

Article

Variation and Selection in the Putative Sperm-Binding Region of ZP3 in Muroid Rodents: A Comparison between Cricetids and Murines

Margarida Alexandra Duarte ^{1,2,3,4,*} , Carlos Rodríguez Fernandes ^{5,6} , Gerald Heckel ^{7,8},
Maria da Luz Mathias ^{3,4} and Cristiane Bastos-Silveira ⁵

- ¹ Champalimaud Centre for the Unknown, Champalimaud Research, Champalimaud Foundation, Avenida Brasília, 1400-038 Lisboa, Portugal
- ² Museu Nacional de História Natural e da Ciência, Departamento de Zoologia e Antropologia, Universidade de Lisboa, Rua da Escola Politécnica, 58, Lisboa, 1250-102 Lisboa, Portugal
- ³ Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal; mlmathias@fc.ul.pt
- ⁴ Centro de Estudos de Ambiente e Mar, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal
- ⁵ cE3c-Centre for Ecology, Evolution and Environmental Changes, Departamento de Biologia Animal, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal; cafernandes@fc.ul.pt (C.R.F.); cbsilveira@gmail.com (C.B.-S.)
- ⁶ Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal
- ⁷ Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland; gerald.heckel@iee.unibe.ch
- ⁸ SIB Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Amphipole, CH-1015 Lausanne, Switzerland
- * Correspondence: masctduarte@yahoo.com



Citation: Duarte, M.A.; Fernandes, C.R.; Heckel, G.; da Luz Mathias, M.; Bastos-Silveira, C. Variation and Selection in the Putative Sperm-Binding Region of ZP3 in Muroid Rodents: A Comparison between Cricetids and Murines. *Genes* **2021**, *12*, 1450. <https://doi.org/10.3390/genes12091450>

Academic Editor: Miguel Arenas

Received: 19 July 2021

Accepted: 16 September 2021

Published: 20 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: In mammals, the zona pellucida glycoprotein 3 (ZP3) is considered a primary sperm receptor of the oocyte and is hypothesized to be involved in reproductive isolation. We investigated patterns of diversity and selection in the putative sperm-binding region (pSBR) of mouse ZP3 across Cricetidae and Murinae, two hyperdiverse taxonomic groups within muroid rodents. In murines, the pSBR is fairly conserved, in particular the serine-rich stretch containing the glycosylation sites proposed as essential for sperm binding. In contrast, cricetid amino acid sequences of the pSBR were much more variable and the serine-rich motif, typical of murines, was generally substantially modified. Overall, our results suggest a general lack of species specificity of the pSBR across the two muroid families. We document statistical evidence of positive selection acting on exons 6 and 7 of ZP3 and identified several amino acid sites that are likely targets of selection, with most positively selected sites falling within or adjacent to the pSBR.

Keywords: zona pellucida glycoprotein 3; sperm receptor; female fertilization protein; positive selection; Cricetidae; Murinae

1. Introduction

Gamete surface proteins can play an important role in reproductive isolation. They maintain species-specific barriers to fertilization and thus contribute to post-mating prezygotic isolation, and potentially to speciation [1–3]. In mammals, zona pellucida and sperm-head interacting proteins have co-evolved rapidly, presumably as a result of natural and sexual selection, leading to species-specific fertilization and genetic isolation [1,4,5]. This intersexual co-evolution is necessary to maintain gametic interaction and has led to amino acid differences between diverging populations [1,6]. Subsequently, gametic incompatibility may arise, promoting the differentiation of genomes and possibly ultimately the formation of new species.

One of the most studied reproductive proteins in mammals, both functionally and evolutionarily, is the zona pellucida glycoprotein 3 (ZP3), the sperm receptor of the oocyte and inducer of the acrosome reaction [7,8]. It consists of a polypeptide chain glycosylated with serine/threonine (O)-linked and asparagine (N)-linked oligosaccharides. ZP3 is a primary receptor during fertilization [9] because it binds directly to sperm, through its glycan chains, and inhibits further binding of sperm to the oocyte [10,11]. The putative sperm-binding region (pSBR), located in exon 7, exhibits considerable amino acid variation between species, which, together with modifications in the structure of the O-linked glycans, may enable a species-specific binding of sperm to the oocyte [7,8,12,13].

In house mice (*Mus musculus*), the best-studied system, sperm-oocyte interactions have been associated in particular to a serine (S) rich region (329–334), including the glycosylation sites S-332 and S-334, within the pSBR [14–17]. The classical model of sperm-oocyte binding proposes that gametic interactions occur via O-linked glycans attached to S-332 and S-334, and that after fertilization these residues are deglycosylated thereby preventing further sperm adhesion [14,17]. Studies using genetically modified mouse models have, however, challenged this classical view of sperm-oocyte binding and proposed alternative scenarios [18,19]. For example, it has been suggested that conserved O-linked glycosylation sites outside exon 7 and the pSBR are also exposed on the same 3D protein surface and constitute additional binding sites that may be involved in sperm-oocyte recognition [20,21], and/or sperm binding specificity may be based on the three-dimensional supramolecular structure of the zona pellucida, a matrix composed of ZP3 plus two additional proteins, ZP1 and ZP2 [22–25]. In fact, strong evidence has accumulated implicating ZP2, through a specific domain near its N-terminus, as a primary sperm receptor in mice [26,27]. It has very recently been suggested that the sperm-binding region may lie at the interface between the ZP2 and ZP3 subunits [28].

Although the molecular basis of sperm–oocyte binding remains incompletely understood, despite decades of investigation, and the exact role of the pSBR of ZP3 remains uncertain, it is clear that this glycoprotein, together with other zona pellucida and sperm head ligands, mediates sperm-oocyte binding, regardless of its specific molecular mechanism of action [2,29]. Moreover, species-specificity seems to be ensured both by the presence of a certain sperm receptor signature and by a particular glycosylation pattern of the glycoproteins of the zona pellucida, particularly ZP3 and ZP2 [19].

Several studies on the evolution of mammalian reproductive proteins have mainly consisted of comparing distantly related species, e.g., [30–32], but many more have focused on shorter evolutionary timescales, since fertilization mechanisms within species and among closely related taxa are more relevant to relate amino acid changes and reproductive isolation, e.g., [2,29,33–38]. This approach was tested in Cetartiodactyla, particularly in wild cattle [38] and cetaceans [37]. Both studies did not detect signatures of positive selection on ZP3 or evidence of its contribution to species specificity of sperm binding and prevention of cross-species fertilization. Data from rodent species, however, are contradictory. Turner and Hoekstra [34] documented positive selection acting on the pSBR of ZP3 in several deer mice (*Peromyscus*) species (Cricetidae, Neotominae), suggesting adaptive divergence within the genus. Analyses on Australian murine rodents (Muridae, Murinae) performed by Swann and colleagues [29] did not reach the same conclusions.

Muroid rodents (Rodentia, Muroidea) are by far the largest extant mammalian superfamily, containing nearly one-third of all mammal species. In this study, we expanded the investigation of evolutionary patterns in the pSBR of ZP3 in its two most diverse families, Cricetidae and Muridae, by performing a comparative analysis of 93 species. Special focus is given to the speciose genus *Microtus* (meadow voles) (Cricetidae, Arvicolinae), an evolutionarily young group that started to radiate 1.2–2 million years ago [39]. It has given rise to 65 extant species [40], many of which are undergoing further diversification, e.g., [41–44].

2. Materials and Methods

2.1. Samples, DNA Extraction, Amplification and Sequencing

We examined 93 species of Cricetidae (N = 50) and Muridae (N = 43). Cricetid samples comprised 25 Arvicolinae (20 *Microtus* spp.), 17 Neotominae, four Cricetinae, two Tylomyinae and two Sigmodontinae species. All analyzed murid species were from the Murinae subfamily (Table S1). Tissue samples were provided by natural history museums and university research institutes (Table S1). Genomic DNA was extracted using standard protocols, with tissue digestion in a buffer containing sodium dodecyl sulfate (SDS) and proteinase K, followed by phenol-chloroform DNA extraction [45].

Exon 6, intron 6, and exon 7 of the *ZP3* gene were amplified using newly designed primers M-ZP3-F2 (5'-ATCACCTGTCATCTCAAAGTCA-3') and M-ZP3-R1 (5'-CATGCCTGCGGTTTCTAGAAGC-3'). All polymerase chain reactions (PCR) contained 100 ng of genomic DNA, 0.3 mM of each primer, 1.25 U of GoTaq Flexi DNA Polymerase (Promega, Madison, WI, USA), 1x PCR buffer (Promega), 2.5 mM MgCl₂, 0.1 µg of bovine serum albumin (BSA; New England Biolabs, Ipswich, United Kingdom), and 0.2 mM of each dNTP (Thermo Scientific, Waltham, MA, USA), and water up to a final volume of 25 µL. PCR amplifications were performed in a MyCycler thermal cycler (Bio-Rad Laboratories Inc., Hercules, CA, USA) and consisted of denaturation at 95 °C for 5 min, followed by 35 cycles of denaturation at 94 °C for 1 min, annealing at 58 °C for 1 min and extension at 72 °C for 1 min, and a final extension step at 72 °C for 10 min. The size of the PCR products was verified by electrophoresis in 1% agarose gels and comparison with GeneRuler™ 100 bp Plus DNA Ladder (Fermentas, Waltham, MA, USA). PCR products were purified with ExoI/FastAP (Fermentas). Sequencing in both directions, with the same primers used for the PCR reactions, was carried out by Macrogen Inc. (South Korea and the Netherlands) using an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Waltham, MA, USA).

Sequences were submitted to GenBank (accession numbers MT226280-MT226326; see Table S1 for details).

2.2. Sequence Analyses

Sequences were aligned using Sequencher 4.8 (Gene Codes Corporation) and BioEdit 7.2.5 [46]. We supplemented our sequence dataset with GenBank *ZP3* sequences of Arvicolinae, Neotominae, and Murinae taxa (Table S1). We included a species representing each of the murine genera analyzed.

In subsequent analyses, we focused on the coding regions of exons 6 and 7 because of their potential importance in *ZP3* for the species specificity of sperm binding. Sequences were collapsed into unphased genotypes using the DNACollapser tool in FaBox 1.5 [47]. Heterozygous positions in the larger intraspecific datasets (*Microtus lusitanicus* and *Microtus duodecimcostatus*) were phased using Phase 2.1.1 [48,49] as implemented in DNAsp 5.10.1 [50]. Five independent runs were conducted using default values, and after checking for concordance a final run with 10 times more iterations (1000 iterations and 1000 burn-in) was performed. Heterozygous positions of smaller intraspecific datasets were phased manually. DNA polymorphism parameters, such as the number of variable sites, number of parsimony-informative sites, number of non-synonymous sites, nucleotide diversity (π), and GC content were calculated in DnaSP. The translation of DNA sequences into amino acid sequences was performed with BioEdit. Amino acid sequence conservation and variation were visualized using the WebLogo application [51,52] via the SIB ExPASy Bioinformatics Resource Portal [53].

JModelTest 2.1.7 [54] was used to select the best-fitting model of nucleotide substitution (TVM+G, [55]) based on the Akaike information criterion (AIC) [56]. There were several species with gaps in the alignment of exon 7 sequences, and we wanted to include these indels in the phylogenetic analyses. Bayesian inference with MrBayes 3.1.2 [57,58] allows the incorporation of gaps coded as binary characters in a separate partition with a phylogenetic mixed model. Binary matrices were constructed with SeqState 1.4.1 [59], using two types of gap-coding: the simple indel coding (SIC, [60]) and modified complex

indel coding (MCIC, [61]). Each Bayesian analysis consisted of two parallel Markov Chain Monte Carlo (MCMC) runs with four chains, one cold and three heated, for four million generations, with every 100th generation sampled. We determined convergence between the two runs when the average standard deviation of split frequencies was <0.01 [57]. The first 25% of trees were discarded as burn-in, and the remaining trees were used to construct a consensus tree and estimate Bayesian posterior probabilities. The consensus tree obtained was drawn using FigTree 1.3.1 [62]).

Since recombination may confound selection analyses [63–65], we tested for its presence using a set of methods implemented in RDP 4 [66]: RDP [67], BOOTSCAN [68,69], GENECONV [70], MAXCHI [71,72], CHIMAERA [72], SISCAN [73], and 3SEQ [74].

We tested for positive selection using the CodeML subroutine of PAML 4.8 [75,76]. Maximum likelihood estimates of ω (nonsynonymous (dN)/synonymous (dS) substitution ratio) across codons were inferred under seven models of variable ω among sites: M0 (one ω), M1a (nearly neutral, one ω , two classes of sites), M2a (positive selection, three classes of sites), M3 (discrete, three classes of sites); M7 (nearly neutral with β distribution approximating ω variation, 10 classes of sites), M8 (positive selection with β distribution approximating ω variation, 11 classes of sites) and M8a (ω distribution follows a mixture between a β distribution and a point mass at $\omega = 1$, 11 classes of sites) [77–83]. The ω ratio is a sensitive measure of selective pressure, with positive selection inferred when $\omega > 1$ [78,79].

Additionally, we used branch-site models that allow ω variation among amino acids in the protein and across branches on the phylogenetic tree in order to detect possible (episodic) positive selection affecting a few sites along particular lineages (foreground branches) [83–86]. In our case, this approach may allow us to detect positive selection affecting only a few amino acid residues in the analyzed fragment of ZP3 in specific lineages of the studied Muroidea. In fact, this strategy can be statistically more powerful than site-based tests, which average over all of the phylogeny [84]. The null (model = 2; NSsites = 2; $\omega = 1$) and neutral M1a (model = 0; NSsites = 1; $\omega = 1$) models were compared to the MA1 (model = 2; NSsites = 2; ω estimated), the alternative model in the branch-site test of positive selection [83,85]. The first comparison is a direct test for positive selection on the foreground lineages and therefore has been designated as the ‘branch-site test of positive selection’ [85], whereas the second test is also sensitive to relaxed purifying selection on the foreground branches [83,85]. Likelihood ratio tests (LRTs) of M0 vs. M3, M1a vs. M2a, M7 vs. M8, M8a vs. M8, null model vs. MA1 and M1a vs. MA1 were performed in order to search for evidence of positive selection [78,80,87]. Twice the log-likelihood difference between models ($2\Delta l$) was compared with a chi-square distribution with the number of degrees of freedom (dF) equal to the difference in the number of estimated parameters between the two models [80]. Positively selected sites under M2a, M3, M8, and MA1 were identified using the Naive Empirical Bayes (NEB) and the Bayes Empirical Bayes (BEB) approaches [83].

The M7-M8 test is the most powerful of the site models LRTs in PAML [82,88], but can also be biased towards false inference of adaptive evolution [38,88]. To further reduce the chances of falsely identifying sites as positively selected, we searched for signatures of positive selection using tests available in Datamonkey 2.0 [89–91], a web interface for the HyPhy package [92]. The tests carried out included individual site models that, unlike those available in CodeML, can incorporate synonymous substitution rate variation: SLAC (single likelihood ancestor counting, [93]), FEL (fixed effects likelihood, [93]), MEME (mixed effects model of evolution, [94]), and FUBAR (fast unconstrained Bayesian approximation, [95]). The other tests performed in Datamonkey were aBSREL (adaptive branch-site random effects likelihood, [96,97], an individual branch model that is an improved version of the branch-site models, and BUSTED (branch-site unrestricted statistical test for episodic diversification, [98]), a gene-wide test of episodic positive selection. All tests were performed with a significance threshold of 0.05.

3. Results

3.1. Genetic Variation and Phylogeny

This study generated new sequences (N = 103, corresponding to 47 new haplotypes for exon 6 and 7 with GenBank accession numbers MT226280–MT226326) for 32 cricetid species. After the addition of previously published sequences of 18 cricetid and 43 murid species (Table S1), analysis of the resulting alignment revealed extensive length and sequence variation in ZP3, not only in intron 6 but also in exons 6 and 7, including the pSBR (Figures 1 and 2 and Figure S1). The final data matrix containing only the coding regions was 228 base pairs (bp) long, corresponding to positions 835–1063 in the reference mouse ZP3 gene. We did not observe length variation between the two alleles of an individual, and no evidence of recombination was found in the dataset by any of the detection methods employed. Twenty DNA sequences, four from GenBank and 16 newly produced herein, had heterozygous positions (16 at one position; three at two positions; one at three positions). The phased dataset contained a total of 78 variable sites, of which 63 were parsimony informative, and the GC content was 54.3%. The polymorphisms defined a total of 111 haplotypes, 40 in the murids and 71 in the cricetids (among the latter, 35 in the Arvicolinae, 23 in the Neotominae, four in the Cricetinae, six in the Tylomyinae, and three in the Sigmodontinae). No haplotypes were shared between families or subfamilies, but there was haplotype sharing among species of the same subfamily (Table S1). In total, there were nine haplotypes (12.7% of the cricetid haplotypes) shared between 14 Cricetidae species, almost all of them congeneric (*Microtus*), and two haplotypes (5% of the murine haplotypes) shared between five Murinae species.

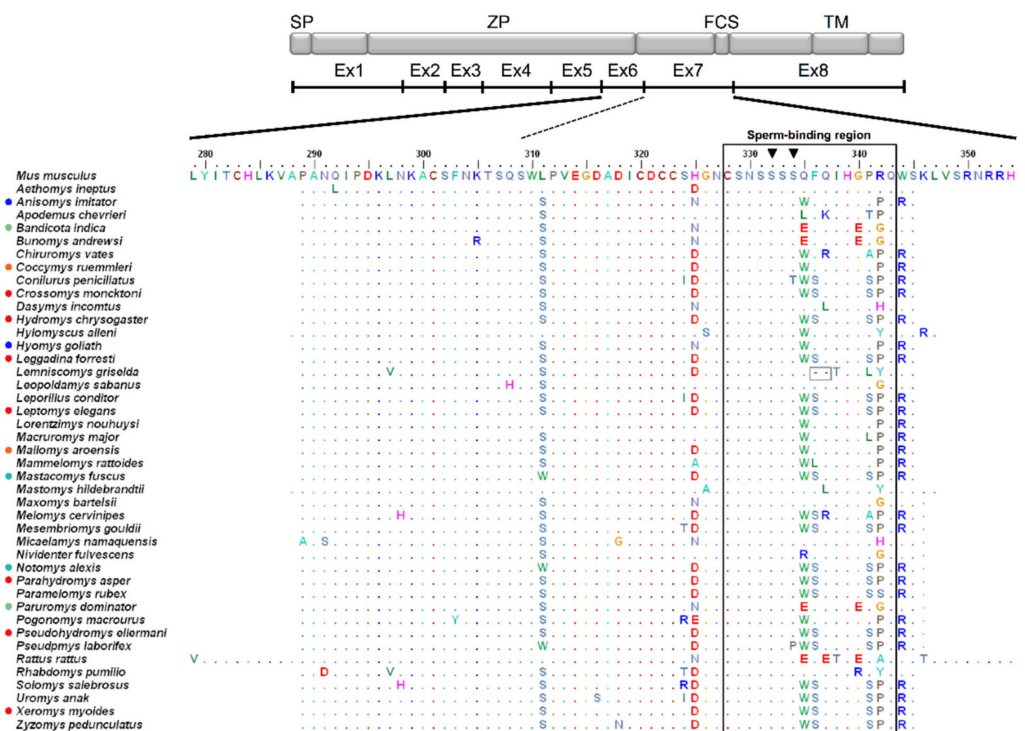


Figure 1. ZP3 (zona pellucida glycoprotein 3) exons 6 and 7 amino acid sequence alignment of Murinae, with a schematic representation of the mouse protein and respective functional domains. Dots represent amino acids identical to the reference *Mus musculus* sequence and colored circles before species names denote shared haplotypes. The black outlined rectangle delimits the putative sperm-binding region according to [15]. Grey outlined squares highlight deletions relative to *M. musculus*. Black inverted triangles indicate glycosylation sites S-332 and S-334. SP = signal peptide, ZP = zona domain, FCS = furin cleavage site, TM = transmembrane domain. Ex1–Ex8: exons 1 to 8.

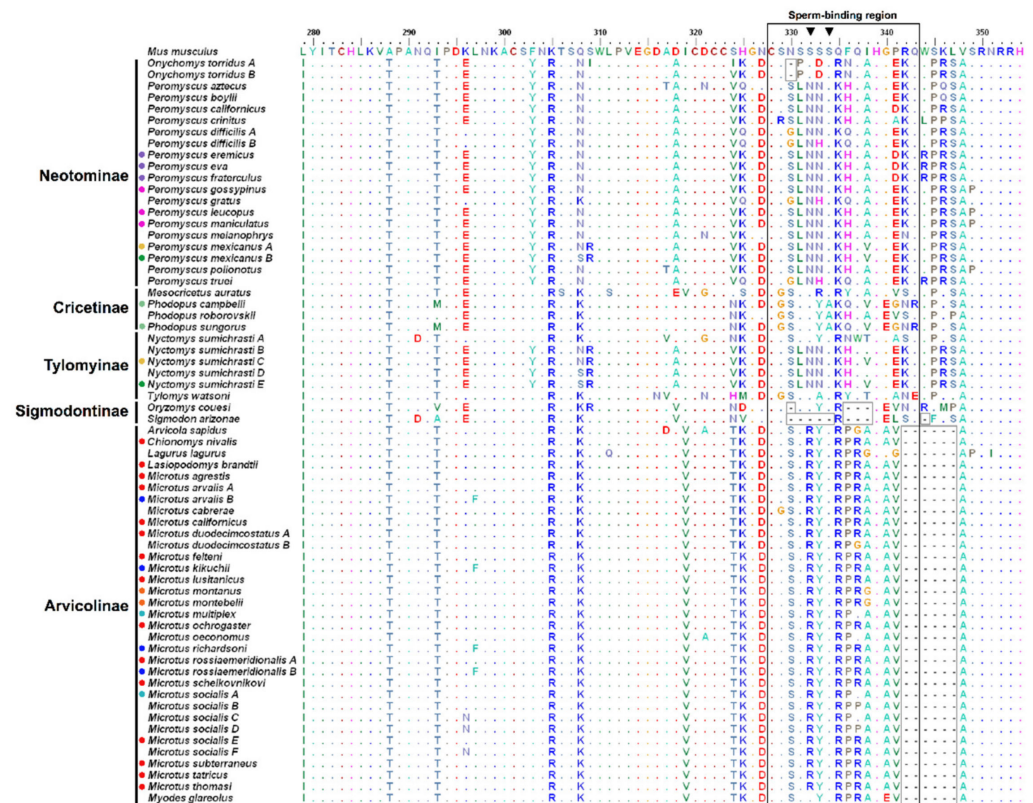


Figure 2. ZP3 exons 6 and 7 amino acid sequence alignment of Cricetidae from the five extant subfamilies. Dots represent amino acids identical to the reference *Mus musculus* sequence and colored circles before species names denote shared haplotypes. The black outlined rectangle delimits the putative sperm-binding region according to [15]. Grey outlined squares highlight deletions relative to *M. musculus*. Black inverted triangles indicate glycosylation sites S-332 and S-334.

No topological differences were observed between the trees from the two replicate MrBayes analyses for each of the three matrices derived from the alignment, excluding gaps or including them coded using SIC or MCIC (here we only present the phylogenetic tree derived using the SIC method, Figure S2). The only differences observed concern branch lengths of some lineages, which are explainable by the different treatment of indels by the three approaches used.

The phylogenetic reconstruction (Figure S2) grouped all species according to their family but yielded a topology within Cricetidae that is not congruent with the phylogeny of its subfamilies [99–101]. Indeed, while the tree obtained here showed essentially an unresolved polytomic relationship between the cricetid subfamilies (Figure S2), the established phylogeny supports two major clades: Arvicolinae + Cricetinae and Neotominae + Sigmodontinae + Tylomyinae [99–101]. The trees obtained from the matrices with gaps either excluded or coded using MCIC also had high support for most nodes (data not shown). The Arvicolinae and Sigmodontinae subfamilies were monophyletic, whereas the Cricetinae and Tylomyinae were not. Neotominae was also monophyletic, but this clade included as well haplotypes found in Tylomyinae taxa (Figure S2). The haplotype of the cricetine *Mesocricetus auratus* did not cluster with any subfamily and the haplotypes of the tylomyines *Tylomys watsoni* and *Nyctomys sumichrasti* grouped with the family Neotominae.

3.2. Amino Acid Variation

The translation of the DNA sequence of exons 6 and 7 yielded 74 amino acids (positions 279–354 according to the reference mouse ZP3 protein; [15]). Forty-five variable amino acid sites (60.8%) and 14 indel positions defined a total of 72 amino acid sequence types (Figures 1 and 2, and Figure S1). Considerable length variation due to amino acid deletions,

mainly in the pSBR (Figure 2, positions 328–343), was observed particularly in Arvicolinae and Sigmodontinae relative to murines (Figure 1). Compared to mouse ZP3, all arvicoline species lacked six amino acids at positions 342–347, and the two studied sigmodontines had amino acid deletions at positions 330 (also present in the neotomine *Onychomys torridus*) and 336–338 (Figure 2). Additional amino acid deletions were detected in *Sigmodon arizonae* (positions 331–334 and 344) (Figure 2). Therefore, the multiple amino acid deletions in the sigmodontines concern the serine-rich region at positions 329–334 and its immediate vicinity, whereas the six amino acid deletion in the arvicolines only involves the last two residues in the pSBR (Figure 2). In contrast to the subfamilies Arvicolinae, Neotominae, and Sigmodontinae, no deletion of amino acids relative to mouse ZP3 was found in Cricetinae and Tylomyinae (Figure 2). In turn, in the examined Murinae, amino acid deletions within the pSBR were only detected in *Lemniscomys griselda* (positions 336–337, Figure 1).

There were amino acid haplotypes shared between species of the same genus and even between genera of the same subfamily (Figures 1 and 2 and Table S1). In cricetid genera represented by multiple species, 20 species of *Microtus* (Arvicolinae) had 11 amino acid haplotypes and 16 species of *Peromyscus* (Neotominae) showed 14 amino acid haplotypes (Figure 2 and Table S1). There were also cases of intraspecific polymorphism, with the presence of more than one amino acid haplotype, in arvicolines (Figure 2 and Table S1).

Considering only the sequences for the pSBR resulted in a decrease in the total number of amino acid haplotypes to 45, 24 for cricetids, and 21 for murines. Within the subfamilies of Cricetidae, there were eight haplotypes in arvicolines (four of them shared among different species), eight in neotomines (three shared between species), three in cricetines (one shared), three in tylomyines, and two in sigmodontines. There was only one case of shared haplotypes between cricetid species of different subfamilies, that between *P. mexicanus* (Neotominae) and *N. sumichrasti* (Tylomyinae). Among the haplotypes found in murines, five were shared by more than one species. In *Microtus* and *Peromyscus*, respectively, there were six (three shared between species) and seven (three shared between species) pSBR amino acid haplotypes.

In the data set for *Microtus*, the three pairs of well-accepted sister species, *M. duodecimcostatus*-*M. lusitanicus*, *M. felteni*-*M. thomasi* and *M. arvalis*-*M. rossiaemeridionalis* [102,103], all have areas of sympatry and share the same respective pSBR amino acid haplotype. In the *Peromyscus* data set, the sister species pairs consistently supported in the literature, *P. gossypinus*-*P. leucopus* [104–106] and *P. gratus*-*P. truei* [35,105,107], have both also areas of sympatry and also share the same respective pSBR amino acid haplotype. Finally, the cricetines *Phodopus campbelli* and *Phodopus sungorus* are sister species [108] with an area of sympatry, and also shared the same pSBR amino acid haplotype.

The greatest variability in the analyzed ZP3 fragment occurred in the pSBR, in which only sites 328 and 339 were invariant in all species of cricetids and murines studied here (Figures 1 and 2 and Figure S3), and adjacent amino acids (Figures 1 and 2). Notably, almost all murine species examined, with the exception of *Conilurus penicillatus* and *Pseudomys laborifex* show conservation of the characteristic serine-asparagine-serine-serine-serine sequence at positions 329–334 (SNSSSS) (Figure 1), whereas in no cricetid species this sequence is present and there is variability within each subfamily (Figure 2).

3.3. Selection Tests

The selection tests indicated that the analyzed ZP3 sequences are under variable selective pressure among sites (Table 1 and Table S2). PAML LRTs rejected the null hypothesis site models M0, M1a, M7, and M8a in favor of the alternative M3, M2a, and M8 ($p < 0.001$ or $p < 0.05$) in tests with the full data set (Cricetidae + Murinae), Murinae only, and in *Microtus*, but only rejected the null hypothesis site model M0 in the analyses in Cricetidae and in *Peromyscus* (Table 1). The fact that for these last two data sets only the alternative model M3 was supported indicates that it was possible to detect variable selective pressure among sites but not positive selection [30,109]. For the different data sets, ω values < 1 in the supported site models indicate that most codons are under purifying selection (Table S2).

For example, for the Cricetidae + Murinae data set the models M2a and M8 respectively estimated 65% and 88% of sites with $\omega < 1$ and 9% and 12% of sites with $\omega > 1$.

Table 1. Results of the likelihood ratio tests (LRT) considering site- and branch-site models implemented by PAML on exon 6 and 7 of the *ZP3* gene of the analyzed.

Type	LRT	2Δl	d.f.	p Value
Site-models: Cricetidae + Murinae	M0 vs. M3	273.646	4	<0.001
	M1a vs. M2a	37.492	2	<0.001
	M7 vs. M8	42.397	2	<0.001
	M8a vs. M8	35.522	1	<0.001
Site-models: Cricetidae	M0 vs. M3	139.759	4	<0.001
	M1a vs. M2a	0.726	2	0.696
	M7 vs. M8	4.4671	2	0.107
	M8a vs. M8	3.490	1	0.062
Site-models: Murinae	M0 vs. M3	102.930	4	<0.001
	M1a vs. M2a	16.674	2	<0.001
	M7 vs. M8	21.164	2	<0.001
	M8a vs. M8	18.286	1	<0.001
Site-models: <i>Microtus</i>	M0 vs. M3	60.304	4	<0.001
	M1a vs. M2a	8.622	2	0.013
	M7 vs. M8	8.350	2	0.015
	M8a vs. M8	8.232	1	0.004
Site-models: <i>Peromyscus</i>	M0 vs. M3	15.034	4	0.004
	M1a vs. M2a	0.000	2	1.000
	M7 vs. M8	0.025	2	0.988
	M8a vs. M8	0.022	1	0.883
Branch-site models: Murinae	null vs. MA1	0.796	2	<0.001
	M1a vs. MA1	1.420	2	<0.001
Branch-site models: Cricetidae	null vs. MA1	1.077	2	<0.001
	M1a vs. MA1	12.208	2	<0.001
Branch-site models: Arvicolinae	null vs. MA1	6.280	2	<0.001
	M1a vs. MA1	6.560	2	<0.001
Branch-site models: Cricetinae	null vs. MA1	0.000	2	1.000
	M1a vs. MA1	0.927	2	0.629
Branch-site models: Neotominae	null vs. MA1	0.000	2	1.000
	M1a vs. MA1	0.000	2	1.000
Branch-site models: Sigmodontinae	null vs. MA1	0.000	2	1.000
	M1a vs. MA1	0.004	2	0.998
Branch-site models: Tylomyinae	null vs. MA1	0.087	2	<0.001
	M1a vs. MA1	0.242	2	0.242

Using the full data set (Cricetidae + Murinae), all selection methods applied identified both positively and negatively selected sites distributed throughout exons 6 and 7 of *ZP3* (Figure 3 and Table S2). With regard to the sites inferred to be under positive selection in the pSBR, site 337 was identified by all methods, site 336 was detected in all tests except SLAC,

and sites 341 and 342 were indicated by all PAML site models (Figure 3 and Table S2). Yet another site in the pBSR, 335, was inferred as positively selected by all PAML site models on the Murinae data set (Figure 3 and Table S2). Outside the pBSR but still in its immediate vicinity, amino acids 311, 325, and 346 were determined to be under positive selection by all PAML site models in analyses of both the full data set and the murine dataset (Figure 3 and Table S2).

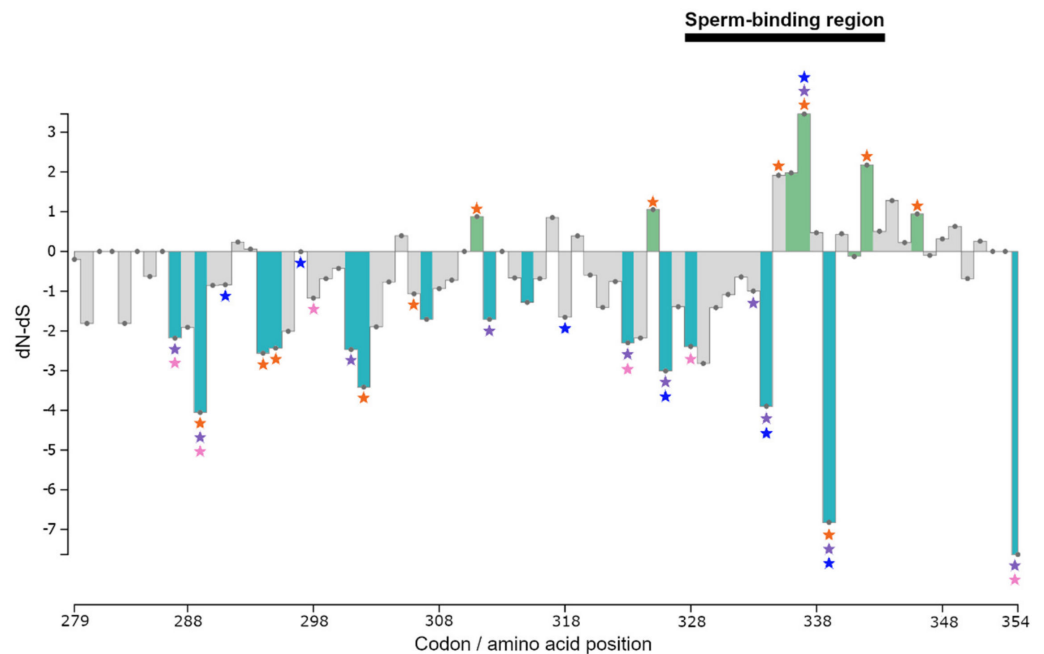


Figure 3. Distribution of amino acid sites under selection in exons 6 and 7 of *ZP3* as identified by PAML site models M2a, M3, and M8 and by HyPhy site tests SLAC, FEL, FUBAR, and MEME ($p < 0.05$). For the Cricetidae + Murinae data set, dN-dS columns corresponding to sites indicated as possibly being under positive selection by either all PAML models or all HyPhy tests are denoted in green, while columns of negatively selected sites in all HyPhy tests are shown in light blue. Grey columns correspond to sites that were not inferred to be under either positive or negative selection in all PAML and/or HyPhy tests. Coloured stars indicate sites selected only in particular data sets: purple = Cricetidae; orange = Murinae; blue = *Microtus*; and pink = *Peromyscus*. The normalized dN-dS per codon was calculated by SLAC.

While in *Microtus*, in addition to site 337, residue 297 was indicated to be under positive selection by all PAML site models, in *Peromyscus* no consistent evidence of positive selection was found (Table S2). Overall, across analyses and data sets, most sites identified as positively selected fall within or adjacent to the pBSR (Figure 3). Fifteen sites were identified to be under purifying selection at p-value threshold 0.05 by HyPhy tests FEL, FUBAR, and SLAC on the Cricetidae + Murinae data set (Table S2). These included the serine-rich site 334 and the two invariant sites in the pBSR (C-328 and H-339) (Figure 3).

The PAML branch-site comparisons of the null model vs. MA1 and M1a vs. MA1 revealed variable selective pressure, depending on the family/subfamily set as the foreground branch (Table 1 and Table S2). The null hypothesis of no positive selection was rejected ($p < 0.001$) for both Murinae and Cricetidae, and for two cricetid subfamilies, the Arvicolinae and Tylomyinae. The MA1 model identified several sites, all outside the pBSR, in Murinae (315 and 322), Cricetidae (287, 307 and 311), Arvicolinae (287), and Tylomyinae (307 and 311) (Table S2) as positively selected. With the exception of site 311, all others were invariant across the entire data set and, thus, likely false positives [93,110].

Sampling and stochastic errors, model misspecification, and assumption violations, and testing of multiple foreground lineages, can lead to false positives [83,85,86,97,111–113]. For instance, it has been noted that the branch-site models in PAML may be sensitive

to small sequences [29]. As noted by Zhang and colleagues [85], identifying sites under positive selection, especially in the case of episodic selection, is intrinsically more difficult than testing whether such sites exist, but it is still useful to be able to detect positive selection acting on a region. Importantly, site 311 was also suggested to be influenced by positive selection in all PAML site model tests with the full data set (Table S2). The HyPhy tests aBSREL and BUSTED found no evidence of episodic diversifying selection across the ZP3 phylogeny.

4. Discussion

We investigated patterns of variation and natural selection at the pSBR and adjacent exonic sequences of the reproductive protein ZP3 in species of Cricetidae and Murinae. We found that murine pSBR is fairly conserved, in particular the serine-rich stretch containing the glycosylation sites, which has been proposed as essential for sperm binding, being invariant in 41 of the 43 murine genera examined. In contrast, the amino acid sequence of the pSBR was much more variable in cricetids, and the serine-rich motif typical of murines was generally substantially modified, implying that in Cricetidae the ZP3-mediated sperm binding does not follow the classical model proposed for the mouse.

The virtual lack of sequence variation in the serine-rich region at positions 329–334 and the relatively high conservation of the entire pSBR in the murine species studied (Figure 1 and Figure S3) suggests a functional conservation of this region in murines. According to [29,36], the reduced intergeneric variability in the murine pSBR seems to indicate a limited role for this region in species-specific sperm-ZP binding, a scenario supported by our results as well.

In our study, of the six sites (311, 325, 335, 337, 342, and 346) that were consistently detected in different selection tests as potentially under positive selection in Murinae, sites 325 [29,30] and 311, 337 and 342 [30] have also been constantly strongly supported as positively selected in previous analyses in murines. Two of these sites (337 and 342) were also identified in a study of closely related species of the mouse genus *Mus* [33]. However, when this analysis was repeated without including outgroup sequences, no evidence for positive selection was found [34]. This case illustrates the advantages of the approach followed in our study. In addition to using multiple different positive selection tests, as recommended by many authors to increase the robustness of the results, e.g., [80,88,109,114], we conducted analyses at various phylogenetic levels using different hierarchical subsets of the full data set.

With regard to Cricetidae, the pSBR and even the serine-rich region were remarkably variable (Figure 2, Figures S1 and S3). In fact, the serine-rich motif SNSSSS typical of murines was generally substantially modified in all cricetid subfamilies (Figure 2 and Figure S3). In particular, most cricetids do not have a serine at position 332, but on the other hand, they have a serine at position 330 (Figure 2 and Figure S3), which is apparently fixed for asparagine in Murinae (Figure 1).

Our survey of the pSBR of ZP3 across all the subfamilies of the Cricetidae indicates that in this family the ZP3-mediated sperm-oocyte binding does not follow the classical model proposed for the mouse. In fact, not even all of the murines examined were conserved for serine at position 334 (Figure 1). According to the classical model suggested for the mouse, this would affect gamete interaction since S-332 and S-334 are hypothesized to carry O-linked glycans that are essential for sperm-oocyte binding. Hence, even within murines, there may be alternative mechanisms and other anchor amino acids crucial for gamete recognition.

The role attributed to ZP3 with regard to a species-specific function in gamete recognition may involve regions of the protein other than those encoded by exons 6 and 7. It has been proposed that sperm binds to ZP3 by interacting with O-linked glycans not linked to S-332 and S-334 [115] or with N-linked glycans and accessible protein regions located within the C-terminal domain of ZP3 [116]. In particular, it has been suggested that two conserved O-linked glycosylation sites (residues T-155 and T-162/S-164/S-165) shared by

mouse and human ZP3, and which are exposed on the same 3D protein surface as the pSBR in exon 7, may be the actual attachment sites of the sperm-binding glycans [20,21]. Moreover, more recently, significant experimental evidence has indicated that the sperm-binding region of the zona pellucida may reside in ZP2 [26,27], and lately, it has even been proposed that it might lie at the interface between the ZP2 and ZP3 subunits [28]. Therefore, uncertainty and debate on this issue remain high [19,117].

The most extreme cases of divergence in the serine-rich region, due to multiple amino acid deletions, were detected in arvicolines and sigmodontines (Figure 2). Given the established phylogeny for the cricetid subfamilies [99–101], supporting the two clades Arvicolinae + Cricetinae and Neotominae + Sigmodontinae + Tylomyinae, the observed deletions do not seem to be associated with the evolutionary relationships among subfamilies, and the few deletions shared between different subfamilies appear to be clear cases of convergence (Figure 4).

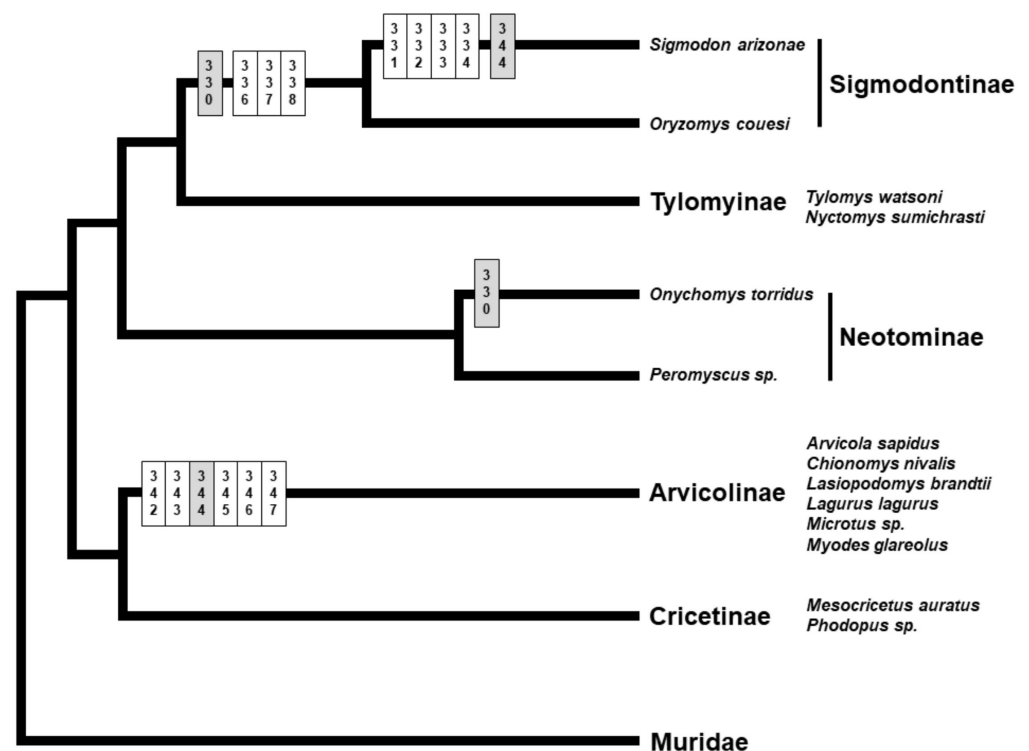


Figure 4. Schematic representation of the phylogenetic tree for the cricetid subfamilies (based on [99–101]), with the amino acid deletions observed in the putative sperm-binding region and its immediate vicinity in the ZP3 of different lineages indicated on the respective branches of the tree. Position numbers are according to the mouse reference sequence for ZP3. Sites in white are amino acid deletions unique to a lineage, while sites in grey are amino acid deletions shared between cricetid subfamilies.

We observed intraspecific amino acid variation in several species of *Microtus* and other cricetid genera (Table S1), similar to previous reports for *Peromyscus* [34,35]. Moreover, the inter-specific and -generic sharing of amino acid haplotypes may not only represent shared ancestral polymorphism but may also be due to convergence as a by-product of (balancing) selection maintaining divergent alleles within species [2,35].

Regarding a possible general involvement of the pSBR in the species specificity of fertilization in cricetids, this is seemingly contradicted by the extensive haplotype sharing between congeners and even among different genera (Figure 2). The fact that sister, or closely related species with sympatric areas of distribution, share pSBR amino acid haplotypes has been considered evidence of lack of selection for pSBR divergence to prevent or reduce hybridization [29,30,35]. However, this conclusion is only valid, and

relevant, if the analyzed sister or closely related species were not sampled from allopatric populations. In this study, data from *Peromyscus* sister taxa *P. gossypinus* and *P. leucopus* [34] and *P. gratus* and *P. truei* [35] were, in both cases, from allopatric populations. Regarding *Microtus*, in the three pairs of sister species sharing haplotypes (*M. duodecimcostatus*-*M. lusitanicus*, *M. felteni*-*M. thomasi*, and *M. arvalis*-*M. rossiaemeridionalis*), only for the first pair the data were from sympatric locations. Genetic data from *M. duodecimcostatus* and *M. lusitanicus* revealed historical introgression of mitochondrial DNA, but low gene flow given the clear differences in nuclear DNA at the sympatry zone [42]. The factors maintaining the genetic integrity of the two taxa in the sympatric region remain unknown. However, this study and recent data from crossbreeding of sympatric individuals [118] indicate the development of prezygotic and postzygotic barriers but not gametic isolation.

5. Conclusions

In conclusion, the results of the present study indicate a general lack of species specificity of the pSBR in muroid rodents. However, our data are consistent with hypotheses and models that describe multiple distinct binding sites in sperm-oocyte recognition. Thus, we suggest that future studies should focus on the complete ZP3 and ZP2 proteins, and looking for signatures of coevolution, also with sperm head proteins, in order to compare and evaluate different proposed sperm-binding regions. Moreover, to clarify their potential role in the development and/or maintenance of reproductive isolation and speciation, efforts should be made to investigate closely related species in areas of sympatric distribution.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/genes12091450/s1>, Figure S1: ZP3 exons 6 and 7 amino acid sequence alignment of Cricetidae, with a schematic representation of the mouse protein and respective functional domains, Figure S2: Bayesian inference phylogenetic tree obtained for exons 6 and 7 of ZP3 of the studied muroid rodents, with indels coded using the SIC method, Figure S3: Amino acid sequence logos showing patterns of conservation and variation in the putative sperm-binding region of ZP3, for different data sets, Table S1: List of species examined in this study, Table S2: Detailed results of selection tests.

Author Contributions: Conceptualization, M.A.D., C.R.F., G.H., and C.B.-S.; validation, M.A.D.; formal analysis, M.A.D.; investigation, M.A.D.; resources, G.H., M.d.L.M., and C.B.-S.; data curation, M.A.D., and C.B.-S.; writing—original draft preparation, M.A.D., G.H., and C.B.-S.; writing—review and editing, C.R.F., M.d.L.M., and C.B.-S.; visualization, M.A.D.; supervision, C.B.-S.; project administration, C.B.-S.; funding acquisition, G.H., M.d.L.M., and C.B.-S. All authors have read and agreed to the published version of the manuscript.

Funding: M. A. Duarte was supported by a Ph.D. grant SFRH/BD/70646/2010 and C. B. Silveira by research grant SFRH/BI/128387/2017, both from the Portuguese Foundation for Science and Technology (FCT) and G. Heckel by grants 31003A-149585 and 176209 from the Swiss National Science Foundation. C. Fernandes thanks the support of cE3c through an assistant researcher contract (FCiência.ID contract #366) and FCT (Fundação para a Ciência e a Tecnologia) for Portuguese National Funds attributed to cE3c within the strategic project UID/BIA/00329/2020; C. Fernandes also thanks FPUL for a contract of invited assistant professor. This research was funded by FCT project grant PTDC/BIA-BEC/103729/2008 (FCT), coordinated by C. Bastos-Silveira. Thanks are due, for financial support, to Centro de Ecologia, Evolução e Alterações Ambientais (UIDB/00329/2020), Centro de Estudos de Ambiente e Mar (UID/AMB/50017), to FCT/MEC through national funds, and the co-funding by the FEDER, within the PT2020 Partnership Agreement and Compete 2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in GenBank in accession numbers MT226280-MT226326. See Table S1 for details.

Acknowledgments: We are grateful to Robert Baker, Heath Garner, and Kathy MacDonald from the Museum of Texas Tech University (Lubbock, TX, USA), Peter Fritzsche from the Institute of Zoology

of the Martin-Luther University (Halle (Salle), Germany), and Isabel Rey from the Museo Nacional de Ciencias Naturales (Madrid, Spain) for providing biological samples. We would also like to thank Susanne Tellenbach for technical assistance and Jonathan Cook for proofreading our manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Swanson, W.J.; Vacquier, V.D. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **2002**, *3*, 137–144. [[CrossRef](#)]
2. Turner, L.M.; Hoekstra, H.E. Causes and consequences of the evolution of reproductive proteins. *Int. J. Dev. Biol.* **2008**, *52*, 769–780. [[CrossRef](#)] [[PubMed](#)]
3. Findlay, G.D.; Swanson, W.J. Proteomics enhances evolutionary and functional analysis of reproductive proteins. *BioEssays* **2010**, *32*, 26–36. [[CrossRef](#)] [[PubMed](#)]
4. Coyne, J.A.; Orr, H.A. *Speciation*; Sinauer Associates: Sunderland, MA, USA, 2004.
5. Seehausen, O.; Butlin, R.K.; Keller, I.; Wagner, C.E.; Boughman, J.W.; Hohenlohe, P.A.; Peichel, C.L.; Saetre, G.-P.; Bank, C.; Brännström, A.; et al. Genomics and the origin of species. *Nat. Rev. Genet.* **2014**, *15*, 176–192. [[CrossRef](#)] [[PubMed](#)]
6. Clark, N.L.; Gasper, J.; Sekino, M.; Springer, S.A.; Aquadro, C.F.; Swanson, W.J. Coevolution of interacting fertilization proteins. *PLoS Genet.* **2009**, *5*, e1000570. [[CrossRef](#)] [[PubMed](#)]
7. Wassarman, P.M.; Litscher, E.S. Sperm-egg recognition mechanisms in mammals. *Curr. Top Dev. Biol.* **1995**, *30*, 1–19.
8. Wassarman, P.M. Mammalian fertilization: Review molecular aspects of gamete adhesion, exocytosis, and fusion. *Cell* **1999**, *96*, 175–183. [[CrossRef](#)]
9. Kinloch, R.A.; Wassarman, P.M. Nucleotide sequence of the gene encoding zona pellucida glycoprotein ZP3—The mouse sperm receptor. *Nucleic Acids Res.* **1989**, *17*, 2861–2863. [[CrossRef](#)]
10. Bleil, J.D.; Wassarman, P.M. Structure and function of the zona pellucida: Identification and characterization of the proteins of the mouse oocyte's zona pellucida. *Dev. Biol.* **1980**, *76*, 185–202. [[CrossRef](#)]
11. Bleil, J.D.; Wassarman, P.M. Autoradiographic visualization of the mouse egg's sperm receptor bound to sperm. *J. Cell Biol.* **1986**, *102*, 1363–1369. [[CrossRef](#)]
12. Wassarman, P.M.; Jovine, L.; Qi, H.; Williams, Z.; Darie, C.; Litscher, E.S. Recent aspects of mammalian fertilization research. *Mol. Cell. Endocrinol.* **2005**, *234*, 95–103. [[CrossRef](#)]
13. Litscher, E.S.; Williams, Z.; Wassarman, P.M. Zona pellucida glycoprotein ZP3 and fertilization in mammals. *Mol. Reprod. Dev.* **2009**, *76*, 933–941. [[CrossRef](#)] [[PubMed](#)]
14. Florman, H.M.; Wassarman, P.M. O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity. *Cell* **1985**, *41*, 313–324. [[CrossRef](#)]
15. Rosière, T.K.; Wassarman, P.M. Identification of a region of mouse zona pellucida glycoprotein mZP3 that possesses sperm receptor activity. *Dev. Biol.* **1992**, *154*, 309–317. [[CrossRef](#)]
16. Kinloch, R.A.; Sakai, Y.; Wassarman, P.M. Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 263–267. [[CrossRef](#)] [[PubMed](#)]
17. Chen, J.; Litscher, E.S.; Wassarman, P.M. Inactivation of the mouse sperm receptor, mZP3, by site-directed mutagenesis of individual serine residues located at the combining site for sperm. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6193–6197. [[CrossRef](#)] [[PubMed](#)]
18. Redgrove, K.A.; Aitken, R.J.; Nixon, B. More than a simple lock and key mechanism: Unraveling the intricacies of sperm-zona pellucida binding. In *Binding Protein*; Abdelmohsen, K., Ed.; InTech: Rijeka, Croatia, 2012; pp. 73–122.
19. Tumova, L.; Zigo, M.; Sutovsky, P.; Sedmikova, M.; Postlerova, P. Ligands and Receptors Involved in the Sperm-Zona Pellucida Interactions in Mammals. *Cells* **2021**, *10*, 133. [[CrossRef](#)]
20. Chalabi, S.; Panico, M.; Sutton-Smith, M.; Haslam, S.M.; Patankar, M.S.; Lattanzio, F.A.; Morris, H.R.; Clarck, G.F.; Dell, A. Differential O-glycosylation of a conserved domain expressed in murine and human ZP3. *Biochem* **2006**, *45*, 637–647. [[CrossRef](#)]
21. Monné, M.; Jovine, L. A structural view of egg coat architecture and function in fertilization. *Biol. Reprod.* **2011**, *85*, 661–669. [[CrossRef](#)]
22. Rankin, T.L.; Tong, Z.B.; Castle, P.E.; Lee, E.; Gore-Langton, R.; Nelson, L.M.; Dean, J. Human ZP3 restores fertility in Zp3 null mice without affecting order-specific sperm binding. *Development* **1998**, *125*, 2415–2424. [[CrossRef](#)]
23. Dean, J. Reassessing the molecular biology of sperm-egg recognition with mouse genetics. *Bioessays* **2004**, *26*, 29–38. [[CrossRef](#)]
24. Clark, G.F.; Dell, A. Molecular models for murine sperm-egg binding. *J. Biol. Chem.* **2006**, *281*, 13853–13856. [[CrossRef](#)] [[PubMed](#)]
25. Gahlay, G.; Gauthier, L.; Baibakov, B.; Epifano, O.; Dean, J. Gamete recognition in mice depends on the cleavage status of an egg's zona pellucida protein. *Science* **2010**, *329*, 216–219. [[CrossRef](#)] [[PubMed](#)]
26. Avella, M.A.; Baibakov, B.; Dean, J. A single domain of the ZP2 zona pellucida protein mediates gamete recognition in mice and humans. *J. Cell Biol.* **2014**, *205*, 801–809. [[CrossRef](#)] [[PubMed](#)]
27. Bianchi, E.; Wright, G.J. Find and fuse: Unsolved mysteries in sperm-egg recognition. *PLoS Biol.* **2020**, *18*, e3000953. [[CrossRef](#)] [[PubMed](#)]
28. Stsiapanava, A.; Xu, C.; Brunati, M.; Zamora-Caballero, S.; Schaeffer, C.; Bokhove, M.; Han, L.; Hebert, H.; Carroni, M.; Yasumasu, S.; et al. Cryo-EM structure of native human uromodulin, a zona pellucida module polymer. *EMBO J.* **2020**, *39*, e106807. [[CrossRef](#)]

29. Swann, C.A.; Cooper, S.J.B.; Breed, W.G. Molecular evolution of the carboxy terminal region of the zona pellucida 3 glycoprotein in murine rodents. *Reproduction* **2007**, *133*, 697–708. [[CrossRef](#)] [[PubMed](#)]
30. Swanson, W.J.; Yang, Z.; Wolfner, M.F.; Aquadro, C.F. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2509–2514. [[CrossRef](#)] [[PubMed](#)]
31. Morgan, C.C.; Loughran, N.B.; Walsh, T.A.; Harrison, A.J.; O'Connell, M.J. Positive selection neighboring functionally essential sites and disease-implicated regions of mammalian reproductive proteins. *BMC Evol. Biol.* **2010**, *10*, 39. [[CrossRef](#)]
32. Swann, C.C.; Hart, M.W. Molecular evolution of mammalian genes with epistatic interactions in fertilization. *BMC Evol. Biol.* **2019**, *19*, 154.
33. Jansa, S.A.; Lundrigan, B.L.; Tucker, P.K. Tests for positive selection on immune and reproductive genes in closely related species of the murine genus. *Mus. J. Mol. Evol.* **2003**, *56*, 294–307. [[CrossRef](#)] [[PubMed](#)]
34. Turner, L.M.; Hoekstra, H.E. Adaptive evolution of fertilization proteins within a genus: Variation in ZP2 and ZP3 in deer mice (*Peromyscus*). *Mol. Biol. Evol.* **2006**, *23*, 1656–1669. [[CrossRef](#)] [[PubMed](#)]
35. Turner, L.M.; Hoekstra, H.E. Reproductive protein evolution within and between species: Maintenance of divergent ZP3 alleles in *Peromyscus*. *Mol. Ecol.* **2008**, *17*, 2616–2628. [[CrossRef](#)]
36. Swann, C.A.; Cooper, S.J.B.; Breed, W.G. The egg coat zona pellucida 3 glycoprotein—Evolution of its putative sperm-binding region in Old World murine rodents (Rodentia: Muridae). *Reprod. Fertil. Dev.* **2017**, *29*, 2376–2386. [[CrossRef](#)] [[PubMed](#)]
37. Amaral, A.R.; Möller, L.M.; Beheregaray, L.B.; Coelho, M.M. Evolution of 2 reproductive proteins, ZP3 and PKDREJ, in cetaceans. *J. Hered.* **2011**, *102*, 275–282. [[CrossRef](#)]
38. Chen, S.; Costa, V.; Beja-Pereira, A. Evolutionary patterns of two major reproduction candidate genes (Zp2 and Zp3) reveal no contribution to reproductive isolation between bovine species. *BMC Evol. Biol.* **2011**, *11*, 24. [[CrossRef](#)] [[PubMed](#)]
39. Chaline, J.; Brunet-Lecomte, P.; Montuire, S.; Viriot, L.; Courant, F. Anatomy of the arvicoline radiation (Rodentia): Palaeogeographical, palaeoecological history and evolutionary data. *Ann. Zool. Fenn.* **1999**, *36*, 239–267.
40. Musser, G.M.; Carleton, M.D. Family Cricetidae. In *Mammal Species of the World: A Taxonomic and Geographic Reference*; Wilson, D.E., Reeder, D.M., Eds.; Smithsonian Institution: Washington, DC, USA, 1993; pp. 955–1189.
41. Fink, S.; Fischer, M.C.; Excoffier, L.; Heckel, G. Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): The utility of AFLPs versus mitochondrial and nuclear sequence markers. *Syst. Biol.* **2010**, *59*, 548–572. [[CrossRef](#)]
42. Bastos-Silveira, C.; Santos, S.M.; Monarca, R.; Mathias, M.L.; Heckel, G. Deep mitochondrial introgression and hybridization among ecologically divergent vole species. *Mol. Ecol.* **2012**, *21*, 5309–5323. [[CrossRef](#)] [[PubMed](#)]
43. Paupério, J.; Herman, J.S.; Melo-Ferreira, J.; Jaarola, M.; Alves, P.C.; Searle, J.B. Cryptic speciation in the field vole: A multilocus approach confirms three highly divergent lineages in Eurasia. *Mol. Ecol.* **2012**, *21*, 6015–6032. [[CrossRef](#)] [[PubMed](#)]
44. Beysard, M.; Heckel, G. Structure and dynamics of hybrid zones at different stages of speciation in the common vole (*Microtus arvalis*). *Mol. Ecol.* **2014**, *23*, 673–687. [[CrossRef](#)]
45. Sambrook, J.; Fritschi, E.F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: New York, NY, USA, 1989.
46. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
47. Villesen, P. FaBox: An online toolbox for fasta sequences. *Mol. Ecol. Notes* **2007**, *7*, 965–968. [[CrossRef](#)]
48. Stephens, M.; Donnelly, P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.* **2003**, *73*, 1162–1169. [[CrossRef](#)]
49. Stephens, M.; Smith, N.; Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **2001**, *68*, 978–989. [[CrossRef](#)] [[PubMed](#)]
50. Librado, P.; Rozas, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **2009**, *25*, 1451–1452. [[CrossRef](#)] [[PubMed](#)]
51. Schneider, T.D.; Stephens, R.M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **1990**, *18*, 6097–6100. [[CrossRef](#)]
52. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)]
53. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [[CrossRef](#)]
54. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [[CrossRef](#)]
55. Posada, D. Using MODELTEST and PAUP* to select a model of nucleotide substitution. *Curr. Protoc. Bioinform.* **2003**, *6*, 6.5.1–6.5.14. [[CrossRef](#)]
56. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* **1974**, *19*, 716–723. [[CrossRef](#)]
57. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [[CrossRef](#)]
58. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [[CrossRef](#)] [[PubMed](#)]
59. Müller, K. SeqState—Primer design and sequence statistics for phylogenetic DNA data sets. *Appl. Bioinform.* **2005**, *4*, 65–69.

60. Simmons, M.P.; Ochoterena, H. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **2000**, *49*, 369–381. [[CrossRef](#)] [[PubMed](#)]
61. Müller, K. Incorporating information from length-mutational events into phylogenetic analysis. *Mol. Phylogenet. Evol.* **2006**, *38*, 667–676. [[CrossRef](#)]
62. Rambaut, A. FigTree. 2010. Available online: <http://tree.bio.ed.ac.uk/software/figtree/> (accessed on 9 March 2016).
63. Arenas, M.; Posada, D. Coalescent Simulation of Intracodon Recombination. *Genetics* **2010**, *184*, 429–437. [[CrossRef](#)] [[PubMed](#)]
64. Arenas, M.; Posada, D. The influence of recombination on the estimation of selection from coding sequence alignments. In *Natural Selection: Methods and Applications*; Fares, M.A., Ed.; CRC Press/Taylor & Francis: London, UK, 2014; pp. 112–125.
65. Del Amparo, R.; Branco, C.; Arenas, J.; Vicens, A.; Arenas, M. Analysis of selection in protein-coding sequences accounting for common biases. *Brief Bioinform.* **2021**, *22*, bbaa431. [[CrossRef](#)]
66. Martin, D.P.; Lemey, P.; Lott, M.; Moulton, V.; Posada, D.; Lefevre, P. RDP3: A flexible and fast computer program for analyzing recombination. *Bioinformatics* **2010**, *26*, 2462–2463. [[CrossRef](#)] [[PubMed](#)]
67. Martin, D.; Rybicki, E. RDP: Detection of recombination amongst aligned sequences. *Bioinformatics* **2000**, *16*, 562–563. [[CrossRef](#)]
68. Salminen, M. Identification of breakpoints in intergenotypic recombinants of HIV type I by bootscanning. *AIDS Res. Hum. Retrovir.* **1995**, *11*, 1423–1425. [[CrossRef](#)] [[PubMed](#)]
69. Martin, D.P.; Posada, D.; Crandall, K.A.; Williamson, C. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res. Hum. Retrovir.* **2005**, *21*, 98–102. [[CrossRef](#)]
70. Padidam, M.; Sawyer, S.; Fauquet, C.M. Possible emergence of new geminiviruses by frequent recombination. *Virology* **1999**, *265*, 218–225. [[CrossRef](#)] [[PubMed](#)]
71. Maynard Smith, J. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **1992**, *34*, 126–129.
72. Posada, D.; Crandall, K.A. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13757–13762. [[CrossRef](#)] [[PubMed](#)]
73. Gibbs, M.J.; Armstrong, J.S.; Gibbs, A.J. Sister-Scanning: A Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **2000**, *16*, 573–582. [[CrossRef](#)]
74. Boni, M.F.; Posada, D.; Feldman, M.W. An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* **2007**, *176*, 1035–1047. [[CrossRef](#)]
75. Yang, Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **1997**, *13*, 555–556. [[CrossRef](#)]
76. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)]
77. Goldman, N.; Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **1994**, *11*, 725–736.
78. 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. [[CrossRef](#)]
79. Yang, Z.; Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **1998**, *46*, 409–418. [[CrossRef](#)] [[PubMed](#)]
80. Yang, Z.; Nielsen, R.; Goldman, N.; Pedersen, A.-M.K. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **2000**, *155*, 431–449. [[CrossRef](#)] [[PubMed](#)]
81. Swanson, W.J.; Nielsen, R.; Yang, Q.F. Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **2003**, *20*, 18–20. [[CrossRef](#)]
82. Wong, W.S.W.; 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. [[CrossRef](#)]
83. Yang, Z.; Wong, W.S.W.; Nielsen, R. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **2005**, *22*, 1107–1118. [[CrossRef](#)]
84. Yang, Z.; Nielsen, R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **2002**, *19*, 908–917. [[CrossRef](#)]
85. Zhang, J.; Nielsen, R.; Yang, Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* **2005**, *22*, 2472–2479. [[CrossRef](#)] [[PubMed](#)]
86. Yang, Z.; dos Reis, M. Statistical properties of the branch-site test of positive selection. *Mol. Biol. Evol.* **2001**, *28*, 1217–1228. [[CrossRef](#)]
87. Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* **1998**, *15*, 568–573. [[CrossRef](#)]
88. Berlin, S.; Smith, N.G. Testing for adaptive evolution of the female reproductive protein ZPC in mammals, birds and fishes reveals problems with the M7-M8 likelihood ratio test. *BMC Evol. Biol.* **2005**, *5*, 65. [[CrossRef](#)]
89. Kosakovsky Pond, S.L.; Frost, S.D.W. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **2005**, *21*, 2531–2533. [[CrossRef](#)]
90. Delport, W.; Poon, A.F.; Frost, S.D.W.; Kosakovsky Pond, S.L. Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **2010**, *26*, 2455–2457. [[CrossRef](#)]
91. Weaver, S.; Shank, S.D.; Spielman, S.J.; Li, M.; Muse, S.V.; Kosakovsky Pond, S.L. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **2018**, *35*, 773–777. [[CrossRef](#)] [[PubMed](#)]

92. Kosakovsky Pond, S.L.; Frost, S.D.K.; Muse, S.V. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* **2005**, *21*, 676–679. [[CrossRef](#)]
93. Kosakovsky Pond, S.L.; Frost, S.D.W. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **2005**, *22*, 1208–1222. [[CrossRef](#)] [[PubMed](#)]
94. Murrell, B.; Wertheim, J.O.; Moola, S.; Weighill, T.; Scheffler, K.; Kosakovsky Pond, S.L. Detecting individual sites subject to episodic diversifying selection. *PLoS Gene.* **2012**, *8*, e1002764. [[CrossRef](#)] [[PubMed](#)]
95. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S.L.; Scheffler, K. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol. Biol. Evol.* **2013**, *30*, 1196–1205. [[CrossRef](#)]
96. Smith, M.D.; Wertheim, J.O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S.L. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **2015**, *32*, 1342–1353. [[CrossRef](#)]
97. Kosakovsky Pond, S.L.; Murrell, N.; Fourment, M.; Frost, S.D.W.; Delpont, W.; Scheffler, K. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* **2011**, *28*, 3033–3043. [[CrossRef](#)] [[PubMed](#)]
98. Murrell, B.; Weaver, S.; Smith, M.D.; Wertheim, J.O.; Murrell, S.; Aylward, A.; Eren, K.; Pollner, T.; Martin, D.P.; Smith, D.M.; et al. Gene-Wide Identification of Episodic Selection. *Mol. Biol. Evol.* **2015**, *32*, 1365–1371. [[CrossRef](#)] [[PubMed](#)]
99. Steppan, S.; Adkins, R.; Anderson, J. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. *Syst. Biol.* **2004**, *53*, 533–553. [[CrossRef](#)] [[PubMed](#)]
100. Fabre, P.-H.; Hautier, L.; Dimitrov, D.; Douzery, E.J.P. A glimpse on the pattern of rodent diversification: A phylogenetic approach. *BMC Evol. Biol.* **2012**, *12*, 88. [[CrossRef](#)]
101. Steppan, S.J.; Schenk, J.J. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS ONE* **2017**, *12*, e0183070. [[CrossRef](#)]
102. Jaarola, M.; Martínková, N.; Gündüz, I.; Brunhoff, C.; Zima, J.; Nadachowski, A.; Amori, G.; Bulatova, N.S.; Chondropoulos, B.; Fraguédakis-Tsolis, S.; et al. Molecular phylogeny of the speciose vole genus *Microtus* (Arvicolinae, Rodentia) inferred from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* **2004**, *33*, 647–663. [[CrossRef](#)]
103. Martínková, N.; Moravec, J. Multilocus phylogeny of arvicoline voles (Arvicolini, Rodentia) shows small tree terrace size. *Folia Zool.* **2012**, *61*, 254–267. [[CrossRef](#)]
104. Hogan, K.M.; Davis, S.K.; Greenbaum, I.F. Mitochondrial-DNA Analysis of the Systematic Relationships within the *Peromyscus maniculatus* Species Group. *J. Mammal.* **1997**, *78*, 733–743. [[CrossRef](#)]
105. Bradley, R.D.; Durish, N.D.; Rogers, D.S.; Miller, J.R.; Engstrom, M.D.; Kilpatrick, C.W. Toward a Molecular Phylogeny for *Peromyscus*: Evidence from Mitochondrial Cytochrome-b Sequences. *J. Mammal.* **2007**, *88*, 1146–1159. [[CrossRef](#)]
106. Platt, R.N., II; Amman, B.R.; Keith, M.S.; Thompson, C.W.; Bradley, R.D. What Is *Peromyscus*? Evidence from nuclear and mitochondrial DNA sequences suggests the need for a new classification. *J. Mammal.* **2015**, *96*, 708–719. [[CrossRef](#)] [[PubMed](#)]
107. Gering, E.J.; Opazo, J.C.; Storz, J.F. Molecular evolution of cytochrome b in high- and low-altitude deer mice (genus *Peromyscus*). *Heredity* **2009**, *102*, 226–235. [[CrossRef](#)]
108. Neumann, K.; Michaux, J.; Lebedev, V.; Yigit, N.; Colak, E.; Ivanova, N.; Poltoraus, A.; Surov, A.; Markov, G.; Maak, S.; et al. Molecular phylogeny of the Cricetinae subfamily based on the mitochondrial cytochrome b and 12S rRNA genes and the nuclear vWF gene. *Mol. Phylogenet. Evol.* **2006**, *39*, 135–148. [[CrossRef](#)]
109. Anisimova, M.; Bielawski, J.P.; Yang, Z. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. *Mol. Biol. Evol.* **2001**, *18*, 1585–1592. [[CrossRef](#)]
110. Suzuki, Y.; Nei, M. False-Positive Selection Identified by ML-Based Methods: Examples from the *Sig1* Gene of the Diatom *Thalassiosira weissflogii* and the *tax* Gene of a Human T-cell Lymphotropic Virus. *Mol. Biol. Evol.* **2004**, *21*, 914–921. [[CrossRef](#)]
111. Suzuki, Y. False-positive results obtained from the branch-site test of positive selection. *Genes Genet. Syst.* **2008**, *83*, 331–338. [[CrossRef](#)] [[PubMed](#)]
112. Anisimova, M.; Yang, Z. Multiple Hypothesis Testing to Detect Lineages under Positive Selection that Affects Only a Few Sites. *Mol. Biol. Evol.* **2007**, *24*, 1219–1228. [[CrossRef](#)] [[PubMed](#)]
113. Nozawa, M.; Suzuki, Y.; Nei, M. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6700–6705. [[CrossRef](#)] [[PubMed](#)]
114. Anisimova, M.; Bielawski, J.P.; Yang, Z. Accuracy and Power of Bayes Prediction of Amino Acid Sites under Positive Selection. *Mol. Biol. Evol.* **2002**, *19*, 950–958. [[CrossRef](#)]
115. Visconti, P.E.; Florman, H.E. Mechanisms of sperm-egg interactions: Between sugars and broken bonds. *Sci. Signal.* **2010**, *3*, pe35. [[CrossRef](#)] [[PubMed](#)]
116. Clark, G.F. The molecular basis of mouse sperm-zona pellucida binding: A still unresolved issue in developmental biology. *Reproduction* **2011**, *142*, 377–381. [[CrossRef](#)]
117. Moros-Nicolás, C.; Chevret, P.; Jiménez-Movilla, M.; Algarra, B.; Cots-Rodríguez, P.; González-Brusi, L.; Avilés, M.; Izquierdo-Rico, M.J. New Insights into the Mammalian Egg Zona Pellucida. *Int. J. Mol. Sci.* **2021**, *22*, 3276. [[CrossRef](#)] [[PubMed](#)]
118. Cerveira, A.M.; Soares, J.; Bastos-Silveira, C.; Mathias, M.L. Reproductive isolation between sister species of Iberian pine voles, *Microtus duodecimcostatus* and *M. lusitanicus*. *Ethol. Ecol. Evol.* **2018**, *31*, 121–139. [[CrossRef](#)]