

RESEARCH ARTICLE

Open Access



# Proteomic and evolutionary analyses of sperm activation identify uncharacterized genes in *Caenorhabditis* nematodes

Katja R. Kasimatis, Megan J. Moerdyk-Schauwecker, Nadine Timmermeyer and Patrick C. Phillips\*

## Abstract

**Background:** Nematode sperm have unique and highly diverged morphology and molecular biology. In particular, nematode sperm contain subcellular vesicles known as membranous organelles that are necessary for male fertility, yet play a still unknown role in overall sperm function. Here we take a novel proteomic approach to characterize the functional protein complement of membranous organelles in two *Caenorhabditis* species: *C. elegans* and *C. remanei*.

**Results:** We identify distinct protein compositions between membranous organelles and the activated sperm body. Two particularly interesting and undescribed gene families—the Nematode-Specific Peptide family, group D and the here designated Nematode-Specific Peptide family, group F—localize to the membranous organelle. Both multigene families are nematode-specific and exhibit patterns of conserved evolution specific to the *Caenorhabditis* clade. These data suggest gene family dynamics may be a more prevalent mode of evolution than sequence divergence within sperm. Using a CRISPR-based knock-out of the NSPF gene family, we find no evidence of a male fertility effect of these genes, despite their high protein abundance within the membranous organelles.

**Conclusions:** Our study identifies key components of this unique subcellular sperm component and establishes a path toward revealing their underlying role in reproduction.

**Keywords:** Spermiogenesis, Proteomics, Molecular evolution, Nematodes

## Background

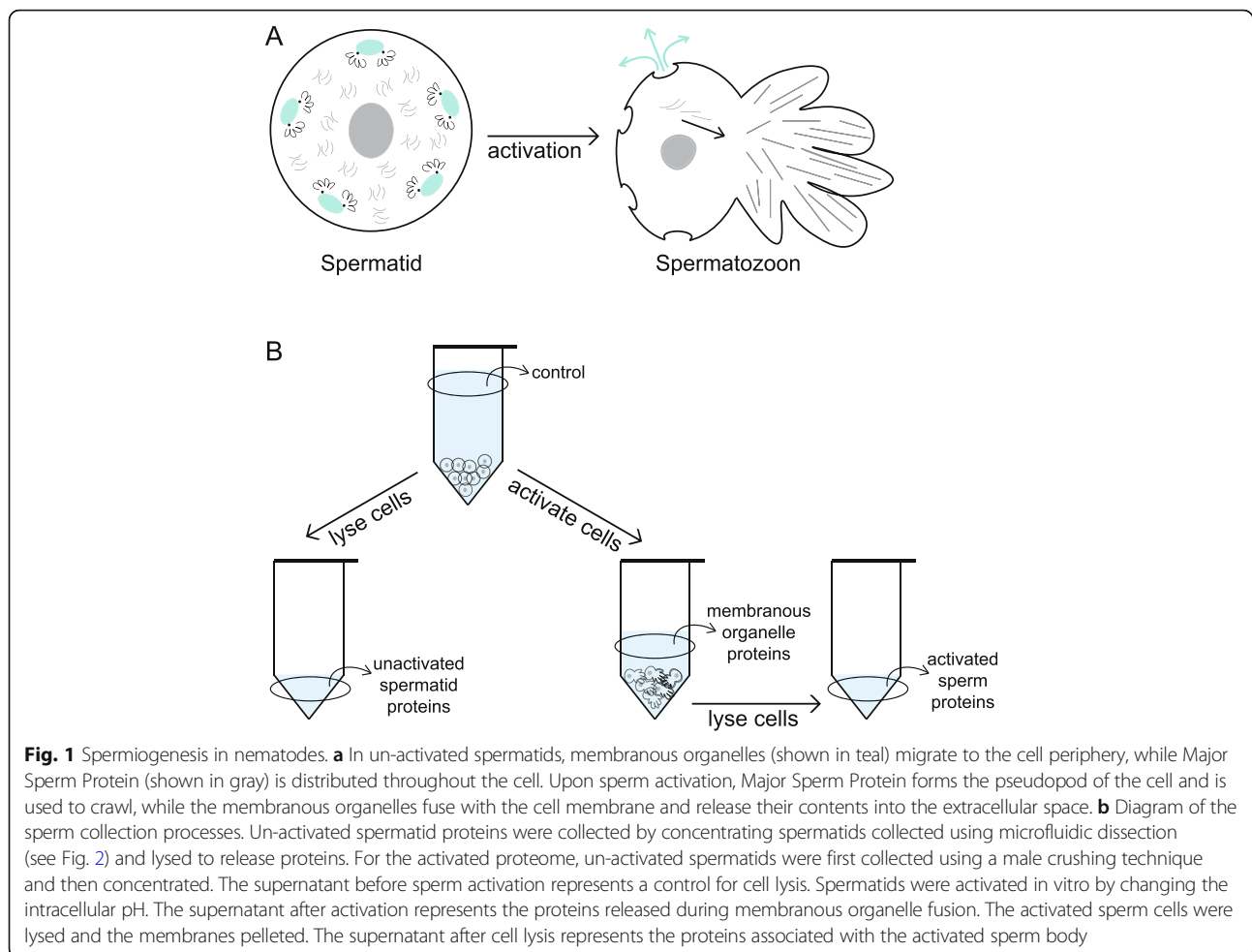
Despite coming in a wide variety of morphologies, sperm exhibit three key cellular traits that are widely conserved across metazoans (reviewed in [1, 2]). First, it appears all sperm undergo a histone-to-protamine chromatin condensation [3]. Second, the vast majority of sperm swim using a flagellum coupled to an actin/myosin cytoskeleton [4]. Third, most sperm contain an acrosome or acrosome-like membrane domain that aids in sperm-egg recognition and fusion [5]. In contrast to other animals, the phylum Nematoda has a distinctly different sperm morphology and molecular biology [6]. Namely, nematodes have large, amoeboid-like sperm cells that use non-actin mediated locomotion [6]. While other species with aflagellate sperm rely on passive diffusion for locomotion [1, 4], nematodes use Major Sperm Protein

(MSP)-mediated motility to crawl [6, 7]. Nematode sperm also lack an acrosome [6], and membrane remodeling during spermiogenesis (sperm activation) is instead largely driven by membranous organelles [8]. Both the use of MSP-mediated motility and the presence of membranous organelles are critical components of nematode sperm biology that are unique to and conserved across this ancient phylum.

Perhaps not surprisingly, these two unique components of nematode sperm interact with one another throughout spermatogenesis. Membranous organelles are membrane bound vesicles derived from the Golgi that are found throughout the dividing cell [8]. During spermatogenesis membranous organelles and MSP associate to form fibrous body membranous organelles. As spermatogenesis concludes, these fibrous body membranous organelles dissociate and the membranous organelles migrate to the cell periphery while the MSP remains distributed throughout the cytoplasm (Fig. 1a) [9]. During spermiogenesis MSP forms branching

\* Correspondence: [pphil@uoregon.edu](mailto:pphil@uoregon.edu)  
Institute of Ecology and Evolution, University of Oregon, 5289 University of Oregon, Eugene, OR 97403-5289, USA





filaments, which structure the pseudopod of motile sperm [10, 11]. Meanwhile, the membranous organelles remain associated with the cell body, fusing with the cell membrane to create cup-like structures reminiscent of secretory vesicles (Fig. 1a) [7, 8]. Unlike an acrosome reaction, however, the membranous organelles fuse prior to any contact with an oocyte. The role of membranous organelles and the function of these fusion events remains unknown, largely because of the challenge of studying subcellular components in single gametes. Nevertheless, mutant screens targeting faulty spermatogenesis have shown that incorrect membranous organelle fusion results in sterility [12–14] and therefore that these organelles must play an important functional role within sperm. One hypothesis for membranous organelle function is that the increased membrane surface area and incorporation of additional proteins is important for membrane microdomain remodeling and fluidity [15, 16]. Since membranous organelles release their contents into the extracellular space, they may have an additional function as a source of seminal fluid proteins and therefore be involved in post-insemination reproductive tract

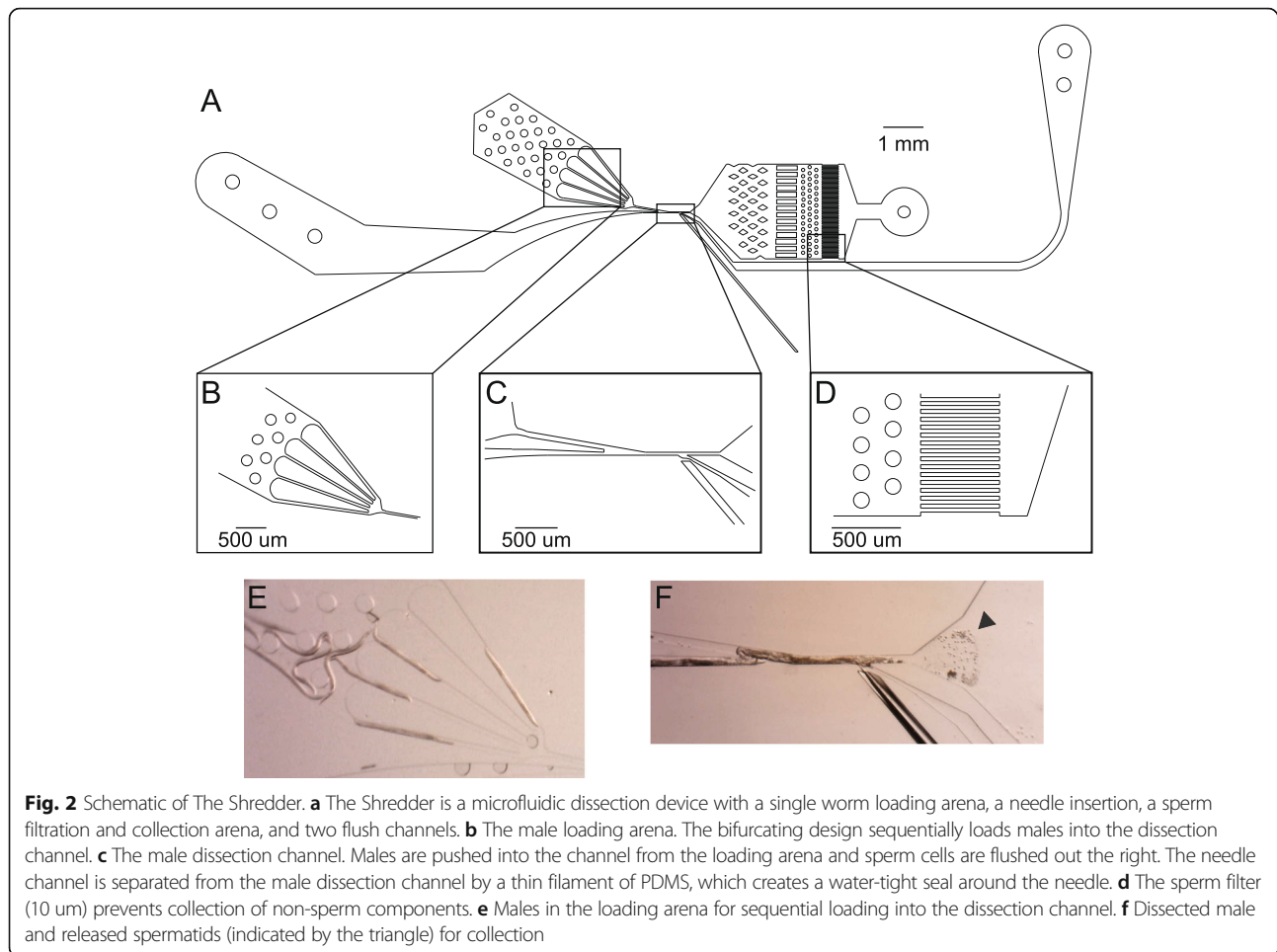
dynamics. However, without information on the composition of membranous organelles, determining the full functional role of their fusion is a challenge.

Here we take a novel approach that co-opts sperm activation events to proteomically characterize membranous organelles within two *Caenorhabditis* species. We identify two particularly interesting gene families—the Nematode-Specific Peptide family, group D and Nematode-Specific Peptide family, group F—that are previously undescribed and use evolutionary analysis and genomic knockouts to more directly probe their function.

## Results

### Proteomic characterization of spermiogenesis in *C. elegans*

Un-activated spermatids were collected from males using a novel microfluidic dissection technique. This male dissection technique utilizes a custom microfluidic device with a fine glass needle to slice through the cuticle and testis of males to release stored spermatids (Fig. 2). The un-activated spermatids were lysed to characterize non-membrane-bound sperm proteins

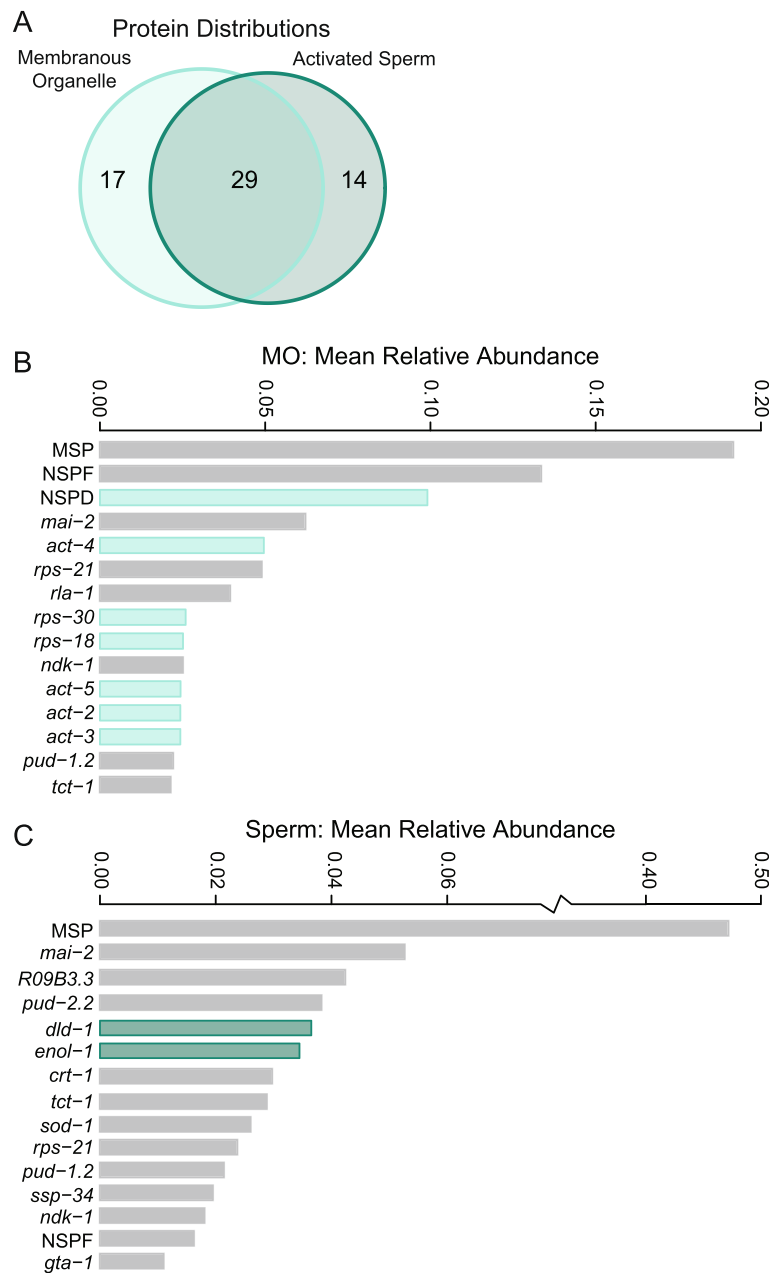


(Fig. 1b). The un-activated spermatid proteome was dominated by the MSP, confirming that pure sperm cell samples were being collected (Additional file 1). The most abundant proteins, however, were from the Nematode-Specific Peptide family, group D (NSPD), which comprised approximately 50% of the total protein abundance. Since mass spectrometry identified a single peptide motif for these proteins, NSPD abundance was described at the gene family level. The NSPD family is uncharacterized, but has been previously shown to exhibit a pattern of male-enriched expression [17]. Actin proteins were also identified at < 1% abundance, which is comparable to previous biochemical estimates [6]. While relatively few total protein calls were made, fully one third of the un-activated spermatid proteome is previously uncharacterized in biological function.

To isolate soluble proteins within the membranous organelle from those associated with the sperm body, we took advantage of natural membranous organelle-membrane fusion during sperm activation. Since this analysis required a higher-throughput, un-activated spermatids were collected using a male crushing technique (modified from [18, 19]).

This method squeezes the testis out of males to release spermatids. Spermatids were then activated *in vitro* by changing the intracellular pH [8] and the proteomes of the membranous organelle secretions and activated sperm fractions were collected via centrifugation (Fig. 1b). Again, the MSP was in high abundance, though now identified in both the membranous organelle and activated sperm proteomes (Fig. 3). Interestingly, our data reveal three previously unannotated genes (Y59E9AR.7, Y59H11AM.1, and ZK1248.4) as MSPs based on high nucleotide sequence identity and presence of the MSP domain [20]. Overall, 62% of the proteins identified in the un-activated spermatid proteome were also identified in either the membranous organelle or activated sperm proteome. The lack of one-to-one correspondence between the un-activated proteome and the two activated components is unsurprising given the low total number of proteins identified and the pseudo-quantitative nature of shotgun proteomics. Nevertheless, all the proteins identified were previously found in the un-activated spermatid proteome collected by Ma et al. [21].

The proteins released from the membranous organelle during activation were distinct from those remaining in



**Fig. 3** Proteomic characterization of the membranous organelle and activated sperm proteomes in *C. elegans*. **a** The two proteomes were distinct, with 17 proteins found only in membranous organelles and 14 proteins found only in activated sperm. **b** The 15 most abundant proteins identified in the membranous organelles. Proteins unique to membranous organelles (highlighted in teal) include the Nematode-Specific Peptide family, group D (NSPD) as well as several housekeeping gene families. **c** The 15 most abundant proteins identified in activated sperm. The proteins unique to activated sperm (highlighted in teal) are predominantly involved in energy production. Protein abundance is shown as the relative mean normalized spectrum abundance frequency

the activated sperm (Fig. 3a). Seventeen proteins were unique to the membranous organelle proteome, including the NSPD family, which comprised 10% of the total membranous organelle protein abundance (Fig. 3b). The actin gene family was also unique to the membranous organelle, as were several other housekeeping-related gene families. Within the activated sperm proteome, we

identified 14 unique proteins, the majority of which were involved in energy production (Fig. 3c). Of noticeable interest were the genes F34D6.7, F34D6.8, and F34D6.9, which again were described using a single abundance measure due to identical mass spectrometry peptide sequence identification. These genes were in fact the most abundant membranous organelle protein after MSP, with a

ten-fold greater abundance in membranous organelles than in activated sperm (Fig. 3b–c). The F34D6.7, F34D6.8, and F34D6.9 genes in *C. elegans*, display male-specific expression [17], consistent with our observations. They are organized distinctly from other genes in this region as an array and have a nucleotide sequence similarity of 93.9%. Given their genomic organization, sequence similarity, and co-localization of expression, these genes appear to be a small gene family that originated via tandem duplication. Additionally, an amino acid blast search of these F34D6 sequences in NCBI reveals that they are nematode-specific. Thus, they comprise a newly identified Nematode-Specific Peptide family, which we designate as NSP group F (NSPF).

### Proteome composition is largely conserved between species

Spermatids were also collected from the obligate outcrossing nematode *C. remanei*. To compare proteome composition between divergent species, we condensed all protein calls to the gene family level. Within *C. remanei*, we identified 64 gene families in the membranous organelle proteome and 94 gene families within the activated sperm proteome, with 51 families being shared between the proteomes (Additional file 2). Of all the proteins identified, eight did not have an annotated *C. elegans* ortholog. However, a BLAST search against the *C. elegans* genome indicates that three of these genes (CRE18007, CRE13415, CRE00499) may have unannotated orthologs. Of the remaining unique genes, three appear to be paralogs (CRE12049, CRE30219, CRE30221), suggesting a potential *C. remanei*-specific sperm protein family. A total of 34 gene families were identified in both *C. elegans* and *C. remanei*, capturing the majority of highly abundant genes identified. However, more proteins of low abundance were identified in *C. remanei*. Three gene families – NSPD, Actin, and Ribosomal Proteins, Large subunit – unique to the membranous organelle proteome in *C. elegans* were identified in low abundance within activated sperm in *C. remanei*, potentially because of differential success in activating *C. remanei* sperm in vitro (Additional file 2). Two noticeable differences between species were the presence of histone proteins and the absence of NSPF orthologs in *C. remanei*.

### Evolutionary analysis of membranous organelle proteins

Proteomic analysis identified NSPD and NSPF proteins as being highly abundant and localized their expression to the membranous organelle. Yet no information exists about the molecular or biological function of these genes. To better understand the nature of these gene families, we analyzed their evolutionary history across the *Elegans* supergroup within *Caenorhabditis*. We made custom annotations of these gene families in 11 species using the annotated *C. elegans* genes (ten NSPD and

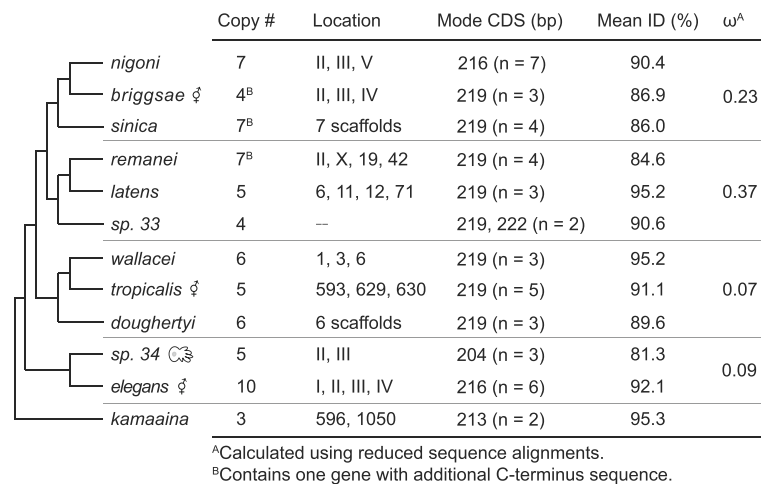
three NSPF) as the query dataset. Our sampling included the three lineage transitions to self-fertilizing hermaphroditism [22, 23] and the single lineage transition to sperm gigantism [24] found within this supergroup.

Across all 12 species we identified 69 NSPD homologs (Additional file 3). The NSPD gene family ranged from three to ten gene copies, with *C. elegans* having the highest copy number and *C. kamaaina* having the lowest (Fig. 4). Coding sequence length was largely conserved between paralogs, but differed across species. Sequence length differences were particularly driven by a 24–30 base pair region in the middle of the gene containing repeating of asparagine and glycine amino acids, which tended to be the same length within a species, but differed across species (Additional file 4). Despite these species-specific repeats, amino acid sequence identity between paralogs was high, ranging from 81.3 to 95.3%. No secondary structure was predicted for these genes and in fact they were biochemically categorized as being 73% intrinsically disordered due to low sequence complexity and amino acid composition biases [25, 26].

The NSPD genes were broadly distributed across the genome, occurring as single copies on multiple chromosomes or scaffolds in each species (Additional file 3). This seemingly independent arrangement of individual genes throughout the genome precluded a robust syntentic analysis. Additionally, phylogenetic analysis showed NSPD genes predominantly cluster within species and thus they do not convey a strong signal of ancestral gene orthology (Additional file 5). Since orthologous genes could not be assigned, the protein coding sequences were analyzed within the four monophyletic clades represented. Even within these shorter evolutionary timescales, orthologous genes were not readily apparent, again suggesting species-specific evolution at the gene family level. To assess variation in evolutionary rate across the gene family, we estimated a single, alignment-wide ratio of non-synonymous to synonymous substitutions ( $\omega$ ) using reduced sequence alignments. Specifically, we removed the species-specific amino acid repeats in the middle of the gene, which were highly sensitive to alignment parameters. The  $\omega$ -values varied widely from 0.07 to 0.37 with the more recently derived clades having higher values (Fig. 4), although none indicate a strong signal of positive selection. Rather, these genes seem to be weakly constrained outside of the species-specific repeats, which was unexpected given their disordered nature.

We identified and annotated 22 NSPF orthologs in ten species (Additional file 3). Like the NSPD family, the NSPF genes do not have a predicted secondary structure and are 40% intrinsically disordered. They are, however, biochemically predicted to be signaling peptides (mean signal peptide score = 0.9) with a predicted cleavage site





	Copy #	Location	Mode CDS (bp)	Mean ID (%)	$\omega^A$
<i>nigoni</i>	7	II, III, V	216 (n = 7)	90.4	
<i>briggsae</i> ♀	4 <sup>B</sup>	II, III, IV	219 (n = 3)	86.9	0.23
<i>sinica</i>	7 <sup>B</sup>	7 scaffolds	219 (n = 4)	86.0	
<i>remanei</i>	7 <sup>B</sup>	II, X, 19, 42	219 (n = 4)	84.6	
<i>latens</i>	5	6, 11, 12, 71	219 (n = 3)	95.2	0.37
<i>sp. 33</i>	4	—	219, 222 (n = 2)	90.6	
<i>wallacei</i>	6	1, 3, 6	219 (n = 3)	95.2	
<i>tropicalis</i> ♀	5	593, 629, 630	219 (n = 5)	91.1	0.07
<i>doughertyi</i>	6	6 scaffolds	219 (n = 3)	89.6	
<i>sp. 34</i> C <sub>3</sub> S	5	II, III	204 (n = 3)	81.3	0.09
<i>elegans</i> ♀	10	I, II, III, IV	216 (n = 6)	92.1	
<i>kamaaina</i>	3	596, 1050	213 (n = 2)	95.3	

<sup>A</sup>Calculated using reduced sequence alignments.  
<sup>B</sup>Contains one gene with additional C-terminus sequence.

**Fig. 4** The evolution of the Nematode-Specific Peptide family, group D (NSPD) across the *Caenorhabditis elegans* Supergroup. Listed for each species are: the number of gene copies annotated, the genomic location (Roman numerals represent chromosome level assemblies and numbers represent scaffolds), the mode coding sequence length in base pairs (n = number of gene copies of said length), the mean amino acid sequence identity between paralogs, and the alignment-wide estimate of the ratio of non-synonymous to synonymous substitutions ( $\omega$ ). The complete gene annotation list is provided in Additional file 3 and the sequence alignments are given in Additional file 4

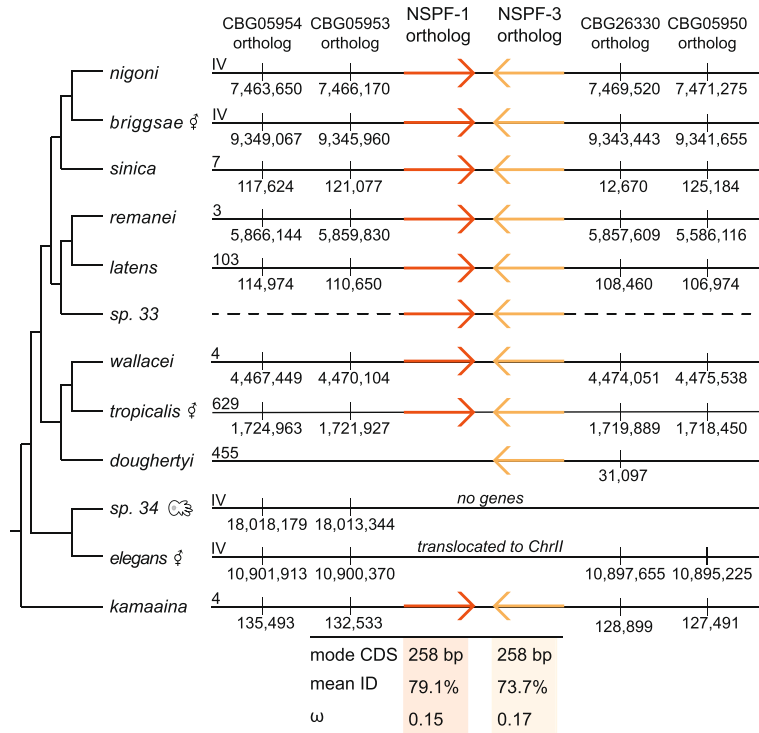
between amino acid residues 20 and 21 (Additional file 6). No genes were located within *C. sp. 34* genome (which is very well assembled). Nine species had two gene copies, while *C. doughertyi* has a single copy and, as mentioned, *C. elegans* has three annotated copies. Examination of 249 sequenced *C. elegans* natural isolates [27] suggests that *nspf-2* arose through a duplication of *nspf-1* as, while all copies of *nspf-1* align to the same position, there is variation in the intergenic space across the isolates. This duplication appears fixed within the *C. elegans* lineage—though one strain (CB4856) has a premature stop codon—and sequence identity is high between duplicates. Additionally, the *C. elegans* NSPF gene family has translocated to Chromosome II while the other species show conserved synteny to Chromosome IV (Fig. 5). Using syntenic relationships coupled with gene orientation and phylogenetic clustering, we were able to assign gene orthology within the family (Additional file 7). Within these orthologous groups, species relationships were largely recapitulated with  $\omega$ -values of 0.53 and 0.26 for the *nspf-1* and *nspf-3* orthologs, respectively. However, when the *C. elegans* lineage was excluded, the  $\omega$ -values sharply decreased to 0.15 for the *nspf-1* and 0.17 for the *nspf-3* orthologs, indicating a pattern of sequence constraint (Fig. 6). We explicitly tested if the *C. elegans* lineage was evolving at a different rate than the other lineages. Indeed, the *nspf-1* ( $\omega = 1.1$ , C.I. of  $\omega = 0.78$ –1.5,  $-2\Delta\ln = 5.11$ ) and to a lesser extent the *nspf-3* ( $\omega = 0.57$ , C.I. of  $\omega = 0.34$ –0.87,  $-2\Delta\ln = 2.34$ ) *C. elegans* lineages showed some evidence of positive selection, although the differences in the likelihoods of the two models were not statistical significant.

#### Functional analysis of the NSPF gene family

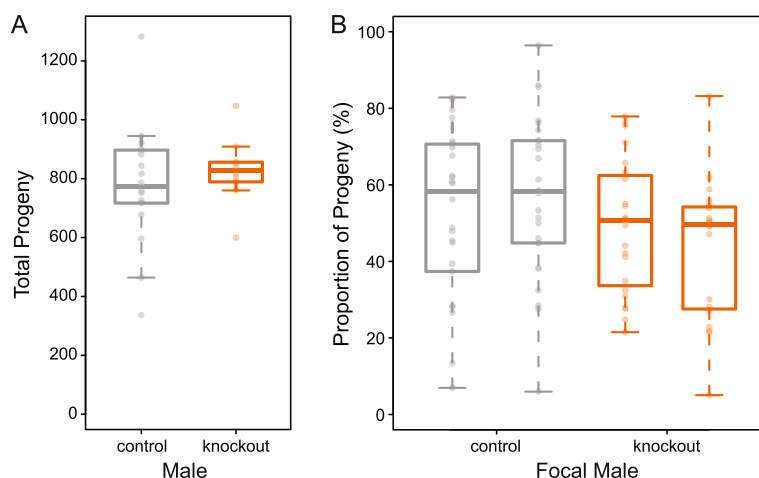
Given the high abundance of the NSPF protein, the conserved nature of these genes, and their potential as signaling peptides, we hypothesized these genes could be important for male fertility either during spermatogenesis or in sperm competition. Using CRISPR, we knocked out the three NSPF genes in the *C. elegans* standard laboratory strain (N2) to directly test the function of this gene family. We quantified male reproductive success, by allowing single males to mate with an excess of females over a 24 h period. Very little difference in progeny production was observed between knockout and wildtype males ( $t = -0.81$ ,  $df = 26$ ,  $p = 0.42$ ; Fig. 6a). Given the size of our experiment and the large sampling variance in individual fecundity, we would have been able to detect a difference between backgrounds of 24% with 80% power, so we possibly missed some effects if they were particularly subtle. We also measured the role of these genes in male competitive success, finding again that knocking out these genes had no effect on male fertility (Fig. 6b). In fact, knockout males were no worse competitors than wildtype males ( $z = -0.12$ ,  $p = 0.90$ ) and produced roughly 50% of the progeny measured (proportions test:  $\chi^2 = 1.27$ ,  $df = 1$ ,  $p = 0.26$ , C.I. of progeny produced = 27.4–55.9%). Overall, then, despite its prevalence within the sperm membranous organelle, the NSPF gene family does not appear to play an important role in male fertilization success.

#### Discussion

We used a proteomic approach coupled with molecular evolution analyses and direct functional assays to characterize the composition and role of membranous



**Fig. 5** The evolution of the Nematode-Specific Peptide family, group F (NSPF) across the *Elegans* Supergroup. The orthologous *nspf-1* and *nspf-3* genes are shown in orange on the chromosome or scaffold to which they locate. The Chromosome IV gene anchors used to determine synteny are shown. For each orthologous group the mode coding sequence length (in base pairs), the mean amino acid sequence identity, and the alignment-wide estimate of the ratio of non-synonymous to synonymous substitutions ( $\omega$ ) are shown. The *C. elegans* orthologs are excluded from the mean identity and  $\omega$  estimates as they show distinctly different patterns of evolution. The complete gene annotation list is provided in Additional file 3 and the sequences alignments are given in Additional file 6



**Fig. 6** Functional assays of the NSPF gene family in *C. elegans* male fertility. **a** In a non-competitive sperm setting, knockout males (orange) do not produce significantly fewer progeny than control males (gray) when given an excess of females with which to mate ( $t = -0.81$ ,  $df = 26.0$ ,  $p = 0.42$ ). **b** In a competitive sperm setting, knockout males (orange) do not produce significantly fewer progeny than control wildtype males ( $z = -0.12$ ,  $p = 0.90$ ) nor do they produce a significant deviation from 50% of the total progeny production (proportions test:  $\chi^2 = 1.27$ ,  $df = 1$ ,  $p = 0.26$ , C.I. of progeny produced = 27.4–55.9%). All fecundity data are provided in Additional file 7

organelles in nematode sperm. Our approach capitalized upon the natural sperm activation process to accurately isolate secreted membranous organelle proteins for the first time. This proteome set captures the most abundant proteins found in sperm and shows that the composition of the membranous organelle proteome is seemingly distinct from that of the activated sperm body. Since the complete proteomes were likely not identified, the abundance values presented are relative and therefore direct comparisons across samples is misleading. Nevertheless, interesting and uncharacterized gene families were identified as some of the most abundant proteins sampled. Unsurprisingly, the most abundant protein in activated sperm was the major sperm protein (MSP). Interestingly, MSPs were also the most abundant proteins in the membranous organelle. Since MSP proteins are important not only for motility, but also for oocyte signaling [28], identifying them as an abundant membranous organelle component implicates membranous organelle fusion as an additional method by which free-floating MSP is added to the seminal fluid (see [29]). There are 31 annotated MSP gene copies in *C. elegans*, with potentially more uncharacterized copies as seen here, and as of yet we do not know if some of them might be subfunctionally located within different parts of the sperm [20]. We also found that sperm proteome composition was largely conserved between *C. elegans* and *C. remanei*, particularly within the activated sperm itself. This is the first investigation of the proteome of a gonochoristic nematode. Although similarity is the rule, we did identify several *C. remanei* proteins lacking *C. elegans* orthologs, which are potentially a unique sperm family and warrant future molecular characterization, including determining if they are gonochoristic-specific genes.

Two gene families identified in the membranous organelles are particularly notable. First, the NSPD gene family was unique to the membranous organelle. This previously uncharacterized gene family shows high sequence similarity between paralogs and low levels of divergence between species. The high degree of similarity between paralogs is particularly interesting as these genes are not organized as a single cluster and therefore sequence similarity is likely not maintained through non-homologous DNA repair (i.e., gene conversion) [30]. Additionally, NSPDs lack secondary structure and are in fact predicted to be intrinsically disordered. This lack of divergence coupled with little biochemical constraint is unusual and suggests NSPD function requires a specific amino acid sequence along its entire length. However, not all regions of the gene appear to be under the same constraint, as evidenced by the short species-specific repeating motif, although the functional relevance of this motif remains unknown. The pattern of

seemingly independent gene copy number expansion and genomic organization despite sequence constraint observed here is strikingly similar to the evolutionary pattern we previously observed in the MSP gene family [20], and suggests lineage-specific gene family evolution rather than preservation of an ancestral gene family structure.

The newly defined NSPF family showed enriched expression in the membranous organelle, as well as sequence conservation across the clade. While the degree of gene family evolution was far more limited, the duplication of *nspf-2* in *C. elegans* isolates combined with apparent gene losses in *C. sp. 34* and *C. doughertyi* suggest that this family is not completely static. The *C. elegans* lineage, in particular, appears to be evolving differently from the rest of the genus, including changes in copy number and genomic organization. Despite their predicted signaling function, we found no compelling evidence that these genes are involved in male reproductive success, though a subtle fertility difference could have been swamped out by the high individual variance in fecundity. These null results suggest that this family could be redundant as is supported by apparent species-specific gene losses, although if true we might expect to see greater sequence divergence across the genus due to genetic drift. Alternatively, this family may play a role in female post-mating physiological response or male re-mating behavior and not on male fertility per se.

One noticeable difference between these nematode-specific gene families is the lack of a signal peptide in NSPD genes, which is puzzling given that membranous organelles are Golgi-derived vesicles and thus proteins are presumably loaded through ER-Golgi signaling pathways. One possibility is that proteins produced in very high abundance—such as NSPDs and MSPs—could passively leak from the ER into membranous organelles [31]. Alternatively, transporters on the surface of membranous organelles could actively or passively transport proteins into the vesicle [32]. An entirely different explanation for identifying non-signaling proteins in the secreted proteomes is that activation releases other exosomes similar to the budding MSP vesicles previously shown in fully activated sperm [29]. However, such exosomes have not yet been identified during spermiogenesis itself. These questions of packaging warrant future studies tagging the NSPD proteins, though such an endeavor may prove challenging given their high sequence similarity, short size, and disordered nature.

While these data represent a foundation for membranous organelle molecular biology, no clear functional role for the soluble proteins within this subcellular component stands out. Nevertheless, two non-exclusive hypotheses suggest themselves. First, membranous organelles may serve as a contributor to the overall composition of the



seminal fluid (although perhaps a minor contributor). The presence of MSP within the organelles supports this hypothesis. Future studies that track where membranous organelle proteins are found after activation—at the female vulva opening, in the spermatheca, or possibly transferred back to the male cloaca—will be valuable in verifying this hypothesis. Alternatively, the membranous organelle could be more important during spermatid stasis and establishing membrane fluidity upon activation [16]. Here, membrane fusion is the more critical functional component, and the release of membranous organelle contents would then represent an incidental “trash dump” as sperm cells move on to the next phase of their life cycle. The presence of actin exclusively in the membranous organelle supports this hypothesis, as activated sperm function is known to be actin-independent. Additionally, the null functional data for the NSPF family support this “trash dump” hypothesis. Both hypotheses warrant continued investigation to further understand the functional role of this unique sperm component.

## Conclusions

Overall, our findings of sequence conservation over such long evolutionary time periods are contrary to observations within many other organisms, where elevated signals of positive selection are detected in seminal fluid proteins [33–35]. From an evolutionary perspective, then, patterns of evolution in secreted membranous organelle proteins do not match expectations for typical seminal fluid proteins. However, this pattern of sequence conservation coupled with lineage-specific gene family evolution observed here has also been previously identified for the MSP gene family [20]. There thus appears to be a “nematode sperm protein evolution syndrome” in which structural rearrangements and copy number variants are a more prevalent mechanism of genetic evolution than sequence divergence per se. Such a pattern could potentially be due to the conserved and unique sperm biology in nematodes, especially the biochemistry of locomotion. These results further support the need for taking a holistic approach when understanding the evolutionary history of genes.

## Methods

### Sperm collection

#### *Worm culture and strains*

Sperm were collected from *Caenorhabditis elegans* (standard laboratory strain N2 and strain JK574: *fog-2*(q71) V on the N2 background) and *C. remanei* (strain EM464). The *fog-2* mutation blocks *C. elegans* hermaphrodite self-sperm production, resulting in a functionally male-female population, thereby increasing the ease with which males could be collected. All strains were raised on NGM-agar plates seeded with OP50 *Escherichia coli* bacteria and raised at

20 °C [36]. Synchronized cultures of larval stage 1 animals were produced through hypochlorite treatment [37]. Males sourced for microfluidic dissection were isolated from females starting as young adults (44 h post-larval stage 1) for 24 h to build up their stored spermatid supply. Males sourced for testis crushing were maintained on mixed sex plates at population densities of approximately 1000 animals until the second day of adulthood (62 h post-larval stage 1).

#### *Microfluidic-based sperm collection*

The Shredder (final design: v5.0; Additional file 8) was designed using CAD software (Vectorworks 2013 SP5, Nemetschek Vectorworks, Inc) to function as a precise method of dissecting the male testis. The design has a single worm inlet that sequentially pushes males past a glass dissection needle, which slices through the cuticle, punctures the testis, and releases stored spermatids (Fig. 2). Two additional liquid channels flush males out of the dissection channel and flush sperm through a filtration system into the sperm outlet. Single layer devices were fabricated from polydimethylsiloxane (PDMS) using soft lithography [38] and bonded to a glass microscopy slide following exposure to air plasma. Dissection needles were made using a laser micropipette puller (Sutter Instrument P-2000) and inserted into each device following bonding.

A single Shredder could be used once to dissect up to 20 males. Each device was first flushed with 20 mM ammonium bicarbonate (pH 7.8), after which 20 virgin males were loaded into the worm inlet. The collected spermatids were concentrated by centrifugation (500 rcf for 15 min) and then lysed in liquid nitrogen. The cell membranes were pelleted, leaving the spermatid proteins in the supernatant for collection. A total of four pooled *C. elegans* replicates (259 males) and five pooled *C. remanei* replicates (265 males) formed the un-activated spermatid proteome for each species.

#### *Testis-crushing sperm collection*

To increase the amount of protein collected, particularly the membranous organelle protein contribution, we also used a male crushing technique to collect spermatids (modified from [18, 19]). Males were raised in mixed sex populations and size separated from females on the second day of adulthood. This developmental time point was optimal for maximizing the difference in diameter between the sexes and minimizing progeny. The sexes were separated using Nitex nylon filters (35 µm grid for *C. elegans* and 30 µm grid for *C. remanei*) with an average male purity of 91%. The filtration set-up was kept within a sterilized box to reduce external contamination.

Males were pelleted and plated between two 6" × 6", silane-coated

(tridecafluoro-1,1,2,2-tetrahydrooctyl-1-trichlorosilane) plexiglass squares. The plexiglass was then placed between two 6" × 6" × 1" wooden blocks. A heavy-duty bench vise was used to apply pressure to males, releasing the testis and spermatids. Spermatids were washed off the plexiglass using 20 mM ammonium bicarbonate (pH 5.6) onto a 10 μm grid Nitex nylon filter. This filter size was large enough to let spermatids freely pass, but not adult carcasses or eggs. Spermatids were concentrated by centrifugation and the supernatant collected (Fig. 1b). Supernatant collected before sperm activation was used to control for proteins released by cell lysis. No protein was measured in the pre-sperm activation supernatant. Spermatids were activated in vitro by adding 100 μL of 70 mM triethanolamine (TEA) to the pelleted volume [8] and were left to activate on a chilled block for 15 min. Our ability to activate sperm was verified by microscopy. The supernatant was collected to provide the membranous organelle proteome (Fig. 1b). The remaining activated cells were lysed as before and the proteins were collected as the activated sperm proteome. Six pooled replicates for *C. elegans* (maximum 19,075 males) and four pooled replicates for *C. remanei* (maximum 13,400 males) formed the membranous organelle and activated sperm proteomes for each species.

### Proteomic characterization of sperm

#### Tandem mass spectrometry

The proteomes were prepared and characterized by the Genome Science Mass Spectrometry Center at the University of Washington. Samples were denatured and digested according to standard protocols [39] and then analyzed on a Thermo Velos-Pro mass spectrometer coupled with a Thermo Easy nano-LC. Analytical replicates were run for each sample. MS/MS data were analyzed using the Comet database search algorithm [40] with either the *C. elegans* (PRJNA13758) or *C. remanei* (PRJNA53967) reference protein database. Peptide q-values and posterior error probabilities were calculated using Percolator [41]. Peptides were assembled into protein identification using ID picker [42] with a 1% false discovery rate cutoff.

#### Proteomic data analysis

Raw MS/MS information for each proteome was processed so as to include the minimum number of proteins that account for the observed peptides (i.e. parsimonious proteins) and filtered to exclude non-nematode proteins. Additionally, we combined isoform calls into a single gene and condensed four classes of genes (MSP family, NSPD family, SAMS family, F34D6 family) to the gene family level because of identical peptide coverage and high overall sequence similarity of paralogs. Overall, then, our final datasets were the most conservative

representation of our data. We then calculated the relative normalized spectrum abundance frequency (measured NSAF divided by the total worm NSAF) for each protein. The two runs were combined by taking the mean relative NSAF of each protein.

Biological functions for each protein were assigned using WormBase when possible [17]. The composition of the membranous organelle and activated sperm proteomes were compared to determine which proteins were shared and which were unique to a given proteome. Since the *C. remanei* genome is not as well functionally annotated, *C. elegans* orthologous gene families were assigned to characterize biological function. Proteome composition between species was compared at the gene family level. All analyses were performed using the R statistical language [43].

### Evolutionary analysis of the membranous organelle

#### Gene annotations

We used the well-annotated *C. elegans* reference genome (PRJNA13758: CEGMA: 100% complete, 0% partial; BUSCO 98% complete,  $n = 982$ ) to compile our query dataset for the NSPD and NSPF (genes F34D6.7, F34D6.8, and F34D6.9) gene families. Genes were annotated in 11 species across the *Caenorhabditis elegans* supergroup: *C. sp.* 33 (from J. Wang), *C. sp.* 34 (PRJDB5687), *C. briggsae* (PRJNA10731), *C. doughertyi* (PRJEB11002), *C. kamaaina* (QG2077\_v1), *C. latens* (PX534\_v1), *C. nigoni* (PRJNA384657), *C. remanei* (PRJNA248909), *C. sinica* (PRJNA194557), *C. tropicalis* (PRJNA53597), and *C. wallacei* (from E. Schwarz). Annotations were generated using custom amino acid blast (tblastn) searches in Geneious v10.2.3 [44]. Blast results were hand-curated for accuracy. In particular, five NSPF sequence motifs found to be conserved between *C. elegans* and *C. briggsae* were used as markers during annotation. We annotated a total of 59 NSPD genes and 19 NSPF family genes (Additional file 3) in the 11 species.

The *Caenorhabditis* Natural Diversity Resource [27] was used to probe the duplication and translocation of the NSPF family across the 249 isotypes identified from whole genome sequencing of 429 natural isolates. The NSPF gene region (II: 2,687,625 – 2,690,180) was extracted using SAMTOOLS. Coverage was calculated and those positions with less than 3× coverage were masked. A consensus sequence for each isotype was created. These sequences were aligned using ClustalW [45] in Geneious.

Synteny of the NSPF family was analyzed to determine gene orthology. The *C. elegans* NSPF family formed a cluster on Chromosome II, however, the *C. briggsae* NSPF family formed a cluster on Chromosome IV. Therefore, additional genes surrounding both the *C. elegans* and *C.*

*briggsae* clusters were identified using the UCSC Genome Browser [44]. These genes served as syntenic Chromosome II and IV anchors, respectively, following the approach outlined in Kasimatis and Phillips [20]. The NSPD family was spread across more than half the chromosomes in *C. elegans* and *C. briggsae*, precluding rigorous syntenic analysis.

Secondary structure was predicted using the Phyre<sup>2</sup> server [46]. Biochemical predications about protein structure and function were made using the Predictors of Natural Disordered Regions Server [25] and the SignalP Server [47].

### Evolutionary rate tests

The gene sequences for the NSPF and NSPD families were aligned using ClustalW. Amino acid sequence identity was calculated for all pairwise gene combinations within a species as well as across the clade. Unrooted maximum likelihood phylogenies were constructed in PhyML [48] of orthologous genes for the NSPF family. Since orthology could not be assigned within the NSPD family, phylogenies were constructed based on monophyletic species trios. Alignment-wide estimates of the non-synonymous to synonymous substitution ratio ( $\omega$ -ratio) were calculated using HyPhy [49] under a GTR mutation model. Selection within the NSPF family was estimated across the genus for orthologous genes. Additionally, orthologous genes were analyzed using a branch-site framework in the package BS-REL [50] within HyPhy to determine if the *C. elegans* branch in particular was evolving differently than the rest of the gene family. The NSPD family was analyzed using reduced alignments of all genes within monophyletic species triplets. Reduced alignments were constructed by removing the species-specific repeating amino acid motifs (~ 8 residues) in the middle of the gene. Here sequence alignment was highly dependent on the gap/extension penalty, thereby potentially confounding evolutionary inference.

### Functional verification of NSPF gene family

#### Strain generation by CRISPR/Cas9

Guide sequences were chosen using the CRISPRdirect [51], MIT CRISPR Design (<http://crispr.mit.edu>) and Sequence Scan for CRISPR [52] tools. For deletion of the *nspf-1*, *nspf-2*, and *nspf-3* genes, cr:tracrRNAs (Synthego) targeting the sequences CAGAGCCCATAATTCAAAGACGG and AGATGAGATTCTAATCAGGTAGG were annealed and pre-incubated with Cas9 (PNA Bio) in accordance with the manufacturer protocol. Young adult N2 individuals were injected in the gonad with a final mix consisting of 1.7  $\mu$ M of each cr:tracrRNA, 1.65  $\mu$ g/ $\mu$ l Cas9 and 50 ng/ $\mu$ l of the oligonucleotide repair template (5'-GTAAGAATACAATTTTCTTTGTGACTTACCGT

CTGGTAGGGTGGCAGATCAGTGTTTCAGAAAGGAAGTGA-3'), along with an additional cr:tracrRNA and oligonucleotide repair template to allow for screening by *dpy-10* co-conversion (see [53]). Individuals from broods containing Roller or Dumpy individuals were screened for the deletion by PCR and confirmed by Sanger sequencing. Individuals with confirmed deletions were then crossed to males with the *him-5* mutation (strain CB4088: *him-5*(e1490) on the N2 background). The *him-5* mutation increases the frequency of X chromosome non-disjunction events during meiosis, resulting in roughly 30% male progeny from self-fertilizing hermaphrodites [54]. Five generation of backcrossing were done to purge potential off-target CRISPR affects. The resulting strain, PX623, (*fxDf1* II; *him-5*(e1490) V) was used for functional analyses of the NSPF genes.

#### Fertility assays

We assayed the fertility of knockout males in both non-competitive and competitive sperm environments. To assess the overall reproductive success of knockout males, we mated a single knockout male with three wildtype, virgin females (strain JK574) for 24 h. As a control, wildtype males (strain JK574) were mated to wildtype females following the same male to female ratio. Matings were done on small NGM-agar plates (35 mm diameter) seeded with 10  $\mu$ l OP50 *E. coli*. After 24 h, each male was removed and the females were transferred to a new plate to continue laying eggs. Females were transferred to new plates every 24 h until progeny production ceased. The total number of progeny was counted as a measure of each male's reproductive success (Additional file 7). To measure competitive ability, individual wildtype, virgin females (strain JK574) were mated with a knockout male and an RFP marked male (strain PX626: *fxIs2*[*P*<sub>*hsp-16.41*</sub>::*PEEL-1::tbb-2* 3' UTR, *P*<sub>*rpl-28*</sub>::*mKate2::unc-54* 3' UTR, *P*<sub>*rps-0*</sub>::*HgrR::unc-54* 3' UTR, I: 2851040]; *fog-2*(q71) V). Again as a control, virgin females were mated to a wildtype male and an RFP marked male. Worms were mated overnight on small NGM-agar plates seeded with 10  $\mu$ l OP50 *E. coli* and then the males were removed. Progeny were collected over the next 24 h, counted, and screened for the number of RFP positive progeny. Two independent biological replicates of the competitive assay were performed (Additional file 9).

The fertility data were analyzed using R, with the significance of non-competitive reproductive success evaluated using Welch's Two Sample t-test and an analysis of the power of the comparison computed using the package *pwr* [55]. Male sperm competitive success was analyzed using a generalized linear model framework with random effects and a Poisson

distribution within the package *lme4* [56]. An equality of proportions test was performed for the competitive sperm assay to determine if wildtype and knockout males sired half of the total progeny.

## Additional files

**Additional file 1:** The un-activated sperm proteome of *C. elegans*. The majority of the proteome is comprised of the Nematode-Specific Peptide family, group D (NSPD) and the Major Sperm Protein (MSP). Protein abundance is shown as the relative mean normalized spectrum abundance frequency. Proteins found to be unique to either the membranous organelle or activated sperm proteomes are highlighted in teal, while proteins found in both proteomes are shown in gray. Proteins shown in white were not identified in the membranous organelle or activated sperm proteomes, but were found in the previously published un-activated spermatid proteome of Ma et al. [21]. (PDF 863 kb)

**Additional file 2:** Proteome data for *C. elegans* and *C. remanei*. Un-activated spermatid, membranous organelle, and active sperm proteome data for both species analyzed, including WormBase gene identifiers, protein abundances, and peptide coverage. (XLSX 64 kb)

**Additional file 3:** Gene annotations for the NSPD and NSPF gene families. Orthologous genes for the Nematode-Specific Peptide family, group D (NSPD) and Nematode-Specific Peptide family, group F (NSPF) family in 11 *Caenorhabditis* species. Annotations are listed by species, along with the gene start position and coding sequence length. (DOCX 29 kb)

**Additional file 4:** Consensus sequence alignments for the Nematode-Specific Peptide family, group D (NSPD). The amino acid sequence is largely conserved, except for the species-specific amino acid repeats in the middle of the gene. (PDF 79 kb)

**Additional file 5:** An unrooted maximum likelihood phylogeny for the Nematode-Specific Peptide family, group D (NSPD). Genes tend to cluster within species and do not recapitulate an evolutionary history of gene orthology. Asterisks denote bootstrap values greater than 80%. (PDF 945 kb)

**Additional file 6:** Sequence alignments for the Nematode-Specific Peptide family, group F (NSPF) orthologous genes. Amino acid sequence is largely conserved across orthologs. (PDF 158 kb)

**Additional file 7:** Unrooted maximum likelihood phylogenies for the Nematode-Specific Peptide family, group F (NSPF) orthologous genes. Overall, gene trees recapitulate species relationships. Asterisks denote bootstrap values greater than 80%. (PDF 867 kb)

**Additional file 8:** The Shredder microfluidic design. The Shredder v5.0 is designed to dissect day 1 adult males. The blueprint is accessible using CAD software. A master height of 35um is recommended. (EPS 2020 kb)

**Additional file 9:** Functional assays of the Nematode-Specific Peptide family, group F (NSPF) gene family. The fecundity data for total reproductive success and competitive reproductive success of NSPF knockout males. (XLSX 60 kb)

## Abbreviations

CRISPR: Clustered regularly interspaced short palindromic repeats; MS/MS: Tandem mass spectrometry; NSAF: Normalized spectral abundance frequency

## Acknowledgements

Biological mass spectrometry was performed by the Genome Sciences Mass Spectrometry Center at the University of Washington and we would especially like to thank Gennifer Merrihew and Michael MacCoss for assistance. Discussions with Willie Swanson regarding proteomic approaches to studying sperm function helped to set the stage for much of this work. Refinements of The Shredder design were aided with contributions from Stephen Banse. John Johnson assisted with microfluidic sperm dissections. Matthew Rockman, Erich Schwarz, John Wang generously shared unpublished genomic data, and Anastasia Teterina assisted with the *Caenorhabditis* natural-isolate annotations.

Finally, we would like to thank William Cresko, Michael Harms, the Harms Lab Group, and three anonymous reviewers for constructive feedback.

## Funding

This work was supported by the National Institutes of Health (training grant T32 GM007413 to KRK and R01 GM102511 and R01 AG049396 to PCP) and the ARCS Foundation Oregon Chapter (KRK). The funders did not supervise or contribute to the design, analysis, or interpretation of the data or in the writing of the manuscript.

## Availability of data and materials

The datasets supporting the conclusions of this article are available in Additional files 2 and 9. The genomes used are publically available from the following open access sources: WormBase ([www.wormbase.org](http://www.wormbase.org)) (*C. elegans*, PRJNA13758; *C. briggsae*, PRJNA10731; *C. sinica*, PRJNA194557; *C. tropicalis*, PRJNA53597), the *Caenorhabditis* Genomes Project (*C. doughertyi*, PRJEB11002; *C. latens*, PX534\_v1; *C. kamaaina*, QG2077\_v1; *C. remanei*, PRJNA248909), and NCBI (*C. sp. 34*, PRJDB5687; *C. nigoni*, PRJNA384657). The genome for *C. wallacei* was provided by E. Schwarz and the transcriptome for *C. sp. 33* was provided by J. Wang. Worm strains N2, JK574, CB4088, EM464, and PX623 are available from the *Caenorhabditis* Genetics Center. Worm strain PX626 is available from the Phillips Lab upon request.

## Authors' contributions

KRK and PCP designed the study; NT designed the original version of The Shredder, which was further refined by KRK; KRK collected samples for MS analysis; KRK analyzed the data; MMS and KRK created strain PX623; KRK performed fecundity analyses; KRK and PCP wrote the manuscript. All authors have read and approved the final version of this manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 4 April 2018 Accepted: 31 July 2018

Published online: 07 August 2018

## References

- Tarín JJ, Cano A, editors. Fertilization in protozoa and metazoan animals. Berlin: Springer; 2012.
- Dunbar BS, O'Rand M, editors. A comparative overview of mammalian fertilization. New York: Plenum Press; 1991.
- Eirín-López JM, Frehlick LJ, Ausió J. Protamines, in the footsteps of linker histone evolution. *J Biol Chem*. 2005;281:1–4.
- Morrow EH. How the sperm lost its tail: the evolution of aflagellate sperm. *Biol Rev*. 2004;79:795–814.
- Tanphaichitr N, Kongmanas K, Kruevaisayawan H, Saewu A, Sugeng C, Fernandes J, et al. Remodeling of the plasma membrane in preparation for sperm-egg recognition: roles of acrosomal proteins. *Asian J Androl*. 2015;17:574–9.
- Nelson GA, Roberts TM, Ward S. *Caenorhabditis elegans* spermatozoan locomotion: amoeboid movement with almost no actin. *J Cell Biol*. 1982;92:121–31.
- Nelson GA, Ward S. Vesicle fusion, pseudopod extension and amoeboid motility are induced in nematode spermatids by the ionophore monensin. *Cell*. 1980;19:457–64.
- Ward S, Hogan E, Nelson GA. The initiation of spermiogenesis in the nematode *Caenorhabditis elegans*. *Dev Biol*. 1983;98:70–9.
- L'Hernault SW. Spermatogenesis. WormBook, ed. The *C. elegans* Research Community, WormBook; 2006. <https://doi.org/10.1895/wormbook.1.85.1>.
- Burke DJ, Ward S. Identification of a large multigene family encoding the major sperm protein of *Caenorhabditis elegans*. *J Mol Biol*. 1983;171:1–29.



11. Bottino D, Mogilner A, Roberts T, Stewart M, Oster G. How nematode sperm crawl. *J Cell Sci.* 2002;115:367–84.
12. Achanzar WE, Ward S. A nematode gene required for sperm vesicle fusion. *J Cell Sci.* 1997;110:1073–81.
13. Washington NL, Ward S. FER-1 regulates Ca<sup>2+</sup>-mediated membrane fusion during *C. elegans* spermatogenesis. *J Cell Sci.* 2006;119:2552–62.
14. Chatterjee I, Richmond A, Putiri E, Shakes DC, Singson A. The *Caenorhabditis elegans* *spe-38* gene encodes a novel four-pass integral membrane protein required for sperm function at fertilization. *Development.* 2005;132:2795–808.
15. Xu XZS, Sternberg PW. A *C. elegans* sperm TRP protein required for sperm-egg interactions during fertilization. *Cell.* 2003;114:285–97.
16. Roberts TM, Ward S. Membrane flow during nematode spermiogenesis. *J Cell Biol.* 1982;92:113–20.
17. Lee RYN, Howe KL, Harris TW, Arnaboldi V, Cain S, Chan J, et al. WormBase 2017: molting into a new stage. *Nucleic Acids Res.* 2017;46:1–6.
18. Klass MR, Hirsh D. Sperm isolation and biochemical analysis of the major sperm protein from *Caenorhabditis elegans*. *Dev Biol.* 1981;84:299–312.
19. Miller MA. Sperm and oocyte isolation methods for biochemical and proteomic analysis. *Methods Mol Biol.* 2006;351:193–201. New Jersey: Humana Press
20. Kasimatis KR, Phillips PC. Rapid gene family evolution of a nematode sperm protein despite sequence hyper-conservation. *G3.* 2018;8:353–62.
21. Ma X, Zhu Y, Li C, Xue P, Zhao Y, Chen S, et al. Characterisation of *Caenorhabditis elegans* sperm transcriptome and proteome. *BMC Genomics.* 2014;15:1–13.
22. Braendle C, Felix M-A. Sex determination: ways to evolve a hermaphrodite. *Curr Biol.* 2006;16:R468–71.
23. Kiontke KC, Felix M-A, Aillon M, Rockman MV, Braendle C, Penigault J-B, et al. A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol Biol.* 2011;11:339.
24. Woodruff GC, Willis JH, Phillips PC. Dramatic evolution of body length due to post-embryonic changes in cell size in a newly discovered close relative of *C. elegans*. *Evol Lett.* 2018; in press.
25. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. *Biochemistry.* 2002;41:6573–82.
26. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol.* 2015;16:18–29.
27. Cook DE, Zdraljević S, Roberts JP, Andersen EC. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* 2017;45:D650–7.
28. Miller MA, Nguyen VQ, Lee MH, Kosinski M, Schedl T, Caprioli RM, et al. A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. *Science.* 2001;291:2144–7.
29. Kosinski M, McDonald K, Schwartz J, Yamamoto I, Greenstein D. *C. elegans* sperm bud vesicles to deliver a meiotic maturation signal to distant oocytes. *Development.* 2005;132:3357–69.
30. Chen J-M, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet.* 2007;8:762–75.
31. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Transport from the ER through the Golgi apparatus. *Mol Biol Cell.* 4th ed. New York: Garland Science; 2002.
32. Beer KB, Wehman AM. Mechanisms and functions of extracellular vesicle release in vivo—what we can learn from flies and worms. *Cell Adhes Migr.* 2017;11:135–50.
33. Swanson WJ, Vacquier VD. Reproductive protein evolution. *Annu Rev Ecol Syst.* 2002;33:161–79.
34. Clark NL, Aagaard JE, Swanson WJ. Evolution of reproductive proteins from animals and plants. *Reproduction.* 2006;131:11–22.
35. Mordhorst BR, Wilson ML, Conant GC. Some assembly required: evolutionary and systems perspectives on the mammalian reproductive system. *Cell Tissue Res.* 2015;363:267–78.
36. Brenner S. The genetics of *Caenorhabditis elegans*. *Genetics.* 1974;77:71–94.
37. Kenyon C. The nematode *Caenorhabditis elegans*. *Science.* 1988;240:1448–53.
38. Qin D, Xia Y, Whitesides GM. Soft lithography for micro- and nanoscale patterning. *Nat Protoc.* 2010;5:491–502.
39. Merrihew GE, Davis C, Ewing B, Williams G, Kall L, Frewen BE, et al. Use of shotgun proteomics for the identification, confirmation, and correction of *C. elegans* gene annotations. *Genome Res.* 2008;18:1660–9.
40. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. *Proteomics.* 2013;13:2–4.
41. Käll L, Storey JD, Noble WS. Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. *Bioinformatics.* 2008;24:42–8.
42. Zhang B, Chambers MC, Tabb DL. Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res.* 2007;6:3549–57.
43. R Core Team. R: A language and environment for statistical computing. Vienna: Foundation for Statistical Computing; 2015. Available from: <https://www.R-project.org/>
44. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* 2012;28:1647–9.
45. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80.
46. Mezulis S, Yates CM, Wass MN, Sternberg MJE, Kelley LA. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10:845–58.
47. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* 2011;8:785–6.
48. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol.* 2003;52:696–704.
49. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics.* 2005;21:676–9.
50. Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol.* 2011;28:3033–43.
51. Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics.* 2015;31:1120–3.
52. Xu H, Xiao T, Chen C-H, Li W, Meyer CA, Wu Q, et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* 2015;25:1147–57.
53. Paix A, Folkmann A, Rasoloson D, Seydoux G. High efficiency, homology-directed genome editing in *Caenorhabditis elegans* using CRISPR-Cas9 ribonucleoprotein complexes. *Genetics.* 2015;201:47–54.
54. Hodgkin J, Horvitz HR, Brenner S. Nondisjunction mutants of the nematode *Caenorhabditis elegans*. *Genetics.* 1979;91:67–94.
55. Champely S. pwr: basic functions for power analysis. R package version 1.2-1; 2017. p. 1–20. Available from: <http://CRAN.R-project.org/package=pwr>
56. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67:1–48.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

