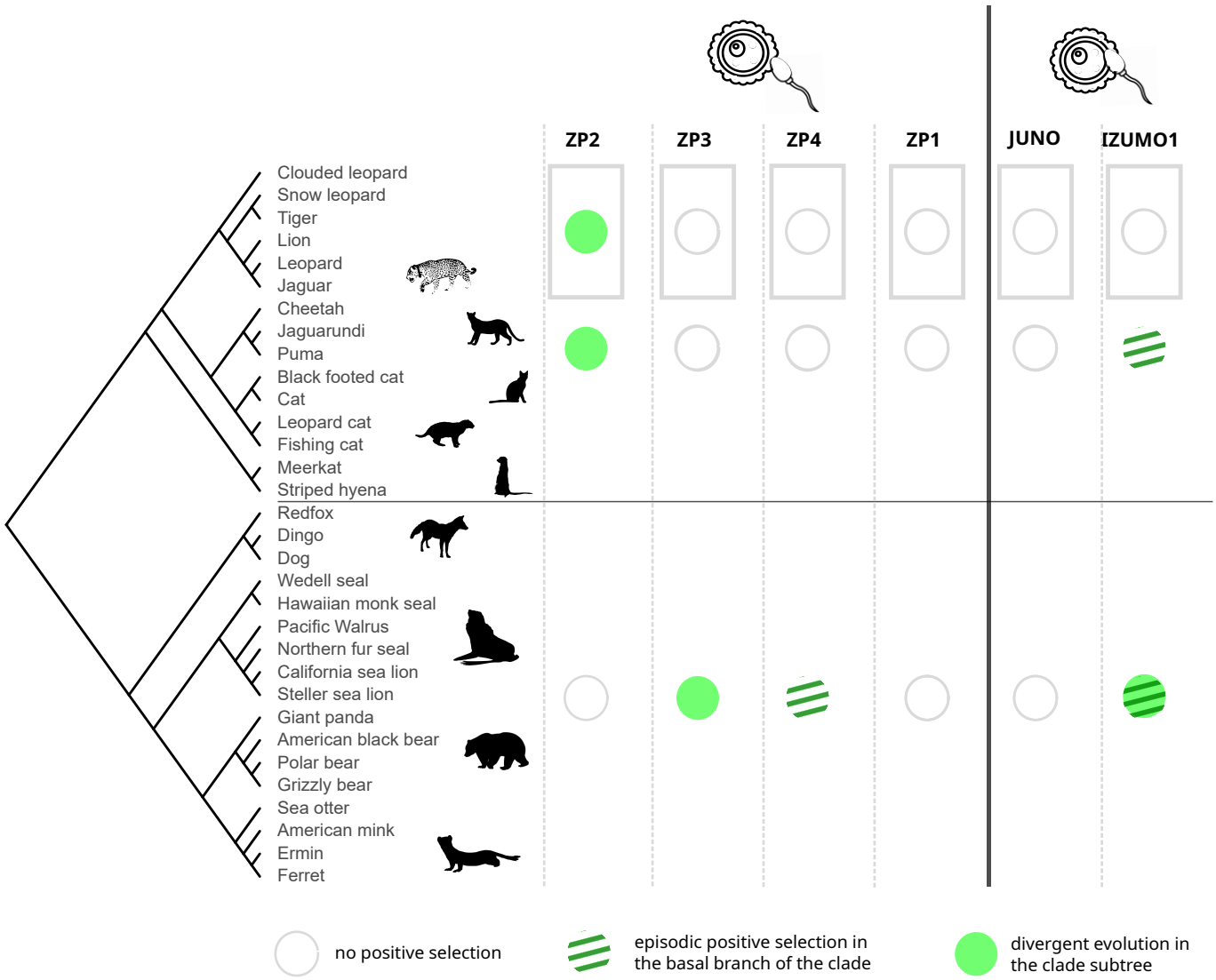

ZP2
ZP3
ZP4
ZP1
JUNO
IZUMO1
Clouded leopard
Snow leopard
Tiger
Lion
Leopard
Jaguar
Cheetah
Jaguarundi
Puma
Black footed cat
Cat
Leopard cat
Fishing cat
Meerkat
Striped hyena
Redfox
Dingo
Dog
Wedell seal
Hawaiian monk seal
Pacific Walrus
Northern fur seal
California sea lion
Steller sea lion
Giant panda
American black bear
Polar bear
Grizzly bear
Sea otter
American mink
Ermin
Ferret
no positive selection
episodic positive selection in the basal branch of the clade
divergent evolution in the clade subtree

Figure 6: Summary of Positive selection analyses

Positive selection analyses results for ZP proteins, JUNO and IZUMO1 in all tested groups: caniforms, feliforms and pantherines. A filled circle with green diagonal stripe pattern indicate episodic positive selection at the basal branch of the clade. A light green circle indicates divergent positive selection in the clade sub-tree. Empty white circle denotes absence of positive selection. Species tree is representative of the full set of carnivore species employed in this study.