Author for correspondence:

Phil Grayson
e-mail: pgrayson@fas.harvard.edu

THE ROYAL SOCIETY PUBLISHING

Izumo1 and Juno: the evolutionary origins and coevolution of essential sperm–egg binding partners

Phil Grayson

Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

Reproductive proteins are among the most rapidly evolving classes of proteins. For a subset of these, rapid evolution is driven by positive Darwinian selection despite vital, well-conserved, reproductive functions. Izumo1 is the only essential sperm–egg fusion protein currently known on mammalian sperm, and its egg receptor (Juno; formerly Folr4) was recently discovered. Male knockout mice for Izumo1 and female knockout mice for Juno are both healthy but sterile. Here, both sperm–egg binding proteins are shown to be evolving under positive selection. Within mammals, coevolution of Izumo1 and Juno is also uncovered, suggesting that similar forces have shaped the evolutionary histories of these binding partners within Mammalia. Additionally, genomic analyses reveal an ancient origin for the Izumo gene family, initially reported as conserved exclusively in mammals. Newly identified Izumo1 orthologues could serve reproductive functions in birds, fish and reptiles. Surprisingly, these same analyses support Juno's presence in mammals alone, suggesting a recent mammalian-specific duplication and neofunctionalization of the ancestral folate receptor. Despite the indispensability of their reproductive interaction, and their apparent coevolution within Mammalia, this binding pair arose through strikingly different evolutionary forces.

1. Introduction

In sexually reproducing species, fertilization occurs through the recognition and fusion of haploid sperm and egg to produce a diploid zygote. Although the processes involved in this essential reproductive stage have been well studied, the molecular basis of mammalian sperm–egg fusion was largely unknown

until recently. The first essential sperm–egg fusion protein to be discovered on mammalian sperm was Izumo1. An Izumo1 knockout results in male mice that are completely sterile, but otherwise healthy, while female mice exhibit a normal phenotype [1]. Recently, the egg receptor for mammalian Izumo1 was discovered to be folate receptor 4 (Folr4), which was renamed Juno (or Izumo1R) [2]. Juno knockouts result in a phenotype opposite to Izumo1 knockouts: healthy, yet sterile female mice, and phenotypically normal males [2]. Izumo1 and Juno are the first essential sperm–egg binding pair identified in any species, and their conserved binding ability has been demonstrated in numerous mammals [1–3]. Additional mammalian egg proteins, including CD9 and CD81, are also known to be essential for sperm–egg fusion [4]. Currently, it is unclear whether these egg proteins interact directly with a binding partner on the sperm, and their precise roles in sperm–egg fusion have yet to be determined [4–7].

Comparative genomic studies have identified reproductive proteins as one of the most rapidly evolving gene-classes across diverse taxa [8]. *A priori*, one might expect positively selected reproductive genes to play secondary roles in reproductive success while proteins central to fertilization remain generally conserved, but positive selection has been identified in many classes of reproductive proteins including those functioning in sperm competition, sperm–egg binding and spermatogenesis [7,9–12]. Although Izumo1 has an indispensable role in fertilization, it too has been identified as evolving rapidly through positive selection in both large exploratory studies and in-depth targeted analyses [7,11,13].

Izumo1 paralogues have also been discovered, resulting in a four-member Izumo family in mammals [14]. Izumo1, Izumo2 and Izumo3 are expressed exclusively in the testes, while Izumo4 lacks a transmembrane domain and is present throughout the body [14]. Molecular evolutionary study revealed that Izumo1 is evolving under positive selection across Mammalia, driven by rapid evolution within the laurasiatherian mammals, while Izumo2 is positively selected within the Primates, and Izumo4 is evolving rapidly in the Glires (Rodentia and Lagomorpha), probably due to a relaxation of selective constraint [13].

Juno, previously Folr4, is also a member of a four-gene family. The mammalian folate receptor family consists of Folr1, Folr2 and Folr4 across Mammalia, and Folr3 within Primates [15–18]. Folr1–3 are known to bind folate by a newly described mechanism, and the heightened expression of Folr1 and Folr2 in disease-causing cells in cancers and inflammatory diseases has led to extensive research into folate receptor-targeted therapies [19–22]. Human and mouse Juno differ from the other family members at a number of essential folate-binding sites and are indeed unable to bind folate [2]. In mice, Folr1 and Folr2 do not interact with Izumo1; similarly, Izumo3 and Izumo4 do not interact with Juno [2].

Molecular evolutionary analyses of gene families not only provide insight into the history of the genes, but can also identify targets for future functional experiments [23]. No such study has been conducted for the folate receptor family. Here, I describe the evolutionary history of the only essential sperm–egg fusion pair currently known, Izumo1–Juno, in the context of their respective gene families. I examine the evolutionary history of the folate receptor family in vertebrates for the first time and conclude that Juno arose through a recent, mammalian-specific duplication and neofunctionalization of an ancestral folate receptor gene. I uncover an ancient origin of the Izumo gene family and Izumo1, which has only been previously described and studied in mammals. These newly discovered Izumo1 orthologues could serve similar reproductive functions in fish, birds and reptiles. I identify positive selection driving the evolution of Juno within mammals and compare these signatures of selection to those previously reported for Izumo1 [13]. Finally, I investigate coevolution between Izumo1 and Juno within Mammalia and identify lineage-specific signals of coordinated evolution between these binding partners. I conclude that Izumo1 and Juno have distinct temporal origins, but coevolve under similar selective pressures within the mammals.

2. Methods and results

2.1. Folate receptor family evolution

Homologous sequences for mammalian Folr1, Folr2, Folr3 and Folr4 were identified from NCBI and Ensembl. These sequences were evaluated through a suite of techniques: reciprocal BLAST searches on NCBI [24,25], visual inspection of alignments in MEGA v. 6.06 using MUSCLE and CLUSTALW [26–28], phylogenetic analysis in MRBAYES and RAxML [29,30], and synteny when the target gene was present on a contig of suitable size [25,31]. Synteny figures were constructed by hand, based on Ensembl's Region in Detail and NCBI's Genomic Region tools. When syntenic genes appeared to be missing for a species of interest, or when the genome browsers provided conflicting results, BLAST searches were carried out to

Figure 1. Phylogeny and synteny of the folate receptor family in vertebrates outlining three synteny groups: the ancestral folate receptor (blue), the derived folate receptor (green) and Juno (pink). Juno is only present in mammals. The phylogenetic tree was constructed using Bayesian and ML methods and support for each node is presented for each method as described in the legend. The Bayesian phylogeny was constructed following JMODELTEST2's [32] AICc-based recommendation of partitioned codon positions with the following models: GTR + G (position 1), HKY + G (position 2) and GTR + I + G (position 3). RAxML was run under the default model (GTR + CAT) partitioned by codon position with 1000 bootstraps. Syntenic relationships are shown with solid arrows connected by dotted lines. Non-syntenic genes are drawn as smaller, dotted arrows. Chromosome colour is representative of synteny group on the phylogeny.

Species	53	55	74	78	102	103	106	135	136	137	138	140	171	174	175	Subs
Zebrafish	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
Fugu	Y	F	D	Y	W	R	R	H	T	G	W	W	W	S	Y	1
LampreyI	Y	F	D	Y	W	R	R	H	S	G	W	W	W	S	Y	1
BurmesePythonI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
AmericanAlligatorIII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
GreenSeaTurtleII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
GreenSeaTurtleI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
PaintedTurtleII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
PaintedTurtleIII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
PaintedTurtleI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
ChineseSoftShellTurtleIII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
ChineseSoftShellTurtleI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
AmericanAlligatorII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
AmericanAlligatorI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
TibetanGroundTitII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	F	1
AmericanCrowII	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
TibetanGroundTitI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
AmericanCrowI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
XenopusI	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
Human2	Y	F	D	Y	W	R	R	H	R	G	W	W	W	S	Y	1
Human3	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	F	1
Mouse2	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
Mouse1	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
Human1	Y	F	D	Y	W	R	R	H	K	G	W	W	W	S	Y	0
BushBaby4	Y	F	A	Y	G	L	R	H	V	G	W	W	W	L	F	6
Chimpanzee4	Y	F	A	Y	G	Q	R	G	G	G	W	W	W	S	F	6
Gorilla4	Y	F	A	Y	W	E	R	R	G	G	W	W	W	S	F	5
Human4	Y	F	A	Y	G	Q	R	G	G	G	W	W	W	S	F	6
Marmoset4	Y	F	A	Y	W	E	R	H	G	S	W	W	W	S	F	5
Macaque4	Y	F	A	Y	G	Q	R	R	G	G	W	W	W	S	F	6
Orangutan4	Y	F	A	Y	W	E	R	R	G	G	W	W	W	S	F	5
Gibbon4	Y	F	A	Y	G	Q	R	R	G	G	W	W	W	S	F	6
SquirrelMonkey4	Y	F	A	Y	G	Q	R	H	G	G	W	W	W	S	F	5
Antelope4	Y	F	A	Y	W	Q	R	L	S	S	W	W	W	S	F	6
BaijiDolphin4	Y	F	A	Y	G	Q	R	H	G	G	W	W	W	S	F	5
Ferret4	Y	F	A	Y	G	P	R	H	-	G	W	R	W	S	F	6
Cat4	Y	F	A	Y	G	P	R	H	G	D	W	W	W	S	F	6
Horse4	Y	F	A	Y	G	Q	R	H	G	G	W	W	W	S	F	5
KillerWhale4	Y	F	A	Y	G	Q	R	H	G	G	W	W	W	S	F	5
Megabat4	Y	F	A	Y	W	D	R	H	S	S	W	W	W	S	F	5
Microbat4	Y	F	A	H	E	Q	R	L	R	G	W	W	W	S	F	7
MinkeWhale4	Y	F	A	Y	G	R	R	H	G	G	W	W	W	S	F	4
Panda4	Y	F	A	Y	G	P	R	H	G	D	W	R	W	S	F	7
Pig4	Y	F	A	Y	G	Q	R	L	G	G	W	W	W	S	F	6
Sheep4	Y	F	A	Y	W	Q	R	L	G	S	W	W	W	S	F	6
WaterBuffalo4	Y	F	A	Y	W	Q	R	L	G	G	W	W	W	S	F	5
Yak4	Y	F	A	Y	W	Q	R	L	G	G	W	W	W	S	F	5
Rat4	F	F	A	Y	Q	E	R	L	H	G	W	W	W	T	F	8
NakedMoleRat4	H	F	A	Y	G	L	Q	Q	G	G	W	L	W	S	F	9
Rabbit4	Y	F	A	Y	G	P	R	H	G	G	W	W	W	T	F	6
Pika4	Y	F	A	Y	W	E	R	H	D	G	W	W	W	T	F	5
KangarooRat4	H	F	A	Y	-	-	R	H	G	D	W	W	W	S	F	7
Mouse4	F	F	A	H	Q	E	R	L	H	G	W	W	W	T	F	9
GuineaPig4	H	F	A	Y	D	L	R	H	G	G	W	W	W	S	F	6
Opossum4	Y	F	V	Y	L	G	S	Y	S	G	W	W	W	S	Y	7

Figure 2. Sites 53, 55, 74, 78, 102, 103, 106, 135–138, 140, 171, 174 and 175 from the mature human Juno protein aligned with representative members from both the extended folate family (blue and green synteny groups) and mammalian Juno (pink synteny group) support the conservation of Juno in only Mammalia. These 15 sites are essential for folate binding in other Folr proteins, while Juno is known to be unable to bind folate. Derived sites are highlighted in pink, and the total number of substitutions present in each species is displayed in the far right column.

determine if automated annotation tools had missed (or mislabelled) the gene. These regions were also visually inspected for large gaps or Ns that could result in a gene being dropped from the final assembly. Additional homologues were identified and validated through this suite of analyses and a final collection of sequences was chosen to provide key phylogenetic positions, with priority given to sequences of high quality contained on large contigs.

The phylogeny and synteny presented here for the folate receptor family (figure 1) support the presence of a single ancestral folate receptor of the blue synteny group (generally adjacent to ANAPC15 and/or INPPL1) for ray-finned fish, lobe-finned fish and snakes, prior to lineage-specific duplications and losses which resulted in one Folr (currently listed as Folr1 on NCBI) in alligators, birds and

Xenopus, two to three Folr genes in turtles, and three in mammals. All members of the blue synteny group have been maintained in this syntenic gene cluster. Two independent duplications with genomic translocation have also occurred, producing additional synteny groups. The green synteny group is known only from archosaurs, with two translocated Folr genes in alligators and one in birds. Juno, the pink synteny group, appears to have split from the ancestral (blue group) folate receptor early in mammalian evolution and has been maintained in its new chromosomal location in both marsupials and placental mammals.

In opossum (*Monodelphis domestica*), Juno's interaction with Izumo1 has been shown to be conserved [2], and its presence in the genome is supported through both synteny (electronic supplementary material, figure S1) and phylogeny. Due to genome quality, the evolutionary history of this gene family could not be fully assessed in platypus (*Ornithorhynchus anatinus*), although flanking genes from the pink syntenic group are present, while Juno appears absent (scaffold NW_001794203), and at least one partial Folr gene is present in the platypus genome (XM_001513334).

The absence of Juno from non-mammalian genomes is supported through phylogeny (figure 1), synteny analysis and examination of sites essential for folate binding (figure 2). Synteny of Juno's flanking genes is conserved throughout this region across most vertebrates, but Juno is only present in mammals (electronic supplementary material, figure S1); in addition to the standard BLAST searches on NCBI, each syntenic region in tetrapods and fish was downloaded and individually examined with BLAST+ using TBLASTN, TBLASTX and BLASTP to confirm Juno's absence [33].

Fifteen sites known to be essential for folate binding [2,21,22] were also examined *in silico*. Although a few sequences belonging to the blue and green synteny groups have one substitution among these 15 essential sites, the vast majority of non-Juno sequences show full conservation across all 15 sites in tetrapods and fish (figure 2). Alternatively, all Juno sequences examined have between four and nine substitutions across these sites. Four of these substitutions at essential folate-binding sites may have been crucial for Juno's neofunctionalization from the ancestral Folr gene, since they appear in all mammalian Juno genes examined with the exception of two species which each have the ancestral amino acid at one of the four sites. These patterns support the hypothesis that all Folr family members of the blue and green synteny groups (figure 1) possess the ability to bind folate and that Juno's loss of functional folate-binding predates the divergence of marsupial and eutherian mammals.

2.2. Izumo family evolution

Orthologous sequences for Izumo1, Izumo2, Izumo3 and Izumo4 were identified using the protocol outlined above for the Folr family. Additionally, Izumo family members share an approximately 150 amino acid 'Izumo domain' near their N-terminus, and all Izumo1 sequences possess an immunoglobulin-like domain [1,14]. These domains also aided in the validation of newly discovered Izumo proteins.

The synteny and phylogeny (figure 3) support the discovery of non-mammalian members of all four Izumo paralogues, thereby placing the origin of this gene family 200–250 million years earlier than previous reports. Izumo2 and Izumo3, which are adjacent on reptilian contigs, appear to share a common ancestor with Izumo1, while Izumo4 is placed most distantly. This result is not surprising in light of Izumo4's absent transmembrane domain and expanded expression pattern [14].

In searching for Izumo genes, an interesting pattern emerged: despite the increasing number of assembled and annotated avian genomes (approx. 60 on NCBI), Izumo1 was difficult to identify in all but the Tibetan ground tit (*Pseudopodoces humilis*—the only bird to have a full Izumo1 gene identified). The other Izumo paralogues were not detected in birds, despite conservation in mammals and reptiles. Additional avian Izumo1 sequences were discovered through SRA BLAST, and the short, exonic sequences identified are reported (electronic supplementary material, table S1).

One common bias in sequence data is underrepresentation of GC- and AT-rich genomic regions [34]. Izumo1's GC content is approximately 45–61% in fish and mammals, approximately 55–68% in reptiles and 71% for Tibetan ground tit. This high GC content within the CDS of the ground tit is also present in the surrounding genomic regions (65% GC, compared to 55% and 57% for human and turtle, respectively). Syntenic genes, including RASIP1 and ISOC2, are also absent from all avian genomes except the ground tit. These findings might explain the absence of Izumo1 sequences for many birds. Non-avian Izumo paralogues share fairly consistent GC content within a given species, a pattern that (if maintained in birds) could suggest that Izumo2–4 are not lost, but simply hidden due to sequencing bias (figure 3). Alternatively, these paralogues may have been removed during the massive avian-specific loss of ancestral genes originally identified in chicken [35].
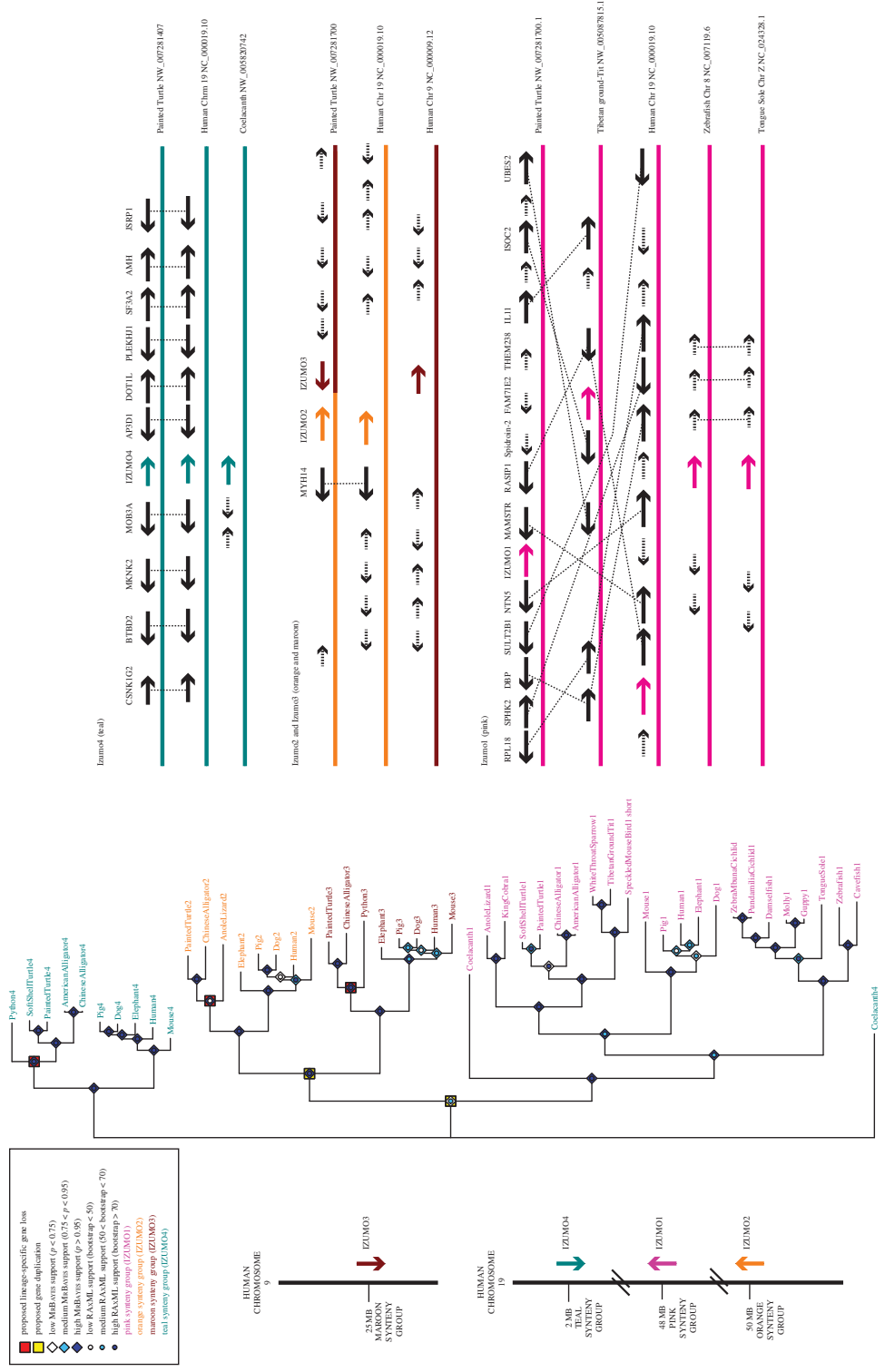
Figure 3. Phylogeny and synteny analysis of the extended Izumo family including the newly discovered, non-mammalian family members. Tree construction, coloration and synteny-mapping were carried out as in figure 1 (except GTR + G was used for all three codon positions during Bayesian reconstruction based on the AICc value in Jmodeltest2). Izumo1 is present in fish, reptiles, birds and mammals, Izumo4 is found in fish, mammals and reptiles, and Izumo2 and Izumo3 are only present in mammals and reptiles. These analyses support an ancient origin for the Izumo family.
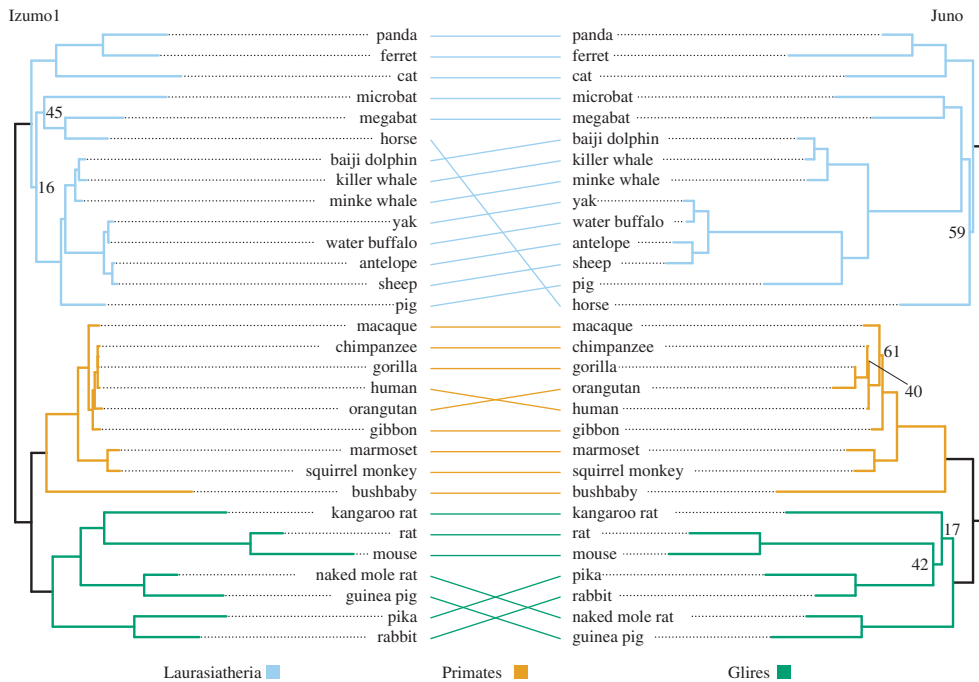
Figure 4. Mirrored gene trees for Izumo1 and Juno containing the 30 common mammals used in PAML (table 1) and Mirrortree (table 2) analyses. Trees were constructed using RAxML under the default model (GTR + CAT) partitioned by codon position with 1000 bootstraps. Bootstrap support values below 70 are indicated. This figure was constructed using the APE [38] and phytools [39] packages in R.

2.3. Positive selection and coevolution in Izumo1 and Juno

For both Izumo1 and Juno, phylogenetic analysis by maximum likelihood (PAML) [36,37] identified signals of positive selection across 30 common mammals with full, annotated genomes and good synteny in the target regions (table 1 and figure 4; see electronic supplementary material, table S1c for accession numbers). These results support those reported for one of the best-studied invertebrate sperm–egg partnerships, the lysin–VERL system of marine abalone. Both lysin on the sperm and VERL, its receptor on the egg, evolve under adaptive evolution [10,40]. Non-mammalian Izumo1 in birds, reptiles and fish appears to be evolving under purifying selection (table 1).

Previous in-depth analysis of molecular evolution within the Izumo family identified that the positive selection driving mammalian adaptive evolution for Izumo1 is localized to the laurasiatherian mammals, being absent from both Glires and Primates [13]. To determine if this pattern is also present in Juno, suggestive of coevolution between the binding partners, each clade was tested independently for both Izumo1 and Juno (table 1). These analyses have never been performed on Juno and are more robust than those previously carried out for Izumo1 in both species number and sequence coverage. Izumo1's signal of selection within laurasiatherians was present once more, and Juno's mammalian signal of positive selection appears to be driven by the Primates alone (table 1).

An additional technique used to investigate protein coevolution, called Mirrortree, compares two gene trees (e.g. a ligand and its receptor) to detect similarities greater than those expected under a standard molecular clock. Strong correlations (defined as more than 0.800) are suggestive of coordinated changes in both genes, resulting from similar evolutionary forces [41]. Mirrortree was used to examine the same 30 mammalian species used in the PAML analysis for coevolutionary signals between the gene trees of Izumo1 and Juno.

Correlation values between Izumo1 and Juno were generated through two separate analyses on the Mirrortree Web Server [42]: (i) using the 'From trees' option and (ii) using the 'From multiple alignments' option (table 2, figure 4). The 'From trees' approach analysed the same RAxML trees used for PAML analyses (figure 4), while the 'From multiple alignments' method examined a CLUSTAL NJ tree produced on the Web Server from provided Izumo1 and Juno protein alignments [42]. Both analyses identified a robust signal of coevolution between Izumo1 and Juno within Primates (table 2). The 'From trees' approach also identified a signal of coevolution within the Glires, but not across all 30 mammals or

Table 1. Positive selection is acting on both members of the essential sperm–egg binding pair (Izumo1/Juno), but is confined to different lineages. PAML and Bayes empirical Bayes (BEB) results for clade-based selection tests. Likelihood ratio test values are presented in columns M7 versus M8 and M8 versus M8a. M7 allows ω to vary between 0 and 1, while M8 adds an additional parameter whereby ω can exceed 1; M8a (M7 with $\omega = 1$) is the null hypothesis for M8 [36]. The total number of BEB sites with probability greater than 0.5 are reported before the semi-colon, followed by those sites identified as having p-values of 0.95 or better [37]. BEB sites were only produced for clades evolving under positive selection.

gene	group	n	M7 versus M8	M8 versus M8a	BEB sites
Juno	mammals	30	12.87***	11.32***	5; 115L***
Juno	Glires	7	1.39	0	n.a.
Juno	Laurasiatheria	14	1.05	0.74	n.a.
Juno	Primates	9	11.17***	7.83***	8; 111T*, 113L**, 123I*
Izumo1	mammals	30	65.03***	43.32***	101; 28P*, 167P***, 261S**, 274A*, 276T*, 303M* 307E**, 308L***, 310T***, 390E**, 391T*, 400R*, 403R***
Izumo1	Glires	7	2.72	0.99	n.a.
Izumo1	Laurasiatheria	14	53.77***	44.68***	65; 28P**, 118H*, 167K***, 173E*, 193I*, 274Q*, 277T*, 278A*, 298V**, 302-*, 303P*, 353-*, 354T*, 382-*, 385-**
Izumo1	Primates	9	0.68	0.37	n.a.
Izumo1	non-mammals	18	0.50	0.30	n.a.

*$p \geq 0.95$, **$p \geq 0.975$, ***$p \geq 0.99$.

Table 2. Mirrortree Web Server [42] identifies lineage-specific signals of coevolution between mammalian Izumo1 and Juno. Correlation values were generated using both Newick trees created in RAxML (figure 4), and multiple sequence alignments (MSA) created with MUSCLE. Italicized correlation values lie above the Mirrortree correlational cut-off of 0.800 and are suggestive of coevolution between Izumo1 and Juno within that group.

group	n	correlation value from trees	p-value	correlation value from MSA	p-value
mammals	30	0.720	<0.000001	*0.817*	<0.000001
Glires	7	*0.866*	<0.000001	0.690	<0.000001
Laurasiatheria	14	0.669	<0.000001	0.775	<0.000001
Primates	9	*0.990*	<0.000001	*0.989*	<0.000001

within the Laurasiatheria (table 2). The 'From multiple alignments' method did not identify a signal in Glires or Laurasiatheria, but did show evidence of coevolution across all 30 mammals analysed (table 2).

These results support the hypothesis that Juno and Izumo1 are experiencing coordinated evolution within the mammals. The strength of the Mirrortree signal appears to vary by clade and by method of analysis. The discrepancies seen between the two methods could be explained by tree support. For example, low bootstrap support within a portion of the Laurasiatheria (figure 4) might factor into the non-significant signal of coevolution detected for that clade. The Izumo1 clade containing horse (*Equus caballus*) has a bootstrap support value of 45 (figure 4), and the horse phylogenetic position is one of the least correlated across the analysed mammals. Additionally, the 'From trees' approach used trees built from nucleotide alignments, while the 'From multiple alignments' method requested protein alignments, which it used to construct each phylogenetic tree. Analysing trees built using nucleotides versus proteins could also affect the signals identified through Mirrortree.

3. Discussion and conclusion

Here, I describe the evolutionary history of Juno and Izumo1 in the context of their gene families. In doing so, new Izumo genes in tetrapods and fish with potential roles in reproduction are described for the first

time, and the mammalian origin of Juno is elucidated. The discovery of disparate evolutionary origins for these proteins raises a number of future research questions. What is the function of Izumo1 outside of Mammalia? How did the interaction between Izumo1 and the ancestral folate receptor first arise? Since sperm, eggs and the modes of fertilization vary across the vertebrate phylogeny, whether non-mammalian Izumo genes possess reproductive functionality is currently uncertain, but merits further investigation. Targeted examination of expression patterns for these genes across the phylogeny and throughout the vertebrate body is one potentially fruitful approach to address these questions.

Through bioinformatic analyses, I present evidence that the first essential sperm–egg binding pair discovered in any organism is evolving under positive selection. I demonstrate that different clades have driven the positive selection identified in each binding partner, and that the mammalian conservation of Izumo1–Juno binding ability and their essential function in sperm–egg recognition have probably resulted in coevolution between these two proteins. These results support the long-standing hypotheses that rapid evolution and positive selection observed in reproductive proteins are often driven by coevolution between male and female proteins, either as an arms race (through sexual conflict) or through adaptations to species-specific gamete recognition systems that are key to avoiding hybridization. For example, positive selection and coevolution between lysin and VERL of marine abalone are believed to have resulted from one or both of these processes [10,43].

It is unlikely that Izumo1 and Juno play major roles in species-specific gamete recognition for two reasons. The first is that Juno localizes on the egg membrane. This membrane sits interior to the zona pellucida (ZP), a major barrier in preventing hetero-specific sperm from binding to eggs in mammals [44]. The second is the recent finding that cross-species Izumo1–Juno binding is possible between human Izumo1 and Juno on the Syrian golden hamster (*Mesocricetus auratus*) egg following removal of the ZP. In this species pair, it appears that Juno can bind hetero-specific Izumo1; however, this trend does not appear to be consistent across mammals [3]. The same manuscript reports that human Izumo1 does not bind to mouse Juno, suggesting that some level of species recognition might be carried out between Izumo1 and Juno after sperm penetrates the ZP [3].

The possibility that Izumo1 and Juno are coevolving under the pressures of sexual conflict is also an interesting one, which would probably be best studied in a small clade of species with varied levels of sperm competition. A recent study using this model identified positive selection and coevolution between Izumo1 on sperm and CD9 on the egg across the *Mus* genus [7]. The influence of sexual selection on the evolution of a protein pair can be examined by comparing lineage-specific ω values (an indicator of selective pressure) to a proxy of sexual selection (e.g. relative testes mass). This method produced negative results for CD9 and Izumo1, but could be applied to Juno and Izumo1 in the future [7].

While it is clear that selection analyses conducted across distantly related species can uncover signals of positive selection at high taxonomic levels (e.g. Mammalia), it is important to note that they often overlook lineage-specific or recent bouts of selection, and can be biased based on the sequenced taxa that are currently available [45,46]. For example, Izumo1 is reportedly evolving under positive selection in *Mus*, despite appearing to be under the influence of purifying selection in Glires [7]. In other instances, the coarse signals identified by analysing previously sequenced species remain unchanged with the addition of new closely related species. When Izumo1 was first studied within Primates, purifying selection was identified using sequence data for just five species from online databases [13]. Recently, both CD9 and Izumo1 were sequenced and analysed in 12 Primates, and neither gene was found to be evolving adaptively in these groups [47]. If there are lineage-specific bouts of positive selection acting on Izumo1 within Primates, they are probably hidden at the genus level, as seen in *Mus*.

Although Izumo1 and CD9 do not appear to be evolving under positive selection in the Primates, they again display signals of coevolution [47]. Izumo1 and CD9 are not known to interact directly, but the coevolutionary signals identified in *Mus* and Primates are suggestive of either functional associations or interaction at the molecular level [7,47]. A similar conclusion can be drawn from the signal of coevolution identified between Izumo1 and Juno in the current paper. Since the Izumo1–Juno signal of coevolution has already been validated through molecular data describing the essential sperm–egg binding behaviour of these proteins, my results lend support to methods such as Mirrortree for candidate gene screening and potential identification of interacting proteins using bioinformatics [2,48]. Understanding the evolutionary histories of gene families and examining patterns of coevolution can help inform functional studies in humans, model and non-model organisms.

References

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

2. Bianchi E, Doe B, Goulding D, Wright GJ. 2014 Juno is the egg Izumo receptor and is essential for mammalian fertilization. *Nature* **508**, 483–487. (doi:10.1038/nature13203)

3. Bianchi E, Wright GJ. 2015 Cross-species fertilization: the hamster egg receptor, Juno, binds the human sperm ligand, Izumo1. *Phil. Trans. R. Soc. B* **370**, 20140101. (doi:10.1098/rstb.2014.0101)

4. Yáñez-Mó M, Gutiérrez-López MD, Cabañas C. 2011 Functional interplay between tetraspanins and proteases. *Cell. Mol. Life Sci.* **68**, 3323–3335. (doi:10.1007/s00018-011-0746-y)

5. Le Naour F. 2000 Severely reduced female fertility in CD9-deficient mice. *Science* **287**, 319–321. (doi:10.1126/science.287.5451.319)

6. Inoue N, Hamada D, Kamikubo H, Hirata K, Kataoka M, Yamamoto M, Ikawa M, Okabe M, Hagihara Y. 2013 Molecular dissection of IZUMO1, a sperm protein essential for sperm–egg fusion. *Development* **140**, 3221–3229. (doi:10.1242/dev.094854)

7. Vicens A, Roldan E. 2014 Coevolution of positively selected IZUMO1 and CD9 in rodents: evidence of interaction between gamete fusion proteins? *Biol. Reprod.* **90**, 1–9. (doi:10.1095/biolreprod.113.116871)

8. Swanson W, Vacquier V. 2002 The rapid evolution of reproductive proteins. *Genetics* **3**, 137–144. (doi:10.1038/nrg733)

9. Haerty W *et al.* 2007 Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* **177**, 1321–1335. (doi:10.1534/genetics.107.078865)

10. Galindo BE, Vacquier VD, Swanson WJ. 2003 Positive selection in the egg receptor for abalone sperm lysin. *Proc. Natl Acad. Sci. USA* **100**, 4639–4643. (doi:10.1073/pnas.0830022100)

11. Dorus S, Wasbrough ER, Busby J, Wilkin EC, Karr TL. 2010 Sperm proteomics reveals intensified selection on mouse sperm membrane and acrosome genes. *Mol. Biol. Evol.* **27**, 1235–1246. (doi:10.1093/molbev/msq007)

12. Wong A. 2011 The molecular evolution of animal reproductive tract proteins: what have we learned from mating-system comparisons? *Int. J. Evol. Biol.* **2011**, 1–9. (doi:10.4061/2011/908735)

13. Grayson P, Civetta A. 2012 Positive selection and the evolution of Izumo genes in mammals. *Int. J. Evol. Biol.* **2012**, 958164. (doi:10.1155/2012/958164)

14. Ellerman D, Pei J, Gupta S. 2009 Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. *Mol. Reprod. Dev.* **76**, 1188–1199. (doi:10.1002/mrd.21092.Izumo)

15. Elwood PC. 1989 Molecular cloning and characterization of the human folate-binding protein cDNA from placenta and malignant tissue culture (KB) cells. *J. Biol. Chem.* **264**, 14 893–14 901.

16. Shen F, Ross JF, Wang X, Ratnam M. 1994 Identification of a novel folate receptor, a truncated receptor, and receptor type .beta. in hematopoietic cells: cDNA cloning, expression, immunoreactivity, and tissue specificity. *Biochemistry* **33**, 1209–1215. (doi:10.1021/bi00171a021)

17. Spiegelstein O, Eudy JD, Finnell RH. 2000 Identification of two putative novel folate receptor genes in humans and mouse. *Gene* **258**, 117–125. (doi:10.1016/S0378-1119(00)00418-2)

18. Petronella N, Drouin G. 2014 Purifying selection against gene conversions in the folate receptor genes of primates. *Genomics* **103**, 40–47. (doi:10.1016/j.ygeno.2013.10.004)

19. Lu Y, Sega E, Leamon CP, Low PS. 2004 Folate receptor-targeted immunotherapy of cancer: mechanism and therapeutic potential. *Adv. Drug Deliv. Rev.* **56**, 1161–1176. (doi:10.1016/j.addr.2004.01.009)

20. Peer D, Karp JM, Hong S, Farokhzad OC, Margalit R, Langer R. 2007 Nanocarriers as an emerging platform for cancer therapy. *Nat. Nanotechnol.* **2**, 751–760. (doi:10.1038/nnano.2007.387)

21. Chen C, Ke J, Zhou XE, Yi W, Brunzelle JS, Li J, Yong E-L, Xu HE, Melcher K. 2013 Structural basis for molecular recognition of folic acid by folate receptors. *Nature* **500**, 486–489. (doi:10.1038/nature12327)

22. Wibowo AS, Singh M, Reeder KM, Carter JJ, Kovach AR, Meng W, Ratnam M, Zhang F, Dann CE. 2013 Structures of human folate receptors reveal biological trafficking states and diversity in folate and antifolate recognition. *Proc. Natl Acad. Sci. USA* **110**, 15 180–15 188. (doi:10.1073/pnas.1308827110)

23. Turner LM, Hoekstra HE. 2008 Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* **52**, 769–780. (doi:10.1387/ijdb.082577lt)

24. Altschup SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)

25. NCBI Resource Coordinators. 2013 Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **41**, D8–D20. (doi:10.1093/nar/gks1189)

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

27. Edgar RC. 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. (doi:10.1093/nar/gkh340)

28. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013 MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729. (doi:10.1093/molbev/mst197)

29. Ronquist F, Huelsenbeck JP. 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574. (doi:10.1093/bioinformatics/btg180)

30. Stamatakis A. 2006 RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690. (doi:10.1093/bioinformatics/btl446)

31. Flicek P *et al.* 2012 Ensembl 2012. *Nucleic Acids Res.* **40**, D84–D90. (doi:10.1093/nar/gkr991)

32. Darriba D, Taboada GL, Doallo R, Posada D. 2012 jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772. (doi:10.1038/nmeth.2109)

33. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009 BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421. (doi:10.1186/1471-2105-10-421)

34. Benjamini Y, Speed TP. 2012 Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Res.* **40**, e72. (doi:10.1093/nar/gks001)

35. Hughes AL, Friedman R. 2008 Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes. *Mol. Biol. Evol.* **25**, 2681–2688. (doi:10.1093/molbev/msn207)

36. Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)

37. Yang Z, Wong WSW, Nielsen R. 2005 Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118. (doi:10.1093/molbev/msi097)

38. Paradis E, Claude J, Strimmer K. 2004 APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290. (doi:10.1093/bioinformatics/btg412)

39. Revell LJ. 2012 phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223. (doi:10.1111/j.2041-210X.2011.00169.x)

40. Swanson WJ, Vacquier VD. 1998 Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**, 710–712. (doi:10.1126/science.281.5377.710)

41. Pazos F, Valencia A. 2001 Similarity of phylogenetic trees as indicator of protein–protein interaction. *Protein Eng.* **14**, 609–614. (doi:10.1093/protein/14.9.609)

42. Ochoa D, Pazos F. 2010 Studying the co-evolution of protein families with the Mirrortree web server. *Bioinformatics* **26**, 1370–1371. (doi:10.1093/bioinformatics/btq137)

43. Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. 2009 Coevolution of interacting

fertilization proteins. *PLoS Genet.* **5**, e1000570. (doi:10.1371/journal.pgen.1000570)

44. Wassarman PM. 1990 Profile of a mammalian sperm receptor. *Development* **108**, 1–17.

45. Grayson P, Civetta A. 2013 Positive selection in the adhesion domain of *Mus* sperm *Adam* genes through gene duplications and function-driven

gene complex formations. *BMC Evol. Biol.* **13**, 217. (doi:10.1186/1471-2148-13-217)

46. Civetta A. 2012 Fast evolution of reproductive genes: when is selection sexual? In *Rapidly evolving genes and genetic systems* (eds R Singh, J Xu, R Kulathinal), pp. 165–175. Oxford, UK: Oxford University Press.

47. Claw KG, George RD, Swanson WJ. 2014 Detecting coevolution in mammalian sperm–egg fusion proteins. *Mol. Reprod. Dev.* **81**, 531–538. (doi:10.1002/mrd.22321)

48. Ochoa D, Juan D, Valencia A, Pazos F. 2015 Detection of significant protein coevolution. *Bioinformatics* **31**, 2166–2173. (doi:10.1093/bioinformatics/btv102)