Entropy-Based Analysis of Vertebrate Sperm Protamine Sequences: Evidence of Dityrosine and Cysteine-Tyrosine Cross-Linking in Sperm Protamines

Christian D. Powell[1,3], Daniel C. Kirchhoff[1], Jason E. DeRouchey[1], and Hunter N.B. Moseley[2,3,4,*]

[1]Department of Chemistry
[2]Department of Molecular & Cellular Biochemistry
[3]Markey Cancer Center
[4]Institute for Biomedical Informatics
University of Kentucky, Lexington, KY 40506, United States
[*]Corresponding author: hunter.moseley@uky.edu

Abstract

Background

Spermatogenesis is the process by which germ cells develop into spermatozoa in the testis. Sperm protamines are small, arginine-rich nuclear proteins which replace somatic histones during spermatogenesis, allowing a hypercondensed DNA state that leads to a smaller nucleus and facilitating sperm head formation. In eutherian mammals, the protamine-DNA complex is achieved through a combination of intra- and intermolecular cysteine cross-linking and possibly histidine-cysteine zinc ion binding. Most metatherian sperm protamines lack cysteine but perform the same function. This lack of dicysteine cross-linking has made the mechanism behind metatherian protamines folding unclear.

Results

Protamine sequences from UniProt's databases were pulled down and sorted into homologous groups. Multiple sequence alignments were then generated and a gap weighted relative entropy score calculated for each position. For the eutherian alignments, the cysteine containing positions were the most highly conserved. For the metatherian alignment, the tyrosine containing positions were the most highly conserved and corresponded to the cysteine positions in the eutherian alignment.

Conclusions

High conservation indicates likely functionally/structurally important residues at these positions in the metatherian protamines and the correspondence with cysteine positions within the eutherian alignment implies a similarity in function. One explanation is that the metatherian protamine structure relies upon dityrosine cross-linking between these highly conserved tyrosines. Also, the human protamine P1 sequence has a tyrosine substitution in a position expecting eutherian dicysteine cross-linking. Similarly, some members of the metatherian Planigales genus contain cysteine substitutions in positions expecting plausible metatherian dityrosine cross-linking. Rare cysteine-tyrosine cross-linking could explain both observations.

Background

The process in which male germ cells develop into sperm cells is called spermatogenesis. During spermatogenesis, DNA undergoes hypercondensation in order to form a smaller nucleus. This is accomplished through the replacement of a vast majority of somatic DNA histones (>90%) with one of three nuclear proteins; sperm-specific histones, protamine-like proteins, or protamines [Balhorn 2007]. In mammals, sperm protamines are small (<60 amino acids), arginine-rich nuclear proteins. Hypercondensation of DNA mediated by protamines result in haploid male germ cell nuclei, which are genetically inactive but just 1/20th the size of a somatic cell nucleus [Steger & Balhorn 2018]. This reorganization of the spermatozoa DNA is also thought to protect the paternal genome against oxidative damage [Balhorn 2007; Bennetts & Aitken 2005; Villani et al. 2010; Enciso et al. 2011].

Genetically the family of sperm protamines is highly diverse, being observed across the tree of life. For example, a single species of fish can contain multiple genes for protamine and protamine-like proteins, whereas birds tend to have two identical copies of a single protamine gene [Balhorn 2007]. While all protamines perform the task of binding and condensing DNA, the sizes and structural components of the protamines can vary greatly from species to species. Despite these differences most all sperm protamines include large arginine-rich DNA binding regions and phosphorylation sites [Balhorn 2007; Biegeleisen 2006; Queralt et al. 1995]. The positively charged arginine residues in these binding regions are able to engage in an electrostatic interaction with the negatively charged DNA phosphate backbone [Biegeleisen 2006]. These interactions form a toroid shaped protamine-DNA complex conforming to an internal hexagonal lattice [Brewer 2011]. The various phosphorylation sites in the protamine sequences are involved in a number of post-translational modifications and are thought to regulate the interactions with DNA.

Some of the simplest protamines are those of fish. Like most sperm protamines, the sequences of fish protamines are populated with a large percentage of arginine residues. However, fish protamines tend to be under 35 amino acids in length and contain increased frequencies of arginine (approximately 70%) in comparison to their mammalian analogs. The secondary structure of the protamines consist of multiple beta turns, with limited CD, NMR, and fluorescence data, indicating the formation of a possible globular structure [Arellano et al. 1988; Cid & Arellano 1982].

Mammals have a relatively conserved set of protamines, with metatherian mammals having only one protamine gene, while eutherian mammals have two to three varieties. These mammalian sequences tend to start with MARYR at the N-terminus, typically followed by a region containing a phosphorylation site (or multiple phosphorylation sites in eutherian sperm protamine P2s), then a DNA binding region comprised of multiple blocks of arginine residues, and ending with a varied C-terminal region [Queralt et al. 1995].

The eutherian protamine P1 is encoded by the PRM1 gene. Alignment of the sequences of eutherian mammal sperm protamine P1 have shown the sequences to be relatively conserved. In eutherian sperm protamine P1 sequences, the arginine-rich DNA binding regions are broken up

by cysteine residues, which are involved in both inter- and intramolecular disulfide cross-linkings [Balhorn et al. 1991; Queralt et al. 1995; Vilfan et al. 2004]. In bull protamine P1, the intra-protamine disulfide bonds were shown to create a hairpin-like structure, with disulfide crosslinks formed between the cysteines in positions 7 and 15 as well as the cysteines at positions 40 and 48. The remaining cysteine positions in bull protamine P1 are involved in inter-protamine bonding. More recently, we have shown that this disulfide mediated secondary structure of the bull protamine is required for proper chromatin remodeling [Hutchison et al. 2017; Kirchhoff et al. 2019].

The other two eutherian sperm protamine types are encoded in the PRM2 gene. These other protamine proteins are longer than the eutherian sperm protamine P1 type protamine and include a number of post-translational truncation sites in the N-terminal tail [Balhorn 2007]. Unlike the eutherian P1 type sperm protamines, the P2 protamines engage in zinc ion binding that is stochiometrically 1:1 for many eutherian mammals [Bench et al. 2000]. This zinc ion binding is achieved with highly conserved cysteine and histidine residues in the P2 protamine sequence. Both eutherian P1 and P2 type protamines engage in intermolecular disulfide cross-linking with one another when forming the DNA protamine complex [Vilfan et al. 2004]. For all eutherian sperm protamine types, it is thought that the cysteine cross-linkages are important for protecting the spermatozoa from oxidative damage [Bennetts & Aitken 2005; Villani et al. 2010; Enciso et al. 2011].

Also in eutherian mammals, a testis-specific variant of glutathione peroxidase (GPx4) is involved in the formation of the thiol cross-linking between and within the protamines and with protecting the sperm cells from oxidative stress due to reactive oxygen species [Pfeifer et al. 2001; Conrad et al. 2005]. In particular, Conrad et al. showed that without GPx4, sperm develop abnormal heads likely due to a lack of stabilizing disulfide cross-linking [Conrad et al. 2005].

In contrast to eutherian sperm protamines, little is known about metatherian sperm protamines, except that metatherian sperm protamines tend to lack cysteine residues with the only exception to this tendency involving species of the *Planigale* genus [Retief et al. 1995]. To our knowledge, there is no consensus on the structure of metatherian sperm protamines, nor is there prior evidence to suggest that inter- and intramolecular cross-linking occurs in metatherian sperm protamines, with the exception of species of the *Planigale* genus where it was suggested [Retief et al. 1995]. Additionally, it is unclear if GPx4 is required for the proper function of metatherian sperm protamines, although it is known that metatherian mammals do express glutathione peroxidase for defense against oxidative stress [Whittington et al. 1995]. Sequence data also exists for the testes specific version of glutathione peroxidase in Tasmanian Devils (G3WAH0_SARHA) [UniProt 2018]. Metatherian spermatozoa are more susceptible to oxidative damage, likely due to a lack of stabilizing disulfide cross-linkages [Bennetts & Aitken 2005; Villani et al. 2010; Enciso et al. 2011]. These current gaps in knowledge prompted the following analyses of multiple sequence alignments (MSAs) of eutherian P1, eutherian P2, metatherian P1, and fish sperm protamine sequences, which provide some insight into the structures of protamines and mechanism behind protamine mediated DNA condensation.

Results

MSAs were generated for 145 eutherian sperm protamine P1, 16 eutherian sperm protamine P2, 95 metatherian sperm protamine, and 34 fish protamine sequences, all retrieved from the UniProt knowledgebase and aligned using MUSCLE 3.8.31 [UniProt 2018; Edgar 2004]. The MSAs were then analyzed using the relative entropy method described.

Fish Protamine

From the fish protamine MSA, the relative entropy-based analysis showed that no position in the alignment had conservation scores above the relative entropy threshold. The most highly conserved positions were those containing only arginine residues. There was in fact a four-way tie for the most highly conserved position with positions 15, 16, 17, and 27 all having a conservation score equal to the conservation threshold of 4.135.

```
                        10         20         30
PRTA_ACIST       -ARRRRRHAS TKLKRRR--- -----RRRRH GKKSHK
PRT_ORYLA        ----MRRQAS LPARRRRRVR RTRVVRRRRR VGRRRH
PRT_PERFV        --PRRRRHAA RPVRRRRRTR RSSRVHRRRR AVRRRR
PRTB_MUGCE       --PRRRRETS RPIRRRRRAR RAPI-RRRRR VVRRRR
PRT1_SCOSC       -MPRRRRRAS RPVRRRRRAR RSTAVRRRRR VVRRRR
PRT2_SCOSC       -MPRRRRRAS RPIRRRRRAR RSTAVRRRRR VVRRRR
PRT_DICLA        --PRRRRQAS RPVRRRRRTR RSTAERRRRR VVRRRR
PRTY_THUTH       --PRRRRQAS RPVRRRRRYR RSTAARRRRR VVRRRR
PRTZ_THUTH       --PRRRRRSS RPVRRRRRYR RSTVARRRRR VVRRRR
PRTZ1_SAROR      --PRRRRRSS RPVRRRRRYR RSTAARRRRR VVRRRR
...
```

Figure 1. Alignment showing a selection of 10 fish protamine amino acid sequences from the alignment of fish protamine sequences. The full multiple sequence alignment used 34 sequences from the 2019_05 release of the UniProt knowledgebase and was aligned with MUSCLE 3.8.31. For full alignment, see Supplemental Figure 1.

Eutherian Sperm Protamine P1

For the eutherian sperm protamine P1 MSA, a total of nine positions were determined to be highly conserved, based on relative entropy scores and conservation threshold described above. All nine highly conserved positions within the alignments were positions comprised primarily of cysteine residues. In descending conservation score, the most highly conserved positions were positions 7, 49, 50, 60, 38, 17, 29, 6, and 37. All highly conserved positions within the alignment were composed of over 69% cysteine residues (excluding gaps) and the most highly conserved

positions tended to all be within one residue of a known intramolecular cross-linking region (positions 7, 49, 50, 60, 17, and 6) [Balhorn et al. 1991].

```
                    10         20         30         40         50         60
HSP1_CAPHI      MARYRCCLTH --SRSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVVRC TRQ-
HSP1_SHEEP      MARYRCCLTH --SRSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVVRC TRQ-
HSP1_BOVIN      MARYRCCLTH --SGSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVIRC TRQ-
HSP1_PIG        MARYRCCRSH --SRSRCR-P ---R-RRRCR -RRRRRCCPR --RRR-AVCC RRY--TVIRC RRC-
HSP1_HORSE      MARYRCCRSQ --SQSRCR-R ---RRRRRCR -RRRRRSVRQ --RR---VCC RRY--TVLRC RRRR
HSP1_ORCOR      MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRIRCCRR --QR--RVCC RRY--TTTRC ARQ-
HSP1_MOUSE      MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
HSP1_RAT        MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTFRC KRY-
HSP1_HUMAN      MARYRCCRSQ --SRSRYY-R ---Q-RQRSR -RRRRRSCQT --RRRAMRCC RPR--YRPRC RRH-
HSP1_GORGO      MARYRCCRSQ --SRSRCY-R ---Q-RQTSR -RRRRRSCQT --QRRAMRCC RRR--NRLRR RKH-
...
```

Figure 2. Alignment showing a selection of sperm protamine P1 amino acid sequences from 10 common eutherian mammals. The full multiple sequence alignment used 145 sequences from the 2019_05 release of the UniProt knowledgebase and was aligned with MUSCLE 3.8.31. Positions with relative entropy score greater than the conservation threshold of 4.135 are highlighted (see Supplemental Table 1). For full alignment see Supplemental Figure 2.

Eutherian Sperm Protamine P2

For the truncated eutherian sperm protamine P2 MSA, a total of eleven positions were determined to be highly conserved, based on relative entropy scores and conservation threshold described above. While intramolecular cross-linking in sperm protamine P2 proteins have yet to be determined [Vilfan et al. 2004], half of the positions identified as highly conserved in the P2 alignment consisted primarily of cysteine residues (positions 59, 75, 83, 93, and 107). The remaining positions are either primarily composed of histidine (positions 68, 89, 53, 85, and 110) or tyrosine (position 54) residues.

```
                    58         68         78         88         98        108
PRM2_RATTU      RG--HHRHRR CSRKRLHRIH KRR-RSCRRR RRHSCCHRRR HRRGCRRSRR RRRCRCRKCR RQCH
PRM2_MOUSE      RGHHHHRHRR CSRKRLHRIH KRR-RSCRRR RRHSCRHRRR HRRGCRRSRR RRRCRCRKCR RHHH
PRM2_RATFU      RG--HHRHRR CSRKRLHRIH KRR-RSCRRR RRHSCCHRRR HRRGCRRSRR RRRCKCRKCR RHCH
PRM2_ALOSE      QGCYGYRRRL CSRRRLYRVH RRQRRSCRRR C---CRYRRR NRRGCRT-RR RT------CR RH--
PRM2_CALJA      QGYSSYRRRR CSRRRRYRIH RRRSRSCRRR RRRSCRYRRR PRRGCRSRRR RR------CR RY--
PRM2_SEMEN      QGYSHHRRRR CSRRRLYRIH RRRHRSCRRR RRRSCRHRRR HRRGCRT-RR RR------CR RY--
PRM2_ERYPA      QGHSHHRRRR CSQRRLHRIH RRRHRSCRRR RRRSCRHRRR HRRGCRT-RR RR------CR RY--
PRM2_MACNE      RGHSHHRRRR CSRRRLHRIH RRRHRSCRRR RRRSCRHRRR HRRGCRT-RR RR------CR RH--
PRM2_MACFU      RGHSHHRRRR CSRRRLHRIH RRRHRSCRRR RRRSCRHRRR HRRGCRT-RR RR------CR RH--
PRM2_MACMU      -GHSHYRRRH CSRRRLHRIH RRRHRSCRRR RRRSCRHRRR HRRGCRT-RR RR------CR RH--
PRM2_GORGO      -GHSHYRRRH CSRRRLRRIH RQQHRSCRRR KRRSCRHRRR HRKGCRT-RR RT------CR RH--
PRM2_PANPA      -GHSHYRRRH CSRRRLRRIH RQQHRSCRRR KRRSCRHRRR HRRGCRT-RR RT------CR KH--
PRM2_PANTR      -GHSHYRRRH CSRRRLRRIH RQQHRSCRRR KRRSCRHRRK HRRGCRT-RR RT------CR RH--
PRM2_HUMAN      -GQSHYRRRH CSRRRLHRIH RRQHRSCRRR KRRSCRHRRR HRRGCRT-RK RT------CR RH--
PRM2_PONPY      -GHSHYRRRH CSRRRLHRIH RQQHRSCKRR RRHSCRHRRK HRRGCRT-RR RT------CR RH--
PRM2_HYLLA      -GHSHYRRRH CSRRRLHRIH RQQHRSCGRR RRRSCRQRRR HRRGCRT-RR RR------CR RH--
```

Figure 3. Alignment showing truncated sperm protamine P2 amino acid sequences from the alignment of 19 eutherian mammals. Sequences were pulled from the 2019_05 release of the UniProt knowledgebase and was aligned with MUSCLE 3.8.31. Positions with relative entropy score greater than the conservation threshold of 4.135 are highlighted (see Supplemental Table 2 for the truncated alignment and see Supplemental Table 3 for the untruncated alignment). For the untruncated alignment see Supplemental Figure 3.

Metatherian Sperm Protamine

When the relative entropy method was applied to the metatherian sperm protamine alignment, similar results are found in the eutherian sperm protamine P1 alignment. Instead of nine positions determined to be conserved, like in the eutherian P1 alignment, the metatherian alignment only has seven highly conserved positions. Of these seven positions, six were found to primarily contain tyrosine (positions 4, 57, 16, 62, 75, and 34). The remaining highly conserved position in the alignment primarily consisted of histidine residues (position 7).

```
                    10         20         30         40         50         60
HSP1_MONDO      MARYRRRSRS RSRSRYGRRR RRSRS----R RRRSRRRRR- ---RRGRR-- ---GRGYHRR
HSP1_DIDVI      MARYRRRSRS RSRSRYGRRR RRSRS----R RRRSRRRRR- ---RRGRR-- ---GRGYHRR
HSP1_PHACI      MARYR-HSRS RSRSRY-QRR RRRRSRYRSQ RRRYRRRRGS RRRRRRGR-- ---RRGY-RR
HSP1_PSECU      MARYR-HSRS RSRSRYRRRR RRRRSRYRGR RRRYRRSRR- RRRRGRRRGN CLGRRGYRRR
HSP1_MACRU      MARYR-HSRS RSRSRY-RRR RRRRSRYRSQ RRRYRGRRR- -RRSRRGR-- ---RRGYSRR
HSP1_MACGI      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_WALBI      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_THYCY      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PLAMS      MARCRRHSRS RSRSRN-QCQ RRRRRRY-NR RRTYRRSRR- HSRRRRGR-- ---RRGCSRR
HSP1_PLAIN      MARSRRHSRS RSRSR--NQC QRRRRRTYNR RRTMREKPR- HSRRRRVR-- ---RRGCSCR
                       70         80         90        100        110        120
HSP1_MONDO      SPHRR---RR RRRR------ -
HSP1_DIDVI      SPHRR---RR RRRR------ -
HSP1_PHACI      RYS-----RR --RRY----- -
HSP1_PSECU      RYS-----RR RRRRYY---- -
HSP1_MACRU      RYS------R RRRRY----- -
HSP1_MACGI      RYS------R RRRRY----- -
HSP1_WALBI      RYS-----RR RRRRY----- -
HSP1_THYCY      RYS-----RR GRRRY----- -
HSP1_PLAMS      RYS-----RR GRRRY----- -
HSP1_PLAIN      RCS-----RR RRRRC----- -
...
```

Figure 4. Alignment showing a selection of sperm protamine P1 amino acid sequences from 10 metatherian mammals. The full multiple sequence alignment used 95 sequences from the 2019_05 release of the UniProt knowledgebase and was aligned with MUSCLE 3.8.31. Positions with relative entropy score greater than the conservation threshold of 4.135 are highlighted (see Supplemental Table 5). For full alignment see Supplemental Figure 4.

Whole Sequence Arginine-Lysine Density Analysis

The arginine-lysine frequencies for each protamine in each homologous protamine group are shown in Figure 5 and in Table 1. It is clear from the differences in these arginine-lysine frequency distributions that the relative proportion of the DNA binding region to the whole protamine sequence is quite different for the protamine groups, especially the eutherian P2 protamines.

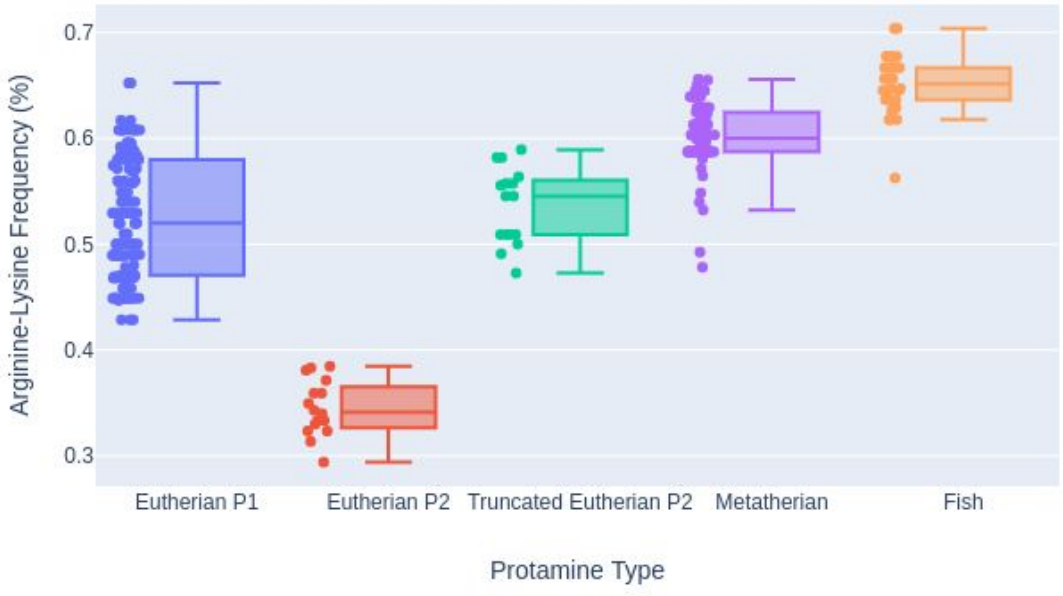

Figure 5. Box plots of the arginine frequency for eutherian protamine P1, truncated eutherian P2, metatherian P1, and fish protamines. Individual arginine frequencies are plotted next to the associated box plot.

Alignment	Min	Q1	Median	Q3	Max
Eutherian P1	0.429	0.471	0.520	0.580	0.652
Eutherian P2	0.294	0.328	0.341	0.362	0.385
Eutherian P2 Truncated	0.473	0.509	0.545	0.559	0.589
Metatherian	0.478	0.587	0.600	0.624	0.656
Fish	0.562	0.636	0.652	0.667	0.704

Table 1. Arginine-Lysine Density of Sperm Protamine Groups. Low, quartile, and max information for eutherian P1, eutherian P2, truncated eutherian P2, metatherian, and fish protamine groups.

DNA Binding Region Arginine-Lysine Density Analysis

Figure 6 and Table 2 show the arginine-lysine frequencies in the hypothesized DNA binding regions for each protamine in the eutherian P1 and metatherian MSAs. FIsh protamines were included in Figure 6 for comparison. The distributions were analyzed using a Welch's *t*-test which showed that each protamine group's arginine-lysine frequency distribution in their hypothesized DNA binding region is statistically discrete from any other groups. Comparing the DNA binding region of the eutherian P1 sperm protamine group to that of the metatherian sperm protamine group yielded a p-value of 2.184e-2. Comparing the DNA binding region of the eutherian P1 sperm protamine group to the whole fish sequence sperm protamine group yielded a p-value of 4.762e-11. Comparing the DNA binding region of the metatherian sperm protamine group to the whole fish sequence sperm protamine group yielded a p-value of 4.987e-8. However, the differences between the medians of these distributions is less than 0.017 (see Table 2) and the possible functional consequences of these relatively small differences in arginine-lysine frequencies are unclear given the relatively high variance of each group. It is possible that the arginine-lysine frequency of the DNA binding region across all of the protamines is just a species- and protein-specific optimization of the DNA binding function.

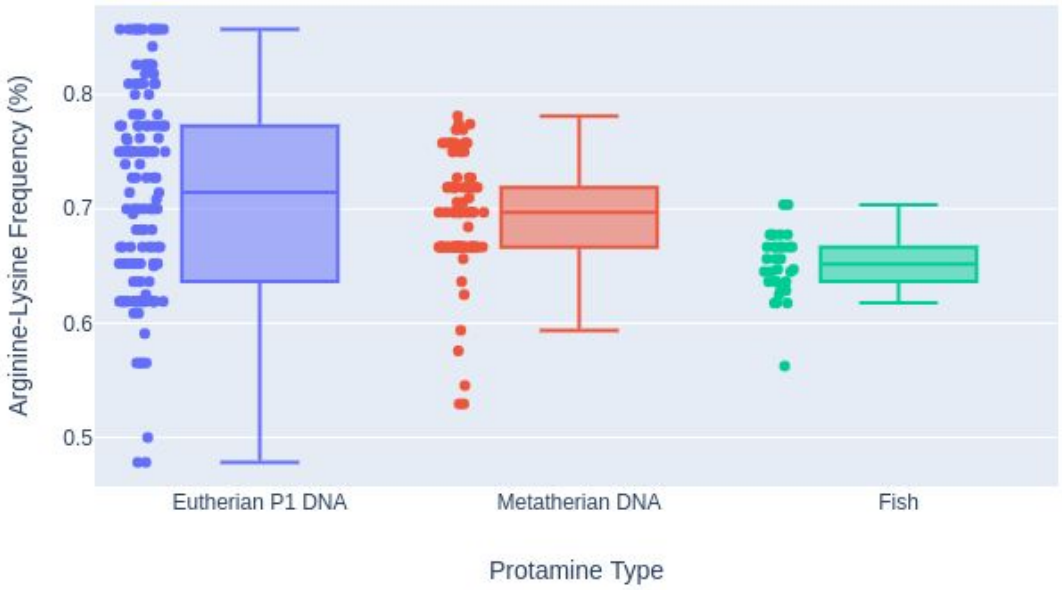

Figure 6. Box plots of the arginine-lysine frequency for the DNA binding regions of eutherian protamine P1, metatherian sperm protamine, and the whole sequences of fish protamines. Individual arginine-lysine frequencies are plotted next to the associated box plot.

Alignment	Min	Q1	Median	Q3	Max
Eutherian P1 DNA	0.478	0.636	0.714	0.773	0.857
Metatherian DNA	0.529	0.667	0.697	0.719	0.781

Table 2. Arginine-Lysine Density of DNA Binding Regions. Low, quartile, and max information for the hypothesized DNA binding region of eutherian P1 and metatherian sperm protamines.

Discussion

The results from the MSA analysis of fish protamines showed no highly conserved residues above arginine. This likely indicates that the only functionally/structurally important residues in the fish protamines are arginine. The arginine-lysine density analysis showed that fish have a greater charge density across their entire protamines sequences than any of the other protamine groups.

The results from the MSA analysis of the eutherian protamine P1 sequences showed that the most highly conserved positions tend to be cysteine containing. The high evolutionary sequence conservation indicates that the positions are of great functional/structural importance. When these highly conserved positions are overlaid onto a proposed schematic structure for bull sperm protamine P1 [Balhorn et al. 1991; Vilfan et al. 2004], it is clear that the conserved positions align with the cysteines involved in intra- and intermolecular bonding in bull sperm protamine P1. It is also notable that the cysteines involved in the intramolecular cross-linkings were shown to be more highly conserved than those involved in the intermolecular cross-linkings. This likely supports the hypothesis that the hairpin-like secondary structure of eutherian sperm protamine P1s is required for proper DNA hypercondensation [Hutchison et al. 2017; Kirchhoff et al. 2019].

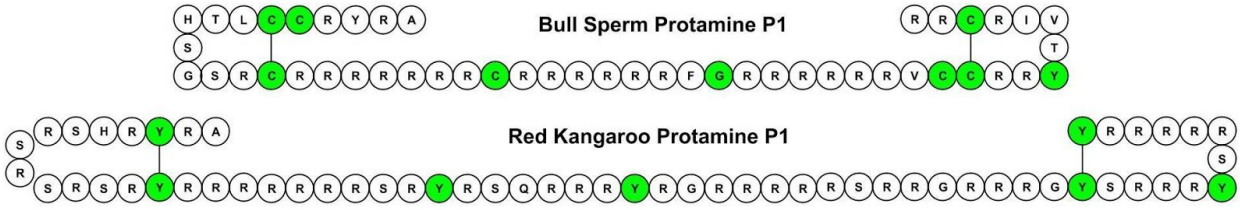

Figure 7.Highly conserved positions highlighted and overlaid onto proposed schematic structures for Bos taurus (Bull) and Macropus rufus (Red Kangaroo) sperm protamine P1s. Schematic structure of Bull P1 sperm protamines based off of Balhorn 1991 [Balhorn et al. 1991]. Red Kangaroo schematic structure assumed from similarity.

Comparing the metatherian P1 MSA to the eutherian P1 MSA, we find a number of commonalities. Both the N-terminal regions contain phosphorylation sites followed by blocks of arginine residues broken up by residues which can engage in cross-linking (cysteine in eutherians and tyrosine in metatherians) [Queralt et al. 1995]. A proposed schematic structure for metatherian sperm protamine P1 with conserved tyrosine positions is shown in Figure 7. The conserved tyrosine positions are visualized interacting in a similar cross-linking pattern as is observed in the cysteine containing eutherian mammal sperm protamines. Due to the similar conserved nature and similar spacing between the tyrosine residues in the metatherian protamine MSA and the cysteine residues in the eutherian protamine MSA, we hypothesize that metatherian protamines take on an analogous structure and folding mechanism as their eutherian counterparts. Folding of the metatherian protamines could possibly be facilitated by an orthologous enzyme to the glutathione peroxidase found in eutheria or an analogous peroxidase. There are specific peroxidases (e.g. certain myeloperoxidases) that are capable of catalysing dityrosine cross-linking in proteins [Bayse et al. 1972; Heinecke 2002; Mai et al. 2011]. However pi-pi stacking of two nearby tyrosines can represent another possible structural motif hypothesis [Lee et al. 2019].

The comparison of arginine-lysine frequencies of the differing protamine groups showed that although metatherian sperm protamines and fish protamines have similarly high charged residue frequencies across their entire sequences, metatherian mammals have an increased density of charged residues towards the middle of the sequences. This trend is also found in eutherian sperm protamine P1s where there is a central DNA binding region. The charged residue frequencies of the hypothesized DNA binding regions of eutherian P1 and metatherian sperm protamines were the most similar distributions by a significant factor.

Additional deviations are also apparent in the MSAs. One example is the tyrosine substitutions (tyr15cys and tyr16cys) of two positions in human sperm protamine P1, which are expected to be involved in intramolecular dicysteine cross-linking in the N-terminal staple fold. It is possible that cysteine-tyrosine cross-linking preserves the cross-linking function in human sperm protamines with these substitutions. Cysteine-tyrosine cross-linking, while not novel, is extremely rare in nature and are known to be mediated by copper metalloprotein enzymes [Martinie et al. 2012]. Also, prior research has shown that human spermatozoa are more susceptible to severe oxidative stress (>= 5 mM H_2O_2) in comparison to other eutherian mammals [Bennetts & Aitken 2005]. These substitutions could be an explanation for these observations.

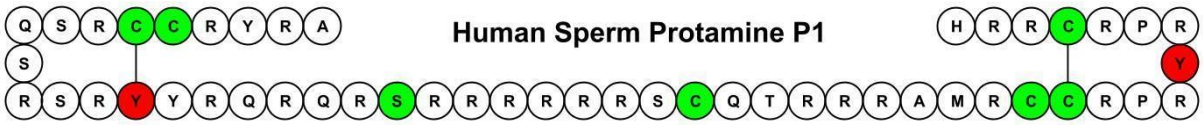

Figure 8. Highly conserved positions highlighted and overlaid onto proposed schematic structures for Homo sapien sperm protamine P1. Tyrosine containing positions highlighted in red and cysteine and other residue containing positions highlighted in green. Human schematic structure assumed from similarity [Balhorn et al. 1991].

Likewise, a counter deviation found in the metatherian alignment is with species of the *Planigale* genus, for which all species but *P. maculata maculata* contain cysteine substitutions in a number of highly conserved positions where tyrosine residues are expected (Supplemental Figure 5). It is notable that the change from tyrosine to cysteine is a single nucleotide substitution and that the arginine-lysine heavy regions are still found in the sequences of the *Planigale* species. What impact that these reciprocal substitutions may have on the fertility of the *Planigale* species and humans is not currently known.

Conclusions

In summary, the common patterns of sequence conservation between eutherian and metatherian protamine P1 sequence families support hypotheses for dityrosine cross-linking in the metatherian P1 protamines and a rare cysteine-tyrosine cross-linking in human sperm protamine P1. The presence of cysteine cross-linking in a number of species of the *Planigale* genus also indicates that metatherian sperm protamines are likely capable of taking on a hairpin-like structure analogous to eutherian sperma protamine P1. This was additionally supported by the finding that metatherian sperm protamines also contain an increased density of arginine and lysine residues in the center of the sequence, which likely represents a large DNA binding region analogous to the DNA region found in eutherian sperm protamine P1 sequences. In addition to directly testing these hypotheses with wet lab and analytical experiments, the next logical steps involve searching for an analogous peroxidase enzyme with expression localized to the testis in metatherian mammals to provide further evidence of dityrosine cross-linking and possibly an analogous peroxidase with a copper binding cofactor or new mechanism to support the formation of a cysteine-tyrosine cross-linking in human sperm protamines. Moreover, these proposed mechanisms and structures may play a role in fertility, particularly human fertility.

Methods

We used an entropy-based method to determine the functionally important residues in the MSAs of various protamine groups. Additionally, the charged residues densities of the protamine sequences in each group were determined. By comparing the conserved residues and the charged residue densities, we made predictions about structural features for related protamine groups and mechanisms behind their ability to bind DNA.

Creation of Homologous Protamine Groups

All entries containing the keyword “protamine” were downloaded from the May 2019 release (release 2019_05) of the UniProt KnowledgeBase (SwissProt/TrEMBL) [UniProt 2017] to create the initial dataset of protamine and protamine-like proteins.

The protamine dataset was then broken down into four homologous groups based on existing UniProt gene name and organism classification annotations. The four groups were eutherian sperm protamine P1, eutherian sperm protamine P2, metatherian sperm protamine P1, and fish protamine. The eutherian sperm protamine P1 group was parsed by collecting all sequences

which contained 'Eutheria' in their organism classification and the gene name of either 'PRM1' or 'Prm1'. The same approach was performed for the eutherian sperm protamine P2 group, but using the gene name of either 'PRM2' or 'Prm2'. The metatherian sperm protamine P1 group was parsed by collecting all sequences which contained 'Metatheria' in their organism classification and with the gene name of either 'PRM1' or 'Prm1'. The fish protamine group was parsed by collecting any sequence which contained 'Actinopterygii' in their organism classification and that did not contain the word 'like' in their description. For the eutherian and metatherian groups only a single entry was allowed per organism per group. If multiple sequences existed in a single group, preference was given to the Swiss-Prot entry since these are reviewed entries. Therefore, each group is composed of orthologous genes, with the exception of the fish protamine group where some species have more than one protamine gene in the group.

Additional Filtering and Truncation of Protamine P2 Sequences

Sperm protamine P2's contain multiple post-translational cleavage sites, which lead to the removal of 40% of the amino terminus of these proteins [Balhorn 2007]. After processing, the protein sequence is slightly longer than that of protamine P1 and the processed protein of P2 also has a higher arginine frequency than that of the unprocessed sequence [Balhorn 2007]. As only the processed version of the protein interacts with DNA, the P2 alignment is truncated to only include the processed versions of these proteins. This is achieved by determining the closest post-translational processing site to each protein's DNA binding region in the eutherian sperm protamine P2 alignment. The post-translating sites were found by using mouse sperm protamine P2 (MOUSE_PRM2) as a reference [Balhorn 2007]. MOUSE_PRM2 residue 44 is the closest post-translational processing site to the protein's DNA binding region in mice. MOUSE_PRM2 residue 44 can be found at position 48 in the eutherian P2 sperm protamine alignment. Position 48 of the alignment was therefore used as the truncation site.

Aberrant sperm protamine P2 sequences (Rat, Boar, Bovin) caused gapping in an initial MSA and were found to lack significant translational expression in prior literature [Bunick et al. 1990; Maier et al. 1990]. Also the sequence for Chinese hamster was severely truncated. Therefore these sequences were removed before final alignment.

Multiple Sequence Alignment and Conservation Analysis

A fasta file was generated for each homologous protamine group (i.e., eutherian P1, eutherian P2, and metatherian P1) and then MUSCLE 3.8.31 [Edgar 2004] was used to create an MSA using default settings.

Relative entropy (Kullback-Leibler divergence) was used to determine residue conservation scores for each position (column of residues) in the alignment. Relative entropy incorporates background frequencies of amino acids to measure the distance between the amino acid frequency in a position of the alignment versus the background frequencies [Capra & Singh 2007; Cover & Thomas 2006; Hannenhalli & Russell 2000].

$$D_{KL}(P||Q) = \sum_{a \epsilon AA} P(a) log(\frac{P(a)}{Q(a)}) \qquad (1)$$

D(P||Q) is calculated for each position in the alignment and uses all 20 of the standard amino acids plus Asx (B) for Asp or Asn, Glx (Z) for Glu or Gln, and Xaa (X) for unknown. P(a) is the frequency of the amino acid in the position. Q(a) is the background frequency of an amino acid. For this analysis, the natural abundance of amino acids determined by the UniProt knowledgebase was used [UniProt Consortium 2017]. Relative entropy has been shown to be one of the most effective algorithms for determining functionally/structurally important residues from alignments and tied for the most effective method for determining positions playing a role in protein-protein interactions [Capra & Singh 1991; Hannenhalli & Russell 2000].

Additionally, a weighting was used to deal with the presence of gaps in the alignment. The gap weighting is incorporated by multiplying the calculated relative entropy measure by the percent of non-gap residues in the position.

$$G_W = D_{KL}(P||Q) * \%non-gap\ residues \qquad (2)$$

To determine which positions are conserved in the alignment, a conservation score threshold equal to a position entirely composed of arginine residues (~4.1354) was used. If the gap weighted conservation score was greater than the threshold, the position was determined to be conserved. Methionine residues at the beginning of protein sequences are ignored.

Arginine-Lysine Density Analysis

The arginine-lysine density of each protamine group was calculated by counting the number of arginine and lysine residues in each protamine sequence in the group. Histidine was left out of the analysis as it is mostly deprotonated at physiological pH. Lysine was included above a simple arginine frequency due to known DNA interactions for lysine residues in a variety of DNA binding proteins, but more importantly, an observed reduction in the severity of lower bound outliers for all protamine groups analyzed. This improvement in the lower bound outliers is greater for lysine than for the inclusion of any other amino acid (see Supplemental Tables 5-9). The quartile ranges for each group was then calculated and graphed using Plotly [Plotly 2015]. For eutherian sperm protamine P2, the truncated sequence was used, as determined by the method mentioned above. Additionally, for each protein in the eutherian P1 and metatherian sperm protamine alignments the charged residue density within the hypothesized DNA binding region was calculated. The hypothesized DNA binding region for the eutherian P1 alignment begins at position 17 and ends at position 46. The hypothesized DNA binding region for the metatherian alignment begins at position 16 and ends at position 56.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CDP performed the analyses. DCK, JED, and HNBM were all involved with the study's conception and design. All authors read, edited, and approved the manuscript.

Funding

This work was supported by funding from NSF 1419282 (Moseley), NSF 1453168 (DeRouchey), and NIH UL1TR001998-01 (Kern).

Availability of Data and Material

All datasets, figures, and scripts are available through FigShare: https://doi.org/10.6084/m9.figshare.10292573.v1

References

Arellano, A., Canales, M., Jullian, C., & Brunet, J. E. (1988). Fluorescence studies on clupein protamines: evidence for globular conformation. *Biochemical and biophysical research communications*, 150(2), 633-639.

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. *Genome biology, 8*(9), 227.

Balhorn, R., Corzett, M., Mazrimas, J., & Watkins, B. (1991). Identification of bull protamine disulfides. *Biochemistry, 30*(1), 175-181.

Bayse, G. S., Michaels, A. W., & Morrison, M. (1972). The peroxidase-catalyzed oxidation of tyrosine. *Biochimica et Biophysica Acta (BBA)-Enzymology*, 284(1), 34-42.

Bench, G., Corzett, M. H., Kramer, C. E., Grant, P. G., & Balhorn, R. (2000). Zinc is sufficiently abundant within mammalian sperm nuclei to bind stoichiometrically with protamine 2. *Molecular Reproduction and Development: Incorporating Gamete Research*, 56(4), 512-519.

Bennetts, L. E., & Aitken, R. J. (2005). A comparative study of oxidative DNA damage in mammalian spermatozoa. *Molecular Reproduction and Development: Incorporating Gamete Research*, *71*(1), 77-87.

Biegeleisen, K. (2006). The probable structure of the protamine–DNA complex. *Journal of theoretical biology*, *241*(3), 533-540.

Brewer, L. R. (2011). Deciphering the structure of DNA toroids. *Integrative Biology*, *3*(5), 540-547.

Bunick, D., Balhorn, R., Stanker, L. H., & Hecht, N. B. (1990). Expression of the rat protamine 2 gene is suppressed at the level of transcription and translation. *Experimental cell research*, *188*(1), 147-152.

Capra, J. A., & Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics, 23*(15), 1875-1882.

Cid, H., & Arellano, A. (1982). Secondary structure prediction of protamines. *International Journal of Biological Macromolecules, 4*(1), 3-8.

Conrad, M., Moreno, S. G., Sinowatz, F., Ursini, F., Kölle, S., Roveri, A., ... & Bornkamm, G. W. (2005). The nuclear form of phospholipid hydroperoxide glutathione peroxidase is a protein thiol peroxidase contributing to sperm chromatin stability. *Molecular and cellular biology*, *25*(17), 7637-7644.

Cover, T. M. & Thomas, J. A. (2006). *Elements of information theory 2nd edition.* Hoboken, New Jersey: John Wiley & Sons.

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

Enciso, M., Johnston, S. D., & Gosálvez, J. (2011). Differential resistance of mammalian sperm chromatin to oxidative stress as assessed by a two-tailed comet assay. *Reproduction, Fertility and Development*, *23*(5), 633-637.

Hannenhalli, S. and Russell, R. (2000). Analysis and prediction of functional sub-types from protein sequence alignments. *Journal of Molecular Biology, 303(1),* 61-76.

Heinecke, J. W. (2002). Tyrosyl radical production by myeloperoxidase: a phagocyte pathway for lipid peroxidation and dityrosine cross-linking of proteins. *Toxicology*, *177*(1), 11-22.

Hutchison, J. M., Rau, D. C., & DeRouchey, J. E. (2017). Role of disulfide bonds on DNA packaging forces in bull sperm chromatin. *Biophysical journal*, *113*(9), 1925-1933.

Kirchhoff, D. C., Powell, C.D., Rhinehart, J., Moseley, H.N.B., and DeRouchey, J.E. Disulfide-mediated secondary structure in protamine is critical for DNA condensation in mammalian sperm chromatin (in preparation).

Lee, J., Ju, M., Cho, O. H., Kim, Y., & Nam, K. T. (2019). Tyrosine‐Rich Peptides as a Platform for Assembly and Material Synthesis. *Advanced Science*, *6*(4).

Mai, K., Smith, N. C., Feng, Z. P., Katrib, M., Šlapeta, J., Šlapetova, I., ... & Norton, R. S. (2011). Peroxidase catalysed cross-linking of an intrinsically unstructured protein via dityrosine bonds in the oocyst wall of the apicomplexan parasite, Eimeria maxima. *International journal for parasitology*, *41*(11), 1157-1164.

Maier, W. M., Nussbaum, G., Domenjoud, L., Klemm, U., & Engel, W. (1990). The lack of protamine 2 (P2) in boar and bull spermatozoa is due to mutations within the P2 gene. *Nucleic acids research*, *18*(5), 1249-1254.

Martinie, R. J., Godakumbura, P. I., Porter, E. G., Divakaran, A., Burkhart, B. J., Wertz, J. T., & Benson, D. E. (2012). Identifying proteins that can form tyrosine-cysteine crosslinks. *Metallomics*, *4*(10), 1037-1042.

Pfeifer, H., Conrad, M., Roethlein, D., Kyriakopoulos, A., Brielmeier, M., Bornkamm, G. W., & Behne, D. (2001). Identification of a specific sperm nuclei selenoenzyme necessary for protamine thiol cross-linking during sperm maturation. *The FASEB Journal*, 15(7), 1236-1238.

Plotly Technologies Inc. Collaborative data science. Montréal, QC, 2015. https://plot.ly.

Queralt, R., Adroer, R., Oliva, R., Winkfein, R. J., Retief, J. D., & Dixon, G. H. (1995). Evolution of protamine P1 genes in mammals. *Journal of molecular evolution*, *40*(6), 601-607.

Retief, J. D., Rees, J. S., Westerman, M., & Dixon, G. H. (1995). Convergent evolution of cysteine residues in sperm protamines of one genus of marsupials, the Planigales. *Molecular biology and evolution, 12*(4), 708-712.

Steger, K., & Balhorn, R. (2018). Sperm nuclear protamines: A checkpoint to control sperm

chromatin quality. *Anatomia, histologia, embryologia*, *47*(4), 273-279.

UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45(D1), D158-D168.

Vilfan, I. D., Conwell, C. C., & Hud, N. V. (2004). Formation of native-like mammalian sperm cell chromatin with folded bull protamine. *Journal of Biological Chemistry*, *279*(19), 20088-20095.

Villani, P., Eleuteri, P., Grollino, M. G., Rescia, M., Altavista, P., Spanò, M., ... & Cordelli, E. (2010). Sperm DNA fragmentation induced by DNAse I and hydrogen peroxide: an in vitro comparative study among different mammalian species. *Reproduction*, *140*(3), 445.

Whittington, A. T., Parkinson, A. L., Spencer, P. B. S., Grigg, G., Hinds, L., Gallagher, C. H., ... & Agar, N. S. (1995). Comparative study of the antioxidant defence systems in the erythrocytes of Australian marsupials and monotremes. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology*, 110(3), 267-272.

Entropy Based Analysis of Vertebrate Sperm Protamines Sequences: Evidence of Dityrosine and Cysteine-Tyrosine Cross-Linking in Sperm Protamines

Supplemental Tables

Table 1 Highly Conserved Positions in Eutherian P1 Sperm Protamine MSA.

Rank	Position in Alignment	Relative Entropy (bit)	Represented Residues
1	7	6.157	C: 98.6% R: 1.4%
2	49	6.12	C: 97.9% G: 1.4% S: 0.7%
3	50	5.87	C: 95.1% R: 2.1% Y: 1.4% G: 0.7% S: 0.7%
4	60	5.439	C: 89.5% R: 4.9% S: 4.2% Y: 1.4%
5	38	5.187	C: 86.9% G: 4.1% Y: 2.8% R: 1.4% S: 1.4% V: 1.4% A: 0.7% F: 0.7% H: 0.7%
6	17	4.885	C: 79.3% R: 13.1% G: 5.5% S: 1.4% Y: 0.7%
7	29	4.813	C: 77.2% Y: 13.1% S: 5.5% G: 1.4% R: 1.4% F: 0.7% N: 0.7%
8	6	4.604	C: 95.5% R: 4.5%
9	37	4.33	C: 69.0% S: 19.3% F: 5.5% R: 5.5% T: 0.7%

Table 2 Highly Conserved Positions in Truncated Eutherian P2 Sperm Protamine MSA.

Rank	Position in Alignment	Relative Entropy (bit)	Represented Residues
1	59	6.381	C: 100.0%
2	75	6.381	C: 100.0%
3	83	6.381	C: 100.0%
4	93	6.381	C: 100.0%
5	107	6.381	C: 100.0%
6	68	5.513	H: 100.0%
7	89	4.721	H: 87.5% N: 6.2% P: 6.2%
8	53	4.635	H: 87.5% G: 6.2% S: 6.2%
9	85	4.543	H: 81.2% Y: 12.5% Q: 6.2%
10	110	4.371	H: 75.0% Y: 18.8% Q: 6.2%
11	54	4.288	Y: 56.2% H: 43.8%

Table 3 Highly Conserved Positions in Eutherian P2 Sperm Protamine MSA.

Rank	Position in Alignment	Relative Entropy (bit)	Represented Residues
1	59	6.381	C: 100.0%
2	75	6.381	C: 100.0%
3	83	6.381	C: 100.0%
4	93	6.381	C: 100.0%
5	107	6.381	C: 100.0%
6	14	5.513	H: 100.0%
7	68	5.513	H: 100.0%
8	48	5.126	H: 93.8% Q: 6.2%
9	44	5.093	Y: 100.0%
10	4	4.836	Y: 93.8% C: 6.2%
11	19	4.725	Q: 100.0%
12	26	4.725	Q: 100.0%
13	34	4.725	Q: 100.0%
14	89	4.721	H: 87.5% N: 6.2% P: 6.2%
15	53	4.635	H: 87.5% G: 6.2% S: 6.2%
16	85	4.543	H: 81.2% Y: 12.5% Q: 6.2%
17	110	4.371	H: 75.0% Y: 18.8% Q: 6.2%
18	54	4.288	Y: 56.2% H: 43.8%
19	47	4.171	T: 100.0%

Table 4 Highly Conserved Positions in Metatherian P1 Sperm Protamine MSA.

Rank	Position in Alignment	Relative Entropy (bit)	Represented Residues
1	7	4.98	H: 92.6% R: 5.3% N: 2.1%
2	4	4.835	Y: 95.8% C: 3.2% S: 1.1%
3	57	4.733	Y: 93.7% C: 4.2% K: 2.1%
4	16	4.67	Y: 94.7% N: 3.2% F: 2.1%
5	62	4.599	Y: 92.6% C: 3.2% P: 2.1% S: 2.1%
6	75	4.565	Y: 94.5% C: 3.3% N: 2.2%
7	34	4.536	Y: 91.6% S: 4.2% Q: 2.1% M: 1.1% X: 1.1%

Table 5 Fish Protamines

Rank	Range		STD	
1	K	0.141	K	0.027
2	H	0.185	H	0.035
3	L	0.222	L	0.039
4	T	0.228	T	0.041
5	A	0.256	E	0.045

Table 6 Eutherian P1 Sperm Protamines

Rank	Range		STD	
1	Q	0.182	Q	0.034
2	K	0.224	P	0.044
3	T	0.224	T	0.046
4	S	0.242	N	0.046
5	N	0.244	A	0.049

Table 7 Eutherian P2 Sperm Protamines

Rank	Range		STD	
1	Q	0.07	Q	0.017
2	L	0.09	T	0.022
3	T	0.09	L	0.023
4	K	0.09	E	0.024
5	C	0.099	V	0.025

Table 8 Truncated Eutherian P2 Sperm Protamines

Rank	Range		STD	
1	Q	0.116	K	0.035
2	K	0.117	Q	0.037
3	T	0.127	T	0.04
4	L	0.145	L	0.044
5	C	0.151	G	0.045

Table 9 Metatherian Sperm Protamines

Rank	Range		STD	
1	K	0.177	N	0.027
2	N	0.179	K	0.03
3	G	0.202	T	0.035
4	A	0.223	A	0.035
5	E	0.235	Q	0.035

```
                        10         20         30         40
PRTA_ACIST       -ARRRRRHAS TKLKRRR--- -----RRRRH GKKSHK
PRT_ORYLA        ----MRRQAS LPARRRRRVR RTRVVRRRRR VGRRRH
PRT_PERFV        --PRRRRHAA RPVRRRRRTR RSSRVHRRRR AVRRRR
PRTB_MUGCE       --PRRRRETS RPIRRRRRAR RAPI-RRRRR VVRRRR
PRT1_SCOSC       -MPRRRRRAS RPVRRRRRAR RSTAVRRRRR VVRRRR
PRT2_SCOSC       -MPRRRRRAS RPIRRRRRAR RSTAVRRRRR VVRRRR
PRT_DICLA        --PRRRRQAS RPVRRRRRTR RSTAERRRRR VVRRRR
PRTY_THUTH       --PRRRRQAS RPVRRRRRYR RSTAARRRRR VVRRRR
PRTZ_THUTH       --PRRRRRSS RPVRRRRRYR RSTVARRRRR VVRRRR
PRTZ1_SAROR      --PRRRRRSS RPVRRRRRYR RSTAARRRRR VVRRRR
PRTB_ACIGU       --ARRRRRSS RPQRRRRR-- -----RRHGR RRRGRR
PRTB_ACIST       --ARRRRRSS RPQRRRRR-- -----RRHGR RRRGRR
PRTY1_CLUHA      --ARRRRSSS RPIRRRRPRR RT---TRRRR AGRRRR
PRTY1_CLUPA      --ARRRRSSS RPIRRRRPRR RT---TRRRR AGRRRR
PRT1_ESOLU       -PRRRRASSG RPVRRRRRP- KMS--RRRRR GGRRRR
PRTZ_CLUPA       ARRRRSRRAS RPVRRRRPR- RVS--RRRR- -ARRRR
PRTZ_CLUHA       ARRRRSRRAS RPVRRRRPR- RVS--RRRR- -ARRRR
PRT4_ONCMY       -MPRRRR-AS RRIRRRRRP- RVS--RRRRG GRRRRR
PRTY2_CLUPA      -PRRRTRRAS RPVRRRRPR- RVS--RRRR- -ARRRR
PRTY2_CLUHA      -PRRRTRRAS RPVRRRRPR- RVS--RRRR- -ARRRR
PRT3A_ONCMY      -PRRRRRSSS RPIRRRRRP- RVS--RRRRR GGRRRR
PRTIB_ONCMY      PRRRRRRSSS RPIRRRRPR- RVS--RRRRR GGRRRR
PRT1B_ONCMY      -MPRRRR-AS RRIRRRRRP- RVS--RRRRR GGRRRR
PRT1A_ONCMY      -MPRRRR-AS RRVRRRRRP- RVS--RRRRR GGRRRR
PRTC3_ONCMY      -MPRRRR-AS RPVRRRRRP- RVS--RRRRR GGRRRR
PRT2C_ONCMY      -MPRRRRSSR RPVRRRRRP- RVSR-RRRRR GGRRRR
PRT14_ONCMY      -MPRRRRSSR PPVRRRRRP- RVSR-RRRRR GGRRRR
PRT2_ONCMY       --PRRRRSSS RPVRRRRAR- RVSR-RRRRR GGRRRR
PRT2A_ONCMY      --PRRRRSSS RPVRRRRRA- RVSR-RRRRR GGRRRR
PRTIA_ONCMY      --PRRRRSSS RPVRRRRRPR RVSR-RRRRR GGRRRR
PRT1_ONCKE       -MPRRRRSSS RPVRRRRRP- RVSR-RRRRR GGRRRR
PRT2B_ONCMY      -MPRRRRSSS RPVRRRRRP- RVSR-RRRRR GGRRRR
PRT16_ONCMY      -MPRRRRSSS RPVRRRRRA- RVSR-RRRRR GRRRRR
PRT17_ONCMY      -MPRRRRSSS RPVRRRRRP- RVSR-RRRRR GRRRRR
```

Supplemental Figure 1. Alignment of fish protamines.

```
                          10         20         30         40         50         60
A0A2K5CXY9_AOTNA MARHRCCRSR --SQSRSR-R --DRQKRRCR -TPRRRSCRR --RTA-RRCG RRR--YKPRC RRN-
HSP1_RHIFE       MARYSCCRSH --SRSRSR-R ---R-RQRCR -RRRRRSCGR --RR--RACY RRYTVRYRRR RRRR
HSP1_HIPCO       MARYRCCRSH --SRSRCR-R ---R-RRRSR -RRRRRSCGR --RR--RAGY RRY---TVRY RRRR
A8IYA4_ANTAM     MARYRCCLTH --SRSRCRPR ---RRRRRCR -KLRRRFCRR -PRR--RVCC RRY--TAIRC TR--
A8IYA7_9CETA     MARYRCRLTH --SRSGCR-R ---RRRRRCR -RRRRRFCRR --RRR-RVCC RRY--TVVRC TRQ-
HSP1_CAPHI       MARYRCCLTH --SRSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVVRC TRQ-
HSP1_SHEEP       MARYRCCLTH --SRSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVVRC TRQ-
A8IYA2_OVIDA     MARYRCCLTH --SRSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVVRC TRQ-
HSP1_BOVIN       MARYRCCLTH --SGSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVIRC TRQ-
A0A068B2A1_BOSIN MARYRCCLTH --SGSRCR-R ---RRRRRCR -RRRRRFGRR --RRR-RVCC RRY--TVIRC TRQ-
A0A193KZW0_9CETA MARYRCCLTH --SGSRCR-R ---RRRRRCR -RRRRRFGRR --RR------ -RR------- ----
C8C436_PLAMN     MARNRCCRSQ --SRSRCR-R ---P-KRGCR -SRRRRCYQR --RR--RVCC RRY--TTIRC ARQ-
A8IYB4_PHYMC     MARNRCCRSQ --SRSRCR-R ---P-RRRCR -SPRRRRYQR --RR--RVCC RRY--TVTRC ARQ-
C8C437_KOGSI     MARNRCCRSQ --SRGRCR-R ---P-RRRCR -SPKRRRYQR --RR--RVCC RRS--ATMRC ASQ-
C8C438_KOGBR     MARNRCCRSQ --SRGRCR-R ---P-RRRYR -SPRRRRYQR --RR--RVCC RRS--TTMRC ASQ-
HSP1_PIG         MARYRCCRSH --SRSRCR-P ---R-RRRCR -RRRRRCCPR --RRR-AVCC RRY--TVIRC RRC-
A8IYC1_POTPR     MARYRCCRSH --SRSRCR-P ---R-RRRCR -RRRRRCCPR --RRR-AVCC RRY--TVIRC RRC-
F7VJK3_FELCA     MARYRCCRSH --SRSRCR-R ---R-RRRCR -RRRRRCCRR -PRK--RVCS RRY--RVGRC RRR-
HSP1_OTOHE       MARYRCCRSR --SRSRCR-R ---R-RRKCY -RRRRRCSRK --RRR-RVCC RRY--TVMRC RRR-
C8C444_HEXLI     MARYRCCRSP --SRSRCR-R ---Q-RRRCR -RRRRRCCRQ --RRR-RVCC RRY--TMVRC TRQ-
A8IYB9_HIPAM     MARYRCCRSP --SRSRCR-R ---Q-RRRCR -RRRRRCCRQ --RRR-RVCC RRY--TMVRC TRQ-
HSP1_EQUAS       MARYRCCRSQ --SQSRCR-R ---RRRRRCR -RRRRRCVRR --RR---VCC RRY--TVLRC RRRR
HSP1_HORSE       MARYRCCRSQ --SQSRCR-R ---RRRRRCR -RRRRRSVRQ --RR---VCC RRY--TVLRC RRRR
HSP1_HYPSA       MARYRRCR-- --SRSRCR-R ---R-RRRCH -RRRRRCCRR --RRRRRACC RRY-----RC RRR-
HSP1_NEOBU       MARYRRCR-- --SRSRCR-R ---R-RRRCH -RRRRRCCRR --RRRRRACC RRY-----RC RRR-
HSP1_RHIHA       MARYRCCRSR --SRSRCRPR -----RRRCR -RRRRRCCRR --RR--RVCC RRY---SARC RRRR
HSP1_MURCY       MARYRCCR-- --SRSRCR-R ---R-RRRCH -RRRRRCSRR --RRR-RVCC RRY--TVIRC RRR-
HSP1_PTEPA       MARYRCCRSP --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RVCC RRY---TVRC RRR-
HSP1_DESRO       MARYRCCRSP --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RVCC RRY---TVRC RRR-
HSP1_MONRE       MARYRCCRSP --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RVCC RRY---TVRC RRR-
HSP1_MORME       MARYRCCRSP --SRSRCR-R ---R-RRRCR -RRRRRSCRR --RR--RVCC RRY---TVRC RRR-
HSP1_MYODA       MARYRCCR-- --SRSRCR-R ---R-RRRCY -RRRRRCCRR --RRRRRVCC RRY----SRC RRR-
F7VJK7_MYOLU     MARYRCCR-- --SRSRCR-R ---R-RRRCY -RRRRRCCRR --RRRRRVCC RRY----SRC RRR-
HSP1_GLABE       MARYRCCR-- --SRSRCR-R ---R-RRRSY -RRRRRCCRR --RRR-RVCC RRY----VRC RRR-
HSP1_NATST       MARYRCCRSQ --SRSRCR-P ---R-RRRCR -TRRRRCCRR -RRR--RVCC RRY--TVVRC RRR-
HSP1_CHIMC       MARYRCCRSQ --SRSRCR-R ---R-RRRCR -TRRRRCCRR -RRR--RVCC RRY--TVVRC RRR-
HSP1_EPTFU       MARYRCCR-- --SRSRCR-R ---R-RRRCY -RRRRRCCRR -RRR--RVCC RRY--TVIRC RRR-
HSP1_EPTBR       MARYRCCR-- --SRSRCR-R ---R-RRRCY -RRRRRCCRR -RRR--RVCC RRY--TVIRC RRR-
HSP1_CORTO       MARYRCCRSQ --SRSRCR-R ---R-RRRCY -RRRRRCCRR --RRR-RVCC RRY----TRY RRR-
HSP1_PLEAU       MARYRCCRSQ --SRSRCR-R ---R-RRRCY -RRRRRCCRR -RRR--RVCC RRY--TVVRC RRR-
HSP1_PTEHP       MARYRCCRSQ --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RVCC RRY---TVRC RRRR
F7VJK4_CANLF     MARYRCCRSQ --SRSRCR-R ---R-RRRCR -RRRRRCCRR RRR---RVCC RRY--TVVRC RRR-
C8C440_EUBGL     MARNRCCRSQ --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARR-
C8C439_EUBAS     MARNRCCRSQ --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARR-
B1ACJ7_BALMY     MARNRCCRSQ --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARR-
B1ACJ8_EUBJA     MARNRCCRSQ --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARR-
B1ACJ6_CAPMR     MARNRCCRSQ --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARQ-
B1ACI8_MEGNO     MARNSCCRSP --SRSRRR-R ---P-RQRRR -SQRRRCCQR --RR--RVCC RRY--TTVRC ARQ-
B1ACI9_BALPH     MARNSCCRSP --SRSRRR-R ---P-RQRRR -SQRRRCCQR --RR--RVCC RRY--TTVRC ARQ-
B1ACJ3_BALBN     MARNSCCRSP --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC AGQ-
B1ACJ4_BALAC     MARNSCCRSP --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC AGQ-
B1ACJ1_BALBO     MARNSCCRSP --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARQ-
B1ACJ5_ESCRO     MARNSCCRSP --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARQ-
B1ACJ2_BALED     MARNSCCRSP --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARQ-
B1ACJ0_BALMU     MARNTCCRSP --SRSRRR-R ---P-RQRCR -SQRRRCCRR --RR--RVCC RRY--TTVRC ARQ-
C8C435_BERBI     MARNRCCRSQ --SQSRRR-R ---P-RRRNR -SRRRQCCQR --RR--RVCC RRY--TAIRC ARQ-
A8IYB7_ZIPCA     MARNRCCRGQ --SQSRRR-R ---P-RRRYR -SRRRQCCQK --RR--RVCC RRY--TATRC ARQ-
C8C432_MESGR     MARNTCCRSQ --SQSRRR-R ---P-RRRYR -SRRKQCCQK --RR--RVCC RRY--TAIRC ARQ-
C8C434_TASSH     MARNRCCRSQ --SQSRRR-R ---P-RRRYR -SRRRQCCQK --RR--RVCC RRY--TAIRC ARQ-
C8C433_MESBI     MARNRCCRSQ --SQSRRR-R ---P-RRRYR -SRRRQCCQK --RR--RVCC RRY--TAIRC ARQ-
C8C431_MESPE     MARNRCCRSQ --SQSRRR-R ---P-RRRYR -SRRKQCCQK --RR--RVCC RRY--TAIRC ARQ-
HSP1_ORCOR       MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRIRCCRR --QR--RVCC RRY--TTTRC ARQ-
C8C425_GLOMA     MARNR-CRSP --SQSRCR-R ---P-RRRYR --RRRQCCRR --RR--RVCC RRY--TTTRC ARQ-
C8C426_GRAGR     MARNR-CRSP --SQSRCR-R ---P-RRRYR --RRRQCCRR --RR--RVCC RRY--TTTRC ARQ-
B4YVM8_PSECS     MARNR-CRSP --SQSRCR-R ---P-RRRYR --RRRQCCRR --RR--RVCC RRY--TTTRC ARQ-
C8C424_FERAT     MARNR-CRSP --SQSRCR-R ---P-RRRYR --RRRRCCRR --RR--RVCC RRY--TTTRC ARQ-
F5CBD0_PENEL     MARNR-CRSP --SQSRCR-R ---P-RRRYR --RRRRCCRR --RR--RVCC RRY--TTTRC ARQ-
C8C423_LAGAC     MARNR-CRSP --SQSRCR-R ---P-RRRCL --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
F5CBC9_CEPCM     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--ATTRC ARQ-
F5CBC8_LAGAL     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC ASQ-
C8C421_TURTR     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRLCCRR --RR--RVCC RRY--TTTRC ARQ-
F5CBC0_TURAD     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRLCCRR --RR--RVCC RRY--TTTRC ARQ-
B4YVN1_STECO     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRLCCRR --RR--RVCC RRY--TTTRC ARQ-
F5CBC1_STELO     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
F5CBC5_LAGHO     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
F5CBB9_DELCA     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
F5CBC3_STEAT     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
F5CBC4_SOUCH     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC ARQ-
F5CBD1_9CETA     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC ARQ-
F5CBC7_LAGOL     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
B4YVM9_STEBR     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC ARQ-
F5CBC2_STEFR     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
F5CBC6_SOTFL     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
C8C422_LISBO     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC ARQ-
B4YVN0_DELDE     MARNR-CRSP --SQSRCR-R ---P-RRRCR --RRRRCCRR --RR--RVCC RRY--TTTRC AR--
C8C428_NEOPH     MARNR-CRSP --SQSRGR-C ---P-RRRYR -SKRRRCCQR --RR--RVCR RRY---TRRC ARQ-
C8C427_9CETA     MARNR-CRSP --SQSRGR-C ---P-RRRYR -SKRRRCCQR --RR--RVCR RRY---TRRC ARQ-
B4YVM6_PHOPH     MARNR-CRSP --SQSRGR-C ---P-RRRYR -SKRRRCCQR --RR--RVCR RRY---TRRC ARQ-
F5CIP3_LIPVE     MARNR-CRSP --SQSRGR-R ---P-RRKYR -SRRRRCCQR --RR--RVCC RRY--TTMRC AKQ-
C8C430_DELLE     MARNR-CRSP --SQSRGR-R ---P-RRRYR -SKRRRCCQR --RR--RVCC RRY--TTTRC ARQ-
C8C429_MONMO     MARTR-CRSP --SQSRGR-R ---P-RRRYR -SKRRRCCQR --RR--RVCC RRY--TTTRC ARQ-
A8IYB2_PONBL     MARNR-CRSP --SQNRGR-R ---P-RRRYR -SRRRRCCQR --RR--RVCC RRY--TSVRC ARQ-
A8IYA9_INIGE     MARNR-CRSP --SQSRGR-R ---P-RRRYR -SRRRRCCQR --RR--RVCC RRY--TTVRC ARQ-
F7VJK6_LOXAF     MARYRCCRSR --SRSRCRSR ---R-RRRSH -RRRRRCARR --RRRTRRGC RRR--YSLRR RRY-
HSP1_PONPY       MARYRCCRSQ --SQSRCC-R ---R-RQRCH -RRRRRCCQT --RRRAMRCC RRR--YRLRC RRH-
H2NQ55_PONAB     MARYRCCRSQ --SQSRCC-R ---R-RQRCH -RRRRRCCQT --RRRAMRCC RRR--YRLRC RRH-
HSP1_GALVR       MARYRCCR-- --SRSRCR-R ---R-RRSCR --RRRRCRRR --RA--RRSC RRR--YSLRC CRRY
HSP1_SAGIM       MARYRCCRSQ --SRSRCY-R ---Q-RRRGR -RRRRRTCRR --RRA-SRCC RRR--YKLTC RRY-
A0A2R8PK40_CALJA MARYRCCRSQ --SRSRCY-R ---Q-RRRGR -RRRRRSCRR --RRA-SRCC RRR--YKLPC RRY-
HSP1_ALOSE       MARYRCCRSR SLSRSRCY-R ---Q-RPRCR -RRRRRSCRR -PRA--SRCC RRR--YRLRR RRY-
HSP1_SAISC       MARYRCCRSR --SRSRCY-R ---R-RRRCR -TRRRRCCRR --RRA-RRCC RRR--YKLRC RRY-
A0A2K6TB39_SAIBB MARYRCCRSR --SRSRCY-R ---R-RRRCR -TRRRRCCRR --RRA-RRCC RRR--YKLRC RRY-
A0A2K5R6E5_CEBCA MARYRCCRSR --SRSRCY-R ---Q-RRRCR -RRRRRCRSR --RA--RRCC RRR--YRLRC RRY-
HSP1_RABIT       MVRYRCCRSQ --SRSRCR-R ---R-RRRCR -RRRRRCCQR --RRV-RKCC RRT--YTLRC RRY-
HSP1_CAVPO       MARYRCCRSP --SRSRCR-R ---R-RRRFY -RRRRRCHRR --RR---RCC RRR--YTRRC KRY-
HSP1_COLGU       MARYRCRRSQ --SRSRCC-R ---Q-RRRCR -RRRRQRFRA --RKRAMRCC HRR--YRLRC RRY-
F7VJK5_ERIEU     MARYRCCRSQ --SRSRCS-R RRYR-RRRCR -RRRRRSCRR --RRRR-ACC RYR--Y---- RRY-
C3U1R1_9MURI     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCQR --RR--RRCC RRRRSYTIRC KKY-
HSP1_MOUSE       MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1R5_MUSSP     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1Q6_MUSMB     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1Q9_MOUSE     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1R0_MOUSE     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1Q8_MUSMC     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1R4_MUSSI     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1Q7_MUSCO     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1R3_MUSMA     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTIRC KKY-
C3U1R2_MUSPA     MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--QRCC RRRRSYTIRC KKY-
HSP1_RAT         MARYRCCRSK --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRRRSYTFRC KRY-
C3U1S7_APOSY     MARYRCCRSK --SRSRCR-R ---R-RRRCR SRRRRRCCRR --RR--RRCC RRRRSYTIRC KRY-
HSP1_NASLA       MAKSRCCGSQ --SRSRCC-R ---P-RQRCR -RRRRRSCRA --RRRAMRCC RRR--YRLRC RRY-
HSP1_TRAJO       MARYRRCRSQ --SRSRCC-R ---P-RRRCR -RRRRQSCRA --RRRATRCC RRR--YRLRC RRY-
HSP1_TRAPH       MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRL-S RRY-
L0CM34_TRAPH     MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRL-S RRY-
HSP1_TRACR       MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRLRS RRY-
HSP1_TRAOB       MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRLRS RRY-
L0CM39_9PRIM     MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRLRS RRY-
L0CNC1_TRAPH     MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRLRS RRY-
HSP1_SEMEN       MARYRRCRSQ --SRSRCC-R ---P-RRRCR --RRRRSCRA --RRRATRCC RRR--YRLRC RRY-
HSP1_TRAVT       MARYRRCRSQ --SRSRCC-R ---P-RRRCR --RRRRSCRA --RRRATRCC RRR--YRLRC RRY-
HSP1_TRAPL       MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRLRC RRY-
HSP1_TRAGE       MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRLRC RRY-
HSP1_TRAFR       MARYRCCRSQ --SRSRCC-R ---P-RRRCR -RRRRRSCRA --RRRATRCC RRR--YRLRC RRY-
S4SNK2_9PRIM     -------RSQ --SRSRCC-R ---P-RRRCR --RRRRSCRA --RRRATRCC RRR--YRLRC RRY-
S4SLN4_9PRIM     -------RSQ --SRSRCC-R ---P-RRRCR --RRRRSCRA --RRRATRCC RRR--YRLRC RRY-
F7VJK8_OTOGA     MARYRCCRSQ --SRSRCR-R ---R-RRRCR -RRRRRCCRR --RR--RRCC RRR--YRLRC RRY-
HSP1_PILBA       MARYRCCRSQ --SRSRCC-R ---R-RRRCR -RRRRQRCRA --RRTAMRCC RRR--YRRRC RRY-
HSP1_MACMU       MARYRCCRSQ --SRSRCC-R ---R-RRRCR -RRRRRRCRA --RRRAMRCC RRR--YRLRC RRY-
HSP1_PAPCY       MARYRCCRSQ --SRSRCC-R ---R-RRRCR -RRRRRRCRA --RRRAMRCC RRR--YRLRC RRY-
A0A096MLL5_PAPAN MARYRCCRSQ --SRSRCC-R ---Q-RRRCR -RRRRRRCRA --RRRAMRCC RRR--YRLRC RRY-
HSP1_HYLLA       MARYRCCRSQ --SRSRCY-R ---R-GQRSR -RRRRRSCQT --RRRAMRCC RPR--YRLRR RRH-
HSP1_HUMAN       MARYRCCRSQ --SRSRYY-R ---Q-RQRSR -RRRRRSCQT --RRRAMRCC RPR--YRPRC RRH-
HSP1_GORGO       MARYRCCRSQ --SRSRCY-R ---Q-RQTSR -RRRRRSCQT --QRRAMRCC RRR--NRLRR RKH-
HSP1_PANTR       MARYRCCRSQ --SRSRCY-R ---Q-RQRSR -RRKRQSCQT --QRRAMRCC RRR--SRMRR RRH-
HSP1_PANPA       MARYRCCRSQ --SRSRCY-R -----QRRSR -RRKRQSCQT --QRRAMRCC RRR--SRLRR RRH-
```


Supplemental Figure 2. Alignment of Eutherian P1 type sperm protamines.

```
                    10         20         30         40         50         60
PRM2_RATTU      MVRYRMRSPS ESPHQGPGQD HESEEQG--- -QGQELNPER VEDYGRTHRG --HHRHRRCS
PRM2_MOUSE      MVRYRMRSPS EGPHQGPGQD HEREEQG--- -QGQGLSPER VEDYGRTHRG HHHHRHRRCS
PRM2_RATFU      MVRYRMRSPS EGPHQGPGQD HEREEQG--- -QGQELSPER VEDYGRTHRG --HHRHRRCS
PRM2_ALOSE      MVRYHVRSPS ERPHREYRQL VNGQEQGRHG QEEQGMSAEG VEGYGRTHQG CYGYRRRLCS
PRM2_CALJA      MVRYRVRSPS ERPHEEYRQL VNWQEQGRNG QEEQGLSAEG GEVYGRTHQG YSSYRRRRCS
PRM2_SEMEN      MVRYRMRSLS ERPHEVHGQQ VYGQEQGHNG QEEQGLSPEH VEVYERTHQG YSHHRRRRCS
PRM2_ERYPA      MVRYRTRSLS ERPHEVHGQQ VHGQDQGHNG QEEQGLSPEH VEVYERTHQG HSHHRRRRCS
PRM2_MACNE      MVRYRMRSLS ERPHEVHGQQ VHGQDQGHNG QEEQGLNPEH VEVYERTHRG HSHHRRRRCS
PRM2_MACFU      MVRYRMRSLS ERPHEVHGQQ VHGQDQGHNG QEEQGLNPEH VEVYERTHRG HSHHRRRRCS
PRM2_MACMU      MVRYRMRSLS ERSHEVHGQQ VHGQDQGHNG QEEQGLNPEH VEVYERTH-G HSHYRRRHCS
PRM2_GORGO      MVRCRVRSPS ERSHEVYRQQ LHGQEQGHHG QEEQGLSPEH VEVYERTH-G HSHYRRRHCS
PRM2_PANPA      MVRYRVRSPS EPSHEVYRQQ LHGQEQGHHG QEEQGLSPEH VEVYERTH-G HSHYRRRHCS
PRM2_PANTR      MVRYRVRSPS EPSHEVYRQQ LHGQEQGHHG QEEQGLSPEH VEVYERTH-G HSHYRRRHCS
PRM2_HUMAN      MVRYRVRSLS ERSHEVYRQQ LHGQEQGHHG QEEQGLSPEH VEVYERTH-G QSHYRRRHCS
PRM2_PONPY      MVRYCVRSLS ERSHEVYGQQ LHGQEQGHHD QEEQGLSPEQ VEVYERTQ-G HSHYRRRHCS
PRM2_HYLLA      MVRYCVRSLS ERSHEVYGQQ LRGQEQGHHG QEEQGLSPED VEVYERTH-G HSHYRRRHCS
                    70         80         90        100        110        120
PRM2_RATTU      RKRLHRIHKR R-RSCRRRRR HSCCHRRRHR RGCRRSRRRR RCRCRKCRRQ CH
PRM2_MOUSE      RKRLHRIHKR R-RSCRRRRR HSCRHRRRHR RGCRRSRRRR RCRCRKCRRH HH
PRM2_RATFU      RKRLHRIHKR R-RSCRRRRR HSCCHRRRHR RGCRRSRRRR RCKCRKCRRH CH
PRM2_ALOSE      RRRLYRVHRR QRRSCRRRC- --CRYRRRNR RGCRT-RRRT ------CRRH --
PRM2_CALJA      RRRRYRIHRR RSRSCRRRRR RSCRYRRRPR RGCRSRRRRR ------CRRY --
PRM2_SEMEN      RRRLYRIHRR RHRSCRRRRR RSCRHRRRHR RGCRT-RRRR ------CRRY --
PRM2_ERYPA      QRRLHRIHRR RHRSCRRRRR RSCRHRRRHR RGCRT-RRRR ------CRRY --
PRM2_MACNE      RRRLHRIHRR RHRSCRRRRR RSCRHRRRHR RGCRT-RRRR ------CRRH --
PRM2_MACFU      RRRLHRIHRR RHRSCRRRRR RSCRHRRRHR RGCRT-RRRR ------CRRH --
PRM2_MACMU      RRRLHRIHRR RHRSCRRRRR RSCRHRRRHR RGCRT-RRRR ------CRRH --
PRM2_GORGO      RRRLRRIHRQ QHRSCRRRKR RSCRHRRRHR KGCRT-RRRT ------CRRH --
PRM2_PANPA      RRRLRRIHRQ QHRSCRRRKR RSCRHRRRHR RGCRT-RRRT ------CRKH --
PRM2_PANTR      RRRLRRIHRQ QHRSCRRRKR RSCRHRRKHR RGCRT-RRRT ------CRRH --
PRM2_HUMAN      RRRLHRIHRR QHRSCRRRKR RSCRHRRRHR RGCRT-RKRT ------CRRH --
PRM2_PONPY      RRRLHRIHRQ QHRSCKRRRR HSCRHRRKHR RGCRT-RRRT ------CRRH --
PRM2_HYLLA      RRRLHRIHRQ QHRSCGRRRR RSCRQRRRHR RGCRT-RRRR ------CRRH --
```

Supplemental Figure 3. Alignment of Eutherian P2 type sperm protamines.

```
                        10         20         30         40         50         60
HSP1_PERGU      MASYR-NSRS RSRSRF-RRR RRGRSRVRGR DARQGRSSR- RRRRGKGR-- --AHSGKKGR
HSP1_ISOMA      MASYR-NSRS RSRSRF-RRR RGRRSRVRGR DARQGRSSR- RRRRGKGR-- --AHSGKKGR
HSP1_MONDO      MARYRRRSRS RSRSRYGRRR RRSRS----R RRRSRRRRR- ---RRGRR-- ---GRGYHRR
HSP1_DIDVI      MARYRRRSRS RSRSRYGRRR RRSRS----R RRRSRRRRR- ---RRGRR-- ---GRGYHRR
HSP1_DROGL      MVRYRRHSRS RSRSRY-RRR RRRR--LRNR RRRYRRSRRG RRRRRRGS-- ---RRGYSRR
HSP1_NOTTY      MARYR-HSRS RSRSRY-RRR RRRRSRYRSQ RRRYRRHRRS GRRRRRGR-- ---RRGY-RR
HSP1_PHACI      MARYR-HSRS RSRSRY-QRR RRRRSRYRSQ RRRYRRRRGS RRRRRRGR-- ---RRGY-RR
HSP1_HYPMS      MARYR-HSRS RSRSRY-RRR RRRRSRYRGR RRRYRRSRR- RRRSRRGR-- ---RGYYRRR
HSP1_LAGHI      MARYR-HSRS RSRSGY-RRQ RRRRSRYRSR RRRYRRR-Q- -RRSRRGR-- ---RRGYSRR
HSP1_DENDO      MARYR-HSRS RXRSRY-RRR RRXRSRYRSX RRRYRGRRR- RRSRRGRR-- ---RRGYSRX
HSP1_AEPRU      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGSRR- RRRSRRRR-- ---RRGYSRR
HSP1_PSECU      MARYR-HSRS RSRSRYRRRR RRRRSRYRGR RRRYRRSRR- RRRRGRRRGN CLGRRGYRRR
HSP1_TRIVU      MARYR-HSRS RSRSRYRRRR RRRRSRYRSR RRRYRRSRR- -RRRRGRR-- ---RRGYSRR
HSP1_CAEFU      MARYR-HSRS RSRSRYRRRR RRRRSRYRSR RRRYRRSRR- --RRRRGR-- ---RRGYSRR
HSP1_POTLO      MARYR-HSRS RSRXRY-RRR RRRRSRYRSR RRRYRGSRR- SRSRRRGR-- ---RRGYSRR
HSP1_LAGFA      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGSRR- SRSRRRGR-- ---RRGYSRR
HSP1_BETPE      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGSRR- RRSRRRGR-- ---RRGYSRR
HSP1_ONCFR      MARYR-HSRS RSRSXY-RRR XRRRSRYRSR RRRYRGRRR- -RRSRRGR-- ---RRGYSRR
HSP1_MACRU      MARYR-HSRS RSRSRY-RRR RRRRSRYRSQ RRRYRGRRR- -RRSRRGR-- ---RRGYSRR
HSP1_PETCN      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRXRR--R- RRSRRGRR-- ---RRGYSRR
HSP1_MACPA      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- -XRSRRGR-- ---RRGYSRR
HSP1_MACEU      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRSRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_MACAG      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRSRGRRR- -RRSRRGR-- ---RRGYSRR
HSP1_DORVA      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRXYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_PETXA      MARYR-HSXS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_THYST      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSCR
HSP1_MACGI      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_DORMU      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_WALBI      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_MACRG      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- RRSRRGRR-- ---RRGYSRR
HSP1_ONYUN      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- -RRSRRGR-- ---RRGYSRR
HSP1_DENGO      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- -RRSRRGR-- ---RRGYSRR
HSP1_SETBR      MARYR-HSRS RSRSRY-RRR RRRRSRYRSR RRRYRGRRR- -RRSRRGR-- ---RRGYSRR
HSP1_DASHA      MARYRRRSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SARHA      MARYRRRSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASMA      MARYRRRSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASSP      MARYRRHSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_NEOLO      MARYRRHSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASVI      MARYRRHSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PHADO      MARYRRHSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASAL      MARYRRHSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASGE      MARYRRHSRS RSRSRY-RRR RRRRSRGR-R RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASRO      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PSEBA      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PARAP      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PSENI      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASBY      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_MYOME      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_MYOWA      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PSEMD      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PSEWO      MARYRRHSRS RSRSRY-RRR RRRRSRHRNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIGR      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRRR-- ---RRGYSRR
HSP1_SMIBI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSIRRRGR-- ---RRGYSRR
HSP1_SMILO      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTFL      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_NINTI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_NINYV      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PARRT      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMILE      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PHATA      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIPS      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTBE      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_MICHA      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTGO      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTMI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTNA      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTSW      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_MURME      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIDO      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIHI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIMA      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIOO      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIYO      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PLAMM      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIAR      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_DASCR      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTLE      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTST      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_MURLO      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_MYRFA      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_NINRI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIGA      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIGI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PHACL      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIMU      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIVI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_THYCY      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIAI      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMICR      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_SMIDL      MARYRRHSRS RSRSRY-RRR RRRRSRHHNR RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_ANTLA      MARYRRHSRS RSRSRY-RRR RRRRSRHHN- RRTYRRSRR- HSRRRRGR-- ---RRGYSRR
HSP1_PLAMS      MARCRRHSRS RSRSRN-QCQ RRRRRRY-NR RRTYRRSRR- HSRRRRGR-- ---RRGCSRR
HSP1_PLAIN      MARSRRHSRS RSRSR--NQC QRRRRRTYNR RRTMREKPR- HSRRRRVR-- ---RRGCSCR
HSP1_PLAGI      MARCRRHSRS RSRSRN-QCQ RRRR-RHYNR RRTYRRSRR- HSRRRRVR-- ---RRGCSCR
HSP1_PLATE      MARCRRHIRS RSRSRN-QCQ RRRRRSHYNR RRTYRRSRR- HSRRRRVR-- ---RRGCSCR
                        70         80         90
HSP1_PERGU      RSG-----SR RRKRNNNTEN K
HSP1_ISOMA      RSG-----SR RRKRNNENK- -
HSP1_MONDO      SPHRR---RR RRRR------ -
HSP1_DIDVI      SPHRR---RR RRRR------ -
HSP1_DROGL      RYQSR---RR RRRRY----- -
HSP1_NOTTY      RYH------S HRRRY----- -
HSP1_PHACI      RYS-----RR --RRY----- -
HSP1_HYPMS      RYS-----RR RRRRYYY--- -
HSP1_LAGHI      RYSRRRYSRR RRRRY----- -
HSP1_DENDO      RYS-----RR RRRRY----- -
HSP1_AEPRU      RYS------R RRRRY----- -
HSP1_PSECU      RYS-----RR RRRRYY---- -
HSP1_TRIVU      RYS-----RR GRRRY----- -
HSP1_CAEFU      RYS-----RR RRRRY----- -
HSP1_POTLO      RYS------R RRRRY----- -
HSP1_LAGFA      RYS-----RR RRRRY----- -
HSP1_BETPE      RYS-----RR RRRRY----- -
HSP1_ONCFR      RYS-----RR RRRRY----- -
HSP1_MACRU      RYS------R RRRRY----- -
HSP1_PETCN      RYS-----RR RRRRY----- -
HSP1_MACPA      RYS-----RR RRRRY----- -
HSP1_MACEU      RYS-----RR RRRRY----- -
HSP1_MACAG      RYS-----RR RRRRY----- -
HSP1_DORVA      RYS-----RR RRRRY----- -
HSP1_PETXA      RYS-----RR RRRRY----- -
HSP1_THYST      RYS-----RR RRRRY----- -
HSP1_MACGI      RYS-----RR RRRRY----- -
HSP1_DORMU      RYS-----RR RRRRY----- -
HSP1_WALBI      RYS-----RR RRRRY----- -
HSP1_MACRG      RYS-----RR RRRRY----- -
HSP1_ONYUN      RYS-----RR RRRRY----- -
HSP1_DENGO      RYS-----RR RRRRY----- -
HSP1_SETBR      RYS-----RR RRRRY----- -
HSP1_DASHA      RYS-----RR GRRR------ -
HSP1_SARHA      RYS-----RR GRRRY----- -
HSP1_DASMA      RYS-----RR GRRRY----- -
HSP1_DASSP      RYS-----RR GRRRY----- -
HSP1_NEOLO      RYS-----RR GRRRY----- -
HSP1_DASVI      RYS-----RR GRRRY----- -
HSP1_PHADO      RYS-----RR GRRRY----- -
HSP1_DASAL      RYS-----RR GRRRY----- -
HSP1_DASGE      RYS-----RR GRRRY----- -
HSP1_DASRO      RYS-----RR GRRRY----- -
HSP1_PSEBA      RYS-----RR GRRR------ -
HSP1_PARAP      RYS-----RR GRRRY----- -
HSP1_PSENI      RYS-----RR GRRRY----- -
HSP1_DASBY      RYS-----RR GRRRY----- -
HSP1_MYOME      RYS-----RR GRRRY----- -
HSP1_MYOWA      RYS-----RR GRRRY----- -
HSP1_PSEMD      RYS-----RR GRRRY----- -
HSP1_PSEWO      RYS-----RR GRRRY----- -
HSP1_SMIGR      RYS-----RR GRRRY----- -
HSP1_SMIBI      RYS-----RR GRRRY----- -
HSP1_SMILO      RYS-----RR -RRRY----- -
HSP1_ANTFL      RYS-----RR GRRRY----- -
HSP1_NINTI      RYS-----RR GRRRY----- -
HSP1_NINYV      RYS-----RR GRRRY----- -
HSP1_PARRT      RYS-----RR GRRRY----- -
HSP1_SMILE      RYS-----RR GRRRY----- -
HSP1_PHATA      RYS-----RR GRRRY----- -
HSP1_SMIPS      RYS-----RR GRRRY----- -
HSP1_ANTBE      RYS-----RR GRRRY----- -
HSP1_MICHA      RYS-----RR GRRRY----- -
HSP1_ANTGO      RYS-----RR GRRRY----- -
HSP1_ANTMI      RYS-----RR GRRRY----- -
HSP1_ANTNA      RYS-----RR GRRRY----- -
HSP1_ANTSW      RYS-----RR GRRRY----- -
HSP1_MURME      RYS-----RR GRRRY----- -
HSP1_SMIDO      RYS-----RR GRRRY----- -
HSP1_SMIHI      RYS-----RR GRRRY----- -
HSP1_SMIMA      RYS-----RR GRRRY----- -
HSP1_SMIOO      RYS-----RR GRRRY----- -
HSP1_SMIYO      RYS-----RR GRRRY----- -
HSP1_PLAMM      RYS-----RR GRRRY----- -
HSP1_SMIAR      RYS-----RR GRRRY----- -
HSP1_DASCR      RYS-----RR GRRRY----- -
HSP1_ANTLE      RYS-----RR GRRRY----- -
HSP1_ANTST      RYS-----RR GRRRY----- -
HSP1_MURLO      RYS-----RR GRRRY----- -
HSP1_MYRFA      RYS-----RR GRRRY----- -
HSP1_NINRI      RYS-----RR GRRRY----- -
HSP1_SMIGA      RYS-----RR GRRRY----- -
HSP1_SMIGI      RYS-----RR GRRRY----- -
HSP1_PHACL      RYS-----RR GRRRY----- -
HSP1_SMIMU      RYS-----RR GRRRY----- -
HSP1_SMIVI      RYS-----RR GRRRY----- -
HSP1_THYCY      RYS-----RR GRRRY----- -
HSP1_SMIAI      RYS-----RR GRRRY----- -
HSP1_SMICR      RYS-----RR GRRRY----- -
HSP1_SMIDL      RYS-----RR GRRRY----- -
HSP1_ANTLA      RYS-----RR GRRRY----- -
HSP1_PLAMS      RYS-----RR GRRRY----- -
HSP1_PLAIN      RCS-----RR RRRRC----- -
HSP1_PLAGI      RCS-----RR RRRRC----- -
HSP1_PLATE      RCS-----RR RRRRC----- -
```


Supplemental Figure 4. Alignment of Metatherian P1 type sperm protamines.