BOR Papers in Press. Published on March 26, 2014 as DOI:10.1095/biolreprod.113.116871

Coevolution of Positively Selected IZUMO1 and CD9 in Rodents: Evidence of Interaction Between Gamete Fusion Proteins?[1]

Alberto Vicens and Eduardo R. S. Roldan[2]
Reproductive Ecology and Biology Group, Museo Nacional de Ciencias Naturales (CSIC), Madrid, Spain

Short title: Coevolution of IZUMO1-CD9
Keywords: Coevolution, gamete fusion, positive selection, IZUMO1, CD9, rodents, speciation.

[1]Supported by grant CGL2011-26341 to E.R.S.R. and postgraduate scholarship BES-2009-029239 to A.V., both from the Spanish Ministry of Economy and Competitiveness.

[2]Correspondence: E.R.S. Roldan, Reproductive Ecology and Biology Group, Museo Nacional de Ciencias Naturales (CSIC), c/Jose Gutierrez Abascal 2, 28006-Madrid (Spain).
E-mail: roldane@mncn.csic.es

ABSTRACT

Proteins involved in sexual reproduction are known to evolve rapidly, often as the result of positive Darwinian selection, although the selective forces driving such adaptive changes are poorly understood. A process of coevolution between proteins in male and female gametes may promote rapid divergence of fertilization proteins. In the mouse, only two proteins have been shown so far to be essential for sperm-egg fusion, IZUMO1 in the sperm cell and CD9 in the egg. The role of these proteins has not been fully elucidated, and it has been suggested that they may act as fusogens, interacting in *trans* with proteins on the other cell, or regulators of fusogens through *cis* interactions. Here we analyze the evolution of IZUMO1 and CD9 in a group of rodent species. To assess possible protein interactions between IZUMO1 and CD9, we examined potential coevolution based on analyses of correlated evolutionary rates. We found evidence that both proteins evolve adaptively, with a more intense signal of positive selection in IZUMO1. In addition, our findings suggest that these proteins may have some form of interaction, although they have not been regarded as fusogens interacting directly with each other. The adaptive divergence of IZUMO1 and CD9 could influence reproductive compatibility and, thus, these proteins may participate in the establishment of specific sperm-egg recognition systems. Further studies are required to uncover the role of IZUMO1 and CD9 during gamete fusion in order to understand the molecular basis of their coevolution, as other selective forces could also lead to general signatures of coevolution.

Summary: Correlated evolution between extracellular domains of IZUMO1 and CD9 has been found, suggesting that they could undergo specific molecular interactions.

INTRODUCTION

The evolutionary forces driving the adaptive divergence of reproductive proteins are still poorly understood. Postcopulatory sexual selection (e.g., sperm competition, sexual conflict) acting on gametes in events leading to and ending in fertilization has been proposed to be a powerful selective force [1-3]. In proteins mediating gamete recognition (also known as "fertilization proteins"), a process of coevolution between sperm and egg proteins may be a

major force driving their adaptive evolution [2-4]. This model of coevolution is based on the idea that selective pressures will favor changes in a fertilization protein to adapt it to changes in its interacting partner, maintaining species-specific fertilization through species divergence. Thus, the coevolution of interacting proteins can contribute to barriers of reproductive isolation that lead to speciation.

Several approaches have been used to predict protein interactions by searching for genes that show a correlation of evolutionary rates [5–8]. These approaches are based on the hypothesis that proteins that have a direct, physical interaction, or simply participate together in the same biological process, would experience similar evolutionary pressures. A potential signature of coevolution is observed when phylogenies for two interacting proteins show correlated evolutionary rates along the branches. This is because, as the evolutionary rate along a phylogenetic lineage is accelerated due to mutations occurring in one protein, its cognate is expected to undergo adaptive changes to maintain a stable interaction in the same lineage. Thus, a correlation of evolutionary rates between branches of two phylogenetic trees is expected.

Sperm-egg interaction involves different events ending with the fusion of membranes [9]. In mammals, sperm cells have to recognize and bind to the specific zona pellucida. Exocytosis of acrosomal contents (the so-called acrosome reaction) is required before spermatozoa penetrate the zona but it is still a matter of debate whether this occurs before or during sperm interaction with the zona pellucida [10,11]. After zona penetration, the sperm-membrane attaches to the egg membrane and this ultimately leads to sperm-egg fusion.

Two well characterized proteins are essential for sperm-egg fusion: IZUMO1 in the sperm cell and CD9 in the egg [12,13]. IZUMO1 is a member of the multiprotein family IZUMO, which belongs to the immunoglobulin superfamily (IgSF) of proteins. Mouse IZUMO1 is a ~56-kDa transmembrane protein containing a single immunoglobulin (Ig) domain in the extracellular region and a N-terminal “IZUMO domain” characteristic of the IZUMO family members [14]. IZUMO1 is expressed exclusively in the testis and is exposed on the sperm surface after acrosomal exocytosis [13,15]. Experiments using knockout mice showed that *Izumo1*$^{-/-}$ males were infertile, despite them being viable and having normal mating behavior. In vitro fertilization (IVF) assays revealed that *Izumo1*-deficient sperm were normal and able to bind and penetrate the zona pellucida, but incapable of fusing with the egg membrane [13]. CD9 is an integral membrane protein belonging to the tetraspanin family, whose members have four transmembrane domains [16]. Expression of *Cd9* has been detected in a wide variety of cells, where it associates to integrins and other membrane proteins to form multimolecular complexes [17]. Nonetheless, CD9 has been identified to have an essential function only in the egg. *Cd9*$^{-/-}$ females were normal but infertile, and IVF experiments revealed that infertility was caused by a failure of sperm-egg fusion [12,18,19].

Although the importance of IZUMO1 and CD9 in mouse sperm-egg fusion has been clearly demonstrated, the underlying molecular mechanisms of fusion have not been fully established. Given that both IZUMO1 and CD9 lack features of fusogenic molecules, it has been proposed that these proteins may function as molecules associated with fusogens. Indeed, several proteins have been observed to associate with IZUMO1 in the sperm membrane [20,21] and with CD9 in the egg membrane [22-24].

The aim of our study was to predict and analyze potential interactions between IZUMO1 and CD9 through the search of similar evolutionary histories in a group of closely related rodent species. We first performed comparative analyses to search for evidence of positive selection in IZUMO1 and CD9, assessed whether signals of similar intensity of positive selection were

identified, and mapped such residues with positive selection. Then, we estimated coevolution between this pair of proteins by correlating the ratios of nonsynonymous substitutions (d_N) and synonymous substitutions (d_S) at each branch of the phylogeny [6]. This approach, comparing protein-coding nucleotide sequences, offers several advantages with regards to previous methods using pairwise amino acid distances [5,7,8]. The principal advantages are: (a) the variation in amino acid substitution rate attributable to the variation in nucleotide mutation rate is corrected; (b) comparison of coding sequences allows estimation of multiple amino acid changes at the same codon, and (c) d_N/d_S ratios do not need to be corrected for the similarity effects due to the underlying species phylogeny [6]. This approach has been applied before to the pair of interacting proteins lysin and VERL in free-spawning marine invertebrates, detecting robust signatures of coevolution and demonstrating the power of this method [2,25]. Since earlier studies focused only on broadcast spawners, we understand that, to the best of our knowledge, this is the first study to address protein coevolution in mammalian gametes. Furthermore, in broadcast spawners, proteins mediating gamete interaction are often responsible for reproductive isolation, and their coevolution is driven by sexual conflict and density-dependent selection [26]. In animals with internal fertilization, underlying mechanisms and selective forces leading to postcopulatory prezygotic reproductive isolation are less clear. Thus, our study provides a framework for future studies that may sed light on gamete protein coevolution.

MATERIAL AND METHODS

Animals and Ethics

Gene sequences were analyzed from the following nine murid species: *Mus musculus musculus, Mus m. castaneus, Mus m. domesticus, Mus spretus, Mus macedonicus, Mus spicilegus, Mus caroli, Mus pahari* and *Mus minutoides.* These species belong to a recently derived genus [27], with a wide range of mating systems and sperm competition levels [28], thus constituting an ideal model group to study the evolution of reproductive proteins.

Animals were purchased from the Institut des Sciences de l'Evolution, CNRS-Université Montpellier 2, France. Animal handling and housing followed the standards of the Spanish Animal Protection Regulation RD1201/2005, which conforms to European Union Regulation 2003/65. This study was approved by the Bioethics Committee of the Consejo Superior de Investigaciones Científicas (CSIC, Spain).

Sequence Data and Phylogenetic Trees

Five males and five females of each species were used as sources of materials for sequencing. Sequence data were obtained by reverse transcriptase-polymerase chain reaction (RT-PCR). Tissue samples were obtained from testis for IZUMO1 and from ovaries for CD9. Approximately 50 mg of sample in 700 µl of TRK buffer were homogenized using a rotor-stator homogenizer and total RNA was isolated using the total RNA Kit I (Omega, Madrid, Spain). During the purification protocol, DNA was digested using the RNase free-DNase set (Omega). Reverse transcription was performed using the SuperScript III First-Strand Synthesis System (Life Technologies, Madrid, Spain). First strand cDNA was synthesized on 1-2 µg of total RNA incubated with 50 µM oligo$(dT)_{20}$ primer followed by RT-PCR with SuperScript III Reverse Transcriptase (Life Technologies).

Izumo1 and *Cd9* coding sequences were amplified from cDNA products. RT-PCR primers were designed based on conserved sequences between *Mus musculus* (NCBI accessions NM_001018013 and NM_007657) and *Rattus norvegicus* (NM_001017514 and NM_053018)

located within untranslated regions (UTR) approximately 50 bp from start and stop codons. PCR mixtures were prepared in a 50 µl volume containing PCR Reaction buffer 1x (Roche, Barcelona, Spain), 2.5 mM $MgCl_2$ (Roche), 0.8 mM dNTPs mix supplying 0.2 mM of each deoxinucleotide triphosphate (Applied-Biosystems, Barcelona, Spain), 0.25 mM of forward and reverse primers (Life Technologies), 2 U of DNA polymerase (Biotools, Madrid, Spain), and 10-100 ng/µl of cDNA template. PCRs were performed in a Veriti thermocycler (Applied-Biosystems). PCR conditions consisted in an initial denaturation of 94ºC for 5 min, followed by 35 cycles with a denaturation step of 94ºC for 30 s, an annealing step at 54-58ºC (depending on primer pair) for 60 s and an elongation step at 72ºC for 75 s, followed by a final extension of 7 min. PCR products were purified with the E.Z.N.A. Cycle Pure Kit (Omega) and fragments were sequenced in an automatic sequencer (Secugen S.L., Madrid) using internal primers.

The identity of the sequenced fragments as *Izumo1* or *Cd9* transcripts was verified using BLASTn algorithm (NCBI). Sequence processing was carried out using BioEdit (http://www.mbio.ncsu.edu/bioedit/bioedit.html). Sequence traces were bound to the coding region using CDS for *Mus musculus* as reference and translated from the first frame. Correct translation was checked by comparing protein sequences to those of *Mus musculus* retrieved from GenBank (NCBI accessions NP_001018013 for *Izumo1* and NP_031683 for *Cd9*). Coding sequences were aligned by ClustalW, varying penalty parameters for gap opening and gap extension to test the robustness of alignment. Consensus sequences for each species were obtained by aligning fragments from at least 3 individuals.

In order to test whether sequences annotated as orthologs belong to the same gene, phylogenies were constructed from *Izumo1* and *Cd9* alignments using Maximum Likelihood (PhyML 3.0) and Bayesian (MrBayes 3.1) methods [29,30]. The reliability of tree branching by Maximum Likelihood was assessed running a bootstrap with 1,000 replicates. In Bayesian analyses, estimates of the posterior probability distribution were obtained in 150,000 generations for *Izumo1* and 100,000 generations for *Cd9*, and 25% of samples were used to summarize the parameters and the trees in each case.

Statistical selection for the best-fit model of DNA substitution was done using JmodelTest 0.1.1 [31], choosing the model with highest likelihood. *Rattus norvegicus* was used as outgroup to establish the root of the phylogenies.

Tests of Positive Selection

The nonsynonymous/synonymous substitutions ratio ($dN/dS = \omega$) is a robust indicator of the selective pressure on a gene, with $\omega = 1$ indicating neutral evolution, $\omega < 1$ purifying selection, and $\omega > 1$ diversifying positive selection [32]. Nonetheless, amino acids in a protein are expected to be under different selective constraints and only a small proportion of residues will be subjected to adaptive evolution. We thus applied models that account for variable ω ratios among amino acid sites allowing to identify residues under positive selection [33,34]. These site-models were applied using the Codeml tool implemented in PAML package v4.4 [35]. Likelihood-ratio tests (LRT) were applied to compare the likelihood of a null model that does not allow sites under selection, with the likelihood of a selection model including an additional class allowing values of $\omega > 1$. Two classes of tests were performed. In the first instance, we compared two models assuming a beta distribution for ω values. In this case, the null model M7 that limits ω between 0 and 1 is compared to the alternative model M8, which adds an extra class of sites with the ω ratio allowed to be greater than 1 [33,34]. A third test compared the likelihood of the model M8 to the likelihood of a null model M8a in which ω was fixed to 1 in order to avoid

detecting false signatures of positive selection as a result of functional relaxation [36]. The LRT compared twice the log-likelihoods of selection and neutral models to critical values from a chi-square distribution with the degrees of freedom equal to the difference in the number of parameters between the two models. If the LRT is significant, this implies that selection models show a better fit and thus positive selection can be inferred. A Bayes empirical Bayes (BEB) approach [37] was used to calculate the probability that each site belong to a particular class in terms of the posterior ω value, and those residues with posterior probabilities higher than 0.95 of having $\omega > 1$ were assumed to be under significant positive selection.

Codeml analyses were performed using *Rattus norvegicus* as outgroup without setting a molecular clock. Arbitrary parameters such as codon frequency and number of gamma categories were optimized for the best fit of data according to the likelihood of each analysis.

Analysis of 3D Protein Structure

Protein 3D structure models for *Izumo1* and *Cd9* were searched using the Protein Model Portal (http://www.proteinmodelportal.org/). Structure models of homolog sequences were provided by the ModBase database (http:// http://modbase.compbio.ucsf.edu/) and downloaded from the PDB database (http:// http://www.pdb.org/pdb/home/home.do). Protein structures were visualized and parsed using the PyMol Molecular Graphics System v1.3.

Analyses of Coevolution

We performed linear regression analyses between ω ratios of *Izumo1* and *Cd9* in order to test whether these two proteins show correlated evolutionary rates across murine species. Values of ω were calculated for each branch in the phylogeny using a free-ratio model implemented in Codeml (PAML v4.4) [38]. The tree topology was determined on the basis of previously reported phylogenies for murine rodents [27,39]. Lineage-specific ω values for each locus were also calculated by averaging d_N/d_S ratios from the root of the tree to the respective terminal branch. Rates of evolution through the phylogeny were calculated using the ETE2 toolkit [40].

Linear regressions were conducted with R statistical software (http://www.r-project.org/) and independent analyses were carried out using *Izumo1* and *Cd9* d_N/d_S ratios as dependent variables in each case.

Effects of Sperm Competition

Relative testes mass has been widely used as a reliable index of male postcopulatory intrasexual competition, as larger testes relative to body mass are indicative of higher levels of sperm competition [41], and it shows direct relationships with direct measures of multiple paternity [42]. Five males per species were sacrificed by cervical dislocation and weighed. Testes were removed and then measured and weighed. Values of relative testes mass were calculated using the corrected regression equation for rodents [43]. Relationships between relative testes mass and lineage-specific ω were performed using phylogenetically-corrected linear models to control for associations between variables due to species relatedness [44]. Phylogenetic generalized least-squares (pGLS) approach was conducted using the CAPER package [45] for the R environment.

RESULTS

Sequences and Phylogenetic Analyses

Complete coding sequences of *Izumo1* and *Cd9* genes were obtained by RT-PCR for nine murine species. Using closely related sequences allowed for robust alignments for any combination of penalty parameters in both genes. Alignment of orthologous *Izumo1* sequences spread a length of 1218 positions, identifying a total of 6 regions with insertions or deletions (indel substitutions) (Supplemental Fig. S1; Supplemental Data are available online at www.biolreprod.org). An average ω ratio of 0.937 was calculated over all sites and lineages, and a portion of pairwise comparisons between sequences showed d_N/d_S values greater than 1 (Fig. 1). *Cd9* gene showed a higher degree of conservation than *Izumo1*, reflected by the absence of indels in the alignment (Supplemental Fig. S2), a more reduced average ω ratio (0.334) and a lower distribution of d_N/d_S pairwise values (Fig. 1).

The nucleotide substitution model GTR was selected for the two genes and both proportion of invariable sites and gamma shape parameter value were estimated in each phylogenetic analyses. Differences in topologies were observed between Maximum Likelihood and Bayesian approaches both for *Izumo1* and *Cd9* phylogenies (Supplemental Figs. S3-S4). Moreover, none of the reconstructed trees showed a topology in agreement with the reconstructed species phylogeny [27,39]. The clade grouping the highly related *Mus musculus* subspecies (*Mus m. musculus* and *Mus m. castaneus)* could not be resolved with significant posterior probability in *Izumo1* Bayesian phylogeny. In the case of *Cd9*, only *Mus caroli*, *Mus pahari* and *Mus minutoides* branches could be well resolved by the two methods. The discordance observed between gene-inferred phylogenies and species phylogeny suggests that both *Izumo1* and *Cd9* are undergoing independent evolutionary trajectories with regards to the historical relationships between the species, likely due to the action of directional selective pressures on these genes.

Tests of Positive Selection

We tested whether our target genes are subjected to positive selection by applying models that account for heterogeneous evolutionary rates among codon sites. Evidence of positive selection was detected for *Izumo1* and *Cd9*. In both cases, the likelihood ratio tests (M8 vs M7 and M8 vs M8a) were significant in favor of the selection model (Table 1). A total of 14 positively selected sites were identified in the *Izumo1* alignment with an average ω ratio = 4.654 (Supplemental Fig. S1). *Cd9* showed only two residues under putative positive selection (Supplemental Fig. S2), but the estimated ω for these sites was far greater (ω = 34.381).

The highest proportion of rapidly evolving residues in IZUMO1 fell within the IZUMO domain (Fig. 2A,B), the extracellular region presumably involved in the formation of homodimers. Three out of the four remaining positively selected sites were identified in the short cytosolic domain that regulates the formation of multimers of higher order complexes (Fig. 2A). These results indicate that the potential targets of positive selection in IZUMO1 are those with an ability to form multiprotein complexes in the sperm membrane. A protein model was found for IZUMO1 based on the crystal structure of the extracellular domain of the growth factor receptor ErbB1 [46] (Fig. 2B). Such model aligned with the IZUMO1 extracellular region (spanning positions 32-294) and showed 16% ~~of~~ identity. On this structure, positively selected sites found along the IZUMO domain form a cluster in the region exposed to the extracellular space.

The two codon sites showing a significant signal of positive selection in CD9 were detected in the large extracellular loop (Fig. 2C). This region is involved in sperm-egg fusion [47]. The search for a structural model of the extracellular loop of CD9 yielded as template the extracellular loop of the related human CD81 [48]. The matching region showed a sequence

identity of 24% and encompassed the region between the positions 109-190 of CD9. The positively selected sites were found in the 3D structure flanking the residues SFQ critical for gamete fusion [47] (Fig. 2D). Therefore, the rapid divergence of residues in the large extracellular loop of CD9 may have important implications in the maintenance of the egg fusion machinery.

Coevolution between IZUMO1 and CD9

We tested whether IZUMO1 and CD9 are undergoing coevolution seeking for evidence of correlation between the proteins' evolutionary rates during divergence of the different species. The aim of these analyses was to predict whether IZUMO1 and CD9 may have a direct interaction based on the idea that a pair of interacting proteins will show similar evolutionary histories due to compensatory changes and mutual selective pressures [49].

We used the ω ratio as a measure of evolutionary rate. Estimates of ω were made for each branch of the phylogenetic tree including the 9 murid species (Fig. 3) and using the free ratio model of *codeml* (PAML). Correlations of ω values between IZUMO1 and CD9 were carried out by regression analyses. We employed different approaches to estimate lineage-specific ω ratios. In a first approach we compared the ω values estimated in each terminal branch of the phylogeny. A disadvantage of using this method is that terminal branches frequently lack synonymous substitutions because they have very short lengths, leading to undefined ω ratios which are assigned as values of 999 by the model. In the case of IZUMO1, the branch leading to *Mus m. castaneus* showed an undefined ω value. In order to avoid the disrupting effect of arbitrary values on the regression analysis, we excluded them from the analysis. The linear regression using only reliable terminal-branch ω values was not significant ($F = 2.739$, $P = 0.149$). In an attempt to include all lineages in the regression, we combined those terminal branches showing extreme ω values with the adjacent internal branch (see Fig. 3) to obtain a more accurate ω ratio from the common ancestor [50]. Evidence of a significant relationship between evolutionary rates was not found using this method ($F = 0.3249$, $P = 0.586$).

Given that using terminal branches in evolutionary analyses has several drawbacks, due to the uncertainty of the ω values and the need to apply some arbitrary corrections to avoid them, we developed an approach based on averaging ω values from the root of the tree to each terminal species tip. The estimation of root-to-tip ω ratios provides some benefits to analyses of coevolution. The entire evolutionary history of the loci is accounted for from the last common ancestor, improving the measure of symmetry between phylogenetic trees, interpreted as an indication of coordinated evolution between a pair of proteins [5]. In addition, estimating the rate of evolution from the last common ancestor forces all branches to have the same length and, therefore, the analysis is not subject to temporal effects [51].

Estimates of lineage-specific ω values of IZUMO1 yielded results close to 1 for all species (Fig. 3). These data indicate that IZUMO1 is a rapidly evolving protein, corroborating the results of the analyses described above. In the case of CD9, ω estimates from the root showed smaller values (Fig. 3), supporting the evidence that this protein is evolving at a lower rate than IZUMO1. A significant positive correlation between IZUMO1 and CD9 ω estimates was obtained ($F = 29.19$, $P = 0.001$) (Fig. 4), showing a signature of coevolution between this pair of proteins. Given that the extracellular IZUMO domain of IZUMO1 and the large extracellular loop of CD9 were the regions showing the strongest signals of positive selection, we evaluated whether the divergence of these domains is potentially driving the coevolution of these two proteins. To this end, we estimated lineage-specific ω values using alignments of the IZUMO

domain and the CD9 extracellular loop. A significant positive correlation was obtained when evolutionary rates of the IZUMO domain were compared with the whole sequence of CD9 (F = 18.272, P = 0.0036). On the other hand, ω estimates of the large extracellular loop of CD9 showed a positive relationship as they were correlated with IZUMO1 (F = 69.569, P = 0.00007). When evolutionary rates of the IZUMO domain and the CD9 extracellular loop were compared, a positive correlation with higher significance level was obtained (F = 50.33, P = 0.0002). We also sought for signatures of coevolution in other regions of IZUMO1 (the Ig-like domain and the cytosolic tail) and CD9 (the small extracellular loop and flanking sequences) but we found no significant correlations.

We tested the robustness of our results of IZUMO1-CD9 coevolution using additional approaches. We first evaluated the influence of using a reconstructed species phylogeny by performing analyses of coevolution with alternative tree topologies. We analyzed topologies of trees derived from an alignment including IZUMO1 and CD9 coding sequences by Maximum-Likelihood (ML) and Bayesian Inference (BI) methods. These two approaches yielded different phylogenies and both showed variations in the branching pattern with regards to the reconstructed tree (Supplemental Fig. S5). Nevertheless, regressions between IZUMO1 and CD9 ω estimates were statistically significant (ML tree F = 9.109, P = 0.019; BI tree F = 33.33, P = 0.0007). We also explored an alternative tree which includes a polytomy between *Mus musculus* subspecies in CD9 phylogenies (Supplemental Fig. S5). This phenomenon is frequently observed in phylogenetic analyses of murine rodents [28,52]. Significant positive relationships were also observed using this alternative tree (F = 37.49, P = 0.0005). These results indicate that the correlation of IZUMO1 and CD9 evolutionary rates is not influenced by phylogenetic relationships between species.

A second approach was used to test the possibility that IZUMO1-CD9 coevolution was a false positive result. We performed regression analyses with evolutionary rates estimated for three proteins involved in other gamete functions: CATSPER1, PKDREJ and PRM2. We obtained sequences of *Catsper1, Pkdrej* and *Prm2* for our 9 rodent species. Regression analyses of these genes with *Izumo1* and *Cd9* showed no significant correlations (Table 2). *Prm2* showed a slightly significant correlation with *Izumo1* (F = 6.612, P = 0.042), but the slope was negative in this case. We also analyzed the evolution of a sperm protein involved in fertilization and thought to interact with IZUMO1, i.e., SPESP1 [53]. *Spesp1* did not show signatures of coevolution with *Izumo1* or *Cd9* (Table 2).

Comparison of Replacements Across Lineages

The relative divergence rates between a pair of interacting proteins can provide information about the underlying selective forces acting on each species at the present time [54]. We calculated the number of nonsynonymous changes in IZUMO1 and CD9 phylogenies and compared these values branch by branch (Fig. 5). A higher number of replacements were inferred for IZUMO1 than for CD9 in almost all branches. Nevertheless, the terminal branches leading to *Mus macedonicus* and *Mus spicilegus* showed changes in the CD9 sequence with none in the IZUMO1 sequence. This suggests that, in spite of IZUMO1 being subjected to a stronger positive selection than CD9, the relative rate between these genes may vary across lineages.

Analyses of Sperm Competition

In order to test whether inter-specific differences in postcopulatory sexual selection are driving the coevolution of IZUMO1 and CD9, we used relative testes mass as proxy of sperm

competition [55,56]. We performed phylogenetically corrected-regression analyses between relative testes mass and lineage-specific ω values for each gene. No significant relationships were observed for IZUMO1 ($F = 1.834$, $P = 0.229$) or CD9 ($F = 0.27$, $P = 0.619$).

DISCUSSION

In this study, we analyzed the evolution of two proteins essential for gamete fusion: IZUMO1, present in sperm cells, and CD9, located in the egg plasma membrane. Comparative analyses in 9 murine species revealed evidence of positive selection promoting adaptive changes in *Izumo1* and *Cd9* coding sequences, with positive selection being more intense in domains exposed to the extracellular space. We also analyzed the coevolution of IZUMO1 and CD9 to assess whether they may be interacting proteins. These analyses were based on the prediction that, when a protein undergoes changes in its sequence, the evolutionary rate of its cognate proteins will increase to maintain their molecular interactions, leading to a correlated evolution. We found a significant and positive correlation in the evolutionary rates of IZUMO1 and CD9 across species which suggests that these two proteins could, in some way, act as interacting partners during the binding and fusion of gamete membranes.

Molecular Mechanisms of IZUMO1-CD9 Coevolution

An open question in studies of coevolution is how much of the correlated evolution is attributable to a physical interaction between two proteins or to shared selective pressures due to functional relationships. An approach that can be used to assess whether the signature of coevolution is driven by compensatory changes between interacting partners is to perform detailed correlation analyses of binding interfaces of proteins [57]. In the case of IZUMO1 and CD9, no evidence of physical interaction has been found for these proteins [58]. Nonetheless, in a hypothetical scenario of a direct binding between IZUMO1 and CD9 during sperm-egg fusion, residues in the extracellular domains would be expected to participate in *trans* interactions. Our analyses detected that IZUMO1-CD9 coevolution is mainly based on a correlated evolution between the IZUMO domain and the large loop of CD9, two extracellular regions. Therefore, our results suggest that, as envisaged by this hypothetical scenario, IZUMO1 and CD9 could potentially function as cognate proteins during gamete interaction. In any case, although IZUMO1 and CD9 may be potential sperm-egg binding or fusion partners, caution must be exerted before concluding that a physical interaction is the potential cause of coevolution. It has been proposed that IZUMO1 and CD9 may be part of large protein complexes required for membrane fusion [14, 47]. This is based on evidence that the IZUMO domain of IZUMO1 associates with other sperm membrane proteins [14] and, similarly, that the extracellular loop of CD9 associates with other egg membrane proteins [47]. It is possible that IZUMO1 and CD9 act as pivotal components in the organization and stability of the fusion machinery in each gamete. Thus, mutations in the extracellular loop of CD9 that is involved in protein interactions in the egg membrane could alter the conformation of the egg fusion complex, reducing the affinity for the sperm fusion complex. Selective pressures could then drive adaptive changes in the proteins of the sperm fusion complex to compensate for changes in their egg counterpart, thus maintaining the interaction between both fusion machineries. Coevolution between two proteins with an indirect role in sperm-egg fusion may therefore occur in the absence of a direct interaction between such proteins.

The involvement of IZUMO1 and CD9 in the same biological process of membrane fusion implies some form of functional association between them. Given that groups of proteins

participating in the same processes are frequently subjected to similar functional constraints, this can lead to the detection of correlated evolution between functionally related proteins [49,57]. However, this may not always be the case, as we found that SPESP1, a fertilization protein presumably involved in relocation of sperm membrane proteins (including IZUMO1) after acrosomal exocytosis [53] did not show evidence of correlated evolution with IZUMO1 or CD9. This leads to the conclusion that proteins participating in the same processes, with potential functional links, may be distinguished depending on whether they experience correlated evolution or not.

In a model of protein-protein interaction, a pair of proteins would coevolve with a similar number of changes. However, we found that IZUMO1 showed a higher rate of evolution, as well as a stronger signal of positive selection, than CD9. It is possible that single changes in a protein may be compensated by multiple mutations in the partner [2]. Another explanation could be that a protein is interacting with more than one molecule, and thus the evolutionary rate of a single protein is the product of accumulated changes between its multiple partners. As mentioned above, the emerging hypothesis is that IZUMO1 and CD9 could function through interaction in *cis* with other proteins to regulate the organization of the membrane during sperm-egg fusion [58]. Several proteins have been identified or have been suggested to associate with IZUMO1 [20,21] and CD9 [22–24]. Therefore, it is possible that IZUMO1 can associate simultaneously with a higher number of proteins (either in *cis* in the sperm membrane or in *trans* with proteins on the egg surface) than just CD9 during the formation of membrane complexes. On the other hand, because CD9 is expressed in multiple tissues, it is also possible that this protein has roles in cell types other than the egg, and this may result in stronger functional constraints, allowing for a lower rate of changes in CD9. Finally, signatures of coevolution between pair of proteins may be linked to correlated levels of expression [59]. Given that IZUMO1 and CD9 show different expression patterns (IZUMO1 is expressed exclusively in testis, whereas CD9 is expressed in a variety of tissues), coexpression seems unlikely to be responsible for coevolution of these proteins.

Correlation of Evolutionary Rates as an Approach to Detect Protein Coevolution

Coevolution of IZUMO1 and CD9 was detected correlating root-to-tip d_N/d_S (ω) point estimates of each mouse lineage through linear regression analyses. Estimating average ω values from the root of the phylogeny to each species tip is more reliable than methods using ω ratios of the terminal branches. Root-to-tip ω ratios have been previously used to test genotype-phenotype relationships improving the fitting of regressions [60,61], but this is the first time that average ω values are used for analyses of coevolution. Recent studies of coevolution between pairs of interacting proteins applied approaches based on likelihood ratio tests between evolutionary models [2,25]. This method is highly reliable because it accounts for the uncertainty of ω estimates in each branch. Our approach does not account for such uncertainty but corrects it and also provides additional benefits as root-to-tip ω estimates reflect the entire genotypic evolution of a locus and is therefore more suitable to compare the evolutionary history of a pair of proteins.

The robustness of the significant positive correlation between IZUMO1 and CD9 was evaluated with analyses using alternative topologies and several non-related genes. Comparisons of ω values estimated from different tree topologies did no alter the significance of results, indicating that IZUMO1 and CD9 coevolution is not influenced by the phylogenetic relationship of the species and thus any (or several) selective forces must be responsible of this concerted evolution. On the other hand, neither evolutionary rates of IZUMO1 nor CD9 showed significant

positive relationships when compared to those of non-related proteins (PKDREJ, CATSPER1 and PRM2), suggesting that there are no genome-wide correlations in the mouse, and thus supporting the coevolution of IZUMO1 and CD9.

Coevolution of interacting fertilization proteins has been previously reported in marine invertebrates [2,25]. In internally fertilizing species, where gametes (especially spermatozoa) are exposed to multiple conditions and undergo several physiological processes before fertilization, one may expect mechanisms of gamete interaction to be less stringent and, thus, to show weak or no signal of coevolution of certain proteins. Our findings revealed clear signals of coevolution of fertilization proteins in internally fertilizing species, which may occur either as a consequence of direct interactions or functional associations.

Our study predicts functional interactions between a pair of fertilization proteins based on the characterization of their coevolution, even though the mechanisms underlying their interaction are not yet well understood. Our findings thus open a very promising area for future work because they underscore the reliability of evolutionary methods to predict functional interactions between reproductive proteins, providing a good alternative to functional analyses [62]. In the particular case of IZUMO1 and CD9, coevolutionary analyses could also be applied to predict interactions of these proteins with other candidate proteins present in gamete membranes. For example, some sperm proteins such as ACE3 and TSSK6 have been detected to interact with IZUMO1 in the sperm surface during fertilization [20,21]. Further studies will be required to assess possible evolutionary relationships of IZUMO1 with these or other potential sperm interacting partners. Using a broader approach, in taxa for which large amounts of sequence and SNP data are available, interacting proteins could perhaps be detected searching for associations between allelic variants within populations [2,63].

Selective Forces Driving IZUMO1-CD9 Coevolution

Several selective forces have been suggested to drive the rapid evolution of reproductive genes [1,3,64]. Among these, sexual conflict is a strong candidate that may explain the coevolution of interacting fertilization proteins. Sexual conflict arises as the two sexes have different optimal fitness for a given trait [65]. In most cases, sexual conflict is enhanced by sperm competition. This occurs because sperm competition leads to the evolution of male characters that increase reproductive success resulting in reductions in female fitness due to polyspermy. The egg must thus evolve traits to reduce polyspermy and, in mammals, one possible way to achieve this is through changes in its surface receptors [3].

Our phylogenetically-corrected regression analyses did not detect a significant relationship between species-specific relative testes mass and IZUMO1 or CD9 ω estimates. Therefore, we did not find evidence of sperm competition driving the evolution of these two proteins. Nonetheless, this does not imply that sperm competition is fully absent in the rapid evolution of these proteins, and there may be several reasons why we were unable to detect its influence. One explanation is that multiple selective forces could be acting simultaneously on the evolution of this pair of proteins, masking the direct effect of sperm competition.

An alternative approach to infer whether sperm competition and sexual conflict are driving the evolution of reproductive proteins is to examine the relative divergence of a pair of gamete interacting proteins [25,54]. During the evolution of promiscuous species we would expect to find some stages in which female proteins evolve faster than their male partners due to the development of egg defensiveness. In situations where a low number of spermatozoa compete to fertilize an egg, sexual conflict may be relaxed and we would thus expect higher

divergence in male proteins. The relative rate of replacements across mouse species examined in this study is in agreement with these predictions.

We have predicted an interaction between a pair of fertilization proteins based on the identification of correlated evolutionary rates. This is the first time in which analyses of protein coevolution are applied to the study of mammalian reproductive proteins. In spite of the fact that IZUMO1 in the male, and CD9 in the female, have been identified as critical players in mouse sperm-egg fusion, the molecular role of these proteins during this process is still little known. We found correlated evolution between extracellular domains of IZUMO1 and CD9, suggesting that these proteins could undergo some form of specific molecular interaction. A better understanding of the mechanisms underlying the interaction between IZUMO1 and CD9 will be necessary to decipher the molecular basis of their co-evolution.

ACKNOWLEDGMENTS

We are grateful to François Bonhomme and Annie Orth (Institut des Sciences de l'Evolution, CNRS-Université Montpellier 2, France) for facilitating purchase of animals. We also thank Juan Antonio Rielo for managing the animal facilities and Esperanza Navarro for animal care at the Museo Nacional de Ciencias Naturales (CSIC), Madrid.

REFERENCES

1. Clark NL, Aagaard JE, Swanson WJ. Evolution of reproductive proteins from animals and plants. Reproduction 2006; 131:11-22.
2. Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. Coevolution of interacting fertilization proteins. PLoS Genet 2009; 5:e1000570.
3. Swanson WJ, Vacquier VD. The rapid evolution of reproductive proteins. Nat Rev Genetics 2002; 3:137-144.
4. Vacquier VD. Evolution of gamete recognition proteins. Science 1998; 281:1995-1998.
5. Pazos F, Valencia A. Similarity of phylogenetic trees as indicator of protein-protein interaction. Protein Eng 2001; 14:609-614.
6. Clark NL, Aquadro CF. A novel method to detect proteins evolving at correlated rates: identifying new functional relationships between coevolving proteins. Mol Biol Evol 2010; 27:1152-1161.
7. Sato T, Yamanishi Y, Kanehisa M, Toh H. The inference of protein-protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. Bioinformatics 2005; 21:3482-3489.
8. Juan D, Pazos F, Valencia A. High-confidence prediction of global interactomes based on genome-wide coevolutionary networks. Proc Natl Acad Sci USA 2008; 105:934-939.
9. Wassarman PM, Gustave O, Place LL. Mammalian fertilization: molecular aspects of gamete adhesion, exocytosis, and fusion. Cell 1999; 96:175-183.
10. Florman HM, Jungnickel MK, Sutton KA. Regulating the acrosome reaction. Int J Dev Biol 2008; 52:503-510.
11. Jin M, Fujiwara E, Kakiuchi Y, Okabe M, Satouh Y, Baba SA, Chiba, K, Hirohashi N. Most fertilizing mouse spermatozoa begin their acrosome reaction before contact with the zona pellucida during in vitro fertilization. Proc Natl Acad Sci USA 2010; 108:4892-4896.

12. Miyado K. Requirement of CD9 on the egg plasma membrane for fertilization. Science 2000; 287:321-324.
13. Inoue N, Ikawa M, Isotani A, Okabe M. The immunoglobulin superfamily protein Izumo is required for sperm to fuse with eggs. Nature 2005; 434:4-8.
14. Ellerman DA, Pei J, Gupta S, Snell WJ, Myles D. Izumo is part of a multiprotein family whose members form large complexes on mammalian sperm. Mol Reprod Dev 2009; 76:1188-1199.
15. Satouh Y, Inoue N, Ikawa M, Okabe M. Visualization of the moment of mouse sperm-egg fusion and dynamic localization of IZUMO1. J Cell Sci 2012; 125:4985-4990.
16. Hemler ME. Tetaspanins functions and associated microdomains. Nat Rev Mol Cell Biol 2005; 6:801-811.
17. Boucheix C, Rubinstein E. Tetraspanins. Cell Mol Life Sci 2001; 58:1189-1205.
18. Kaji K, Oda S, Shikano T, Ohnuki T, Uematsu Y, et al. The gamete fusion process is defective in eggs of Cd9-deficient mice. Nat Genet 2000; 24:279-282.
19. Le Naour F. Severely reduced female fertility in CD9-deficient mice. Science 2000; 287:319-321.
20. Inoue N, Kasahara T, Ikawa M, Okabe M. Identification and disruption of sperm-specific angiotensin converting enzyme-3 (ACE3) in mouse. PLoS One 2010; 5:e10301.
21. Sosnik J, Miranda P V, Spiridonov NA, Yoon S, Fissore RA, Johnson GR, Visconti PE. Tssk6 is required for Izumo relocalization and gamete fusion in the mouse. J Cell Sci 2009; 122:2741-2749.
22. Rubinstein E, Ziyyat A, Prenant M, Wrobel E, Wolf J, Levy S, Le Naour F, Boucheix C. Reduced fertility of female mice lacking CD81. Dev Biol 2006; 290:351-358.
23. Ziyyat A, Rubinstein E, Monier-gavelle F, Barraud V, Kulski O, et al. CD9 controls the formation of clusters that contain tetraspanins and the integrin α6β1 in human and mouse gamete fusion. J Cell Sci 2006; 119:416-424.
24. Glazar A, Evans JP. Immunoglobulin superfamily member IgSF8 (EWI-2) and CD9 in fertilisation: evidence of distinct functions for CD9 and a CD9-associated protein in mammalian sperm-egg interaction. Reprod Fert Dev 2009; 21:293-303.
25. Hellberg ME, Dennis AB, Arbour-Reily P, Aagaard JE, Swanson WJ. The *Tegula* tango: a coevolutionary dance of interacting, positively selected sperm and egg proteins. Evolution 2012; 66:1681-1694.
26. Levitan DR, Ferrell DL. Selection on gamete recognition depends on sex, density, and genotype frequency. Science 2006; 312:267-269.
27. Suzuki H, Shimada T, Terashima M, Tsuchiya K, Aplin K. Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. Mol Phylogenet Evol 2004; 33:626-646.
28. Martín-Coello J, Dopazo H, Arbiza L, Ausió J, Roldan ERS and Gomendio M. Sexual selection drives weak positive selection in protamine genes and high promoter divergence, enhancing sperm competitiveness. Proc Biol Sci 2009; 276:2427-2436.
29. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 2003; 19:1572-1574.
30. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 2010; 59:307-321.

31. Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol 2008; 25:1253-1256.
32. Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 1994; 11:725-736.
33. Yang Z, Nielsen R, Goldman N, Pedersen A. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 2000; 155:431-449.
34. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 1998; 148:929-936.
35. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 2007; 24:1586-1591.
36. Wong WSW, Yang Z, Goldman N, Nielsen R. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 2004; 168:1041-1051.
37. Yang Z, Wong WSW, Nielsen R. Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 2005; 22:1107-1118.
38. Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 1998; 46:409-418.
39. Fabre P-H, Hautier L, Dimitrov D, Douzery EJ. A glimpse on the pattern of rodent diversication: a phylogenetic approach. BMC Evol Biol 2012; 12:88.
40. Huerta-Cepas J, Dopazo J, Gabaldón T. ETE: a python environment for tree exploration. BMC Bioinformatics 2010; 11:24.
41. Birkhead TR, Hosken DJ, Pitnick S. Sperm Biology: An Evolutionary Perspective. Oxford: Academic Press; 2009.
42. Soulsbury CD. Genetic patterns of paternity and testes size in mammals. PLoS One 2010; 5:e9581.
43. Kenagy GJ, Trombulak SC. Size and function of mammalian testes in relation to body size. J Mammal 1986; 67:1-22.
44. Freckleton RP, Harvey PH, Pagel M. Phylogenetic analysis and comparative data: a test and review of evidence. Am Nat 2002; 160:712-726.
45. Orme D. The caper package: comparative analysis of phylogenetics and evolution in R. 2012.
46. Bouyain S, Longo PA, Li S, Ferguson KM, Leahy DJ. The extracellular region of ErbB4 adopts a tethered conformation in the absence of ligand. Proc Natl Acad Sci USA 2005; 102:15024-15029.
47. Zhu G, Miller BJ, Boucheix C, Rubinstein E, Liu CC, et al. Residues SFQ (173-175) in the large extracellular loop of CD9 are required for gamete fusion. Development 2002; 129:1995-2002.
48. Kitadokoro K, Bordo D, Galli G, Petracca R, Falugi F, Abrignani S, Grandi, G, Bolognesi, M. CD81 extracellular domain 3D structure: insight into the tetraspanin superfamily structural motifs. EMBO J 2001; 20:12-18.
49. Pazos F, Valencia A. Protein co-evolution, co-adaptation and interactions. EMBO J 2001; 27:2648-2655.
50. Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT. Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. Nat Genet 2004; 36:1326-1329.

51. Wolf JBW, Künstner A, Nam K, Jakobsson M, Ellegren H. Nonlinear dynamics of nonsynonymous (dN) and synonymous (dS) substitution rates affects inference of selection. Genome Biol Evol 2009; 1:308-319.
52. Sandstedt SA, Tucker PK. Male-driven evolution in closely related species of the mouse genus Mus. J Mol Evol 2005; 61:138-144.
53. Fujihara Y, Murakami M, Inoue N, Satouh Y, Kaseda K, Ikawa M, Okabe M. Sperm equatorial segment protein 1, SPESP1, is required for fully fertile sperm in mouse. J Cell Sci 2010; 123:1531-1536.
54. Civetta A. Shall we dance or shall we fight? Using DNA sequence data to untangle controversies surrounding sexual selection. Genome 2003; 46:925-929.
55. Gomendio M, Harcourt AH, Roldan ERS. Sperm competition in mammals. In: Birkhead TR, Moller AP, (eds.), Sperm Competition and Sexual Selection. London: Academic press; 1998: 667-751.
56. Birkhead TR, Moller AP. Sperm Competition and Sexual Selection. San Diego, CA: Academic press; 1998.
57. Hakes L, Lovell SC, Oliver SG, Robertson DL. Specificity in protein interactions and its relationship with sequence diversity and coevolution. Proc Natl Acad Sci USA 2007; 104:7999-8004.
58. Evans JP. Sperm-egg interaction. Annu Rev Physiol 2012; 74:477-502.
59. Clark NL, Alani E, Aquadro CF. Evolutionary rate covariation reveals shared functionality and coexpression of genes. Genome Res 2012; 22:714-720.
60. Montgomery SH, Capellini I, Venditti C, Barton RA, Mundy NI. Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. Mol Biol Evol; 2011; 28:625-638.
61. Lüke L, Vicens A, Serra F, Luque-Larena JJ, Dopazo H, Roldan ERS, Gomendio, M. Sexual selection halts the relaxation of protamine 2 among rodents. PLoS one 2011; 6:e29247.
62. Findlay GD, Swanson WJ. Proteomics enhances evolutionary and functional analysis of reproductive proteins. BioEssays 2010; 32:26-36.
63. Rohlfs R V, Swanson WJ, Weir BS. Detecting coevolution through allelic association between physically unlinked loci. Am J Hum Genet 2010; 86:674-685.
64. Vacquier VD, Swanson WJ. Selection in the rapid evolution of gamete recognition proteins in marine invertebrates. Cold Spring Harb Perspect Biol 2011; 3:a002931.
65. Parker GA. Sexual conflict over mating and fertilization: an overview. Phil Trans R Soc Lond B Biol Sci 2006; 361:235-259.

FIGURE LEGENDS

FIG. 1. Evolutionary rates of IZUMO1 (**A**) and CD9 (**B**). Each point represents the respective ω value (d_N/d_S ratio) calculated between two species. The line was drawn to show identical d_N and d_S values, resulting in $\omega = 1$.

FIG. 2. IZUMO1 and CD9 protein structure. **A**) Scheme of IZUMO1 representing the known domains, with arrowheads indicating the position of amino acids under positive selection. **B**) Crystal structure of the extracellular region of ErbB4 (PDB ref. 2AHX), which serves as model of the extracellular region of IZUMO1. The purple region represents the IZUMO domain and the

orange region represents the immunoglobulin domain. Residues under positive selection are shown as red spheres. **C**) Scheme of CD9 representing the distribution of positively selected residues (arrowheads) in the large extracellular loop. **D**) 3D-structural model based on the crystal structure of human CD81 extracellular domain (PDB ref. 1G8Q). Active residues required for gamete fusion SFQ (173-175) are represented as pink sticks and sites under positive selection are shown as red spheres.

FIG. 3. Evolutionary rates of IZUMO1 and CD9. Lineage-specific ω ratios were calculated for each terminal branch and averaged from the last common ancestor to each species tip. *Rattus norvegicus* was used as outgroup to establish the position of root tree. Tree topology was reconstructed based on reported rodent phylogenies [27,39].

FIG. 4. Coevolution between IZUMO1 and CD9. The ω values estimated for nine murine rodents are plotted. Note the differences in scale between the two proteins revealing a higher divergence for IZUMO1.

FIG. 5. Number of nonsynonymous changes in IZUMO1 (male symbol) and CD9 (female symbol) along a tree topology of 9 mice species [27,39].

TABLE 1. Tests of positive selection for *Izumo1* and *Cd9* genes.[a]

Gene	N[b]	Ls[c]	$L_{(M7)}$	$L_{(M8)}$	$L_{(M8a)}$	$LRT_{(M8vsM7)}$	$LRT_{(M8vsM8a)}$	$\omega_{(M8)}$
Izumo1	10	376	-3222,934	-3198.834	-3222.552	48.201**	47.437**	4.654
Cd9	9	225	-1153.406	-1145.802	-1153.318	15.208**	15.033**	34.381

[a]Likelihood values of site models (L) and likelihood ratio tests (LRT) are presented.
[b]N, number of compared lineages.
[c]Ls, sequence length (in codons) after alignment gaps are removed.
**Differences between log-likelihood values of models with 99% statistical significance level for 2 degrees of freedom are shown.

TABLE 2. Analyses of genes presumably not involved in sperm-egg interaction or apparently not interacting with IZUMO1 or CD9.

Gene	Putative interacting gene	Intercept	Slope	R	*F*	*P*
Catsper1[a]	*Izumo1*	3.019	-1.181	0.245	2.265	0.176
	Cd9	1.116	-0.536	0.333	3.489	0.104
Pkdrej[b]	*Izumo1*	0.681	1.145	0.038	0.236	0.644
	Cd9	0.04	0.554	0.078	0.508	0.503
Prm2[c]	*Izumo1*	1.886	-1.448	0.445	6.612	0.042
	Cd9	0.494	-0.451	0.356	4.866	0.069
Spesp1[d]	*Izumo1*	1.225	-0.179	0.006	0.04	0.848
	Cd9	0.494	-0.355	0.147	1.202	0.309

[a]The first exon of *Catsper1* (~925 bp) was amplified by standard PCR from genomic DNA.
[b]The region coding for the extracellular REJ domain of *Pkdrej* (~1250 bp) was amplified by standard PCR from genomic DNA.
[c]Whole *Prm2* coding sequence (~445 bp) was amplified by standard PCR from genomic DNA.
[d]Whole *Spesp1* coding sequence (~1197 bp) was amplified by RT-PCR from genomic cDNA.

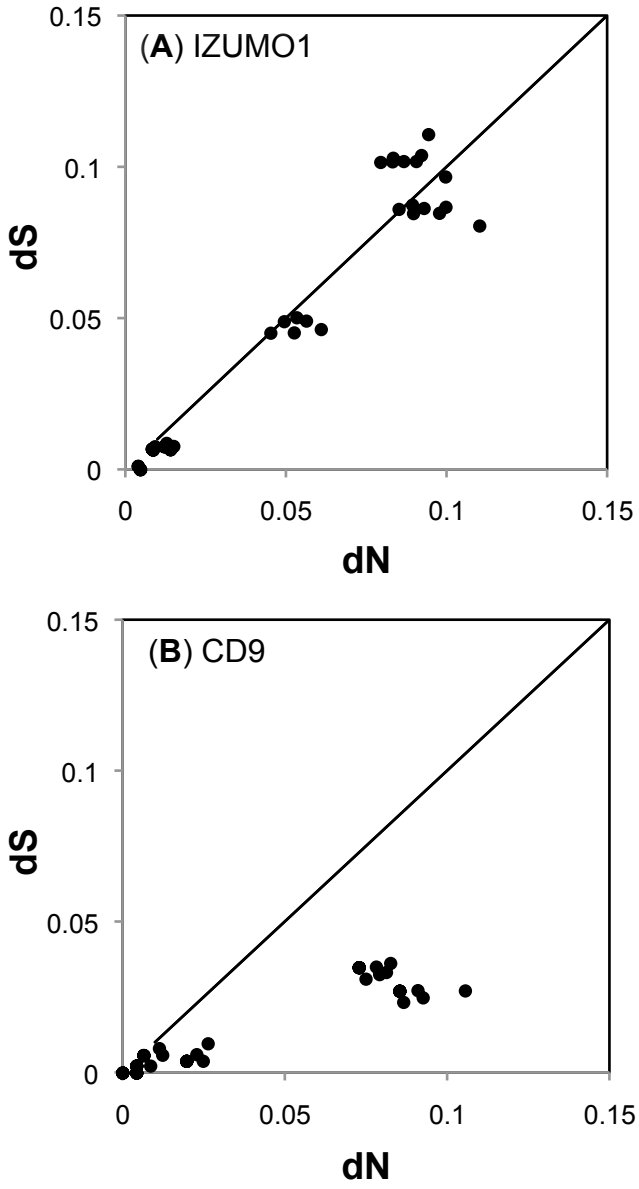
(A) IZUMO1
0.15
0.1
0.05
0
dS
0
0.05
0.1
0.15
dN
(B) CD9
0.15
0.1
0.05
0
dS
0
0.05
0.1
0.15
dN

Figure 1

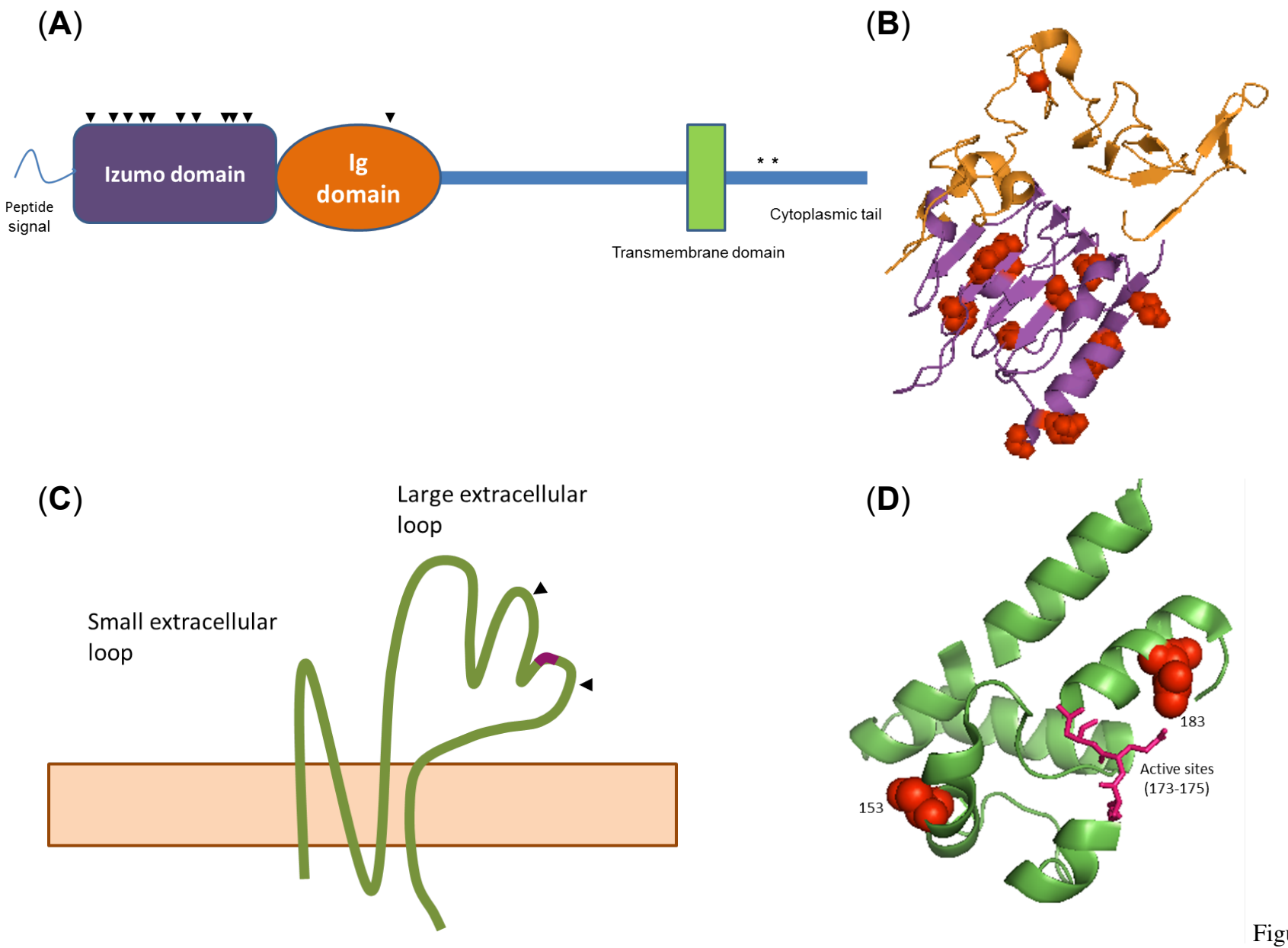
(A)
Izumo domain
Ig domain
Peptide signal
Cytoplasmic tail
Transmembrane domain
(B)
(C)
Large extracellular loop
Small extracellular loop
(D)
183
Active sites (173-175)
153

Figure 2

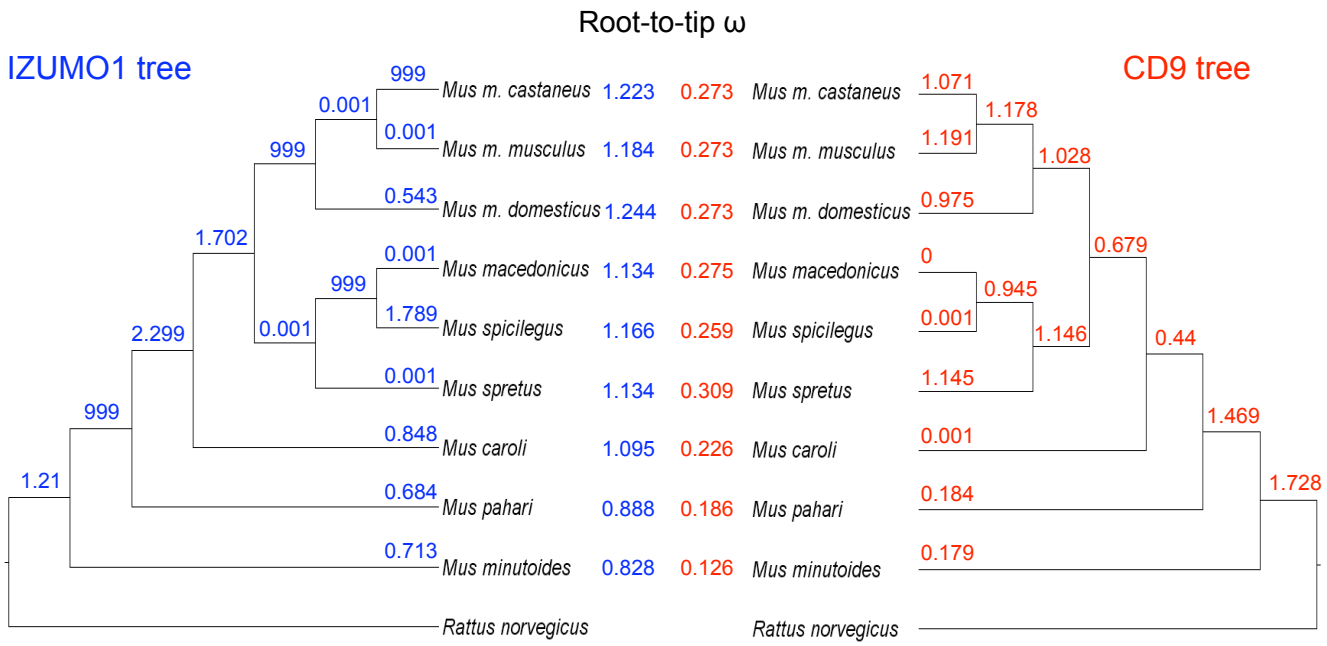
Root-to-tip ω
IZUMO1 tree
CD9 tree
999
0.001
0.001
999
0.543
1.702
0.001
999
0.001
1.789
2.299
0.001
999
0.848
1.21
0.684
0.713
Mus m. castaneus
Mus m. musculus
Mus m. domesticus
Mus macedonicus
Mus spicilegus
Mus spretus
Mus caroli
Mus pahari
Mus minutoides
Rattus norvegicus
1.223
1.184
1.244
1.134
1.166
1.134
1.095
0.888
0.828
0.273
0.273
0.273
0.275
0.259
0.309
0.226
0.186
0.126
Mus m. castaneus
Mus m. musculus
Mus m. domesticus
Mus macedonicus
Mus spicilegus
Mus spretus
Mus caroli
Mus pahari
Mus minutoides
Rattus norvegicus
1.071
1.191
1.178
0.975
1.028
0
0.945
0.001
1.146
1.145
0.679
0.001
0.44
0.184
1.469
0.179
1.728

Figure 3

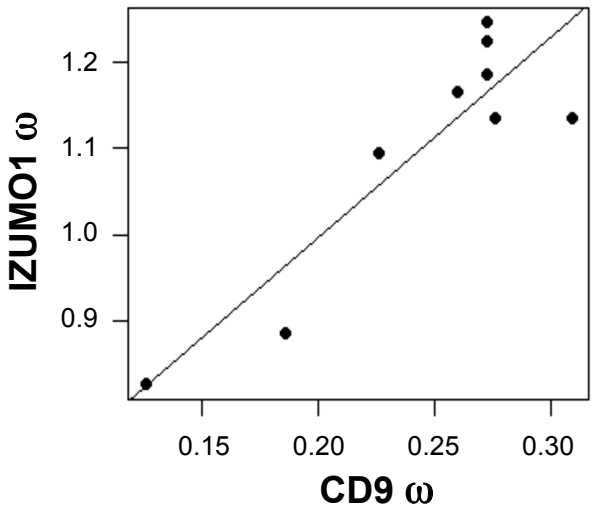

Figure 4

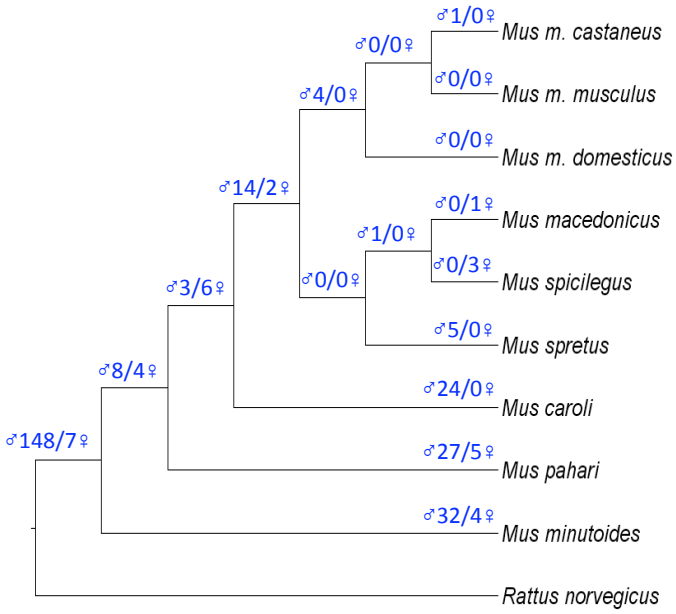
♂1/0♀
Mus m. castaneus
♂0/0♀
♂0/0♀
Mus m. musculus
♂4/0♀
♂0/0♀
Mus m. domesticus
♂14/2♀
♂0/1♀
Mus macedonicus
♂1/0♀
♂0/3♀
Mus spicilegus
♂3/6♀
♂0/0♀
♂5/0♀
Mus spretus
♂8/4♀
♂24/0♀
Mus caroli
♂148/7♀
♂27/5♀
Mus pahari
♂32/4♀
Mus minutoides
Rattus norvegicus

Figure 5