

Interacting Proteins on Human Spermatozoa: Adaptive Evolution of the Binding of Semenogelin I to EPPIN

Erick J. R. Silva[‡], Katherine G. Hamil, Michael G. O'Rand

Department of Cell Biology & Physiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

Abstract

Semenogelin I (SEMG1) is found in human semen coagulum and on the surface of spermatozoa bound to EPPIN. The physiological significance of the SEMG1/EPPIN interaction on the surface of spermatozoa is its capacity to modulate sperm progressive motility. The present study investigates the hypothesis that the interacting surface of SEMG1 and EPPIN co-evolved within the Hominoidea time scale, as a result of adaptive pressures applied by their roles in sperm protection and reproductive fitness. Our results indicate that some amino acid residues of SEMG1 and EPPIN possess a remarkable deficiency of variation among hominoid primates. We observe a distinct residue change unique to humans within the EPPIN sequence containing a SEMG1 interacting surface, namely His92. In addition, Bayes Empirical Bayes analysis for positive selection indicates that the SEMG1 Cys239 residue underwent positive selection in humans, probably as a consequence of its role in increasing the binding affinity of these interacting proteins. We confirm the critical role of Cys239 residue for SEMG1 binding to EPPIN and inhibition of sperm motility by showing that recombinant SEMG1 mutants in which Cys239 residue was changed to glycine, aspartic acid, histidine, serine or arginine have reduced capacity to interact to EPPIN and to inhibit human sperm motility *in vitro*. In conclusion, our results indicate that EPPIN and SEMG1 rapidly co-evolved in primates due to their critical role in the modulation of sperm motility in the semen coagulum, providing unique insights into the molecular co-evolution of sperm surface interacting proteins.

Citation: Silva EJR, Hamil KG, O'Rand MG (2013) Interacting Proteins on Human Spermatozoa: Adaptive Evolution of the Binding of Semenogelin I to EPPIN. PLoS ONE 8(12): e82014. doi:10.1371/journal.pone.0082014

Editor: David H Volle, Inserm, France

Received: September 9, 2013; **Accepted:** October 28, 2013; **Published:** December 2, 2013

Copyright: © 2013 Silva et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding provided by grants HD048843 and HD060494 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and by grant D43TW-HD00627 from Fogarty International Center Program for Training and Research in Population and Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: ejrsilva81@gmail.com

‡ Current address: Section of Experimental Endocrinology, Department of Pharmacology, Universidade Federal de São Paulo, São Paulo, São Paulo, Brazil

Introduction

In primates, semen is a complex biological fluid containing spermatozoa bathed by the seminal plasma, which nurtures spermatozoa while providing the appropriate conditions for the last steps of post-testicular sperm maturation, and protecting them from pathogenic threats during their journey towards fertilization in the female reproductive tract [1]. The seminal vesicle-secreted protein semenogelin I (SEMG1) is the major component of the seminal plasma [2]. Immediately after ejaculation, human semen undergoes a coagulation process forming a gelatinous mass that contains SEMG1 as its structural element [3]. SEMG1 is found in the semen coagulum and on the surface of spermatozoa bound to EPPIN, a member of the whey-acidic protein (WAP)-type four-disulfide core (WFDC) family [4,5]. After ejaculation, the hydrolysis of SEMG1 by activated prostate-specific antigen (PSA), a serine protease,

results in the liquefaction of the semen coagulum allowing spermatozoa to acquire progressive motility [2,6].

The physiological significance of the SEMG1/EPPIN interaction on the surface of spermatozoa is its capacity to modulate sperm progressive motility upon ejaculation in a finely time-regulated manner [5,7]. In parallel, EPPIN inhibits the digestion of SEMG1 by PSA, which results in the modulation of semen liquefaction, further prolonging the inhibitory effects of SEMG1 on sperm motility [8]. This effect is thought to be important for reproduction because it allows spermatozoa to achieve their full fertilizing capacity at the appropriate moment. Men whose semen coagulum fails to liquefy spontaneously have spermatozoa with poor motility and are infertile [9]. Additionally, EPPIN and some SEMG1-proteolytic peptides derived by PSA cleavage have strong antibacterial activity *in vitro* and may protect the integrity of spermatozoa against microorganisms present in the vaginal environment [10-13]. The critical roles of SEMG1 and EPPIN in reproduction are

highlighted by the fact that male monkeys immunized with human recombinant EPPIN that developed high titers of anti-EPPIN antibodies lacked semen coagulum upon ejaculation and became reversibly infertile [14]. Subsequent studies demonstrated that anti-EPPIN antibodies isolated from immunized monkeys blocked the SEMG1 binding site on EPPIN's C-terminal region and mimicked SEMG1 binding by inhibiting progressive motility of human spermatozoa [15,16]. Because of these observations, the SEMG1 binding interface on EPPIN has been acknowledged as a potential target for male contraception [5].

The WFDC locus on human chromosome 20q13, containing the *SEMG1* (in the centromeric cluster) and *EPPIN* (in the telomeric cluster) genes [17,18], has undergone strong adaptive pressure [19] and the *SEMG1* gene in particular has undergone rapid adaptive evolution [18-21], suggesting positive selection driven by their functions in natural immunity and reproductive success [19]. Genes for serine protease inhibitors within the WFDC locus may have been progenitors to the *SEMG* genes [17,22]. In previous studies on the interaction of SEMG1 with EPPIN on the sperm surface, we demonstrated that EPPIN's Cys102, Tyr107, and Phe117 were necessary for SEMG1 binding [23] and that SEMG1's Cys239 was required for binding to EPPIN with subsequent inhibition of sperm motility [24,25]. The SEMG1 amino acid residue Cys239 has been shown to be under positive selection [18]. Consequently we have asked whether EPPIN and SEMG1 have undergone co-adaptive evolution into a receptor ligand relationship that provides protection for spermatozoa and regulates the acquisition of progressive sperm motility prior to capacitation. Here we demonstrate that the Hominoidea underwent a distinct change in specific EPPIN and SEMG1 amino acid residues that appears to be a directional selection throughout the Hominidae (human and great apes) and Hylobatidae (gibbon) Families, resulting in a high affinity binding site between EPPIN and SEMG1 and a functional gain regarding its ability to inhibit sperm motility.

Results

Phylogenetic Analysis of EPPIN and SEMG1

The Catarrhini (old world monkeys, great apes, gibbons, and humans) emerged 43.5 million years ago (MYA) from a common ancestor and subsequently diverged into the Hominoidea and Cercopithecoidea 31.6 MYA [26]. The Homininae emerged 8.3 MYA and underwent rapid speciation into three genera (*Gorilla*, *Pan*, *Homo*); the *Gorilla* branching rapidly away from the *Homo-Pan* genera [26]. We examined the molecular phylogeny of *EPPIN* and *SEMG1* within the Hominoidea and of *Macaca mulatta* and *Papio anubis* within the Cercopithecoidea and found 2.2 more substitutions per site in the *SEMG1* lineage (Figure 1A, B). The Maximum Likelihood method gave the same tree for *EPPIN* (Figure 1A) as all other methods and conformed to the expected evolutionary tree for primates [26]. However, using the same methodology, all the methods gave an unexpected tree for the *SEMG1* sequence (Figure 1B) showing the *Gorilla* branching from the common ancestor of the Cercopithecoidea rather than the common

ancestor of the Homininae. A test of the homogeneity of substitution patterns between sequences revealed that the *Gorilla EPPIN* sequence exhibits heterogeneity with *Homo*, *Pan* and *Pongo* (Table 1, asterisks in row 1 vs. column 3, row 2 vs. column 3, and row 3 vs. column 4), whereas the *Gorilla SEMG1* sequence exhibits heterogeneity with *Macaca mulatta* (Table 2, asterisks in row 3 vs. column 6) and, although not statistically significant, a trend towards heterogeneity with *Homo sapiens* (Table 2, trend symbol in row 1 vs. column 3, row 3 vs. column 6). Although there was an overabundance of non-synonymous substitutions in several codons in each lineage, only *SEMG1* showed positively selected codons that were statistically significant (Table S1 and Table S2).

We found no evidence of positive selection in any site within the primate phylogenetic tree of *EPPIN* (Table S1). On the other hand, detailed examination of the full-length EPPIN sequence revealed that when the Hominoidea and Cercopithecoidea Superfamilies emerged from the Catarrhini 31.6 MYA the Hominoidea underwent a distinct change in seven amino acid residues (Lys35, Lys50, Asp51, Gln55, Lys81, Tyr89, Leu91, Met104, and Lys120) that appears to be a directional selection throughout the Hominidae and Hylobatidae Families (Table 3), separating them from the Papionini. An examination of the EPPIN sequence within the SEMG1 binding site in the Kunitz domain [23] revealed only two residue changes that were different between *Homo* and all the other Hominoidea, namely His92 and Asp99. Only His92 projects into the binding pocket [23] and would represent a significant change from Arg92 in charge and size.

Analysis of the SEMG1 sequence in fragment 74-8 (R165-Q247; Figure 2), a fragment still retaining EPPIN binding and sperm motility inhibitory activities (see below), revealed that the Hominoidea underwent a distinct change in four amino acid residues (Trp167, His169, Ser193, and Gln215), which might be indicative of a directional selection in Hominidae (Table 4). Additionally, Bayes Empirical Bayes (BEB) analysis [27] revealed eleven positively selected codons in this *SEMG1* sequence (Table 4; Table S2); only two (Trp167, His169) were throughout the Hominoidea. Although there were no amino acid residue differences between *Homo sapiens* and all other Hominoidea, residues Asp225 and Leu234 in the *Gorilla* were distinct from other Hominidae and Hylobatidae residues, Glu225 and Val234 respectively, but identical to *Macaca mulatta* and *Papio anubis* residues (Table 4). Residue His239 was unique in the *Gorilla SEMG1* in comparison to SEMG1 from other Hominidae and Hylobatidae (Cys239), as well as *Macaca mulatta* and *Papio anubis* (Arg239). Residue Cys239 and His169 were reported previously to be sites under positive selection [18] and our BEB analysis confirmed that the Cys239 codon underwent positive selection (Table 4; Table S2). *Homo sapiens* Cys239, which is necessary for SEMG1 binding to EPPIN [24,25], underwent a distinct change when the Hominidae and Hylobatidae Families separated from the Cercopithecoidea (R239C). Consequently, based on our phylogenetic analysis, it is likely that the positive selection of eleven of *SEMG1*'s codons within the WFDC locus of the Hominoidea and the co-evolution of EPPIN's His92 resulted in the high affinity binding of SEMG1 to EPPIN in humans.

Silva et al. - Figure 1

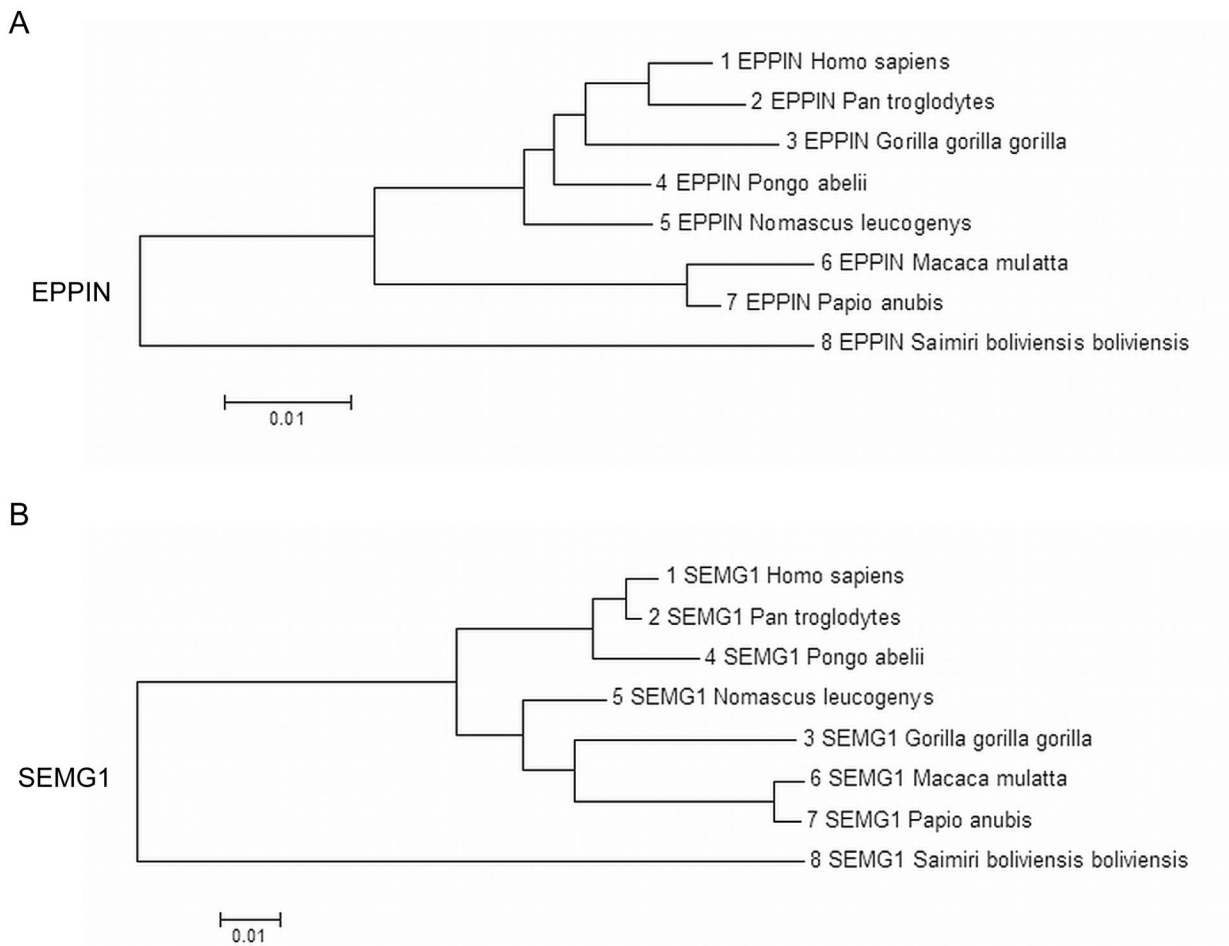


Figure 1. Molecular phylogenetic analysis for EPPIN and SEMG1 within the primate lineage. EPPIN (A) and SEMG1 (B) trees were drawn to scale with branch lengths measured in the number of substitutions per site. GenBank accession number for each DNA sequence used for EPPIN and SEMG1 analyses are shown in Table 6.

doi: 10.1371/journal.pone.0082014.g001

Binding of SEMG1 fragments and Cys239 mutants to EPPIN

As we reported previously [23], the binding of human recombinant SEMG1²¹⁴⁻⁴² fragment to EPPIN was time-dependent (Figure S1) and saturable (Figure 3A). Experiments performed using increasing concentrations of SEMG1²¹⁴⁻⁴² (0.04 - 7.5 μ M) in the presence of a constant concentration of EPPIN indicated a concentration-dependent increase in the signal up to 300 nM SEMG1²¹⁴⁻⁴² (Figure 3A and Figure S1). At higher SEMG1²¹⁴⁻⁴² concentrations, we observed a progressive drop in the signal due to oversaturation of Ni-NTA-chelate donor beads with SEMG1²¹⁴⁻⁴² (the hook effect [28], Figure S1). In order to determine the smallest region within SEMG1²¹⁴⁻⁴² critical for EPPIN binding, we cloned and expressed six SEMG1 truncated fragments lacking N-terminal and/or C-

terminal sequences flanking Cys239 residue, namely SEMG1⁷⁴⁻⁴², SEMG1⁷⁴⁻⁸, SEMG1³²⁻⁴², SEMG1¹⁹⁻²², SEMG1¹⁰⁻³⁰, and SEMG1³²⁻⁸ (Figure 2B). We designed these truncations because they correspond to SEMG1 fragments containing the Cys239 residue generated by PSA digestion [6] and are part of the 15-kDa SEMG1 fragment protected from PSA cleavage by EPPIN [7]. Thus, it is likely they are naturally generated during the semen liquefaction. Among these fragments, only SEMG1⁷⁴⁻⁴² and SEMG1⁷⁴⁻⁸ bound to EPPIN, albeit at much higher concentrations (1 - 7.5 μ M) in comparison to the largest fragment SEMG1²¹⁴⁻⁴² (Figure 3A). We observed no saturation when SEMG1⁷⁴⁻⁴² and SEMG1⁷⁴⁻⁸ fragments were titrated in the presence of EPPIN under similar conditions (Figure 3A).

To confirm these results, we performed competition experiments to investigate the capacity of SEMG1 fragments to

Table 1. Test of the homogeneity of substitution patterns between sequences for *EPPIN*.

Species	1	2	3	4	5	6	7	8
1 <i>Homo sapiens</i>	-	0.000	0.005*	0.002*	0.000	0.000	0.000	0.000
2 <i>Pan troglodytes</i>	1.000	-	0.005*	0.000	0.000	0.000	0.000	0.000
3 <i>Gorilla gorilla gorilla</i>	0.314	0.320	-	0.017*	0.000	0.000	0.000	0.000
4 <i>Pongo abelii</i>	0.442	1.000	0.172	-	0.000	0.015*	0.010*	0.000
5 <i>Nomascus leucogenys</i>	1.000	1.000	1.000	1.000	-	0.000	0.000	0.000
6 <i>Macaca mulatta</i>	1.000	1.000	1.000	0.338	1.000	-	0.000	0.000
7 <i>Papio anubis</i>	1.000	1.000	1.000	0.324	1.000	1.000	-	0.000
8 <i>Saimiri boliviensis boliviensis</i>	1.000	1.000	1.000	1.000	1.000	1.000	1.000	-

MEGA molecular evolutionary genetic analysis software package (version 5.1) [36] used a Monte Carlo test (500 replicates) to estimate the *p*-values [38], which are shown above the diagonal (bold). * *p*-values < 0.05 were considered significant. Estimates of the disparity index per site are shown for each sequence pair below the diagonal (italic). There were a total of 402 positions in the final dataset. GenBank accession number for each DNA sequence used for EPPIN and SEMG1 analyses are shown in Table 6.

doi: 10.1371/journal.pone.0082014.t001

Table 2. Test of the homogeneity of substitution patterns between sequences for *SEMG1*.

Species	1	2	3	4	5	6	7	8
1 <i>Homo sapiens</i>	-	0.000	0.051#	0.011*	0.000	0.000	0.000	0.000
2 <i>Pan troglodytes</i>	1.000	-	0.075	0.011*	0.000	0.000	0.000	0.000
3 <i>Gorilla gorilla gorilla</i>	0.192	0.102	-	0.192	0.141	0.019*	0.000	0.123
4 <i>Pongo abelii</i>	0.234	0.226	0.022	-	0.000	0.000	0.039*	0.000
5 <i>Nomascus leucogenys</i>	1.000	1.000	0.024	1.000	-	0.000	0.000	0.000
6 <i>Macaca mulatta</i>	1.000	1.000	0.280	1.000	1.000	-	0.000	0.000
7 <i>Papio anubis</i>	1.000	1.000	1.000	0.202	1.000	1.000	-	0.000
8 <i>Saimiri boliviensis boliviensis</i>	1.000	1.000	0.174	1.000	1.000	1.000	1.000	-

MEGA molecular evolutionary genetic analysis software package (version 5.1) [36] used a Monte Carlo test (500 replicates) to estimate the *p*-values [38], which are shown above the diagonal (bold). * *p*-values < 0.05 were considered significant. Estimates of the disparity index per site are shown for each sequence pair below the diagonal (italic). There were a total of 738 positions in the final dataset. # indicates a trend towards heterogeneity, although it was not significant. GenBank accession number for each DNA sequence used for EPPIN and SEMG1 analyses are shown in Table 6.

doi: 10.1371/journal.pone.0082014.t002

Table 3. Amino acid residue changes in the full-length EPPIN sequence.

EPPIN Residues (position #)	Species						
	7	6	5	4	3	2	1
8	S	S	S	<i>N</i>	S	S	S
14	I	I	I	<i>V</i>	<i>V</i>	<i>V</i>	<i>V</i>
17	V	A	A	A	<i>V</i>	A	A
35	T	T	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>
43	K	K	K	K	<i>Q</i>	<i>Q</i>	<i>Q</i>
50	R	R	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>
51	H	H	<i>D</i>	<i>D</i>	<i>D</i>	<i>D</i>	<i>D</i>
55	P	P	<i>Q</i>	<i>Q</i>	<i>Q</i>	<i>Q</i>	<i>Q</i>
78	E	E	E	E	E	<i>K</i>	E
81	N	N	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>
89	F	F	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
91	I	I	<i>L</i>	<i>L</i>	<i>L</i>	<i>L</i>	<i>L</i>
92	R	R	R	R	R	R	<i>H</i>
99	N	N	N	N	N	N	<i>D</i>
101	T	T	T	T	<i>S</i>	T	T
104	T	T	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>	<i>M</i>
107	Y	H	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>	<i>Y</i>
111	Q	Q	<i>P</i>	<i>Q</i>	<i>Q</i>	<i>Q</i>	<i>Q</i>
120	E	E	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>	<i>K</i>
128	K	K	K	K	<i>R</i>	K	K
129	N	N	N	<i>K</i>	N	N	N

N-terminal WAP domain and C-terminal Kunitz domain are present within positions 1-74 and 75-133 of EPPIN.

^ indicates the divergence of the Catarrhini into the Hominoidea (species 1-5) and Cercopithecoidea (species 6 and 7), 31.6 MYA; Changes after the split are shown in bold/italics. Residues 101 (Serine) and 128 (Arginine) in the *Gorilla* sequence represent changes unique to *Gorilla*. Residues 92 (Histidine) and 99 (Aspartic acid) in the *Homo sapiens* sequence represent changes unique to *Homo sapiens*. Species: 1, *Homo sapiens*; 2, *Pan troglodytes*; 3, *Gorilla gorilla*; 4, *Pongo abelii*; 5, *Nomascus leucogenys*; 6, *Macaca mulatta*; 7, *Papio anubis*.

doi: 10.1371/journal.pone.0082014.t003

compete with biotinylated (bt)-SEMG1²¹⁴⁻⁴² for its binding site on EPPIN. SEMG1²¹⁴⁻⁴² decreased bt-SEMG1²¹⁴⁻⁴²/EPPIN interaction in a concentration-dependent manner (Figure 3B) with a calculated IC₅₀ value (95% confidence interval) of 8.1 nM (6.8-9.7 nM). Both SEMG1⁷⁴⁻⁴² and SEMG1⁷⁴⁻⁸ inhibited the binding of bt-SEMG1²¹⁴⁻⁴² to EPPIN at higher concentrations in comparison to SEMG1²¹⁴⁻⁴² (Figure 3B). The calculated IC₅₀ values were 123.7 nM (92.4-165.7 nM) and 384.7 nM (104.2-1421 nM) for SEMG1⁷⁴⁻⁴² and SEMG1⁷⁴⁻⁸, respectively. Curiously, SEMG1³²⁻⁸ competed with bt-SEMG1²¹⁴⁻⁴² for binding EPPIN at even higher concentrations (Figure 3B), with an IC₅₀ of 5.1 μM (0.3-75.2 μM).

Next, we measured the binding of the SEMG1²¹⁴⁻⁴² Cys239 point mutants to EPPIN in the AlphaScreen assay. All point mutations negatively affected SEMG1/EPPIN binding as demonstrated by right-shifted concentration-response curves (Figure 3C) and consequently increased EC₅₀ values (Table 5) in comparison to the wild-type when SEMG1²¹⁴⁻⁴² mutants were incubated with EPPIN under similar conditions. We observed a

Silva et al. - Figure 2

A

Human Semenogelin-1 (SEMG1; NP_002998)

```

1  MKPNIIFVLS  LLLILEKQAA  VMGQKGGSKG  RLPSEFSQFP  HGQKGQHYSG  QKGKQQTESK
61  GSFSIQYTYH  VDANDHDQSR  KSQQYDLNAL  HKTTKSQRHL  GGSQQLLHNK  QEGRDHDKSK
121  GHFHRVVIHH  KGGKAHRGTQ  NPSQDQGNP  SGKGISSQYS  NTEERLWVHG  LSKEQTSVSG
181  AQKGRKQGGG  QSSYVLQTEE  LVANKQQRET  KNSHQKNGHY  QNVVEVREEH  SSKVQTSLCP
241  AHQDKLQHGS  KDIFSTQDEL  LVYNKNQHQT  KNLNQDQQHG  RKANKISYQS  SSTEERLHY
301  GENGVQKDVS  QSSIYSQTEE  KAQKGSQKQI  TIPSQEQEHS  QKANKISYQS  SSTEERLHY
361  GENGVQKDVS  QRSIYSQTEK  LVAGKSQIQA  PNPKQEPWHG  ENAKGESGQS  TNREQDLLSH
421  EQKGRHQHGS  HGGLDIVIIE  QEDDSRHLA  QHLNNDRNPL  FT

```

B

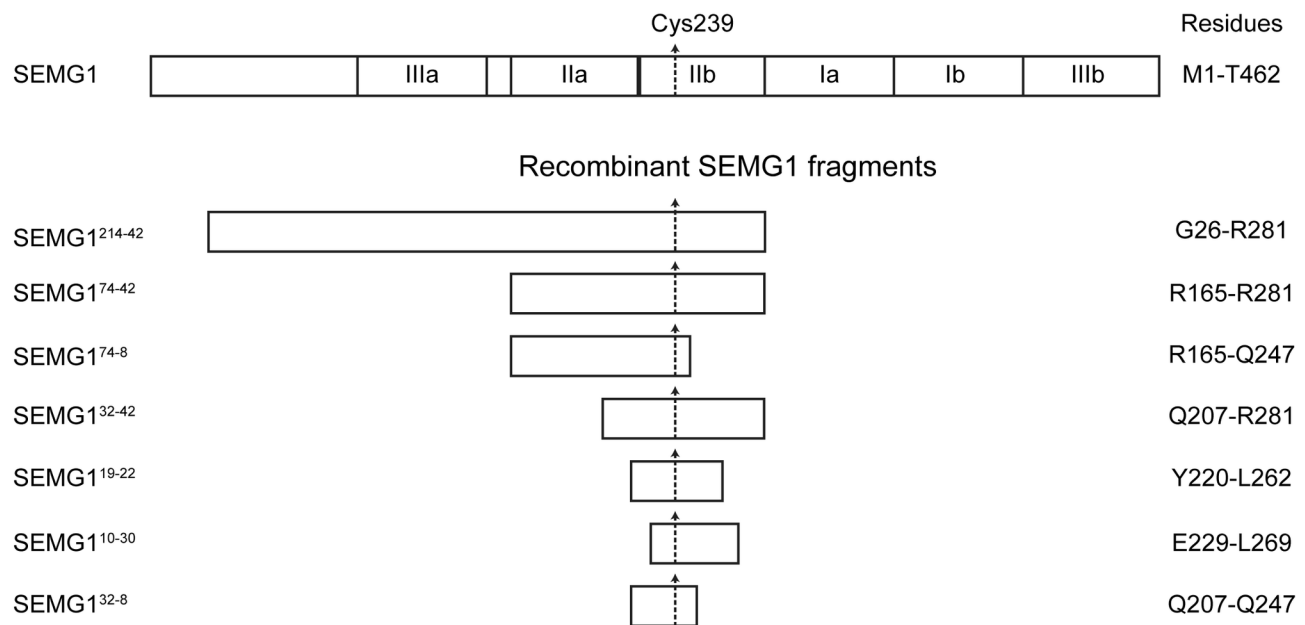


Figure 2. Human semenogelin I (SEMG1) sequence. **A)** SEMG1 primary sequence. Signal peptide is indicated by a dashed underline. **B)** Schematic representation of SEMG1 containing repeats Ia, Ib, IIa, IIb, IIIa, and IIIb. Recombinant SEMG1 fragments used in AlphaScreen assay and CASA experiments: G26-R281 (SEMG1²¹⁴⁻⁴²), R165-R281 (SEMG1⁷⁴⁻⁴²), R165-Q247 (SEMG1⁷⁴⁻⁸), Q207-R281 (SEMG1³²⁻⁴²), Y220-L262 (SEMG1¹⁹⁻²²), E229-L269 (SEMG1¹⁰⁻³⁰), and Q207-Q247 (SEMG1³²⁻⁸) are shown. Cysteine residue at position 239 is indicated by a box in (A) and a vertical arrow in (B).

doi: 10.1371/journal.pone.0082014.g002

~4-fold, ~8-fold and ~20-fold increase in the EC_{50} values of SEMG1/EPPIN binding when glycine/aspartic acid, histidine or serine/arginine residues replaced the Cys239 residue, respectively (Table 5). Altogether, we confirmed previous observations that the Cys239 residue within human SEMG1 primary sequence plays a critical role in the SEMG1/EPPIN interaction. Additionally, our results indicated that sequences within SEMG1 repeats IIIa, IIa and IIb flanking Cys239 residue

may provide interacting interfaces that significantly enhance SEMG1/EPPIN interaction.

Effect of SEMG1 fragments and Cys239 mutants on sperm motility

We have previously demonstrated that the recombinant SEMG1²¹⁴⁻⁴² fragment inhibited different sperm movement-related parameters, including motility, progressive motility and

Table 4. Amino acid residue changes in the SEMG1 sequence (positions 165-247).

SEMG1 Residues (position #)	Species						
	7	6	5	4	3	2	1
			^				
167	R	R	W	W	W	W	W*
169	R	R	H	H	H	H	H*
176	A	A	A	T	A	T	T
178	A	A	A	V	A	V	V
184	G	G	G	G	R	G	G*
186	T	T	T	T	K	K	K*
192	R	R	S	R	S	S	S*
193	N	N	S	S	S	S	S
194	Y	Y	Y	Y	H	Y	Y*
203	A	A	A	A	V	A	A*
204	N	K	N	N	N	N	N
206	Q	Q	R	Q	Q	Q	Q
208	R	R	G	R	R	R	R*
211	Q	Q	Q	K	K	K	K*
215	R	R	Q	Q	Q	Q	Q
225	D	D	E	E	D	E	E
234	L	L	V	V	L	V	V
239	R	R	C	C	H	C	C*†
242	H	H	Y	H	H	H	H
244	H	H	D	G	D	D	D
245	K	K	R	N	R	K	K*

^ indicates the divergence of the Catarrhini into the Hominoidea (species 1-5) and Cercopithecoidea (species 6 and 7), 31.6 MYA; Changes after the split are shown in bold/italics. Residues 225, 234 and 239 in the *Gorilla* are unique to *Gorilla* within the Hominoidea. † Residue 239 is necessary for high affinity binding of SEMG1 to human EPPIN. * Indicates BEB analysis for positively selected codons ($P > 95\%$). Species: 1, *Homo sapiens*; 2, *Pan troglodytes*; 3, *Gorilla gorilla*; 4, *Pongo abelii*; 5, *Nomascus leucogenys*; 6, *Macaca mulatta*; 7, *Papio anubis*.

doi: 10.1371/journal.pone.0082014.t004

VSL (straight line velocity) of human ejaculate spermatozoa in a time- and concentration-dependent manner [24]. To determine the minimum sequence within SEMG1²¹⁴⁻⁴² that inhibits sperm motility, we incubated spermatozoa with different recombinant SEMG1 fragments (Figure 2) and assessed by Computer-assisted sperm analysis (CASA). Based on observations that both sperm motility and velocity are directly correlated to the fertilizing ability of human spermatozoa [29] and that SEMG1 forms a gelatinous mass after ejaculation restricting sperm movement, we developed an index (iRMI) correlating both %motility and VSL parameters, which allowed the normalization of each data set against its respective control, reducing inter-assay variation due to differences in sperm quality in different semen samples.

As expected, SEMG1²¹⁴⁻⁴² at a concentration of 1.5 fmol/sperm (2 μ M) significantly decreased the normalized iRMI in comparison to control group (Figure 4A). Surprisingly, all SEMG1 fragments tested significantly affected sperm motility, including those that did not bind EPPIN in the AlphaScreen assay, as demonstrated by a decrease in the normalized iRMI equivalent to that observed with SEMG1²¹⁴⁻⁴² at the same

concentration (Figure 4A). To investigate the nature of these results, we performed concentration-response experiments in which spermatozoa were incubated with increasing concentrations of recombinant SEMG1²¹⁴⁻⁴², SEMG1¹⁷⁴⁻⁴² and SEMG1³²⁻⁸ fragments under similar conditions and the normalized iRMI was then calculated for each concentration point (Figure 4B). All fragments tested decreased the normalized iRMI in a concentration-dependent fashion (Figure 4B). The calculated EC₅₀ values (95% confidence interval) were 1.69 μ M (0.77 - 3.7 μ M), 3.56 μ M (1.67 - 7.63 μ M), and 1.75 μ M (0.50 - 6.06 μ M) for SEMG1²¹⁴⁻⁴², SEMG1¹⁷⁴⁻⁴² and SEMG1³²⁻⁸ fragments, respectively. Based on these results, we conclude that a region within the sequence E229-Q247 of SEMG1 contains a domain responsible for the inhibition of sperm motility.

CASA experiments performed with SEMG1²¹⁴⁻⁴² Cys239 mutants demonstrated that only C239H-SEMG1 mutant (His239 in the *Gorilla*, Table 4) significantly decreased the normalized iRMI in a similar extent to the wild-type at a concentration of 0.5 fmol/sperm (Figure 4C, left panel). When the concentration was increased to 1.5 fmol/sperm all the SEMG1 mutants, except C239G-SEMG1, significantly decreased the normalized iRMI in comparison to control (Figure 4C, right panel). To confirm the concentration-dependent sperm motility inhibition of the SEMG1 Cys239 mutants, we performed concentration-response experiments using C239H-SEMG1 and C239R-SEMG1 (Figure 4D). Both mutants decreased the normalized iRMI in a concentration-dependent manner (Figure 4D). The calculated EC₅₀ values (95% confidence interval) were 1.08 μ M (0.41 - 2.84 μ M) and 2.21 μ M (1.14 - 4.29 μ M) C239H-SEMG1 and C239R-SEMG1, respectively.

Discussion

Although previous reports indicate that different *WFDC* and *SEMG* genes within the *WFDC* locus are subject to rapid adaptive evolutionary changes [18,19,22,30], little is known about the molecular co-evolution of interacting proteins within this locus. The present study investigated the hypothesis that the interacting surface of SEMG1 and EPPIN co-evolved within the Hominoidea time scale, as a result of adaptive pressures applied by their roles in sperm protection and reproductive fitness. The binding of SEMG1 to EPPIN on the surface of spermatozoa is an important step for modulating sperm motility and providing antimicrobial protection upon ejaculation [5,7]. Our results indicate that distinct amino acid residues of SEMG1 (Trp167, His169, Ser193, and Gln215) and EPPIN (Lys35, Lys50, Asp51, Gln55, Lys81, Tyr89, Leu91, Met104, and Lys120) possess a remarkable deficiency of variation among hominoid primates. Using CodeML software in the PAML4 package [31] our analysis revealed that *SEMG1* codons within the fragment that binds to EPPIN underwent positive selection, confirming that there has been selective pressure on SEMG1's EPPIN binding domain. We confirmed that SEMG1's Cys239 residue within the SEMG1/EPPIN interacting interface is under positive selection. Although other factors such as premature stop codons and gene deletions or homogenizations may have

Silva et al. - Figure 3

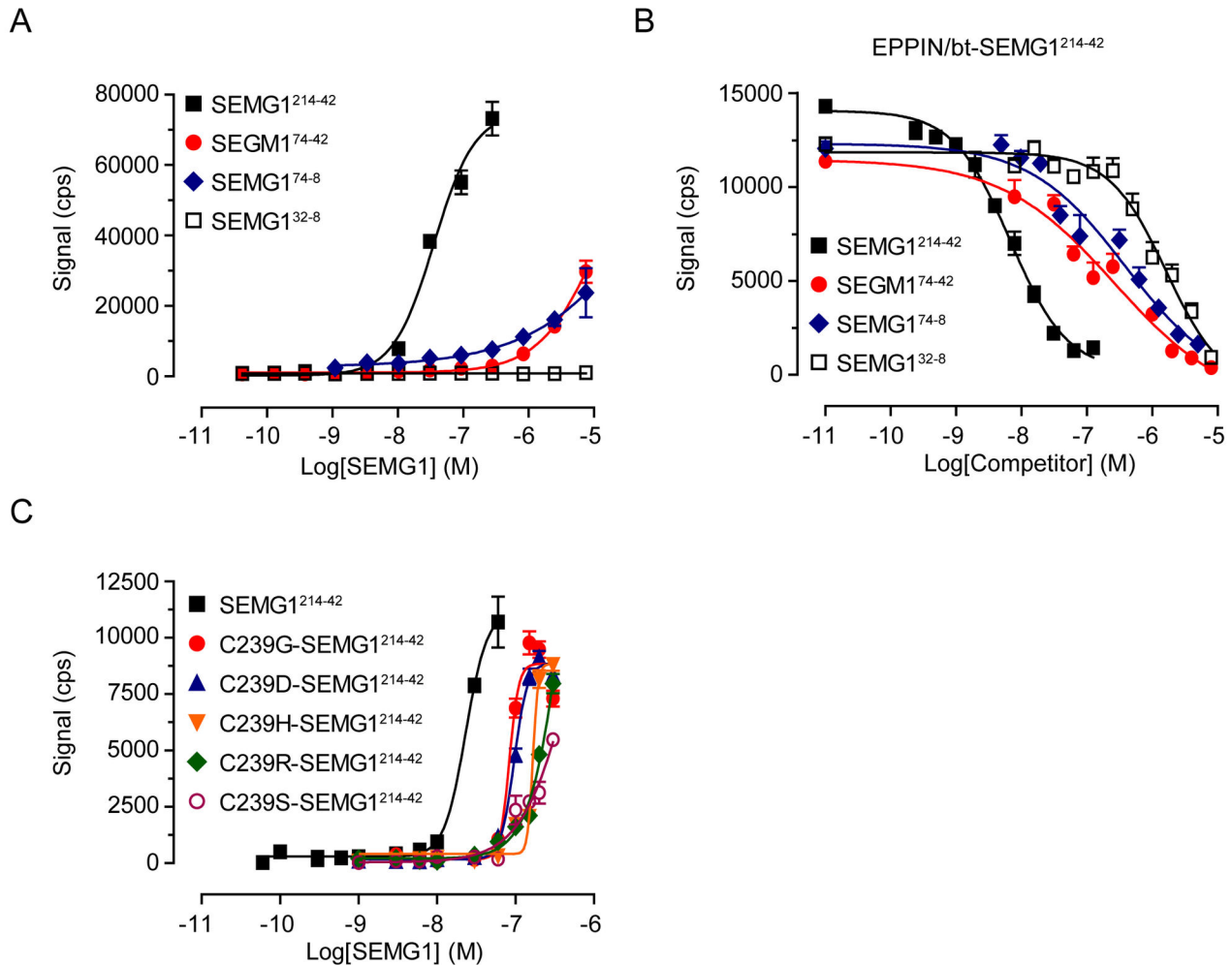


Figure 3. Binding of recombinant SEMG1 fragments to EPPIN in the AlphaScreen assay. **A)** Concentration-response curve for SEMG1²¹⁴⁻⁴², SEMG1⁷⁴⁻⁴², SEMG1⁷⁴⁻⁸, and SEMG1³²⁻⁸ in the presence of a constant concentration of EPPIN. **B)** Displacement of bt-SEMG1²¹⁴⁻⁴² from its binding site on EPPIN by SEMG1 constructs. Specific signal for each data point was normalized as percentage of specific signal in the absence of SEMG1 competitors. Data are expressed as mean \pm SD from two independent experiments, each performed in four replicates. **C)** Concentration-response curve for SEMG1²¹⁴⁻⁴² (wild-type) and Cys239 mutants (C239G, C239D, C239H, C239R, C239S) in the presence of a constant concentration of EPPIN. Data points in **(A)** and **(C)** represent mean \pm SD of specific signal from a representative experiment of **(A)** four (SEMG1²¹⁴⁻⁴² and SEMG1⁷⁴⁻⁴²) and two (SEMG1⁷⁴⁻⁸) experiments; and **(C)** three experiments, each performed in four replicates. cps = counts per second.

doi: 10.1371/journal.pone.0082014.g003

accounted for positive selection during the evolutionary history of SEMG1 [18], in humans Cys239's role in increasing the binding affinity of SEMG1 for EPPIN results in a functional gain in the modulation of sperm motility in the semen coagulum at lower concentrations. In fact, our binding and functional studies in vitro demonstrate that SEMG1 truncations flanking the Cys239 residue or point mutations at this position decrease the binding of SEMG1 to EPPIN and reduce its sperm motility inhibitory effect in a concentration-dependent fashion. Although none of EPPIN's codons underwent positive selection, which

may indicate the conserved stability of EPPIN itself, there were two distinct changes in humans compared to all other Hominoidea; residues His92 and Asp99. Altogether, our results provide unique insights into the co-evolution of binding partners on the surface of ejaculate spermatozoa.

Previously, we demonstrated that blocking human SEMG1's unique cysteine residue (Cys239) either by reduction and carboxymethylation or by a point mutation to a glycine blocked its capacity to bind EPPIN and to inhibit sperm motility [24,25], indicating that this residue played a major role in SEMG1

Table 5. EC₅₀ values for the binding of each indicated SEMG1 construct to EPPIN by AlphaScreen assay.

SEMG1 isoform	EC ₅₀ (95% conf. interval)
SEMG1 ²¹⁴⁻⁴²	23.1 nM (21.3 - 25.0 nM)
C239G-SEMG1 ²¹⁴⁻⁴²	83.3 nM (76.1 - 91.3 nM)
C239D-SEMG1 ²¹⁴⁻⁴²	96.1 nM (91.5 - 101.0 nM)
C239H-SEMG1 ²¹⁴⁻⁴²	167.0 nM (160.0 - 174.0 nM)
C239R-SEMG1 ²¹⁴⁻⁴²	461.0 nM (97.4 - 2180 nM)
C239S-SEMG1 ²¹⁴⁻⁴²	564.4 nM (8.3 - 3836 nM)

EC₅₀ calculations were performed with data showed in Figure 3C.

doi: 10.1371/journal.pone.0082014.t005

biological activities. In an effort to further characterize SEMG1 sequences responsible for binding EPPIN and inhibiting sperm motility we tested the capacity of multiple recombinant SEMG1 fragments and Cys239 mutants to bind EPPIN and to inhibit human sperm motility. It is noteworthy that all smaller SEMG1 constructs tested correspond to SEMG1 fragments generated by PSA digestion upon ejaculation [6], making them endogenous to the liquefied semen. Our results demonstrated that the whole G26-R281 sequence in SEMG1²¹⁴⁻⁴² was necessary for optimal SEMG1/EPPIN interaction, since this construct had the lowest EC₅₀ and IC₅₀ values among the SEMG1 fragments tested in concentration-response and competition experiments, respectively.

Interestingly, micromolar concentrations of the SEMG1³²⁻⁸ construct were able to inhibit the interaction between EPPIN and bt-SEMG1 in competition experiments, although it did not directly bind to EPPIN in concentration-response experiments. Besides the reduced capacity of this fragment to bind EPPIN at higher concentrations, it is possible that aggregation between SEMG1³²⁻⁸ and bt-SEMG1 may have masked the EPPIN binding site in bt-SEMG1 in our assay conditions, thereby leading to a concentration-dependent reduction in the signal. In addition, we cannot discard the possibility that the N-terminal 6X His tag present in the recombinant proteins may disturb the binding capacity of the smallest SEMG1 constructs to EPPIN due to steric/folding issues, which could lead to the absence of interaction in the direct binding assay at the concentrations tested.

Despite the importance of the Cys239 residue to SEMG1 biological activity, our phylogenetic analysis demonstrated that this residue is not highly conserved among primates; positively charged residues histidine or arginine replace the cysteine in SEMG1 from *Gorilla gorilla* and Cercopithecoidea species *Macaca mulatta* and *Papio anubis*, respectively. Based on these observations, we decided to further investigate the physiological significance of this change to SEMG1 function by mutating Cys239 (cysteine = polar) residue in human SEMG1²¹⁴⁻⁴² to histidine (basic), arginine (hydrophilic/basic), as well as glycine (hydrophobic), aspartic acid (negatively charged) and serine (hydrophilic) residues. Our results demonstrated that all SEMG1 mutants required much higher concentrations than the wild-type to bind to EPPIN, as indicated by higher EC₅₀ values in the concentration-response experiments. An interesting observation was the fact that the

C239H-SEMG1 mutant possessed a ~4-fold higher EPPIN binding activity than the C239R-SEMG1 mutant based on EC₅₀ values, suggesting that the size of the side-chain of the residue at this position may hamper SEMG1/EPPIN interaction by inducing steric effects. Taken together, our experimental data support the observation that human SEMG1's Cys239 residue is a hot site subject to positive selection, probably because of its major role in increasing the affinity of SEMG1 for EPPIN on the sperm surface, which can result in an effective gain of function by allowing an effective binding at lower concentrations. Our results further indicate that the sequence G26-E164, which is part of the repeat IIIa, provide a docking surface that is critical for the SEMG1-EPPIN interaction in addition to the Cys239 residue, leading us to hypothesize that SEMG1 has multiple EPPIN interaction sites within repeats IIIa and IIb of the molecule. This hypothesis is further supported by the observation that residues Cys102, Tyr107, and Phe117 in EPPIN's SEMG1 binding pocket are independently important for SEMG1 interaction [23].

We and others have demonstrated by computer analyses of sperm motility that both native and recombinant SEMG1 decreased motility parameters, such as percentage motility and straight line velocity of human ejaculate spermatozoa in a concentration- and time-dependent fashion [15,16,24,32]. Recent studies from our laboratory have begun to unravel the basic mechanisms underlying the SEMG1-mediated inhibition of sperm motility triggered by its interaction with EPPIN on the sperm surface, namely the decrease of intracellular pH leading to a drop in intracellular calcium levels by blocking extracellular calcium uptake [15,16]. These observations demonstrated that the effects of SEMG1 on sperm motility are finely regulated by a cascade of molecular events rather than by simply trapping spermatozoa by physical constraint in the semen coagulum. In our study, an intriguing observation was the fact that all smaller SEMG1 constructs tested inhibited sperm motility to a similar extent as SEMG1²¹⁴⁻⁴², including those that did not bind EPPIN in the AlphaScreen assay. Differences in the nature of these assays may explain our results: while the AlphaScreen assay measures the interaction between EPPIN and SEMG1 molecules in vitro with high sensitivity (nanomolar range), CASA analysis reflects a biological effect in a live cell, namely loss of sperm motility, which is dependent on a cascade of molecular events initiated by the same interaction, such as changes in intracellular pH and extracellular calcium uptake [15,16]. It is possible that the saturation of EPPIN binding sites with SEMG1 on the surface of spermatozoa, which would induce the formation of large macromolecular complexes, may be required for their immobilization.

It has been widely recognized that genes with roles in innate immunity and reproduction are evolving rapidly by positive Darwinian selection, particularly those associated with the male reproductive tract [33-35]. Consistently, the evolution of different sperm- and semen-associated proteins, such as *SEMG* and *WFDC* genes in the primate lineage, is rapid and driven by positive selective forces, especially in relation to sexual function and sperm competition [19]. In fact, the nearly significant heterogeneity between *Gorilla* and *Homo sapiens* *SEMG1* further suggests an ongoing divergent evolutionary

Silva et al. - Figure 4

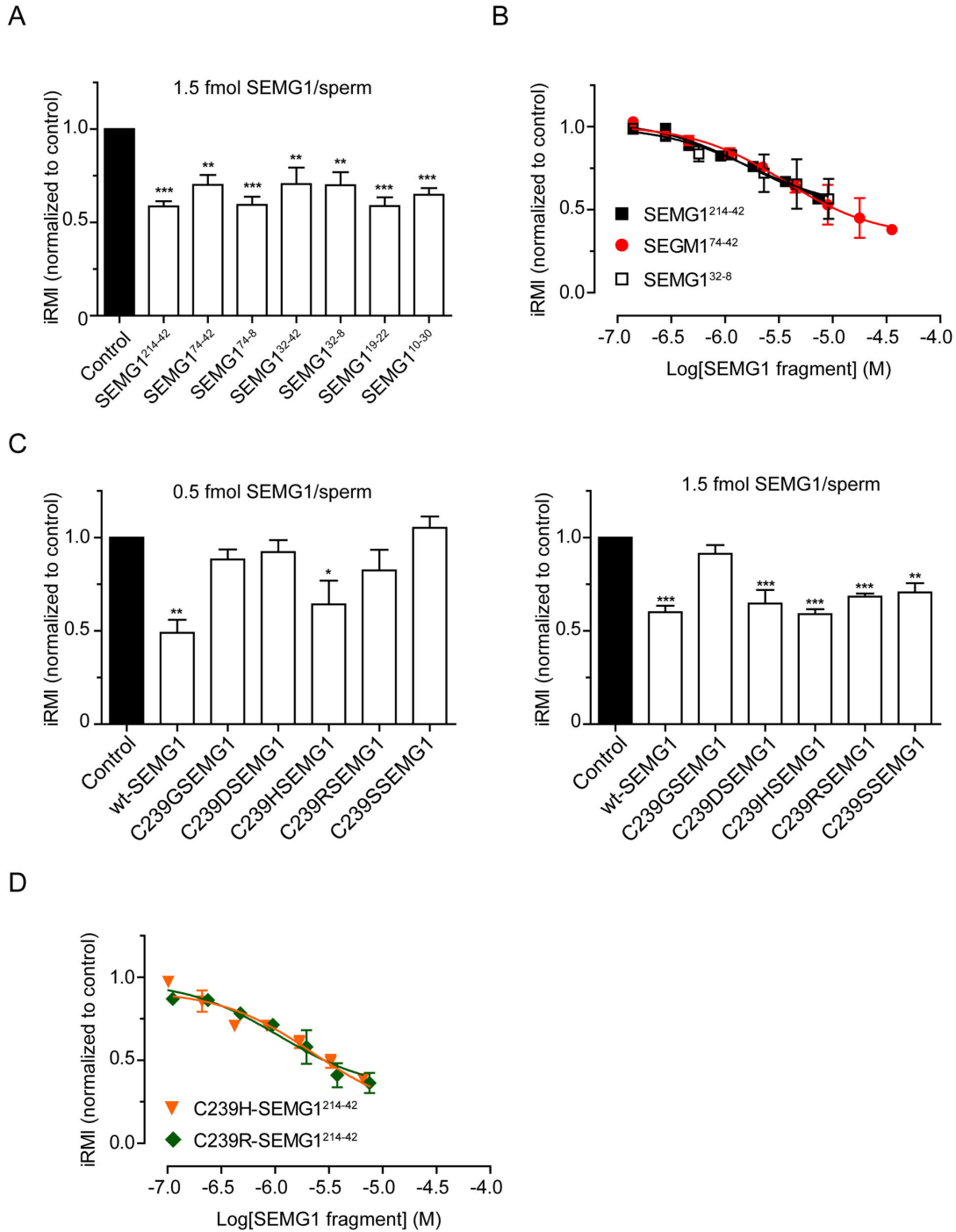


Figure 4. Effect of recombinant SEMG1 constructs on human sperm motility. Spermatozoa were incubated with the same concentration of recombinant SEMG1 fragments (A) or Cys239 mutants (C) and assessed by CASA. Concentration-response curves for spermatozoa incubated in the presence of increasing concentrations of SEMG1²¹⁴⁻⁴², SEMG1⁷⁴⁻⁴², and SEMG1³²⁻⁸ fragments (B) or C239H-SEMG1²¹⁴⁻⁴² and C239R-SEMG1²¹⁴⁻⁴² (D) were obtained. Results are presented as normalized iRMI (%motility*VSL of sample/%motility*VSL of control). Control untreated experiments were performed with buffer only. Data are representative of mean \pm SD of independent experiments performed with ejaculate spermatozoa from six (A) or three (B, C and D) donors. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$ when compared to control group (ANOVA followed by Bonferroni test).

doi: 10.1371/journal.pone.0082014.g004

Table 6. List of species and sequences used in the phylogenetic analysis of EPPIN and SEMG1.

Species	GenBank Accession Number	
	EPPIN	SEMG1
<i>Homo sapiens</i>	NM_020398.3	NM_003007.3
<i>Pan troglodytes</i>	XM_003316966.1	NM_001110538.1
<i>Gorilla gorilla gorilla</i>	XM_004062257.1	XM_004062223.1
<i>Pongo abelii</i>	XM_002830353.1	XM_002830350.1
<i>Nomascus leucogenys</i>	XM_003253642.1	gij328833490:8997382-9005052
<i>Macaca mulatta</i>	NM_001032841.1	XM_002798120.1
<i>Papio anubis</i>	XM_003904690.1	NM_001169022.1
<i>Saimiri boliviensis boliviensis</i>	XM_003936444.1	gb DP000047.1 :154056-155465

doi: 10.1371/journal.pone.0082014.t006

trend after speciation. Our results suggest that the adaptive molecular coevolution of the interacting surface between SEMG1 and EPPIN may have rapidly changed their biochemical properties resulting in an increase in their binding affinity and a gain of function for both proteins at the interface of innate immunity and reproduction.

Materials and Methods

Phylogenetic analysis

We aligned nucleotide sequences of *EPPIN* and *SEMG1* from *Homo sapiens*, *Pan troglodytes*, *Gorilla gorilla gorilla*, *Pongo abelii*, *Nomascus leucogenys*, *Macaca mulatta*, *Papio anubis*, and *Saimiri boliviensis boliviensis* (GenBank accession numbers in Table 6) with ClustalW. We used MEGA5 molecular evolutionary genetic analysis software package (version 5.1) [36] to obtain phylogenetic trees using the Maximum Likelihood method based on the Tamura-Nei model [37] and to perform the test of homogeneity of substitution patterns between sequences for *EPPIN* and *SEMG1* using the neighbor-joining method [38]. Codon positions included were 1st+2nd+3rd. All positions containing gaps and missing data were eliminated. For technical details of these tests, please refer to their original descriptions [36–38]. We chose the phylogenetic trees with the highest log likelihood (*EPPIN*) -926.2396 or (*SEMG1*) -2108.0861 and used them for further analysis. We investigated evidence of positive selection codon-by-codon in the primate phylogenetic tree of *EPPIN* and *SEMG1* using the CodeML software in the PAML 4 package [27,31]. In this case, we compared the likelihood of a nearly-neutral model M1a with that of a selection model M2a. Next, we compared the likelihood of a neutral model M7 with that of a selection model M8. In both cases we used a Likelihood Ratio Test (LRT) to determine if the selection model was a better fit to the data than the neutral model by comparing the $(-2[\text{Log}(\text{neutral}) - \text{Log}(\text{positive})])$ value between the two models with the χ^2 -distribution [18,27]. When the calculations suggested positive selection, we used the BEB method to calculate the probability that each codon experienced positive selection under models M2a and M8 [27].

Production and Expression of Recombinant Proteins

Human recombinant SEMG1 (SEMG1²¹⁴⁻⁴², wild-type, residues G26-R281, Figure 2) and EPPIN (residues P22-P133) were previously cloned into pET100/D-TOPO vector (Invitrogen, Carlsbad, CA) [23]. Additionally, C-terminal and N-terminal truncated fragments flanking the residue Cys239 from the SEMG1 primary sequence, corresponding to residues R165-R281 (SEMG1⁷⁴⁻⁴²), R165-Q247 (SEMG1⁷⁴⁻⁸), Q207-R281 (SEMG1³²⁻⁴²), Y220-L262 (SEMG1¹⁹⁻²²), E229-L269 (SEMG1¹⁰⁻³⁰), and Q207-Q247 (SEMG1³²⁻⁸) (Figure 2B) were produced and cloned into the same vector. All constructs were verified by DNA sequencing. Recombinant proteins were expressed either in BL21 Star™ (DE3) One Shot® *E. coli* (Invitrogen; SEMG1 constructs) or *E. coli* Rosetta-Gami 2™ (DE3) (Novagen, Madison, WI; EPPIN construct) and purified under denaturing conditions using nickel-nitrilotriacetic acid agarose (Ni-NTA) beads (Qiagen, Valencia, CA) as described previously [23].

Removal of N-terminal 6X His-tag from recombinant EPPIN was performed using recombinant enterokinase cleavage/capture kit (Novagen), according to the manufacturer's instructions. For removal of undigested 6X His-tag-EPPIN, the cleavage reaction was incubated with Ni-NTA beads at room temperature for 30 min with gentle mixing and tag-free EPPIN was then eluted in PBS buffer containing 1 M urea. Recombinant SEMG1²¹⁴⁻⁴² was biotinylated as described previously [23].

In Vitro Site-Directed Mutagenesis

Site-directed mutagenesis of wild-type SEMG1²¹⁴⁻⁴² of residue Cys239 to glycine (C239G-SEMG1), aspartic acid (C239D-SEMG1), histidine (C239H-SEMG1), serine (C239S-SEMG1), or arginine (C239R-SEMG1) was performed using the Gene Tailor site-directed mutagenesis system (Invitrogen), according to the manufacturer's instructions. Mutagenesis products were transformed into DH5a-T^{1R} *E. coli* and positive clones selected, followed by DNA sequencing to confirm the mutation. Recombinant SEMG1 mutants were expressed and purified as described above.

AlphaScreen® Assay

The AlphaScreen® assay was carried as previously described [23], with the following modifications. Briefly, recombinant EPPIN was pre-incubated with anti-EPPIN Q20E antibody and Protein A acceptor beads for 30 minutes. In parallel, increasing concentrations of each recombinant SEMG1 construct was incubated with Ni-NTA-chelate donor beads under the same conditions. Equal volumes of each EPPIN/Q20E/Protein A acceptor beads and SEMG1/Ni-NTA-chelate donor beads were pipetted into in white opaque 384-well microplates (OptiPlate-384; PerkinElmer, Waltham, MA) in a final volume of 30 μ l. Plates were covered with top seal and transferred to a Synergy 2 Multiplex automated plate reader (Biotek, Winooski, VT). After shaking for 2 min, plates were read every 2 h during 16 h: excitation using a 680/30 filter, emission using a 570/100 filter and data acquired using Gen5 software (Biotek). A total of 9 time points were generated during each experiment. Each set of samples was pipetted in 4

replicates. The final concentration of assay components was 29 nM EPPIN, 2 nM Q20E antibody, 0.04-7.5 μ M SEMG1, and 10 μ g/ml beads. Negative controls were performed under the same conditions in the absence of EPPIN or SEMG1 and in the presence of beads only. A specific signal for each time point was calculated by subtracting the background signal (obtained in the absence of SEMG1) from its respective total signal. EC₅₀ values for the binding of EPPIN to each SEMG1 construct were calculated by non-linear regression curve fitting using the specific signal obtained after 16 h of incubation.

Competition experiments using AlphaScreen assay platform were performed as previously described [23]. Briefly, EPPIN (10 nM) and biotinylated (bt-)SEMG1²¹⁴⁻⁴² (1 nM) were incubated in the presence of increasing concentrations of SEMG1²¹⁴⁻⁴², SEMG1⁷⁴⁻⁴², SEMG1⁷⁴⁻⁸, or SEMG1³²⁻⁸ (10 pM – 8 μ M) in a 30 μ l reaction. The bead concentration was 15 μ g/ml. The specific signal for each competitor concentration point was calculated as described above and the IC₅₀ values were calculated by nonlinear regression curve fitting.

Preparation of spermatozoa and analysis of sperm motility

Human semen samples were obtained from the Department of Obstetrics and Gynecology, University of North Carolina Memorial Hospital, Chapel Hill, NC. This study was approved by the Committee on the Protection of the Rights of Human Subjects at the University of North Carolina, School of Medicine Chapel Hill, NC. The Institutional Review Board determined that this study did not constitute human subjects research as defined under U.S. federal regulations [45 CFR 46.102 (d or f) and 21 CFR 56.102(c)(e)(I)] and the need for informed consent was waived. All samples were de-identified before further processing. Semen samples were allowed to liquefy for 30 min and subjected to standard semen analysis to determine acceptability and either used fresh or stored in liquid nitrogen. An isolate gradient (Irving Scientific, Irving, CA, cat. #99264) was used to prepare spermatozoa for further analysis [15].

Aliquots of sperm suspension (1500 cells/ μ l) were incubated with increasing concentrations of each SEMG1 construct (0 - 35.6 μ M; diluted in modified M16 medium no phenol red (Millipore, Billerica, MA, cat. #MR/010/D) supplemented with 50 μ M ZnCl₂ in 12- X 75-mm glass tubes at 37°C and 5% CO₂ for 1-2 h. We repeated each experiment with spermatozoa from at least three different ejaculates and evaluated at least 100 cells (range 100-1000 cells) from each data point.

The analysis of sperm motility after incubation with SEMG1 constructs was carried out using computer-assisted sperm analysis (CASA; Ceros version 14.8 software, Hamilton-Thorne), as previously described [24]. In each case, the percentage sperm motility (%motility) and straight-line velocity

(VSL; the average velocity measured in a straight line from the beginning to the end of track in μ m/sec) were used to calculate the index of relative motility inhibition (iRMI) as: %motility*VSL. Normalized iRMI was calculated by dividing the iRMI of each experimental condition by its respective control (absence of SEMG1).

Statistical analysis

Results from AlphaScreen assays and CASA experiments were expressed as mean \pm standard deviation (SD) or standard error of the mean (SEM) from the indicated number of independent experiments. For statistical significance, one-way analysis of variance (ANOVA) followed by the Bonferroni test was performed using GraphPad Prism 5.0 software. $p < 0.05$ was considered statistically significant.

Supporting Information

Table S1. Likelihood ratio test between different evolutionary models for EPPIN (A) and SEMG1 (B). (PDF)

Table S2. Log-likelihood values and parameter estimates for SEMG1. (PDF)

Figure S1. Characterization of the interaction between EPPIN and his-SEMG1 (SEMG1²¹⁴⁻⁴² fragment) in the AlphaScreen Assay. A) Time-course experiment showing the interaction between SEMG1²¹⁴⁻⁴² and EPPIN. Background signal was detected when beads were incubated in the absence of SEMG1²¹⁴⁻⁴². B) Concentration-response curve for SEMG1²¹⁴⁻⁴² in the presence of a constant concentration of EPPIN. A reduction in the signal (hook effect, arrow) was observed with his-SEMG1 concentrations higher than 300 nM. Negative control was performed in the absence of EPPIN. Specific signal for each data point was determined by subtracting the background signal from total signal. Data points represent mean \pm SD of specific signal from a representative experiment of four experiments, each performed in four replicates. cps = counts per second. (PDF)

Author Contributions

Conceived and designed the experiments: EJRS KGH MGO. Performed the experiments: EJRS KGH MGO. Analyzed the data: EJRS KGH MGO. Contributed reagents/materials/analysis tools: EJRS KGH MGO. Wrote the manuscript: EJRS MGO.

References

- Juyena NS, Stelletta C (2012) Seminal plasma: An essential attribute to spermatozoa. *J Androl* 33: 536-551. doi:10.2164/jandrol.110.012583. PubMed: 22016346.
- Robert M, Gagnon C (1999) Semenogelin I: a coagulum forming, multifunctional seminal vesicle protein. *Cell Mol Life Sci* 55: 944-960. doi:10.1007/s000180050346. PubMed: 10412373.
- Malm J, Hellman J, Magnusson H, Laurell C-B, Lilja H (1996) Isolation and characterization of the major gel proteins in human semen, semenogelin I and semenogelin II. *Eur J Biochem* 238: 48-53. doi:10.1111/j.1432-1033.1996.0048q.x. PubMed: 8665951.
- Richardson RT, Sivashanmugam P, Hall SH, Hamil KG, Moore PA et al. (2001) Cloning and sequencing of human Eppin: a novel family of protease inhibitors expressed in the epididymis and testis. *Gene* 270: 93-102. doi:10.1016/S0378-1119(01)00462-0. PubMed: 11404006.
- O'Rand MG, Widgren EE, Hamil KG, Silva EJ, Richardson RT (2011) Epididymal protein targets: A brief history of the development of epididymal protease inhibitor as a contraceptive. *J Androl* 32: 698-704. doi:10.2164/jandrol.110.012781. PubMed: 21441428.
- Robert M, Gibbs BF, Jacobson E, Gagnon C (1997) Characterization of prostate-specific antigen proteolytic activity on its major physiological substrate, the sperm motility inhibitor precursor/semenogelin I. *Biochemistry* 36: 3811-3819. doi:10.1021/bi9626158. PubMed: 9092810.
- O'Rand MG, Widgren EE, Wang Z, Richardson RT (2006) Eppin: An effective target for male contraception. *Mol Cell Endocrinol* 250: 157-162. doi:10.1016/j.mce.2005.12.039. PubMed: 16423450.
- Wang Z-J, Zhang W, Feng N-H, Song N-H, Wu H-F et al. (2008) Molecular mechanism of epididymal protease inhibitor modulating the liquefaction of human semen. *Asian J Androl* 10: 770-775. doi:10.1111/j.1745-7262.2008.00393.x. PubMed: 18645681.
- Robert M, Gagnon C (1995) Sperm motility inhibitor from human seminal plasma: association with semen coagulum. *Hum Reprod* 10: 2192-2197. PubMed: 8567872.
- Bourgeon F, Evrard B, Brillard-Bourdet M, Collet D, Jégou B et al. (2004) Involvement of semenogelin-derived peptides in the antibacterial activity of human seminal plasma. *Biol Reprod* 70: 768-774. PubMed: 14613901.
- Edström AML, Malm J, Frohm B, Martellini JA, Giwercman A et al. (2008) The major bactericidal activity of human seminal plasma is zinc-dependent and derived from fragmentation of the semenogelins. *J Immunol* 181: 3413-3421. PubMed: 18714013.
- Yenugu S, Richardson RT, Sivashanmugam P, Wang Z, O'Rand MG et al. (2004) Antimicrobial activity of human EPPIN, an androgen-regulated, sperm-bound protein with a whey acidic protein motif. *Biol Reprod* 71: 1484-1490. doi:10.1095/biolreprod.104.031567. PubMed: 15229136.
- McCrudden MT, Dafforn TR, Houston DF, Turkington PT, Timson DJ (2008) Functional domains of the human epididymal protease inhibitor, eppin. *FEBS J* 275: 1742-1750. doi:10.1111/j.1742-4658.2008.06333.x. PubMed: 18331357.
- O'Rand MG, Widgren EE, Sivashanmugam P, Richardson RT, Hall SH et al. (2004) Reversible immunocontraception in male monkeys immunized with Eppin. *Science* 306: 1189-1190. doi:10.1126/science.1099743. PubMed: 15539605.
- O'Rand MG, Widgren EE (2012) Loss of calcium in human spermatozoa via EPPIN, the semenogelin receptor. *Biol Reprod* 86: 55: 51-57. PubMed: 22075473.
- O'Rand MG, Widgren EE, Beyler S, Richardson RT (2009) Inhibition of Human Sperm Motility by Contraceptive Anti-Eppin Antibodies from Infertile Male Monkeys: Effect on Cyclic Adenosine Monophosphate. *Biol Reprod* 80: 279-285. doi:10.1095/biolreprod.108.072942. PubMed: 18945989.
- Clauss A, Lilja H, Lundwall A (2002) A locus on human chromosome 20 contains several genes expressing protease inhibitor domains with homology to whey acidic protein. *Biochem J* 368: 233-242. doi:10.1042/BJ20020869. PubMed: 12206714.
- Hurle B, Swanson W, Program NCS, Green ED (2007) Comparative sequence analyses reveal rapid and divergent evolutionary changes of the WFDC locus in the primate lineage. *Genome Res* 17: 276-286. doi:10.1101/gr.6004607. PubMed: 17267810.
- Ferreira Z, Seixas S, Andrés AM, Kretzschmar WW, Mullikin JC et al. (2013) Reproduction and immunity-driven natural selection in the human WFDC locus. *Mol Biol Evol* 30: 938-950. doi:10.1093/molbev/mss329. PubMed: 23292442.
- Carnahan SJ, Jensen-Seaman MI (2008) Hominoid seminal protein evolution and ancestral mating behavior. *Am J Primatol* 70: 939-948. doi:10.1002/ajp.20585. PubMed: 18561295.
- Jensen-Seaman MI, Li W-H (2003) Evolution of the hominoid semenogelin genes, the major proteins of ejaculated semen. *J Mol Evol* 57: 261-270. doi:10.1007/s00239-003-2474-x. PubMed: 14629036.
- Clauss A, Lilja H, Lundwall A (2005) The evolution of a genetic locus encoding small serine proteinase inhibitors. *Biochem Biophys Res Commun* 333: 383-389. doi:10.1016/j.bbrc.2005.05.125. PubMed: 15950183.
- Silva EJ, Hamil KG, Richardson RT, O'Rand MG (2012) Characterization of EPPIN's semenogelin I binding site: A contraceptive drug target. *Biol Reprod* 87: 56: 51-58. PubMed: 22699487.
- Mitra A, Richardson RT, O'Rand MG (2010) Analysis of recombinant human semenogelin as an inhibitor of human sperm motility. *Biol Reprod* 82: 489-496. doi:10.1095/biolreprod.109.081331. PubMed: 19889947.
- Wang Z, Widgren EE, Sivashanmugam P, O'Rand MG, Richardson RT (2005) Association of Eppin with semenogelin on human spermatozoa. *Biol Reprod* 72: 1064-1070. doi:10.1095/biolreprod.104.036483. PubMed: 15590901.
- Perelman P, Johnson WE, Roos C, Seuáñez HN, Horvath JE et al. (2011) A Molecular phylogeny of living primates. *PLOS Genet* 7: e1001342.
- Yang Z, Wong WSW, Nielsen R (2005) Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol Biol Evol* 22: 1107-1118. doi:10.1093/molbev/msi097. PubMed: 15689528.
- Quinn AM, Bedford MT, Espejo A, Spannhoff A, Austin CP et al. (2010) A homogeneous method for investigation of methylation-dependent protein-protein interactions in epigenetics. *Nucleic Acids Res* 38: e11. doi:10.1093/nar/gkp899. PubMed: 19897549.
- Hirano Y, Shibahara H, Obara H, Suzuki T, Takamizawa S et al. (2001) Relationships between sperm motility characteristics assessed by the computer-aided sperm analysis (CASA) and fertilization rates in vitro. *J Assist Reprod Genet* 18: 215-220. doi:10.1023/A:1009420432234. PubMed: 11432113.
- Dorus S, Evans PD, Wyckoff GJ, Choi SS, Lahn BT (2004) Rate of molecular evolution of the seminal protein gene SEMG2 correlates with levels of female promiscuity. *Nat Genet* 36: 1326-1329. doi:10.1038/ng1471. PubMed: 15531881.
- Yang Z (2007) PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 24: 1586-1591. doi:10.1093/molbev/msm088. PubMed: 17483113.
- Yoshida K, Kawano N, Yoshiike M, Yoshida M, Iwamoto T et al. (2008) Physiological roles of semenogelin I and zinc in sperm motility and semen coagulation on ejaculation in humans. *Mol Hum Reprod* 14: 151-156. doi:10.1093/molehr/gan003. PubMed: 18203809.
- Torgerson DG, Kulathinal RJ, Singh RS (2002) Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Mol Biol Evol* 19: 1973-1980. doi:10.1093/oxfordjournals.molbev.a004021. PubMed: 12411606.
- Torgerson DG, Singh RS (2003) Sex-Linked Mammalian Sperm Proteins Evolve Faster Than Autosomal Ones. *Mol Biol Evol* 20: 1705-1709. doi:10.1093/molbev/msg193. PubMed: 12832636.
- Wyckoff GJ, Wang W, Wu C-I (2000) Rapid evolution of male reproductive genes in the descent of man. *Nature* 403: 304-309. doi:10.1038/35002070. PubMed: 10659848.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739. doi:10.1093/molbev/msr121. PubMed: 21546353.
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10: 512-526. PubMed: 8336541.
- Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* 101: 11030-11035. doi:10.1073/pnas.0404206101. PubMed: 15258291.