Beyond DNA Compaction:
An Emerging View of the Diverse Functions of *Drosophila* Sperm Nuclear Basic Proteins

Takuo Yamaki

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Barbara T. Wakimoto, Chair

Steven Henikoff

Willie Swanson

Program Authorized to Offer Degree:

Biology

University of Washington

Abstract

Beyond DNA Compaction:
An Emerging View of the Diverse Functions of *Drosophila* Sperm Nuclear Basic Proteins

Takuo Yamaki

Chair of the Supervisory Committee:
Barbara T. Wakimoto
Department of Biology

While sperm of different animals exhibit morphologically distinct features, the overall themes of sperm nuclei are surprisingly similar among species. Sperm nuclei are extremely condensed and transcriptionally inert, and it has long been realized that sperm DNA in such condensed nuclei interacts with a unique set of proteins known as sperm nuclear basic proteins (SNBPs). Collectively referred to as SNBPs, this class of proteins includes evolutionarily and biochemically distinct chromosomal proteins, from nucleosomal histones to sperm-specific chromatin proteins such as protamines. The prevailing idea has been that SNBPs as a whole allow for extreme nuclear condensation and that the condensed nucleus plays an important role in efficient migration of sperm towards eggs for successful fertilization. However, this idea remains debatable due to the lack of supporting experimental evidence. Alternatively, it may be

that the histone-SNBP transition serves as a mechanism to erase the epigenetic memory on sperm chromosomes so that the zygote can initiate embryogenesis with developmentally totipotent paternal chromosomes. In addition, SNBPs themselves may play active roles in sperm formation and function. For my dissertation, I used a model organism, Drosophila melanogaster, to investigate how distinct aspects of spermatogenesis and embryogenesis are regulated by MST-HMG box proteins, a group of sperm-specific chromatin proteins that derived from HMG box proteins. Also, using mutations in an MST-HMG box protein that specifically affect paternal chromosome stability in the early embryo, I studied how embryos respond to aberrant DNA structures at developmental stages that have long been thought to lack cell cycle checkpoints. Lastly, I describe a proteomic-based approach to comprehensively discover previously unidentified proteins that are likely enriched in the sperm head, including one new member of MST-HMG box family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

For Christine to live in the UW library forever.

Chapter 1: Introduction

It was still in the early days of genetic research when parental contributions to embryogenesis were noted to be unequal between male and female gametes. The first clear case of such parental origin-specific influences on development was reported in 1923 when Sturtevant postulated that shell coiling of the freshwater snail *Limnaea* is determined by the genotype of the mother regardless of the genotype of offspring (STURTEVANT 1923). As for paternal contributions to development, experimental evidence was first documented in 1945 in a study using flies. Mampell discovered that a strain of *Drosophila pseudoobscura* known as the *Mutator* produced offspring prone to genomic instability when the *Mutator* males were crossed to wild-type females (MAMPELL 1945). The reciprocal cross did not yield the same phenotype in offspring, demonstrating that the *Mutator* has a paternal effect on offspring genome stability. Various model organisms have contributed to the identification of actual molecules that mediate maternal effects on development. In contrast, there is a considerably smaller number of genetic screens for paternal effect mutants to identify sperm-derived factors that are required for successful initiation and completion of embryogenesis. Previous work in the lab has recovered such mutants in *Drosophila melanogaster* by identifying mutant males with reduced fertility or complete sterility despite the fact they make motile sperm that fertilize eggs (WAKIMOTO *et al.* 2004). My dissertation stemmed from one of the mutants, *deadbeat*, that affect paternal chromosome stability in the early embryo. Deadbeat is a protein that is evolutionarily related to previously characterized sperm-specific chromatin proteins in *Drosophila.* The characterization of Deadbeat inspired the further investigation of the dynamic evolution of a gene family that encodes sperm-specific chromatin proteins in *Drosophila* and their functional roles in regulating various steps of sperm formation and function.

Evolutionary origins of Sperm Nuclear Basic Proteins

In a wide range of animals including humans, spermatids undergo extensive nuclear transformation from spherical to highly condensed nuclei. This process not only changes nuclear shape but also reduces nuclear volume significantly, and its reduction can be as extreme 200-fold as it is in *Drosophila* (TATES 1971). Nuclear transformation is often accompanied by large-scale biochemical transformation inside the nucleus. In this process, histones are post-translationally modified, the majority of histones are replaced by transition proteins, and transition proteins are ultimately exchanged with sperm-specific chromatin proteins such as protamines (RATHKE *et al.* 2014; CHAMPROUX *et al.* 2016). Selectively retained histones and sperm-specific chromatin proteins as a whole constitute a unique set of chromatin proteins known as sperm nuclear basic proteins (SNBPs) (EIRIN-LOPEZ AND AUSIO 2009). As with histones in somatic cells, SNBPs are distinct from other chromosomal proteins such as transcription factors in that one needs a high salt or acid-based extraction buffer because they interact with DNA tightly. The first discovery of SNBP, which was named protamin, dates back to 1874 when Friedrich Meischer realized that there is a basic component in salmon sperm in addition to the acidic component of what we know today as DNA (MIESCHER 1874) . Since then, extensive biochemical studies of vertebrate and invertebrate sperm have revealed the widespread prevalence of sperm-specific chromatin proteins in the animal kingdom. Unlike nucleosomal histones, these proteins vary tremendously in amino acid sequence and length even in closely related species (EIRIN-LOPEZ AND AUSIO 2009).

With the accumulating evidence that SNBP types vary among animals, Bloch first proposed to categorize SNBPs based on their amino acid compositions (BLOCH 1969). Early studies relied mostly on chemicals that react with certain amino acids such as the Sakaguchi test,

which uses chemicals that specifically react with arginine to yield bright red color. Bloch came up with five different categories based on animals in which the respective types were found: 1) salmon, 2) mouse/grasshopper, 3) *Mytilus*, 4) *Rana*, and 5) crab. The salmon- and mouse/grasshopper-type proteins have in common that they are highly enriched in arginine, and the mouse/grasshopper-type proteins are additionally enriched in cysteine. The *Mytilus*-type proteins are basic but more complex in composition, likely containing two or three different dibasic amino acids such as lysine in addition to arginine. The *Rana* and crab-type proteins are histones, with the crab-type proteins showing less basicity than regular histones. This categorization scheme was instrumental in clarifying that there is no pattern of specific SNBP types appearing in certain animal lineages. In fact, two fish species in the teleost lineage can have completely different SNBP types (SHIMIZU *et al.* 2000; WU *et al.* 2011).

Because arginine-enriched basic proteins (i.e. salmon- and mouse/grasshopper-type SNBPs) are distinct from histones, which are enriched in lysine, several ideas were proposed to explain where these arginine-enriched DNA-binding proteins came from evolutionarily. For instance, partial duplication of a protein that already contained arginine clusters was proposed to explain the birth of protamines (BLACK AND DIXON 1967). To account for sporadic distributions of protamines among animal species, horizontal transfer of the protamine-encoding gene by retroviral transmission was proposed (JANKOWSKI *et al.* 1986). While these ideas were logical and appealing considering the limited amount of sequence information available at that time, they failed to stand the test of time.

In 1973, Subirana proposed an idea that histones and protamines are evolutionarily related proteins (SUBIRANA *et al.* 1973). In this proposal, protamine-like (PL) proteins were introduced for the first time to account for the transition between histones and protamines.

However, it still remained a mystery as to how lysine-rich histones made transitions to arginine-rich protamines through PL-type proteins. Ausio and colleagues later on made a seminal discovery by comparing sequences of SNBPs found in tunicate species that a single frameshift mutation can transform lysine clusters into arginine ones (LEWIS *et al.* 2004). Based on a variety of vertebrates and invertebrates for their SNBP sequences, they proposed that the linker histone H1 gave rise to protamines in a step-wise fashion. H1 with its characteristic winged-helix fold domain (WHD) first accumulated clusters of positively charged residues in its N- and C- tails, resulting in the formation of a PL-type protein. In some lineages, this intermediate form subsequently lost the WHD but retained the C terminus. The resulting smaller protein made transition from lysine to arginine and is now widely referred to as protamine (P). Because some lineages did not make the transition between H1 and PL or between PL and P, this elegantly explains why we see sporadic distributions of P-type SNBPs in the animal kingdom (EIRIN-LOPEZ AND AUSIO 2009).

At the functional level, the general consensus is that SNBPs are required for sperm nuclear condensation. This idea is consistent with earlier studies of mouse transition proteins and protamines. In the mouse, there are two transition proteins (TP1 and TP2) and two protamines (PRM1 and PRM2). The reduced dosage of TPs and PRMs results in decreased male fertility and complete male sterility, respectively (YU *et al.* 2000; CHO *et al.* 2001; ZHAO *et al.* 2001). In all cases, spermatid nuclei do not condense properly (YU *et al.* 2000; CHO *et al.* 2001; ZHAO *et al.* 2001). Interestingly, mouse sperm nuclei in the absence of PRM2 but not PRM1 show signs of extensive DNA breaks (CHO *et al.* 2003). These findings together suggest that while PRM1 and PRM2 are evolutionarily related proteins, they have likely acquired non-overlapping functions in addition to their general role in sperm nuclear condensation.

In *Drosophila*, three SNBPs (Mst35Ba, Mst35Bb, and Mst77F) were first identified from a set of *male-specific transcript* (*Mst*) genes (RUSSELL AND KAISER 1993; JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005). Based on amino acid sequence characteristics, Mst35Ba/b and Mst77F were first thought to be related to mammalian protamines and histone H1, respectively (RUSSELL AND KAISER 1993; JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005). It was naturally logical to link *Drosophila* SNBPs to vertebrate SNBPs because other invertebrate SNBPs had been suggested to be derivatives of histone H1 as in vertebrate protamines. However, this idea became questionable with the discovery of Tpl94D. The temporal profile of this protein is similar to that of mammalian transition proteins in that it only appears transiently in spermatid nuclei during nuclear transformation (RATHKE *et al.* 2007). Interestingly, Tpl94D was predicted to be folded into the high mobility group (HMG) box domain (RATHKE *et al.* 2007). HMG box is characterized by three alpha helices separated by loops and is found in many DNA-binding proteins of all eukaryotic branches. Later characterizations of Mst35Ba/b as well as Mst77F showed that these proteins also form the HMG box (DORUS *et al.* 2008; RATHKE *et al.* 2010). Because HMG box proteins are evolutionarily and biochemically distinctly from H1, this means that not all SNBPs are H1-related in the animal kingdom and that evolutionarily distinct origins gave rise to proteins with similar characteristics: small basic DNA-binding proteins. In Chapters 2 and 3 of my dissertation, I present evidence that the *Drosophila* male-specific HMG box protein family has expanded in terms of the number of family members and that the spatiotemporal behaviors of a few of these proteins as well as their functions have diversified in sperm formation and function.

Large-scale molecular transformation during spermiogenesis

The analysis of mammalian and *Drosophila* spermatogenesis has been fundamental in elucidating the molecular basis of extensive chromatin remodeling during nuclear transformation. These studies have revealed surprisingly similar molecular steps that spermatid nuclei take to replace most histones with non-histone SNBPs. For instance, histone H4 becomes hyper-acetylated in post-meiotic nuclei, presumably relaxing the interaction between histones and DNA. Around the same time with histone H4 hyper-acetylation, double-stranded DNA breaks are become abundant, which likely promotes DNA uncoiling to facilitate the removal of histones. As histones are gradually removed, transition proteins accumulate in spermatid nuclei (RATHKE *et al.* 2014). A recent study has shown that transition proteins are required at least in the mouse to recruit and process protamines so that the transition from histones-based chromatin to SNBP-based chromatin can be completed (BARRAL *et al.* 2017).

Although sperm chromatin folding is largely organized by sperm-specific chromatin proteins such as protamines, it has been known that mature sperm retain some histones. The degree of histone retention varies among animals and has been reported to be ~1% in the mouse and ~15% in the human (CHAMPROUX *et al.* 2016). Molecular mapping of these retained histones in the mouse sperm showed that they are enriched around developmentally important genes such as the *Hox* gene locus and imprinted loci (CARRELL AND HAMMOUD 2010). In another example, centromeric histones escape the large-scale chromatin remodeling process to remain present in mature sperm of bulls and those of flies (PALMER *et al.* 1990; RAYCHAUDHURI *et al.* 2012). At least in flies, this transmission of centromeric histones via sperm is essential for production of viable offspring, as the failure to do so results in development of haploid embryos that propagate maternal chromosomes only (RAYCHAUDHURI *et al.* 2012). The presence of retained histones at

specific loci indicates that the histone-SNBP transition is a highly regulated process. Thus, it has been of great interest and importance to understand how histones are selectively retained while the rest of the genome undergoes the remodeling process. Genome-wide characterization of endonuclease sensitive regions in mature sperm showed that these regions are enriched for CTCF-binding sites (ARPANAHI *et al.* 2009). Because protamine-packaged chromatin is resistant to endonuclease treatment, endonuclease-sensitive regions are indicative of histone retention. Extrapolating from CTCF functions in somatic cells, it is tempting to speculate that CTCF creates a protective chromatin environment for certain loci to escape the histone-SNBP transition.

In addition to histones, there are other chromosomal proteins that are retained in mature sperm such as transcription factors and other chromosomal proteins. In *Drosophila*, there is one particular chromosomal protein called K81 that is retained at telomeres in mature sperm (DUBRUILLE *et al.* 2010; GAO *et al.* 2011). In Chapter 2, I present evidence that the retention of this specific chromosomal protein requires an SNBP that functions regionally, which was a completely surprising result given what we knew about SNBP functions at that time.

Establishment of paternal chromosomes for successful initiation of embryogenesis

During fertilization, SNBPs are evicted from sperm chromatin and replaced by histones as the nucleus becomes decondensed. This process is facilitated by maternally supplied histone chaperones to load histones in a replication-independent manner (LOPPIN *et al.* 2015; YANG *et al.* 2015). Because sperm chromosomes are heterogeneously organized by different classes of proteins, it is important to understand whether specific features of sperm chromosomes such as retained histones play an active role in paternal chromosome assembly during the SNBP-histone

transition. A recent study provided evidence that heterochromatin is established and maintained in a parental origin-specific manner in the early human embryo. van de Werken *et al*. employed immuno-FISH to demonstrate that mature sperm chromatin retains a specific histone modification, H3K9me3, near heterochromatic knobs and centromeric regions. During fertilization, certain histone modifications present on sperm chromosomes are believed to serve as a platform to recruit proteins such as HP1 for re-establishment of heterochromatin on paternal chromosomes (VAN DE WERKEN *et al.* 2014). It remains to be investigated how the loss of H3K9me3 in sperm impacts embryogenesis quality in humans; however, based on a similar study using mouse embryos, the failure to re-establish paternal heterochromatin leads to upregulation of major satellite sequences, which is likely detrimental for embryo health.

In *Drosophila*, a paternal effect mutant, *ms(3)k81,* has revealed that the proper maintenance of paternal telomere identity in the early embryo requires the inheritance of K81 from sperm (FUYAMA 1984; YASUDA *et al.* 1995; DUBRUILLE *et al.* 2010; GAO *et al.* 2011). In this species, telomeres in most cells are maintained by telomere-capping complexes (TCCs), including the HipHop protein (ref). During spermatogenesis, TCCs replace HipHop with K81, a paralogous protein of HipHop that is specifically expressed in the testis (DUBRUILLE *et al.* 2010; GAO *et al.* 2011). The failure to switch from HipHop to K81 results in the transmission of sperm chromosomes without the TCCs (DUBRUILLE *et al.* 2010; GAO *et al.* 2011). In the absence of paternal TCCs, paternal chromosomes will participate in embryogenesis as if their telomeres were broken ends rather than natural ends. Such abnormal DNA structures in somatic cells elicit DNA damage response (O'SULLIVAN AND KARLSEDER 2010), but it was largely assumed that the early embryo lacked cell cycle checkpoints to detect and respond to uncapped telomeres accordingly. In Chapter 2, I describe the consequence of naked paternal telomeres on cell cycle

progression in the early embryo and provide an alternative insight into cell cycle regulation in the early embryo.

In the following chapters, I present three distinct approaches that converged and led to new insights into the complexity of SNBPs in *Drosophila.* Functional characterizations of some of these proteins have revealed their various functions in sperm formation and function, extending the widely accepted view of SNBPs as largely responsible for sperm DNA compaction.

Chapter 2: The Deadbeat Paternal Effect of Uncapped Sperm Telomeres on Cell Cycle Progression and Chromosome Behavior in *Drosophila melanogaster*

This chapter has been published as:

YAMAKI, T., G. K. YASUDA AND B. T. WAKIMOTO, 2016 The *Deadbeat* Paternal Effect of Uncapped Sperm Telomeres on Cell Cycle Progression and Chromosome Behavior in *Drosophila melanogaster*. Genetics 203**:** 799-816.

Abstract

Telomere-capping complexes (TCCs) protect the ends of linear chromosomes from illegitimate repair and end-to-end fusions and are required for genome stability. The identity and assembly of TCC components have been extensively studied, but whether TCCs require active maintenance in non-dividing cells remains an open question. Here we show that *Drosophila melanogaster* requires Deadbeat (Ddbt), a sperm nuclear basic protein (SNBP) that is recruited to the telomere by the TCC and is required for TCC maintenance during genome-wide chromatin remodeling that transforms spermatids to mature sperm. Ddbt-deficient males produce sperm lacking TCCs. Their offspring delay the initiation of anaphase as early as cycle 1 but progress through the first two cycles. Persistence of uncapped paternal chromosomes induces arrest at or around cycle 3. This early arrest can be rescued by selective elimination of paternal chromosomes and production of gynogenetic haploid or haploid mosaics. Progression past cycle 3 can also occur if embryos have reduced levels of the maternally provided checkpoint kinase Chk2. The findings provide insights into how telomere integrity affects the regulation of the earliest embryonic cell cycles. They also suggest that other SNBPs, including those in humans, may have analogous roles and manifest as paternal effects on embryo quality.

Introduction

Telomeres are the natural ends of linear chromosomes and have distinct properties from broken ends. Multiple proteins are enriched predominantly, if not exclusively, at telomeres and form a capping complex that protects telomeric DNA from engaging in aberrant DNA repair activities. Protein components of the telomere- capping complex (TCC) vary among organisms, in part because species differ in telomeric sequences and whether telomeres are maintained by telomerase or alternative mechanisms (RAFFA *et al.* 2011; MASON *et al.* 2015). Nonetheless, the TCC's essential functions are well conserved. Failure to assemble TCCs results in telomeric DNA degradation, telomere fusions, and genomic instability.

Mutations in at least a dozen loci of *Drosophila melanogaster* lead to telomere fusions in neuroblasts. Their analysis has led to the identification of telomere-enriched and telomere-exclusive proteins required for telomere elongation or TCC assembly, maintenance or function (CENCI *et al.* 2005; PIMPINELLI 2006). Absence of any one component results in telomere fusions but components have distinct activities (PIMPINELLI 2006). For example, Heterochromatin Protein 1a (HP1a) binds modified histone H3-MeK9, and represses transcription of telomeric retrotransposons and telomere elongation. HP1a also binds DNA and this activity is required for its capping function. The TCC protein HOAP binds DNA and HP1a. Although HOAP is required for capping, it does not affect retrotransposon transcription or telomere elongation. A third protein, HipHop, binds both HP1a and HOAP. HOAP and HipHop are recruited to telomeres by DNA damage checkpoint /repair proteins. Interactions between HP1a, HOAP, and HipHop are required to form stable and functional TCCs (GAO *et al.* 2010). Similar to yeast and mammalian cells (STEWART *et al.* 2012), each round of DNA replication in *Drosophila* somatic cells provides

the opportunity to assemble and maintain TCCs. However, a relatively unexplored question is how TCCs, once assembled are stably maintained in the absence of DNA replication.

The male germ line provides unique opportunities to study telomere dynamics through mitosis, meiosis, and spermiogenesis, which is the post-meiotic period of spermatid differentiation. Telomere maintenance during spermiogenesis is particularly interesting because it is prolonged relative to other spermatogenic stages, lasting about 5.5 days in *D. melanogaster* (LINDSLEY and TOKUYASU 1980) and 3.4 weeks in humans (AMANN 2008). Moreover, extensive chromatin remodeling occurs. Transformation of round spermatid nuclei to highly condensed sperm heads typically involves histone modification or nearly whole-scale histone replacement by sperm nuclear basic proteins (SNBPs) (EIRIN-LOPEZ and AUSIO 2009).

Several studies have been informative for revealing TCC composition in the *Drosophila* male germ line. In many species, TCCs contain HP1a, HOAP, and HipHop. However, in the melanogaster group species, HipHop has a testis-specific paralog called K81 (DUBRUILLE *et al.* 2010; GAO *et al.* 2011). In *D. melanogaster,* K81 replaces HipHop in early meiosis and *k81* mutant males produce sperm lacking TCCs (DUBRUILLE *et al.* 2010; GAO *et al.* 2011; DUBRUILLE and LOPPIN 2015). The consequence is male sterility due to a perplexing paternal effect (FUYAMA 1984). Most embryos of *k81* fathers arrest by mid-cleavage but a few survive to late embryogenesis as gynogenetic haploids with only the maternal genome (FUYAMA *et al.* 1988; YASUDA *et al.* 1995).

Here we describe the *Drosophila deadbeat* (*ddbt)* gene and show that Ddbt is an SNBP that acts downstream of K81. Ddbt is recruited to telomeres and ensures that TCCs are maintained through spermiogenesis and transmitted to offspring. We provide evidence that uncapped telomeres delay the onset of the first embryonic anaphase and that the embryo's ability

to deal with uncapped telomeres and newly generated breaks changes as it progresses to later cycles. Our findings provide new insights into *ddbt*'s and *k81*'s paternal effect defects and the impact of uncapped telomeres on cell cycle regulation in early embryos.

Materials and Methods

Drosophila mutations, transgenic strains, and fertility assays

The $ddbt^{Z4344}$ and $k81^{2}$ mutations were isolated by Wakimoto *et al*. (2004) and Yasuda *et al*. (1995) respectively. Y. Rong kindly provided the *gfp::k81; k81*^{2} stock (GAO *et al.* 2011); W. Theurkauf provided the grp^{fs1} and mnk^{P6} stocks (SIBON *et al.* 1997; BRODSKY *et al.* 2004); and A. Royou provided the *gfp::bubR1* stock (ROYOU *et al.* 2010). Other strains were obtained from the Bloomington *Drosophila* Stock Center.

We isolated the $ddbt^{+}$ gene from BAC clone *CH322-142G7* (BACPAC Resources) in a 3.84 kb *Eco*RI-*Kpn*I fragment. This fragment contains 1.69 kb upstream and 1.45 kb downstream of the longest transcript annotated for the *CG34264* gene by the Berkeley *Drosophila* Genome Project. Because this fragment allowed for transgenic rescue of the $ddbt^{Z4344}$ sterility (Table 2.1), we used it to construct all modified transgenes. To make a *egfp::ddbt* transgene, EGFP coding sequence and a portion of the multiple cloning site encoding amino acids YSDLELKL was isolated from pEGFP-C3 (Clontech) and inserted in-frame and immediately upstream of the *ddbt* open reading frame. The *mCherry::ddbt* versions were identical except the *egfp* sequence was replaced with a PCR-amplified fragment containing *mCherry* coding sequences from pmCherry (Clontech). Sequences containing portions of the *Mst35Ba* and *Mst77F* genes and the $ddbt^{R31D}$ mutation (CGC to GAC codon change) were created by custom synthesis (Genewiz) and replaced the corresponding sequence in *eGFP::ddbt. Eco*RI-*Kpn*I fragments containing $ddbt^{+}$ or

modified versions were inserted into the pBDP transformation vector (PFEIFFER *et al.* 2008) (Addgene plasmid 17566). Transgenes were targeted to the *attP40* site by PhiC31 integrase (GROTH *et al.* 2004), introduced into the *ddbt* mutant background, and tested for ability to rescue male fertility in two doses. A chromosome two carrying *mCherry-ddbt* and *gfp-k81* was created by recombination, then introduced into a double mutant strain to create the stock *w*; *P{w*$^{+mC}$ *mCherry::ddbt} P{w*$^{+mC}$ *gfp::k81}; ddbt* ^{Z4344} *k81*^{2} */TM6B, Sb Tb*.

For fertility assays, we set up 10 crosses of single males to 4 wild-type *Sevelin* females and compared average progeny yields to those of sibling control males.

Comparative analysis of the ddbt gene and protein

Sequences were downloaded from FlyBase (http://flybase.org) or NCBI (http://ncbi.nlm.nih.gov/nucleotide). We used MacVector software (MacVector, Inc.) for sequence analysis, PSIPRED v3.3 for protein secondary structure prediction (BUCHAN *et al.* 2013), Clustal Omega for multiple sequence alignment (SIEVERS *et al.* 2011), and PAML4 for d_N/d_S analysis (YANG 2007). A subset of FlyBase reported protein annotations were corrected or modified to optimize matches among species (Table S2.1). Specifically, the *D. sechellia* protein required selecting alternative splice sites based on *D. melanogaster* and *D. simulans* annotations. For *D. persimilis*, we noted an error in the assembled genome sequence, so report the predicted amino acid sequence after correction. For *D. grimshawi*, we propose the amino acid sequence based on similarity with *D. virilis* and *D. mojavensis* proteins. Newly annotated protein sequences were based on best fit with those of most closely related species. We did not detect a *D. willistoni* ortholog using tBLASTn or by careful inspection across 100 kb of the syntenic region. The tBLASTn searches also failed to identify orthologs outside of Schizophora. Selected

pairwise comparisons in Figure 2.2A were depicted using format of Eisman and Kaufman (2013). Datasets used to derive the NFLR consensus motifs in Figure 2.1D consisted of 28 Ddbt proteins from 28 insect species and 17 Mst35Ba/b and 9 Mst77F proteins from 12 *Drosophila* species (DROSOPHILA 12 GENOMES *et al.* 2007) and were depicted using WebLogo (CROOKS *et al.* 2004).

We used sequences from *D. melanogaster, D. simulans, D. sechellia, D. erecta, and D. yakuba* to test for positive selection of *ddbt* in the melanogaster species group. Clustal Omega yielded 109 aligned codons, which we used to estimate synonymous and nonsynonymous substitution rates as the d_N/d_S ratio and detect positive selection using the PAML4 CODEML program. We used a log likelihood ratio test for three comparisons of neutral and selection models (M1 vs. M2, M7 vs. M8, and M8a vs. M8). CODEML analysis was run with several different initial values of d_N/d_S to check for convergence. We calculated the $-2\Delta lnL$ and determined p values from a χ^2 distribution. We used Bayes empirical Bayes analysis implemented in PAML4 (YANG *et al.* 2005) to predict positively selected sites. Sites with posterior probability (P) of greater than 70% and their location on *D. melanogaster* Ddbt are reported in Figure 2.2B and Table 2.2.

Protein localization during spermatogenesis

We used protein localization methods of Wilson *et al*. (2006). For localization during spermatogenesis, a minimum of 3 males and 5 cysts from each male were examined to assess fluorescent signal intensity and variation. Antibodies used were: primary rabbit anti-HOAP (1:1000, a gift of W. Theurkauf; KLATTENHOFF *et al*. 2009), mouse anti-HP1 C1A9 (1:20, DSHB), rabbit anti-DsRed (1:500, Clontech), and secondary goat anti-rabbit Alexa Fluor 568

(1:200, Molecular Probes), and goat anti- rabbit or anti-mouse DyLight 650 (1:1000, Thermo Scientific). Preparations were mounted in Vectashield (Vector Laboratories) with the DNA dye DAPI (4',6-diamidino-2- phenylindole). Images were captured on a DeltaVision Elite imaging system run by softWoRx package (Applied Precision) and equipped with sCMOS camera (PCO) and a 100X/NA 1.4 oil immersion objective. Optical sections of 0.20 μm were acquired and deconvolved images were analyzed using NIH ImageJ (SCHINDELIN *et al.* 2012).

To compare location and intensity of mCherry-Ddbt foci to GFP-Ddbt or GFP- K81 foci in spermatids, we examined spermatids of experimental *w*; *P{w*$^{+mC}$ *mCherry::ddbt} P{w*$^{+mC}$ *gfp::k81}/+; ddbt* ^{Z4344} *k81*^{2} males and control *w*; *P{w*$^{+mC}$ *mCherry::ddbt}/ P{w*$^{+mC}$ *egfp::ddbt}; ddbt* ^{Z4344} males. Reliable detection of mCherry-Ddbt expression required immunostaining, whereas EGFP-Ddbt and GFP-K81 could be detected by GFP fluorescence. Testes from 3 different males were used to examine the smallest focus from a total of 100 canoe stage nuclei for each genotype. Images were captured on the DeltaVision system, typically with fifteen 0.20 μm sections. Deconvolved images were used to draw a line across each focus. Pixel intensities along the line were collected by ImageJ Plot Profile. Among foci, the number of pixels with signals above background levels varied from 6 to 8 pixels. To compare signal intensity distributions, location of peaks and degree of mCherry and GFP overlap, the ImageJ Coloc2 plugin was used to calculate the Pearson's coefficient for each focus. One hundred foci were analyzed and mean Pearson's coefficient and standard deviations are reported.

Analysis of ddbt-induced paternal effects

To characterize parental origin of defective chromosomes in *ddbt*-derived embryos, we collected embryos of mothers homozygous for *T(2;3) lt*^{x13} (WAKIMOTO and HEARN 1990) and

fathers that were *ddbt*$^{+}$ or *ddbt*^{Z4344} for 45 minutes, then aged embryos for 45 minutes. Embryos were bleach dechorionated, devitellinized in octane and methanol (1:1), methanol fixed, then incubated for 5 min in 0.5% sodium citrate before they were transferred to poly-L-lysine-coated slides. They were treated briefly with 45% acetic acid, squashed under a coverslip, snap frozen, dehydrated in 95% ethanol, then rehydrated and mounted in Vectashield with DAPI.

For each timed series, we used crosses to obtain control (*ddbt*$^{+}$ *st /TM6, Sb e*) and mutant (*ddbt*^{Z4344} *st / ddbt*^{Z4344} *e)* brothers from the same cross to minimize differences in genetic background. The brothers were mated in parallel to *Sevelin* females. We showed that 90.1% of the embryos (n=604) produced by the control cross hatched but no embryos (n=690) hatched from the experimental cross. Embryo processing involved dechorionation in bleach, devitellinization in a 1:1 mixture of octane and methanol, then fixation in methanol before DAPI staining and/or immunostaining as described by Rothwell and Sullivan (2000). Primary antibodies were mouse monoclonal anti-α-tubulin (1:250, Amersham), rabbit anti-CNN (1:1000, gift of T. Kaufman), rabbit anti-GFP (1:1000, Torrey Pines), rabbit anti-BubR1 (1:1000, gift of Claudio Sunkel), and mouse monoclonal anti-γ-H2V (1:1000, DSHB). Fluorescent secondary antibodies were goat anti-mouse Alexa Fluor 488 (1:200, Molecular Probes), goat anti-rabbit Alexa Fluor 568 (1:1000, Molecular Probes), and goat anti-rabbit DyLight 650 (1:1000, Thermo Scientific). For DNA staining, embryos were mounted on slides in Vectashield with DAPI (Vector Laboratories) or stained with propidium iodide before mounting in SeeDB (KE *et al.* 2013). Images were acquired on a Leica TCS SP5 II confocal microscope with 40X/NA1.25 or 63X/NA1.4 oil immersion objectives or 63X NA1.2 water immersion objective, with 0.42 μm optical sections. Z-series stacks were assembled using Image J and figures were edited using Photoshop (Adobe).

For the time series, developmental stage and cell cycle phase were determined for each embryo with scoring blinded to genotypes. The data were used to compute descriptive statistics, including mean, standard deviation, and percentages of embryos in each stage and phase for each time period. The first three sets (spanning 0 to 45 minutes AED) comprised the early series. For this series, we assessed association of paternal genotype with embryonic cycle using the χ^2 test. Komogorov-Smirnoff tests were used to compare cycle distributions for each time set. We fit a multinomial logistic regression model with cycle at each time as the outcome. The model included paternal genotype, time AED, and genotype by time interaction. To determine if there was an association of paternal genotype with cell cycle phase, we defined anaphase (A) as a single time point then determined if differences were observed in frequencies of embryos at stages prior to A or at or beyond A. We applied Fisher exact tests to compare frequencies. The last six time sets (spanning 45-135 minutes AED) comprised the late series. We recorded the number of embryos with features listed in Table 2.4. We used a scoring scheme that categorized features as a binary outcome (*e.g.* synchronous or asynchronous phases within the embryo). To compare each binary outcome between time intervals, we created a logistic regression model. If a significant difference was observed, the Tukey-Kramer method of multiple comparisons was used to determine which time differed from its previous interval with respect to the outcome. For the number of combinations, we used a non-parametric Krukal-Wallis test to examine any differences in time. We used SAS version 9.4 statistical software package (SAS Institute) and a significance level of 0.05 for all statistical tests.

To study the effects of maternal cytoplasm on *ddbt*-induced cell cycle arrest, we crossed mothers that were heterozygous or homozygous for the chk1 mutation (grp^{fs1}) or the chk2

mutation (mnk^{P6}) to males that were $ddbt^{Z4344}$ *st* / $ddbt^{Z4344}$ *e.* We used the w^{1118} strain as the control for maternal genotype (BRODSKY *et al.* 2004) and processed embryos as described above.

Data availability

The *ddbt* mutant strains are available upon request. The newly annotated sequences noted in Table S2.1 have been submitted to FlyBase.

Results

A paternal effect gene, deadbeat (ddbt), encodes a sperm nuclear basic protein

We recovered the first *ddbt* mutation from a screen for recessive male sterile mutations (WAKIMOTO *et al.* 2004). We found that it belongs to a rare class of paternal effect mutations in which mutant males produce fertilization-competent sperm but no viable offspring. We mapped *ddbt* to the 61B-C cytogenetic interval and found that $ddbt^{Z4344}$*/ Deficiency* flies were male sterile (Table 2.1) but showed no other detectable abnormality. Of the 61B-C genes, *CG34264* was a prime candidate because publicly available expression profiles reported testis-enriched expression (DOS SANTOS *et al.* 2015). We confirmed gene identity by sequencing $ddbt^{Z4344}$ and identifying a G/C to A/T transition in the *CG34264* open reading frame that changes the Gln48 codon to a premature stop. In addition, a transgene containing *CG34264* in a 3.8 kb genomic fragment was sufficient to fully rescue of $ddbt^{Z4344}$ male sterility *(*Figure 2.1A, Table 2.1).

The predicted Ddbt protein, supported by cDNAs, has 117 amino acids, of which 31% are positively charged and mostly clustered in the C-terminal half. Predicted secondary structure consists of four alpha helices spanning 60% of the protein (Figure 2.1A). The structural features

and testis expression are characteristic of Sperm Nuclear Basic Proteins (SNBPs) (EIRIN-LOPEZ and AUSIO 2009). This protein class includes Mst35Ba and Mst35Bb, which are the products of duplicate genes, and Mst77F (RATHKE *et al.* 2014). Consistent with the proposed role of SNBPs, males mutant for Mst35Ba and Mst35Bb or Mst77F produce spermatids or sperm with abnormal nuclear shapes (RATHKE *et al.* 2010; TIRMARCHE *et al.* 2014). However, we did not detect abnormalities in sperm (n=3050) of $ddbt^{Z4344}/Df$ males, as assayed by staining with the DNA dye DAPI.

To study Ddbt conservation across species, we searched for orthologs and identified single copy orthologous loci in other Dipterans within the Schizophora (Table S2.1) but no similar sequences in non-Schizophoran insects. Species containing Ddbt sequences span an estimated divergence time of 65 MY (WIEGMANN *et al.* 2011). Comparisons among predicted proteins revealed the following features. First, they are enriched in Arginine, which comprises 14 to 30% of the residues. Second, three motifs of five to eleven amino acids, each located in a predicted alpha helix, are conserved in sequence and order. Third, spacing between motifs is highly conserved, suggesting that a central core of 51 to 52 amino acids is critical for function (Figures 2.1B and 2.2). There is more variability in the length of N and C tails. The Ddbt motif, PYLNFLRFLKR, is the most highly conserved and is particularly interesting since similar sequences were found in Mst35Ba, Mst35Bb and Mst77F (Figure 2.1, C and D). The presence of the motif among these proteins supports classification of Dbbt as an SNBP and suggests an important role for SNBP function.

In a comparison of human and rodent orthologs, Swanson and Vacquier (2002) discovered that SNBPs are among the most rapidly evolving proteins. Their finding motivated us to ask whether Ddbt shares this property with mammalian SNBPs. To estimate rates of

nonsynonymous to synonymous nucleotide substitutions within the *melanogaster* subgroup, we analyzed sequences from five species using PAML4 (YANG 2007). The results yielded a d_N/d_S (ratio of nonsynonymous to synonymous substitutions) of 0.60 averaged over all sites (Table 2.2), a value that meets the cutoff criteria of a $d_N/d_S > 0.5$ as indicating a high likelihood of being a target of positive selection (SWANSON *et al.* 2004). To determine the fraction of sites under positive selection, we estimated d_N/d_S variation between sites and compared the results of models that assume neutral evolution or positive selection. The comparisons yielded statistically significant signatures of positive selection ($p < 0.05$), indicating 15.8% of amino acid sites analyzed with a d_N/d_S ratio of 3.9. These amino acids reside outside of the central core (Figure 2.2B), supporting our conclusion from analysis of 28 Dipteran species that Ddbt has a constrained central core. Moreover, the data suggest that sequence variation in N and C tails is important and has been driven by advantageous consequences rather than relaxed selection.

Ddbt localizes to telomeres in post-meiotic stages of spermatogenesis

To determine Ddbt's subcellular localization, we constructed *enhanced green fluorescent protein-ddbt* and *mcherry-ddbt* transgenes driven by the endogenous *ddbt* promoter and inserted the transgenes into the *attP40* genomic site. Expression of *egfp- ddbt* and *mcherry-ddbt* rescued $ddbt^{Z4344}$ male sterility to 90% and 100% respectively of control levels (Table 2.1), allowing their use to assay Ddbt localization. We detected EGFP-Ddbt and mCherry-Ddbt only in post-meiotic spermatogenic stages. Progressively later stages of spermatid development can be distinguished by nuclear morphology as round, elongating, canoe, and fully condensed nuclei (LINDSLEY and TOKUYASU 1980). EGFP-Ddbt was first detected at early canoe stage. Each nucleus had two to four distinct foci, with one or two typically larger and brighter than the other foci (Figure 2.3A).

Ddbt appearance preceded that of Mst35Bb, which appeared broadly distributed in late canoe stage nuclei. Like Mst35Bb, EGFP-Ddbt persisted through spermiogenesis and foci were reliably detected in mature sperm. However, we were unable to detect EGFP-Ddbt in decondensing sperm chromatin shortly after sperm enter the egg (n=15) or in the male pronucleus (n=31) by EGFP fluorescence or immunostaining, perhaps due to rapid removal of Ddbt by factors in the egg cytoplasm.

Ddbt foci in spermatids and mature sperm resembled those reported for telomere capping proteins (DUBRUILLE *et al.* 2010; GAO *et al.* 2011). To determine if Ddbt was telomeric, we co-expressed mCherry-Ddbt and GFP-K81 and observed identical patterns in canoe stage nuclei (Figure 2.3B). Because large K81 foci represent aggregated telomeres (WESOLOWSKA *et al.* 2013), we selected the smallest focus in each spermatid nucleus for a precise comparison of signal location (Figure 2.3, B and C). Analysis of the smallest foci of mCherry-Ddbt and GFP-K81 from spermatid nuclei yielded a Pearson Correlation Coefficient of 0.92 ± 0.04 (n=100 nuclei). These values were comparable to those obtained from comparing signals in spermatids expressing mCherry-Ddbt and EGFP-Ddbt (Pearson Correlation Coefficient = 0.89 ± 0.04, n=100). We conclude that Ddbt is telomeric but cannot distinguish whether it resides in or in close proximity (within the 200 nm limit resolution of microscopy) to the TCC.

Ddbt maintains the telomere-capping complex through spermiogenesis

To study the relationship of Ddbt to the TCC, we asked if distribution and dynamics of TCC components were disrupted in $ddbt^{Z4344}$ spermatids. We assayed K81, which is present on meiotic telomeres and retained postmeiotically (GAO *et al.* 2011; DUBRUILLE and LOPPIN 2015). We observed the expected pattern of GFP-K81 foci in control spermatids. GFP-K81 signals were

also present in *ddbt*^{Z4344} spermatids at elongating and early canoe stages, but signals were no longer detectable at late canoe stage (Figure 2.4A). We also tracked HP1a (Figure 2.4B) and HOAP (Figure 2.4C) by immunostaining and obtained results consistent with those for GFP-K81 in control and *ddbt*^{Z4344} spermatids. Therefore, Ddbt is required between early and late canoe stages for telomeric retention of at least three TCC components in spermatids. Hereafter, we describe *ddbt* ^{Z4344} sperm telomeres as uncapped.

Since Ddbt was first detected at early canoe stage, we asked if its localization depended on K81. We used EGFP-Ddbt to monitor the protein in spermatids of males with a *k81*^{2} null mutation (YASUDA *et al.* 1995). In the absence of K81, EGFP-Ddbt was not detected (Figure 2.4D). Thus, Ddbt is recruited to spermatid telomeres in a K81- dependent manner.

As previously described, Ddbt, Mst35B, and Mst77F share the NFLR motif (Figure 2.1C). To test whether Ddbt's motif is important for localization, we replaced its PYLNFLRFLKR with the Mst35Ba and Mst77F motifs in the context of the *egfp-ddbt* transgene. The replacements altered 4/11 and 5/11 amino acids respectively but preserved the invariant R (Figure 2.1C). Both changes led to male sterility (Table 2.1). Spermatids lacked EGFP signals as monitored by GFP fluorescence or immunostaining, suggesting that PYLNFLRFLKR may be important for Ddbt stability. We also created a transgene with the invariant R at amino acid position 7 replaced with aspartic acid (Ddbt^{R31D}). This change also resulted in male sterility (Table 2.1). We tested whether EGFP-Ddbt^{R31D} was expressed and localized to telomeres by monitoring EGFP fluorescence and HOAP immunostaining. In early canoe stage *egfp-ddbt*^{R31D} spermatids, we detected 0-4 foci which were weaker in intensity compared to the 2-4 foci observed in control spermatids. Of the 273 foci observed in 100 early canoe stage *egfp-ddbt*^{R31D} spermatids, 50% were EGFP and HOAP positive, 43% were only HOAP positive, and 7% were

only EGFP positive (Figure 2.4E). EGFP signals were not observed in late canoe stage spermatids (n=500) or mature sperm (n= 700). The detection of EGFP-Ddbt^{R31D} in 57% of the foci indicated that the mutant protein was imported into nuclei and recruited to telomeres but its recruitment and ability to maintain TCCs were compromised.

Loss of Ddbt induces a paternal effect on the earliest embryonic cell cycles

Since *ddbt* males produce sperm lacking TCCs, we characterized *ddbt's* paternal effect to compare it to *k81's* paternal effect. Previous studies showed that embryos of *k81* fathers arrest either in mid-cleavage or late in embryogenesis (FUYAMA 1984; YASUDA *et al.* 1995). Early lethality was proposed to result from efficient paternal telomere fusions in cycle 1 leading to anaphase bridges and chromosome breaks that accumulate and cause arrest (DUBRUILLE *et al*. 2010; GAO *et al.* 2011). Late lethality was attributed to failure of paternal chromosomes to participate in the first division, with maternal chromosomes proceeding through nuclear divisions and yielding gynogenetic haploids that develop to cuticle formation (FUYAMA 1984).

To compare paternal versus maternal chromosome behavior in *ddbt*'s embryos, we crossed *ddbt*$^{+}$ control and *ddbt*^{Z4344} males to females that were homozygous for *T(2;3)lt*^{x13}, a cytologically visible reciprocal translocation (WAKIMOTO and HEARN 1990). We aged their embryos for 45-90 minutes after egg deposition (AED) and examined their karyotypes. In the control group, we observed the expected developmental stages from mid-cleavage to syncytial blastoderm and the expected karyotype, with 36% of the embryos in prometaphase or metaphase (n=60). In contrast, *ddbt*'s embryos had variably sized clumps of chromatin, with an average number of seven and up to 16 per embryo. In addition, 80% of *ddbt*'s embryos were in prometaphase or metaphase (n=60). Abnormal mitotic figures were frequent and included

anaphase configurations in which maternal chromosomes segregated to the poles but entangled paternal chromosomes lagged (Figure 2.5, A and A'). We observed only two clear cases of karyotypes with circular chromosomes indicative of intrachromosomal telomere fusions (Figure 2.5A'', n=327 clumps with prometaphase or metaphase chromosomes analyzed from 45 embryos). Chromosome fragmentation was frequent and involved all chromosomes of the paternal and maternal sets (Figure 2.5A''').

To assess when *ddbt*-induced defects arose, we examined offspring of wild type mothers and $ddbt^{Z4344}$ or $ddbt^{+}$ fathers. We collected newly fertilized eggs over a 15-minute interval, aged them for set times before fixation and assayed developmental stage, morphology, and cell cycle phase. We collected nine sets to capture events from 0 to 135 minutes AED. The first three sets (0-15, 15-30, and 30-45 min AED) constitute the early time series and the next six constitute the late time series.

In *Drosophila,* sperm enter the egg at its anterior end. Within a few minutes, the sperm nucleus decondenses, forms the male pronucleus, moves to a central position in the egg (PM) and replicates. Apposition of male and female pronuclei (PA) occurs at a characteristic internal position in the egg, about one-third the egg's length. Paternal and maternal chromosomes condense and align as separate groups on first spindle, which is called the gonomeric spindle. Parental genomes remain as separate groups until mixing at anaphase. Cycle 1 is completed within 17 minutes of sperm entry (FOE *et al.* 1993). Cycles 2 through 7, each lasting about 8 to 9 minutes, follow in rapid succession internally within the embryo. Cycles through 13 are characterized by synchronous nuclear behavior. Loss of synchrony and cellularization of blastoderm nuclei by membrane invaginations occurs in cycle 14.

In our early time series, observed stages were mostly PM/PA, cycle 1, cycle 2, and cycle 3 (Figure 2.5B). In the first 0-15 min set, stage distributions of control and *ddbt's* embryos were indistinguishable. Over 50% in both sets were at the PM or PA stage. Pronuclear morphology and placement in the egg appeared normal, indicating that Ddbt is not required pre-cycle 1.

Stage distributions differed markedly between the groups at 15-30 min AED ($p<0.0001$) and 30-45 min AED ($p<0.001$), with *ddbt* offspring showing developmental delay in cycle 1. Since our previous analysis of older embryos indicated a high proportion of *ddbt*'s embryos in prometaphase or metaphase, we asked whether the cycle 1 delay could be in the onset of anaphase. We compared the proportion of embryos in phases before anaphase to the proportion in anaphase or later phases for control and experimental groups (Table 2.3). In the 15-30 min AED set, a clear difference was evident, with 37.7% of control embryos at phases prior to cycle 1 anaphase compared to 51.5 % of *ddbt*'s embryos ($p<0.0001$). Developmental delay was also observed as the increased number of *ddbt*'s embryos in cycle 2 and cycle 3 in both 15-30 and 30-45 min AED sets ($p<0.05$ in all cases). Thus, defective paternal chromosomes result in a delay in the onset of anaphase as early as cycle 1 and with a continuing impact in the next two cycles.

Cytological observations provided morphological evidence of defects in the earliest cell cycles. Control embryos appeared normal at cycles 1, 2, and 3 (n=604, Figure 2.5, C and D) but *ddbt*'s embryos showed defects in all three cycles (n=584). We observed an occasional difference in the degree of condensation of parental chromosome sets as early as cycle 1 prometaphase (5.5%, n=128) and metaphase (7.9%, n=114, Figure 2.5E). Spindle morphology appeared normal, as assayed by α- tubulin and Centrosomin (CNN) immunostaining, and maternal and paternal chromosome sets initiated anaphase. Abnormalities were striking and consistent at anaphase (Figure 2.5, F-H) and telophase (Figure 2.5I). By late anaphase, one set

clearly proceeded to segregate to the spindle poles while the other lagged in the mid-zone. This defect was observed in all embryos in anaphase and telophase of cycle 1 (n=54), cycle 2 (n=44) and cycle 3 (n=22) indicating consistent delay or failure in paternal chromosome segregation. We also noted configurations in which distal regions of chromosome arms in the lagging set remained at the mid-zone (Figure 2.5 A' and J). These appeared to be associations between extended telomeric regions of chromatids, as indicated by intense DAPI staining. We refer to these as telomere associations (TAs) rather than telomere end-to-end fusions, which are best documented as a chromatin bridge extending between segregating chromosomes. By the end of Cycle 1 and Cycle 2, embryos did show a thin chromatin bridge extending between well-separated daughter nuclei (Figure 2.5, K and L) (n=20), providing evidence that at least some telomere associations can lead to telomere fusions. We conclude from the early time series that uncapped telomeres cause delayed anaphase onset, frequent telomere associations, delay or failure in paternal chromosome segregation, and chromatin bridging in the earliest cell cycles.

The majority of ddbt offspring show early arrest and chromosomal catastrophes

To understand how *ddbt*-induced early defects lead to embryonic lethality, we examined the late time series, which consisted of embryos aged from 45 to 135 min AED. By this time, 83% of control embryos had developed beyond cycle 6 (Figure 2.6A). In contrast, *ddbt*'s embryos showed a range of phenotypes that warranted a first-level classification into two categories based on developmental progression. The majority had only a few chromatin clumps indicative of early arrest and a minority had progressed further but showed haploid karyotypes (Table 2.4). As noted previously, *ddbt*'s embryos in cycles 1, 2, and 3 showed 100% penetrance of the TA phenotype. Analysis of the late time series confirmed that nearly all of *ddbt*'s embryos

completed cycle 1 (98.7%) and cycle 2 (90.6%) (Figure 2.6A). Thus, TAs were apparently resolved to permit completion of these cycles. Surprisingly, only 22.8% completed cycle 3 and none finished cycle 6. Across the entire late time series, embryos produced an average of six nuclei (Table 2.4). Thus, there is a fundamental difference in response to uncapped telomeres in the first two cycles compared to later cycles.

Although *ddbt's* embryos arrest in or shortly after cycle 3, they undergo changes that are informative for identifying which nuclear or cytoplasmic processes are influenced by uncapped telomeres (Table 2.4). For instance, chromatin clumps became more dispersed in embryos over time. By 120 min AED, nearly 50% of the embryos show a dispersed distribution of clumps along the anterior-posterior axis. This indicates that the cytoskeleton-dependent axial expansion of nuclei, which takes place in normal cycle 4-6 embryos at 45-60 min AED (BAKER *et al.* 1993), can occur in *ddbt's* embryos albeit with delayed timing.

We observed a suite of abnormalities, including loss of nuclear synchrony, hypercondensation, variations in ploidy, and chromosome fragmentation. Defects generally increased in frequency with age but there were differences in when they appeared. Although embryo age varied as much as 15 minutes within a time set due to the egg collection interval, we observed abrupt increases in occurrence of some defects in a narrow 15-30 min interval. For example, only 6% of *ddbt*'s embryos in cycle 3 showed asynchronous nuclear behavior by 45 min AED. By 60 min AED, percentage increased to nearly 50% and by 90 min AED, 93.1% of embryos exhibited asynchrony. Centrosomal defects and disorganized spindles were common, reflecting disruption in coordination of chromosomal and centrosomal cycles (Figure 2.6, B and C). Asynchrony decreased significantly by 105 min AED due to accumulation prometaphase or metaphase configurations. Non-uniform chromatin condensation or hypercondensation were

observed within nuclei (Figure 2.6D), with most significant increase in frequency occurring 75-90 min AED. Marked rise in aneuploidy in this same interval was consistent with continued chromatin loss from spindles. While average number of nuclei per embryo did not change, multiple rounds of DNA replication continued so that nearly all embryos had at least one hyperploid chromatin clump by 90-105 min AED. Strikingly, in the 75-105 min AED interval, there was a sharp increase in the frequency of extensive chromosome fragmentation, involving both parental chromosome sets.

In summary, analysis of the late series identified critical transition points that dictated how nuclei dealt with uncapped telomeres and ultimately led to defects commonly referred to as mitotic catastrophes. An important transition occurred in cycle 3 that prevented most nuclei from continuing through division. Variation in nuclear behavior may be due to differences in the number of uncapped chromosomes inherited by daughter nuclei but embryos in this category responded by developmental arrest no later than cycle 6. Maternal and paternal chromosomes continued to show dynamic changes in condensation and replication until they arrested in a prometaphase- or metaphase-like state then fragmented due to an unknown mechanism.

Uncapped telomeres in the earliest embryonic stages do not show detectable levels of γH2Av or BubR1 markers typical of DNA breaks

Because uncapped telomeres are recognized as double-strand DNA breaks (DSBs) in somatic cells (RONG 2008), we asked whether uncapped telomeres are similarly recognized as DSBs in the earliest embryonic cycles. We monitored the phosphorylation status of the DSB marker H2Av (γH2Av) by immunostaining (LAKE *et al.* 2013). We did not detect γH2Av in *ddbt*'s embryos that were in cycles 1-3 (n=64). However, γH2Av was reproducibly detected in

late stage embryos (n=10) that showed chromosome fragmentation (Figure S2.1A). As a comparison, we collected embryos from fathers that had been treated with 15,000 rads of γ-rays. This dose of irradiation results in early cell cycle arrest in offspring of treated males (SCRIBA 1964). We did not observe γH2Av foci in these embryos in cycles 1-3 (n=28). However, γH2Av foci were detected in embryos aged 45-90 min AED when they exhibited extensive chromosome fragmentation (Figure S2.1A).

BubR1, a component of the spindle assembly checkpoint at the kinetochore, has also been shown to localize to HOAP-depleted telomeres and sites of DNA breaks in larval neuroblasts (MUSARO *et al.* 2008; ROYOU *et al.* 2010). We asked whether BubR1 localized to uncapped telomeres in *ddbt*'s embryos. We assayed for GFP fluorescence in embryos of GFP-BubR1 mothers and *ddbt* fathers. We did not detect GFP-BubR1 telomeric signals in cycle 1-3 embryos (n=26) (Figure S2.1B). As expected, GFP-BubR1 was detected at kinetochores in prometaphase chromosomes in embryos in cycles 1-3 (n=10).

Although we did not detect γH2Av or BubR1 at uncapped telomeres in *ddbt's* embryos in cycles 1-3, we note that γH2Av was not detected in early embryos with heavily irradiated paternal chromosomes that likely carry many DSBs. In both cases, γH2Av was detected at broken ends in later stages. More sensitive assays may be needed to detect these proteins in the early embryo. Alternatively, uncapped telomeres and DNA breaks may not be processed in the early embryo as they are in neuroblasts and so these proteins are not localized to these chromosome ends.

Selective elimination of paternal chromosomes creates lethal haploid and haploid mosaics

As predicted from *k81* studies, a minority of *ddbt's* embryos progressed passed the early cycles and showed haploid karyotypes. Careful inspection of these embryos yielded two unexpected findings. First, their frequency increased over time (Table 2.4). Second, only a subset uniformly exhibited haploid karyotypes or interphase nuclei typical of true haploids that result from cycle 1 elimination of the paternal genome. The remaining embryos exhibited a mixture of haploid, hyper- and hypo-haploid karyotypes (Figure 2.6E-E''). This larger class of haploid mosaics and their increased frequency over time can be accounted for by loss of paternal chromosomes in cycle 2 or later. We observed variability in chromosome content of nearby spindles (Figure 2.6E'') indicating that loss events continued within a lineage and involved a subset or all paternal chromosomes. Haploid and haploid mosaics comprised 7% of all embryos at 45-60 min AED but up to 40.6% by 120 min AED. We observed mosaics that had progressed to syncytial blastoderm stage in the late series. However, most died before late embryogenesis. This conclusion is based on an assay of *ddbt*'s embryos at 24 hrs AED. We found that only 7.2% (n=681) survived long enough to produce the defective denticle belt pattern and mouthparts characteristic of haploid embryos (FUYAMA 1984).

Our discovery of *ddbt*-induced haploid mosaics showed that paternally inherited chromosomes were subject to continued elimination, even after several rounds of DNA replication. Similarly, we observed that up to 20% of the embryos of *k81* fathers are haploid or haploid mosaics at 90 min AED (Table S2.2 and Figure S2.2). Consistent with our results with *ddbt*'s embryos, only a few of *k81*'s embryos survive to late embryogenesis (YASUDA *et al.* 1995; LANGLEY *et al.* 2011). In these haploid and haploid mosaic embryos, a sufficient number of

paternal chromosomes must have been eliminated from most spindles to avoid the cycle 3 arrest and allow haploid or aneuploid nuclei to populate the embryo.

Reduced level of maternal Checkpoint Kinase 2 relieves ddbt-induced cycle 3 arrest.

Our studies showed that embryos produced by wild type mothers and *ddbt* fathers delayed initiation of anaphase at cycle 1 and the majority arrested in or shortly after cycle 3. We asked whether either phenotype was regulated by Chk1 or Chk2 kinases, which are best characterized as regulator of checkpoints at later stages of embryogenesis (FOGARTY *et al.* 1997; TAKADA *et al.* 2003). We analyzed embryos produced by *ddbt* fathers and mothers that were homozygous or heterozygous for grp^{fs1} or mnk^{P6}, and therefore produce eggs with reduced levels of Chk1 or Chk2 respectively. In contrast to females from our wild type strain, *grp* and *mnk* mutant females produced fewer embryos, and even short 15-minute egg collections yielded embryos at a wide variety of developmental stages. Because of these features, we were unable to quantitatively assess the effect of maternal genotype on timing of cycle 1 anaphase initiation. However, we could examine effects on the *ddb*t-induced cycle 3 arrest. Stage distribution of embryos (n=57) produced by *ddbt* fathers and *grp*/+ mothers were similar to that of embryos (n=49) of *grp*/*grp* mothers with over 60% of the embryos in cycle 1-3 at 30-45 min AED (data not shown). This comparison indicates that reduced levels of Chk1 did not have a striking effect on the cycle 3 arrest phenotype.

In contrast, the Chk2 mutation, mnk^{P6}, profoundly affected developmental progression of *ddbt's* embryos. The effect was observed in offspring of mothers with reduced levels of Chk2, consistent with its known haploinsufficiency (IAMPIETRO *et al.* 2014). As expected, the majority of embryos produced by control mothers fail to progress past cycle 3 by 45 min AED. However,

the majority of embryos produced by *mnk*/+ or *mnk*/*mnk* mothers and *ddbt* fathers progressed beyond cycle 3 by 45 min AED (Figure 2.7A). Moreover, embryos in the later cycles frequently showed lagging chromosomes, anaphase bridges, and nuclei with variable chromosome content (Figure 2.7B). These features showed that nuclear cycles were allowed to continue even in the presence of damaged chromosomes. We conclude that the *ddbt*-induced cycle 3 arrest is Chk2 dependent.

Discussion

Sperm nuclear basic proteins comprise an ancient and widespread protein class in animals. Their origin has long been the topic of speculation but current data suggest derivation from histone H1 (EIRIN-LOPEZ and AUSIO 2009). Molecular and temporal profiles support the notion that SNBPs function in spermatid nuclear condensation. However, only a few SNBPs have been knocked out to critically assess function (RATHKE *et al.* 2014). Our study reveals a different role for one SNBP. We show that Ddbt is a telomere-enriched SNBP. It achieves its localization after meiotic replacement of the TCC HipHop with its sperm-specific version K81, but before spermatids incorporate Mst35Ba, Mst35Bb and Mst77F. We propose from its temporal profile and mutant phenotypes that Ddbt is unlikely to be a capping protein. Instead, it requires TCCs for telomeric recruitment but functions to guard against TCC loss in spermatids during replacement of chromosomal proteins with other SNBPs.

Spermatid remodeling is typically envisioned as bulk histone removal and replacement by SNBPs, but there is evidence for regional retention of proteins at specific loci. In human sperm, sites of retained histones are enriched in developmentally important genes (HAMMOUD *et al.* 2009). There is long-standing evidence for retention of a centromeric protein in bovine sperm

(PALMER *et al.* 1990). We expand the list of retained proteins to include TCCs in *Drosophila* sperm with retention achieved through action of a regionally restricted SNBP. To our knowledge, Ddbt is the first reported example of an SNBP that retains chromosomal proteins through spermiogenesis and ensures normal behavior of paternal chromosomes during embryogenesis. We predict that other SNBPs may be similarly specialized and when disrupted, could cause paternal effect defects.

The view of Ddbt as a specialized SNBP raises questions about its origin and conservation. Ddbt's function has only been assessed in *D. melanogaster*, but conservation of its core region suggests conserved function. If so, then the lineage that gave rise to Schizophoran species recruited Ddbt to meet the challenge of TCC maintenance during spermiogenesis over 65 MYA.

Previous studies reported that mammalian SNBPs (SWANSON and VACQUIER 2002) and *Drosophila* TCC-restricted proteins (RAFFA *et al.* 2011) are rapidly evolving. Ddbt is a telomere enriched SNBP with rapidly evolving N and C tails. Its overall d_N/d_S places it in the 99.9 percentile when compared to 8510 single copy orthologs in the melanogaster subgroup whose d_N/d_S ratios were similarly calculated (DROSOPHILA 12 GENOMES *et al.* 2007). Rapid evolution of reproductive proteins has been attributed to a variety of selective forces, including sperm competition and fertilization competency (SWANSON and VACQUIER 2002). Rapid evolution of TCC restricted proteins has been attributed to relaxed constraints on telomere sequences in the Dipteran lineage, after the loss of telomerase (estimated 260 MYA) and with adoption of retrotransposon-based telomere elongation mechanisms (estimated 65-230 MYA) (MASON *et al.* 2015). We propose that Ddbt may be co-evolving with TCCs to enable its telomere recruitment.

It may also be co-evolving with interacting egg proteins that remove it from sperm chromatin or use it to assemble new TCCs.

Ddbt may prevent TCC loss by directly interacting with the capping complex or by creating a protective chromatin environment. Whatever the mechanism, we expect consequences of its loss to be as severe or more severe than K81 loss. Our results confirm those from *k81* studies in demonstrating that loss of K81, HP1a, and HOAP is not problematic for spermiogenesis. Although we cannot eliminate the possibility that Ddbt has additional non-telomeric functions, we did not observe defects in sperm nuclear condensation, production of fertilization competent sperm, or male pronuclear behavior. Instead, $ddbt^{Z4344}$ offspring show striking defects. High penetrance of the paternal effect defects is likely due to complete TCC loss during spermiogenesis and complete failure to assemble TCCs on newly replicated paternal chromosomes, an expected outcome because *Drosophila* telomere capping is epigenetically determined rather than sequence dependent (RONG 2008).

Persistence of uncapped telomeres in *ddbt*'s embryos allowed us to investigate how early embryos respond to uncapped telomeres. Quantitative analysis of a carefully timed series identified responses that were not previously detected in *k81* studies but account for the unusual paternal effect shared by *ddbt* and *k81* mutants. As summarized by Figure 2.8, we account for observed frequencies of early arrest and haploid embryo categories. Importantly, we show that uncapped telomeres are detected in the earliest cycles and elicit different responses in cycles 1 and 2 compared to later cycles. We provide evidence that uncapped telomeres delay anaphase onset as early as cycle 1. We speculate that this delay may be due to the existence of a telomere checkpoint in the earliest cell cycles. Once anaphase initiates, the entire set of *ddbt* paternal chromosomes while present on the gonomeric spindle only infrequently fails to participate in the

first division, resulting in true gynogenetic haploids. Most frequently, both maternal and paternal sets initiate anaphase, but uncapped paternal chromosomes are relatively delayed in segregation. Delay in anaphase segregation may be due to prevalent and persistent telomere associations. These associations (TAs) or telomere fusions may mechanically hinder chromosome movement to the poles yet *ddbt*'s embryos complete cycles 1 and 2 and enter cycle 3. In or shortly after cycle 3, uncapped telomeres are perceived as a Chk2-dependent signal to stop division. Signal strength may depend on the number of uncapped telomeres and newly generated breaks. Most cycle 3 nuclei halt division in or shortly after cycle 3. However, nuclei that lose a sufficient number of uncapped chromosomes undergo one or more additional divisions, generating embryos that may either proceed as far as cycle 6 before arresting or continue longer as haploid mosaics. Precedence for a threshold mechanism of signal strength is provided by the elegant studies of Kaul *et al*. (2012). They showed that normal human cells in culture spontaneously produce dysfunctional non-fusogenic telomeres and the accumulation of five such telomeres is sufficient to induce cell cycle arrest.

It is unclear whether embryos in cycles 1 through 3 are proficient at perceiving uncapped telomeres as either dysfunctional telomeres *per se* or more generically as double-strand DNA breaks (DSBs). Detection as dysfunctional and irreparable telomeres and the response to delay anaphase onset are consistent with findings of Musaro *et al.* (2008) who showed that HOAP-depleted telomeres reduce the frequency of anaphases in larval neuroblasts. Detection of uncapped telomeres as DSBs is supported by the largely similar phenotypes of *ddbt*'s offspring and offspring of males treated with high doses of X-rays. Scriba (1964) X-ray irradiated males with 15,000 rads. The dose was high enough to ensure that offspring were invariably lethal but the extent of damage among sperm could not be controlled. Nonetheless, Scriba noted that

embryos showed asynchronous nuclei and enrichment at prometaphase and metaphase. He concluded that paternal DNA breaks delay mitotic progression but he could not pinpoint when delay occurred. He noted early embryonic arrest before cycle 6, which is similar to that observed in *ddbt*'s and *k81*'s offspring (YASUDA *et al.* 1995). As pointed out by Counce (1973), mid-cleavage crisis can result from a variety of experimental or genetic interventions that damage nuclei or cytoplasm in the early embryo.

In yeast and in human and *Drosophila* somatic cells, uncapped telomeres induce a DNA damage response. In *Drosophila,* mutants that lose TCC components in neuroblasts typically show high frequencies of telomere fusions (RAFFA *et al.* 2011). Early embryos may show a weaker response to TCC loss. Our observation of frequent telomere associations is similar to the rather unusual effect of UbcD1 loss in larval neuroblasts (CENCI *et al.* 1997). We suggest that uncapped telomeres induced by *ddbt* may not be efficient substrates for DNA repair and ligation. Alternatively, the maternal machinery required to recognize and join uncapped telomeres may not be fully functional in the earliest embryonic cycles. Either explanation is consistent with our inability to detect γH2Av and BubR1 at the telomeres of *ddbt's* embryos. Notably, in their description of maternal effect mutants affecting telomere capping in embryos, Gao *et al.* (2009) reported that only a few telomere fusions occurred in embryos prior to cycle 7 but their frequency increased at later stages.

Our hypothesis of a checkpoint sensitive to telomere integrity in the first embryonic cycle may be considered surprising as it is often stated that early embryos lack cell cycle checkpoints. In their seminal paper on checkpoints, Hartwell and Weinert (1989) suggested that compared to yeast and somatic cells, embryos may lack some cell cycle checkpoints and respond differently to incomplete DNA replication and DNA breaks. They reasoned that by allowing cycle

progression, embryos might sacrifice mitotic fidelity during rapid, synchronous divisions, then compensate later by eliminating damaged nuclei. Their ideas were based in part from the behavior of checkpoints and responses in *Drosophila* embryos at the syncytial blastoderm stages (O'FARRELL *et al.* 2004). Our work adds to the few existing studies that have considered checkpoints and responses in *Drosophila's* earliest cell cycles. For example, Brunk *et al*. (2007) showed that early cell cycle progression was delayed in embryos of *microcephalin* mutant mothers. These embryos exhibited uncoordinated nuclear and centrosomal cycles as early as cycle 1. Similar to the *ddbt*-induced cycle 3 defect, microcephalin-induce mitotic defects were partially relieved by reduced levels of the Chk2 checkpoint kinase (RICKMYRE *et al.* 2007).

As noted by Rickmyre *et al*. (2007), the regulation of *Drosophila*'s earliest cycles can be challenging to study. These cycles occur before or just after egg deposition and deep within the egg, so they are refractory to live imaging and treatments with inhibitors that have been instrumental in documenting checkpoints in later stages. Maternal effect mutations have been especially useful for analyzing checkpoints in syncytial blastoderm embryos (FARRELL and O'FARRELL 2014). However, few maternal effect mutations affect cell cycle regulation only in embryos. They must allow egg production and so may produce a residual amount of protein product. Hence, their impact on early cycles may be minimal and difficult to detect. Our studies of *ddbt* had the advantage of analyzing embryos with completely normal maternal contributions and severely impaired and irreparable paternal defects. We contend that continued studies of *ddbt* and other strict paternal effect mutations will provide additional insights into cell cycle checkpoints and responses in early embryos.

Acknowledgements

We dedicate this paper to the memory of Gerold Schubiger, our mentor and friend. We thank Y. Rong, W. Theurkauf, A. Royou and the BDSC for stocks, W. Theurkauf, T. Kaufman, and C. Sunkel for antibodies, K. Fitch and E. Pak for their initial studies of *ddbt*, V. Gildengorin and P. Wakimoto for statistical expertise, W. Swanson for guidance on protein evolution analysis, Y. Rong, M. Schubiger, A. Sustar, and members of the Paredez Lab for helpful discussions. We also thank A. Paredez for generously allowing use of his microscope. This work was supported by M. J. Murdock Charitable Trust and Seattle University research funding to GKY and University of Washington Royalty Research Fund to BTW.

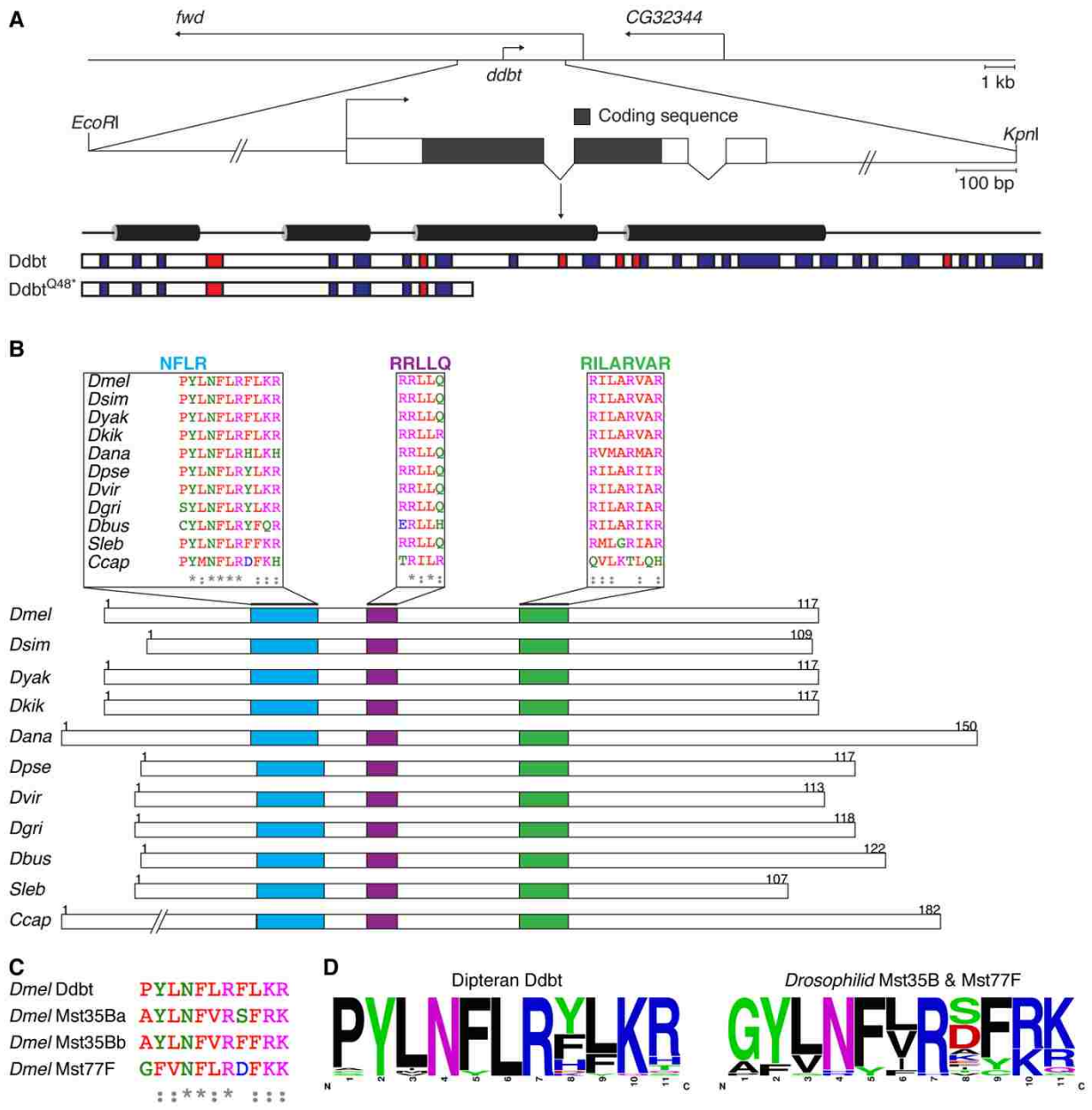


Figure 2.1 *ddbt* Encodes a Conserved Sperm Nuclear Basic Protein.
(A) Diagram depicts *ddbt* gene structure and location in the *fwd* intron, transcript structure, and positions of predicted Ddbt alpha helices, basic (R, K, H; blue) and acidic (D, E; red) amino acids in wild type and truncated proteins. (B) Alignment of orthologs and conserved motifs (colored). See Table S2.1 for species abbreviations. (C) Alignment of NFLR motifs in *D. melanogaster* SNBPs. (D) Logos show consensus motifs derived from 28 Dipteran Ddbts and 26 Drosophilid Mst35Bs and Mst77Fs

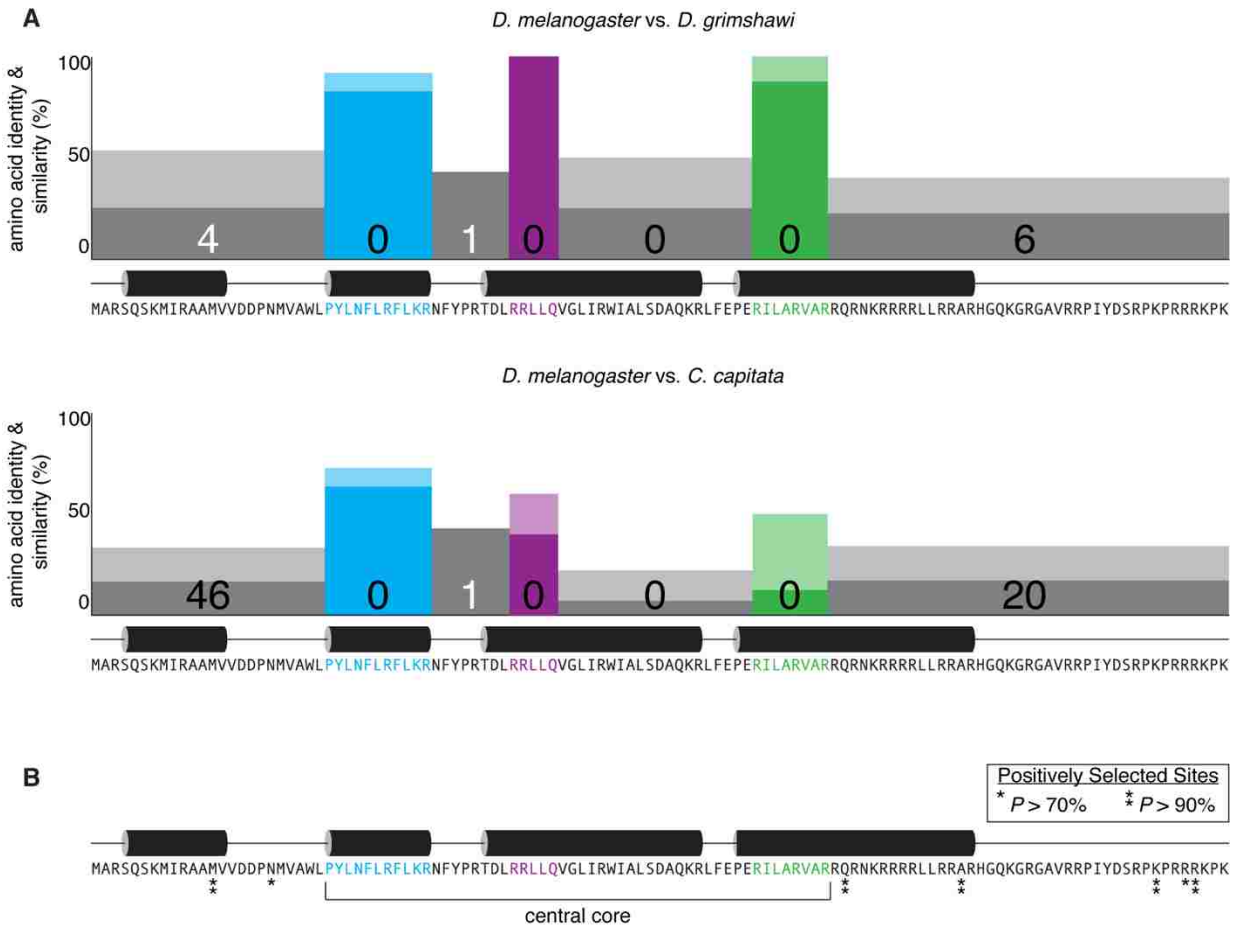


Figure 2.2 Conservation of Ddbt's Central Core.
(A) Pair-wise amino acid comparisons of orthologous proteins revealed a conserved central core. Representative examples compare *D. melanogaster* Ddbt (horizontal line) with the Hawaiian fruitfly *D. grimshawi* and the medfly *C. capitata* proteins. The protein was divided into seven subregions based on Figure 2.1 data and subregion lengths were compared. Numbers indicate if the *D. melanogaster* region has more (white), fewer or the same (black) number of amino acids. Percent similarity is depicted by height of the boxes, inclusive of darker colored boxes whose heights show percent identity with optimal alignment. (B) Location of positively selected amino acids on *D. melanogaster* Ddbt was based on analysis of five melanogaster subgroup species. Asterisks mark sites under positive selection at the noted posterior probability (*P*). See also Tables 2.2 and S2.1.

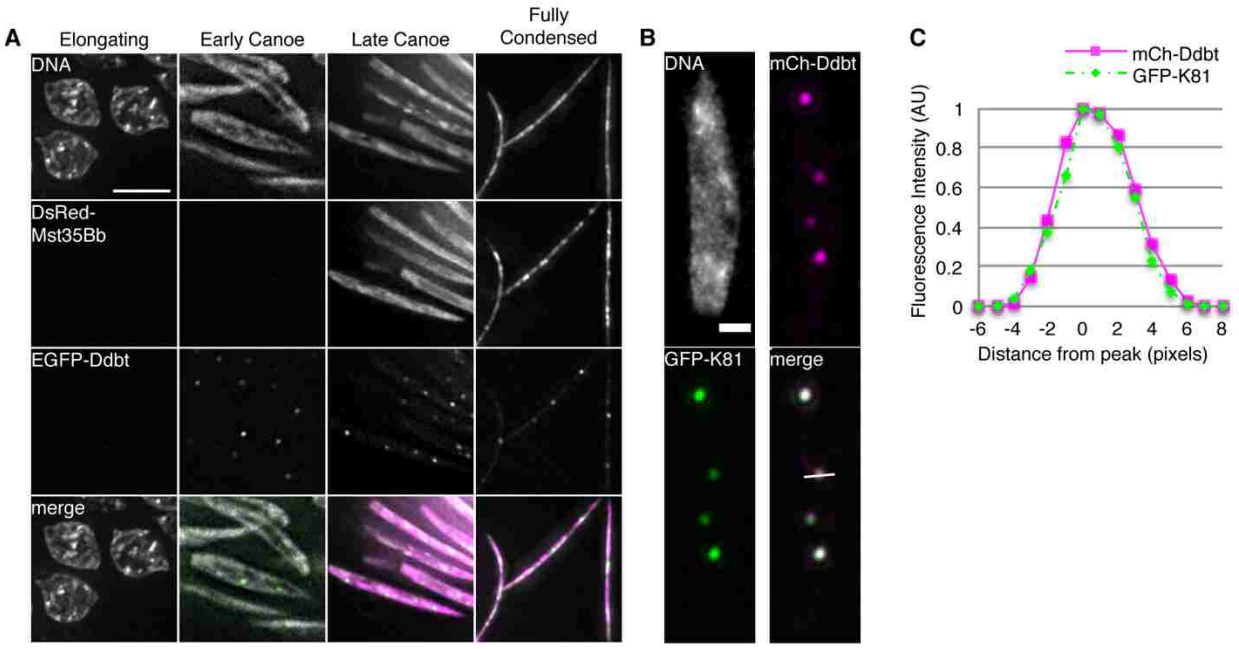


Figure 2.3 Ddbt Localizes to Telomeric Foci.
(A) Stages of spermiogenesis were identified by nuclear shape (DNA, DAPI staining) and presence of DsRed-Mst35Bb. EGFP-Ddbt foci were detected at early canoe and later stages. (B) Canoe stage nucleus with mCherry-Ddbt immunostaining (magenta) and GFP-K81 fluorescence (green). In the merged image, a line is drawn over the focus analyzed in (C). (C) Signal intensities were collected along the line, normalized to the brightest intensity value and plotted as relative values (AU, Arbitrary Units). Graph shows locations of peak intensity values and changes in pixel intensities. Scale bar, 5 μm except 1μm in (B).

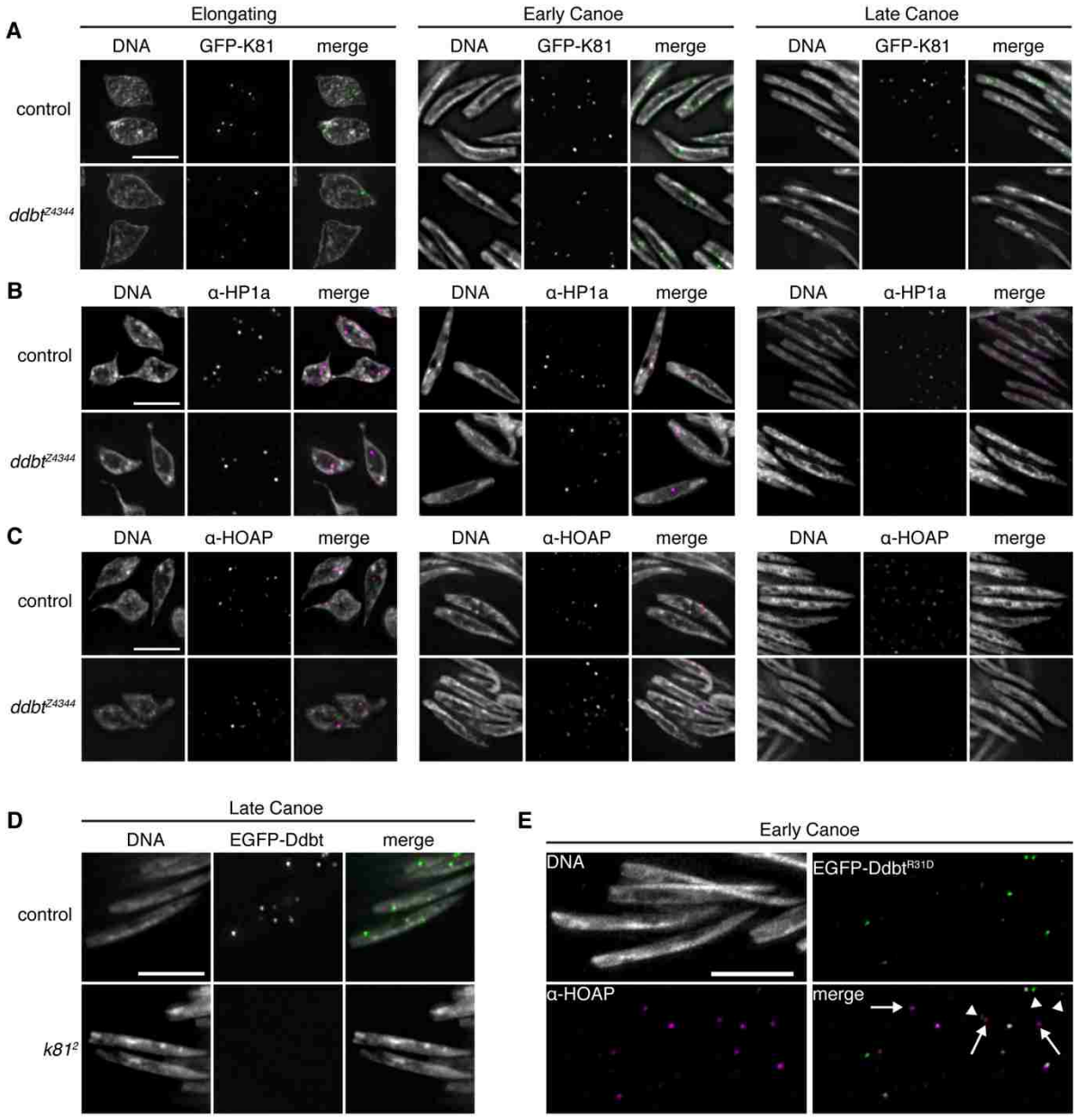


Figure 2.4 Ddbt Requires K81 for its Telomeric Recruitment and Maintains TCCs Through Spermiogenesis.

TCC components in control and *ddbt*^{Z4344} spermatids: (A) GFP-K81 as assayed by GFP fluorescence, and (B) HP1 and (C) HOAP, as assayed by immunostaining. (D) EGFP- Ddbt in control and *k81*^{2} spermatids. (E) Spermatids expressing EGFP-Ddbt^{R31D} assayed for EGFP fluorescence (green) and HOAP (magenta). The merged image shows foci containing both proteins (white), predominantly EGFP-Ddbt^{R31D} (arrowheads) or predominantly HOAP (arrow). Scale bar, 5μm.

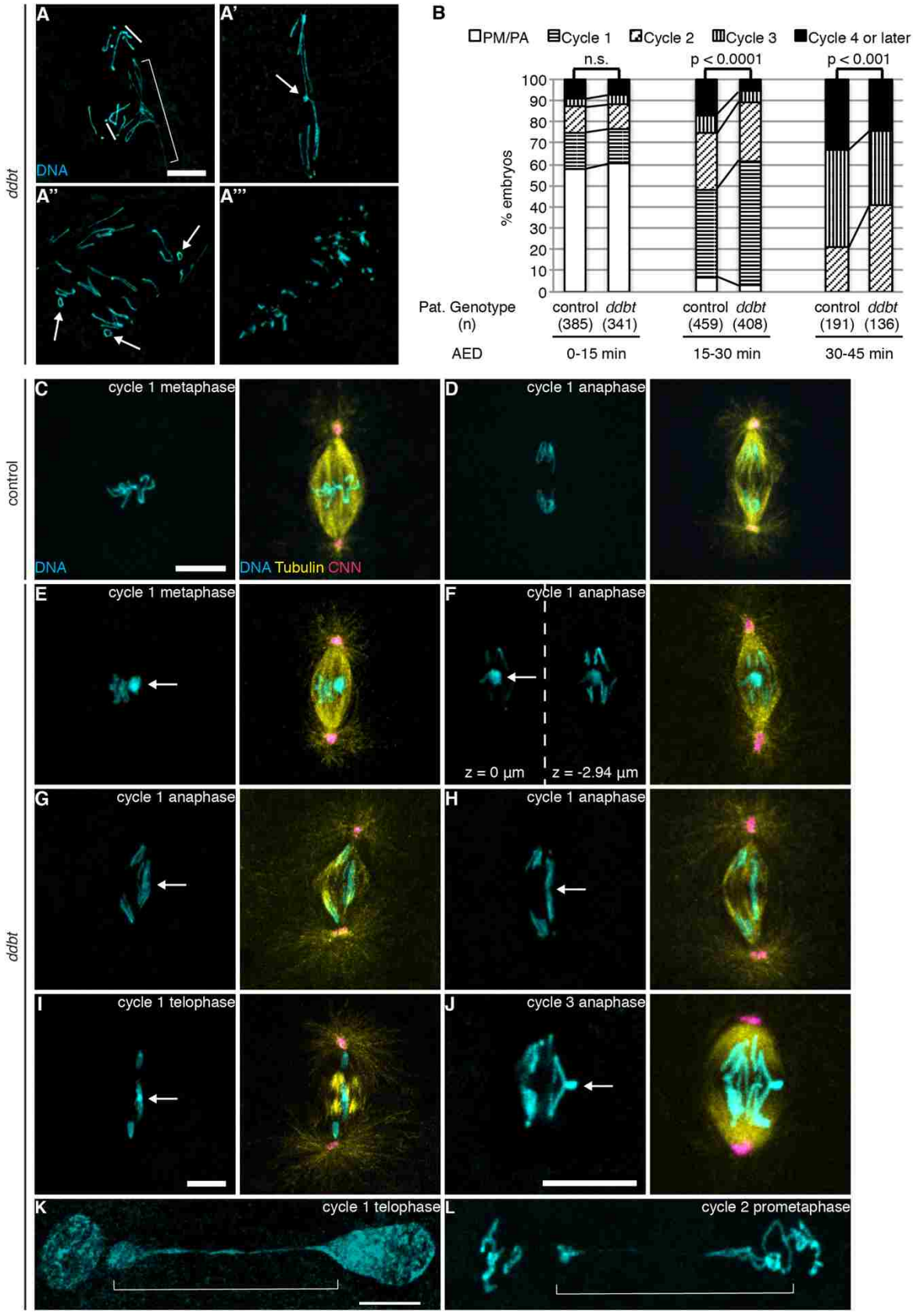

ddbt
A
A'
A''
A'''
DNA
B
PM/PA
Cycle 1
Cycle 2
Cycle 3
Cycle 4 or later
n.s.
p < 0.0001
p < 0.001
% embryos
100
90
80
70
60
50
40
30
20
10
0
Pat. Genotype
(n)
control ddbt
(385) (341)
control ddbt
(459) (408)
control ddbt
(191) (136)
AED
0-15 min
15-30 min
30-45 min
control
C
cycle 1 metaphase
DNA
DNA Tubulin CNN
D
cycle 1 anaphase
E
cycle 1 metaphase
F
cycle 1 anaphase
z = 0 µm
z = -2.94 µm
ddbt
G
cycle 1 anaphase
H
cycle 1 anaphase
I
cycle 1 telophase
J
cycle 3 anaphase
K
cycle 1 telophase
L
cycle 2 prometaphase

Figure 2.5 Uncapped Paternal Telomeres Disrupt Chromosome Behavior and Delay Progression of the Earliest Embryonic Cycles.
Confocal images of chromosomes in embryos of *T(2;3)lt*^{x13} mothers and *ddbt* fathers. (A) Maternal chromosomes, including *T(2:3)* (line), are advanced in anaphase segregation while entangled paternal chromosomes (bracket) lag. Karyotypes also include (A') telomere associations (TAs, arrow), (A'') ring chromosomes (arrows) and (A''') extensive fragmentation. (B) Comparison of stage distributions of control and *ddbt*'s embryos in the early time series. Embryo age is reported as minutes after egg deposition (AED). Embryo stage is classified as pronuclear migration /pronuclear apposition (PM/PA), or nuclear cycle. *p* values, χ^2 test. (C, D) Normal gonomeric spindles in control embryos. (E-L) Abnormalities (arrows or brackets) in *ddbt'*s embryos were: (E) hypercondensed paternal metaphase chromosomes that were spatially distinct from maternal chromosomes; (F-J) lagging anaphase chromosomes with TAs; and (K, L) chromatin bridge. DNA (cyan); tubulin (yellow); Centrosomin (magenta); z=focal plane. Scale bars,10μm.

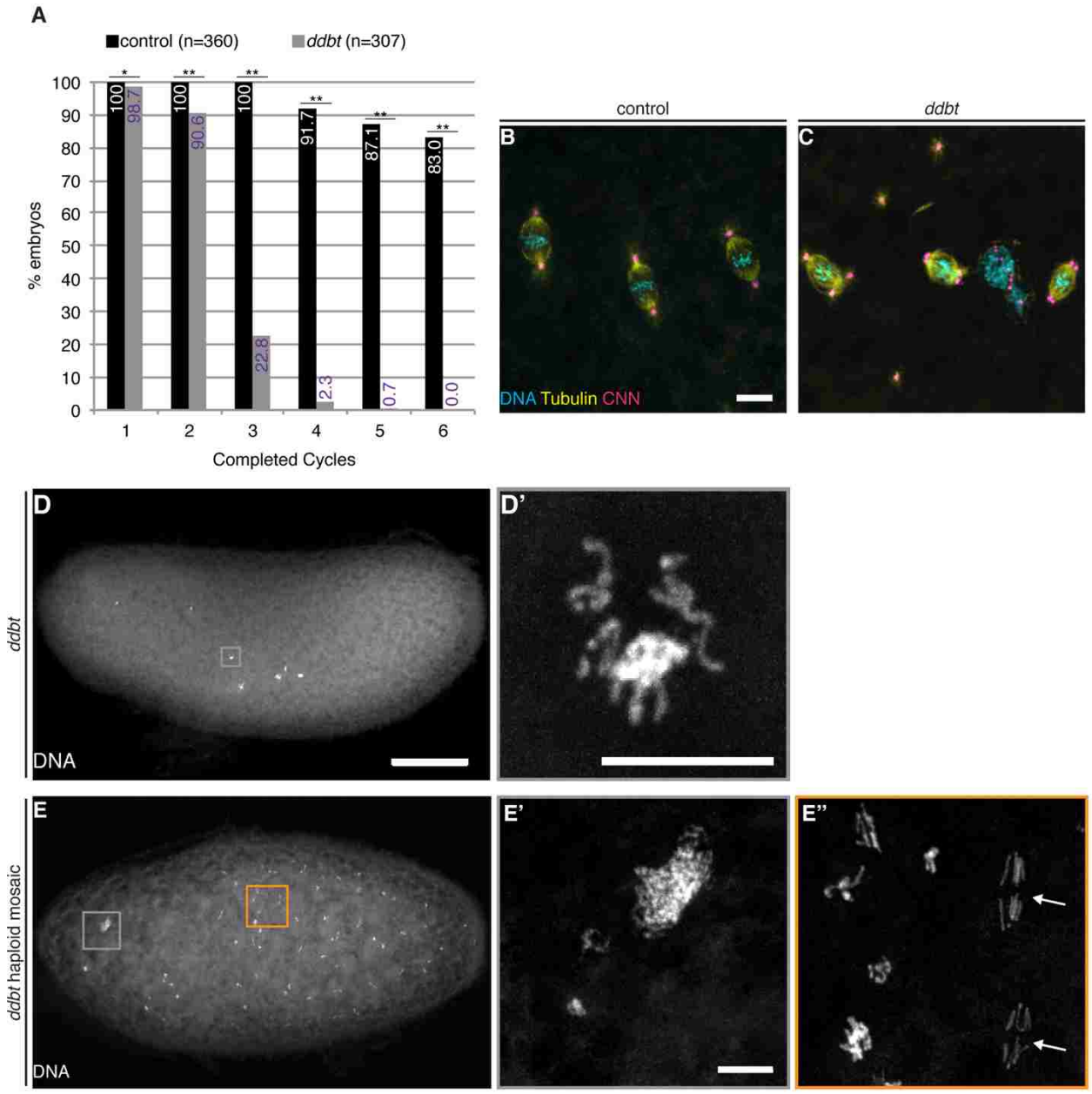


Figure 2.6 Uncapped Paternal Telomeres Cause Lethality Due to Cell Cycle Arrest or Production of Haploids or Haploid Mosaics.
(A) Comparison of cycle completion of control and *ddbt*'s embryos from analysis of the late time series. *$p < 0.05$, **$p < 0.001$, χ^2 test (B) Region from a control cycle 6 embryo showing synchronous metaphases. (C) Region from an embryo of a *ddbt* father showing asynchronous phases, excessive or free centrosomes, and chromatin hyperploidy. (D, E) Two representative embryos of *ddbt* fathers. (D) Embryo arrested at cycle 3 with a mix of prometaphase chromosomes and hypercondensed chromatin. (E) Haploid mosaic that progressed past cycle 7 with (E') hyperploidy and aneuploidy, and (E'') mitoses with different chromosome content, including haploid anaphases (arrows). Scale bars, 10μm except 100μm in D.

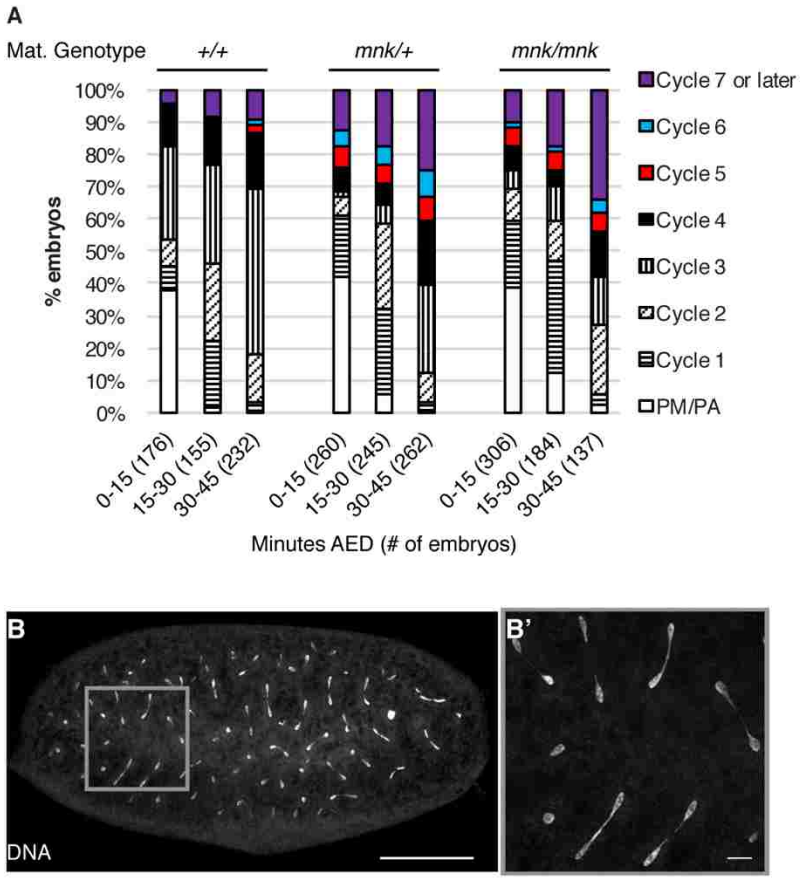


Figure 2.7 Reduced Levels of Maternal Chk2 Permit *ddbt*'s Embryos to Progress Past Cycle 3. (A) Comparison of stage distributions of embryos produced by *ddbt* fathers and control (w^{1118}) mothers, to those of mnk^{P6} heterozygous or homozygous mothers. (B) Cycle 7 embryo of a *ddbt* father and mnk^{P6} homozygous mother showing asynchronous phases with prevalent chromatin bridges visible at telophase. Scale bars, 100µm in B and 10µm in B'.

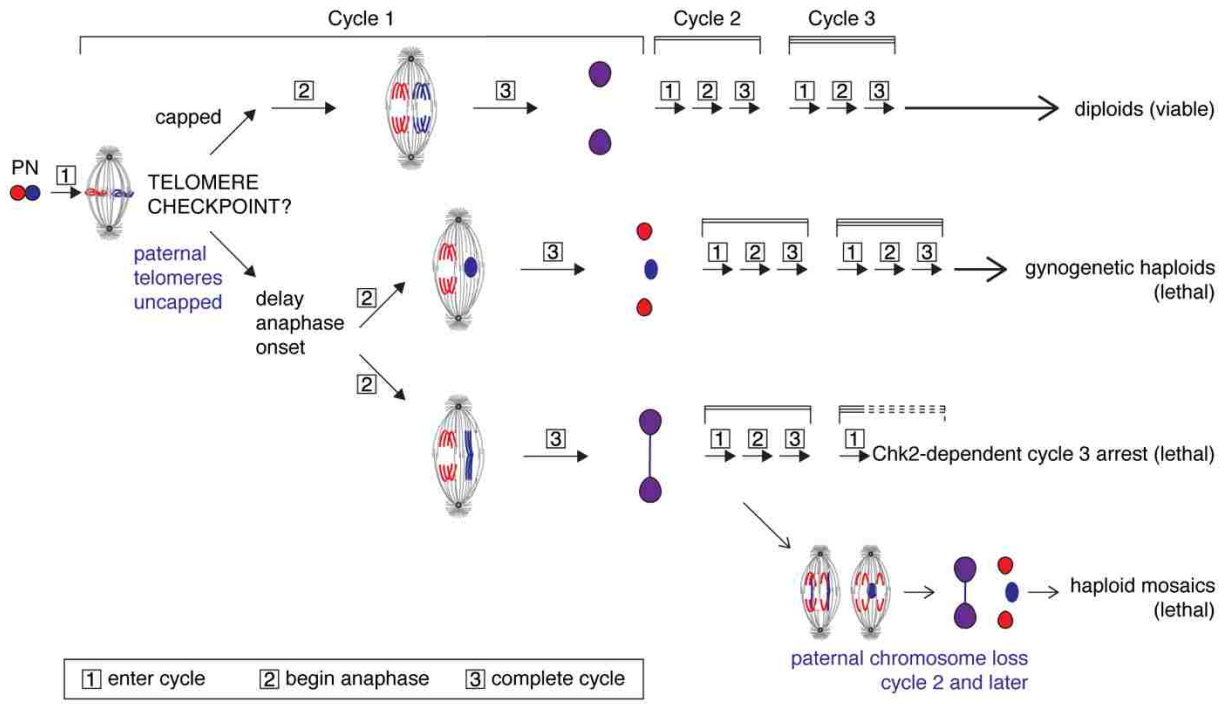


Figure 2.8 The *ddbt*-induced Paternal Effects on Early Embryogenesis.
This sequence begins with the apposition of female and male pronuclei (red and blue PN). Parental genomes enter cycle 1 and align as separate groups on the first spindle. Normally, telomeres of maternal and paternal chromosomes are capped (top sequence). Progression through cell cycles occurs in a timely fashion, with both sets of chromosomes synchronously entering each cycle, beginning anaphase, and completing each cycle (boxed numbers 1-3 respectively), ultimately yielding viable diploids. Embryos of *ddbt* fathers have uncapped paternal telomeres. Their defects may suggest the existence of cycle 1-3 checkpoints that monitor uncapped telomeres or new breaks generated after telomere fusions (middle and bottom sequences). When paternal chromosomes are uncapped, embryos delay onset of cycle 1 anaphase. In fewer than 10% of the embryos, the entire paternal genome fails to segregate in cycle 1 anaphase (middle sequence) but the maternal genome proceeds, resulting in gynogenetic haploids. In the majority of *ddbt*'s embryos (bottom sequences), paternal chromosomes enter anaphase and exhibit frequent telomere associations. Chromatin bridges are observed in telophase. These embryos complete cycles 1 and 2 and most will arrest in or shortly after cycle 3 in a Chk-2 dependent manner. However, in some embryos, one or more nuclei may lose paternal chromosome in cycle 2 or later. Loss of a sufficient number of paternal chromosomes allows these embryos to bypass cycle 3 arrest and develop as haploid mosaics. Although uncapped paternal chromosomes may be eliminated at different times, the result is always embryonic lethality.

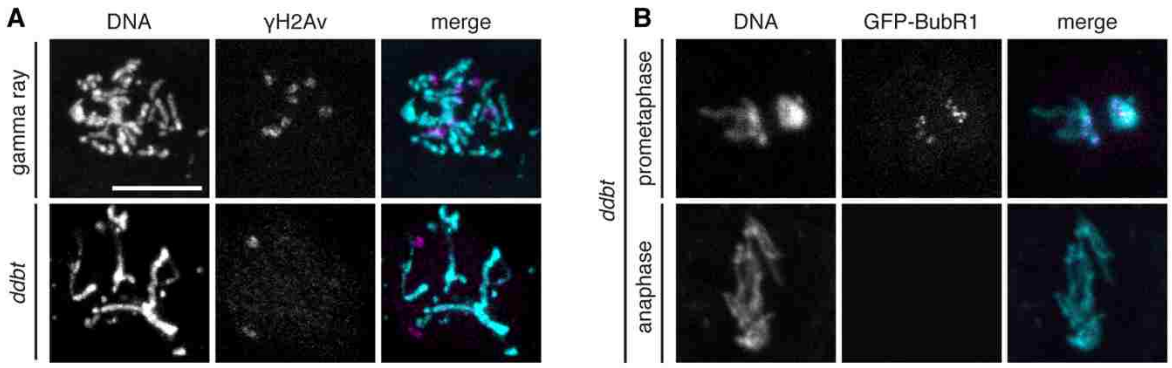


Figure S2.1 Phosphorylated H2Av and GFP-BubR1 in Early Embryos.
(A) Representative images showing detection of γH2Av foci (magenta) on extensively fragmented chromosomes of an embryo fathered by a heavily irradiated male (top) and an embryo fathered by a *ddbt* male (bottom). Both embryos were arrested in cycle 3 and were assayed between 45 and 90 min AED. (B) In these cycle 1 embryos produced by *ddbt* fathers and mothers expressing GFP-BubR1, GFP-BubR1 signals were detected at the kinetochore during prometaphase (top) but no signals were detected at either kinetochores or telomeres at anaphase (bottom). Scale bar, 10μm.

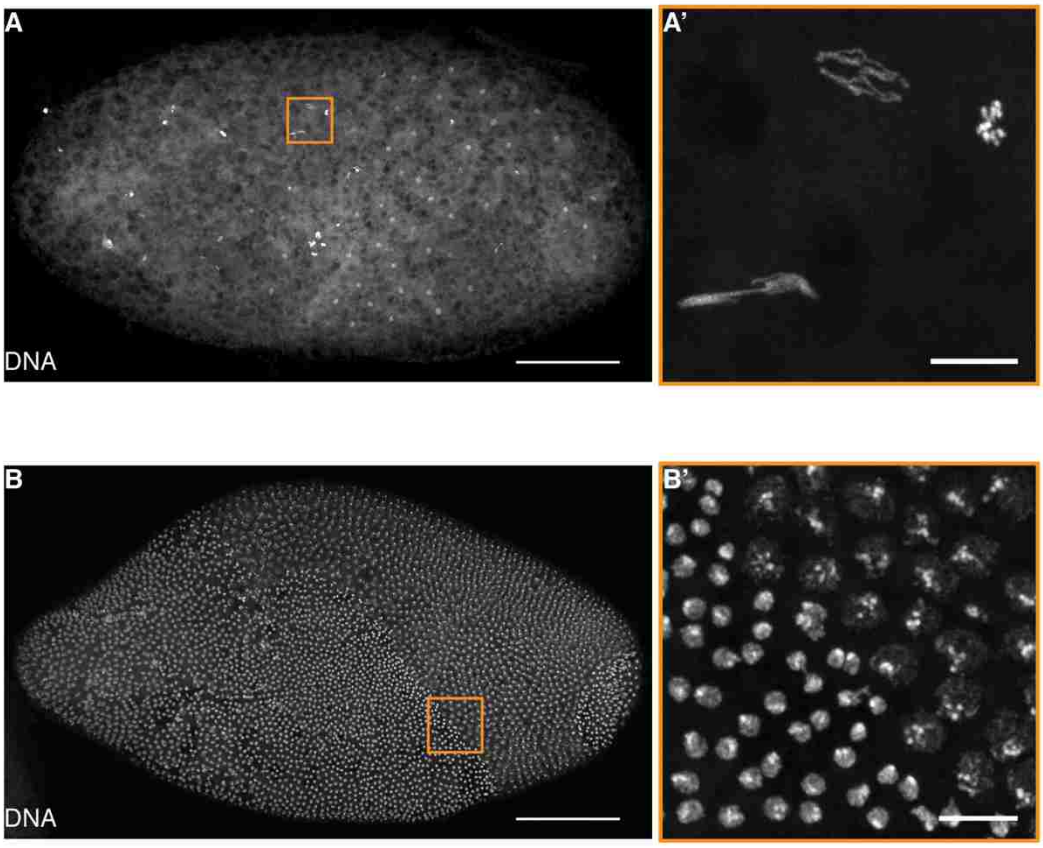


Figure S2.2 *k81*-induced Paternal Chromosome Loss Occurs After Earliest Embryonic Cycles and Creates Haploid Mosaics.
Representative haploid mosaics showing (A) chromatin bridges at cycle 7 and (B) large patches of different nuclear sizes and chromosome lagging and bridging in individual nuclei at syncytial blastoderm. Scale bars, 100µm in A and B, 10 µm in A' and B'.

Table 2.1 Transgenic rescue of male sterility

Male genotype	Relative fertility[a]
ddbt^{Z4344} *st*/*TM6*	1.00
ddbt^{Z4344} *st*/*Df(3L)7C*	0.00
w; ddbt^{Z4344} *e*	0.00
w; *P{w*$^{+mC}$ *ddbt*$^{+t3.8}$*}*; *ddbt*^{Z4344} *st*/*Df(3L)7C*	1.07
w; *P{w*$^{+mC}$ *egfp-ddbt}*; *ddbt*^{Z4344} *st*/*Df(3L)7C*	0.90
w; *P{w*$^{+mC}$ *mCherry-ddbt}*; *ddbt*^{Z4344} *e*	0.95
w; *P{w*$^{+mC}$ *egfp-ddbt*^{Mst35Ba}*}*; *ddbt*^{Z4344} *e*	0.00
w; *P{w*$^{+mC}$ *egfp-ddbt*^{Mst77F}*}*; *ddbt*^{Z4344} *e*	0.00
w; *P{w*$^{+mC}$ *egfp-ddbt*^{R31D}*}*; *ddbt*^{Z4344} *e*	0.00

[a] Based on average progeny yields from 10 single male crosses.

Table 2.2 d_N/d_S analysis of *ddbt* for positive selection[a]

Models compared	*-2ΔlnL*	d_N/d_S estimate	% sites[b] with estimated d_N/d_S	Positively selected sites[c]
M1 (neutral) vs M2 (selection)	8.32 (p<0.05, d.f. = 2)	4.01	14.9	*13M*, *78Q*, *90A*, **110K**, *114R*
M7 (neutral) vs M8 (selection)	8,88 (p<0.05, d.f. = 2)	3.90	15.8	**13M**, 19N, **78Q**, **90A**, **110K**, 113R, **114R**
M8a (neutral with $d_N/d_S = 1$) vs M8 (selection)	8.31 (p<0.05, d.f. = 1)	3.90	15.8	

[a] This analysis used *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. erecta*, and *D. yakuba* sequences (Table S1)
[b] Percentage of 109 codons aligned after removal of 8 gaps from species comparisons
[c] Predicted with $P > 90\%$ (bold), $P > 80\%$ (italics), or $P > 70\%$ (underlined)

Table 2.3 The *ddbt*-induced developmental delay relative to anaphase

	Minutes AED					
	0-15		15-30		30-45	
Paternal Genotype	control	ddbt	control	ddbt	control	ddbt
Number of Embryos	385	341	459	408	191	136
Percentage of Embryos						
Earlier than Cycle 1A	72.5	73.3	37.7	51.5	0	0
Cycle 1A or later	26.5	26.7	62.3	48.3	100	100
	ns		*p*=5.26e-5			
Earlier than Cycle 2A	85.2	87.1	70.6	86.8	6.8	16.9
Cycle 2A or later	14.8	12.9	29.4	13.2	93.2	83.1
	ns		*p*=6.24e-9		*p*=0.0065	
Earlier than Cycle 3A	90.7	90.9	81.5	92.2	58.1	69.1
Cycle 3A or later	9.3	9.1	18.5	7.8	41.9	30.9
	ns		*p*=3.77e-6		*p*=0.049	

p value, Fisher's Exact Test; ns not significant

Table 2.4 Time course of *ddbt*-paternally induced embryonic abnormalities

	Minutes AED					
	45-60	60-75	75-90	90-105	105-120	120-135
Number of Embryos	57	75	76	51	64	66
Early Arrest Category (%)	93.0	96.0	76.3	66.7	59.4	83.3
Median number of clumps (min, max)	6 (3,21)	6 (1,32)	6 (2,13)	6 (1,14)	6 (1,29)	5 (2,11)
With axial expansion of nuclei (%)	7.6	11.1	6.9	20.6	51.1	47.3
With asynchronous nuclei (%)	45.3	72.2	93.1* OR5.2 CL 1.7-16.2	58.8* OR 0.11 CL 0.03-0.4	53.3	10.9* OR 0.11 CL 0.04-0.3
With hypercondensed chromatin (%)	5.7	13.9	70.7* OR 15.0 CL 6.2-35.9	88.2	77.8	92.7
With aneuploidy (%)	30.1	38.9	74.1* OR 4.5 CL 2.2-9.6	64.7	51.1	30.9
With hyperploidy (%)	26.4	51.4	60.3	91.2* OR 6.8 CL 1.9-24.8	84.4	100
With fragmented chromosomes (%)	17.0	12.5	43.1* OR 5.3 CL 2.2-12.7	85.3* OR 7.7 CL 2.6-22.6	80.0	100
Haploid/Haploid Mosaic Category (%)	7.0	4.0	23.7	33.3	40.6	16.7

*significantly different from preceding time, *$p<0.05$, Tukey-Kramer method, OR Odds Ratio, CL 95% Confidence Limits

Table S2.1 Ddbt proteins in the Schizophora

Species	NCBI Gene ID or Genomic location	Amino acids (n)	% Arginine
D. melanogaster	5740226	117	24
D. simulans	6734413	109	24
D. sechellia	6610171	110[a]	24
D. erecta	6545471	117	23
D. yakuba	6532222	117	26
D. eugracilis	scaffold7180000409184:21609..21210	113[b]	21
D. biarmipes	scaffold7180000300910:19879..19473	117[b]	22
D. suzukii	scaffold3:11900223..11900629	117[b]	21
D. takahashii	Contig5689_22333..22740	117[b]	23
D. elegans	scaffold7180000491249:101628..101233	113[b]	20
D. rhopaloa	scaffold7180000777885:14618..14214	116[b]	24
D. ficusphila	scaffold7180000454105:135288..134885	117[b]	24
D. ananassae	6507379	150	20
D. bipectianta	scaffold7180000396390:32060..131545	152[b]	18
D. pseudoobscura	6899976	117	29
D. persimilis	6601101	117[a]	30
D. miranda	chrom XR:21661070..21660663	117[b]	30
D. virilis	6631187	113	23
D. mojavensis	6572658	108	14
D. busckii	scaffold4827:10905..10483	122[b]	21
D. grimshawi	scaffold15110:65546..65956	118[b]	16
Scaptodrosphila lebanonensis	scaffold23557:7171..7550	107[b]	23
Phortica variegata	scaffold12527:3144..3077	125[b]	19
Musca domestica	105261511	178	16
Bactrocera cucurbitae	105212239	181	18
Bactrocera dorsalis	105226460	181	20
Ceratitis capitata	101459550	182	20

[a] Modified from FlyBase reported annotation in this study

[b] Annotated in this study

Table S2.2 Time course of *k81*-paternally induced embryonic abnormalities

	Minutes AED					
	45-60	60-75	75-90	90-105	105-120	120-135
Number of Embryos	82	104	100	75	93	99
Early Arrest Category (%)	84.1	86.5	80.0	82.7	81.7	85.0
Median number of clumps (min, max)	6 (2,11)	6 (2,12)	6 (1,13)	6 (1,10)	6 (1,10)	6 (1,12)
Haploid/Haploid Mosaic Category (%)	15.8	13.5	20.0	17.3	18.2	15.0

Chapter 3: Spatiotemporally-distinct Male-Specific HMG Box Proteins Ensure Production of Functional Sperm in *Drosophila melanogaster*

Abstract

Sperm DNA is associated with histones and non-histone chromosomal proteins that are low molecular weight, enriched in basic amino acids, and collectively known as sperm nuclear basic proteins (SNBPs). Although the prevailing view of SNBP function is to drive the large-scale chromatin remodeling that transforms round spermatid nuclei into highly-condensed sperm nuclei, the function of only a few SNBPs is known. In *Drosophila melanogaster,* a subset of SNBP genes encode male-enriched proteins with a domain that is structurally related to the HMG box. Here we add four proteins to the MST-HMG box family increasing its membership to 12, and identify a characteristic NFLR motif in helix 1 of the MST-HMG domain. Interspecific comparisons document gain, loss, and expansion of family members among drosophilids. The temporal profile and spatial distribution of the newly identified *D. melanogaster* MST-HMG66A during spermatid nuclear transformation and fertilization distinguish it from other characterized family members. Elimination of MST-HMG66A alone did not affect male fertility, but with deletion of MST-HMG35Ba/b, a surprisingly late stage of sperm production was disrupted. From our observations, we suggest that at least three MST-HMGs function to provide condensed sperm heads with the structural integrity needed to withstand mechanical forces experienced during spermatid individualization and sperm tail coiling. Comparisons of the Drosophila MST-HMG class to the mammalian histone H1-derived SNBP class, which includes the true protamines, suggest that the demands of sperm formation

and function have driven structural and functional diversification of distinct classes of opportunistic DNA-binding proteins in different animal lineages.

Introduction

A highly-conserved feature of spermatogenesis is the striking reduction of nuclear volume that occurs during the transformation of round spermatids to mature sperm. The reduction has been reported to be 20-fold in the mouse (BALHORN 2007) and 200-fold in *Drosophila* (TATES 1971). The morphological transformation is accompanied by genome-wide chromatin remodeling in the absence of both DNA replication and widespread transcription (RATHKE *et al.* 2014).

Typically, spermatid chromatin remodeling involves the post-translational modification of histones by acetylation, ubiquitinylation, or phosphorylation, prior to eviction of most histones and their replacement by Transition Proteins (TPs). TPs are subsequently replaced by proteins that are specific to elongated and condensed spermatid nuclei. At the end of nuclear transformation, the mature sperm nucleus contains hundreds of different proteins, of which the most abundant are sperm nuclear basic proteins (SNBPs) (reviewed in (CASTILLO *et al.* 2014)). The term SNBP is inclusive of the retained histones and newly acquired non-histone chromosomal proteins found in mature sperm. Friedrich Miescher reported discovery of the first SNBP, which he called protamin, in salmon sperm in 1874, a landmark discovery because this was also the first description of a nuclear protein (MIESCHER 1874). We now know that SNBPs exist in vertebrate and invertebrate sperm with remarkable species variation in the types and relative abundance of different SNBPs. The consensus view of SNBP function is to organize chromatin folding, whether nucleosomal by retained histones or non-nucleosomal by non-histone

proteins (EIRIN-LOPEZ AND AUSIO 2009). Cytological studies have documented the dynamic temporal profiles of TPs and SNBPs during nuclear transformation and chromatin condensation (RATHKE *et al.* 2014; CHAMPROUX *et al.* 2016). While it has been attractive to view the acquisition of certain SNBPs as the driving force for sperm nuclear condensation, the relationship has remained largely correlative because only a few definitive functional tests have been reported.

The large-scale remodeling that occurs during spermatid differentiation has been most extensively studied in the mouse (review in (CHAMPROUX *et al.* 2016). Mouse spermatids evict over 90% of their histones, then temporarily acquire two functionally redundant transition proteins TP1 and TP2. The final step is TP replacement by Protamine 1 (PRM1) and Protamine 2 (PRM2) which are related by gene duplication. Decreasing the dosage of TPs or PRMs results in reduced or complete male sterility and defective spermatid nuclear condensation(YU *et al.* 2000; CHO *et al.* 2001; ZHAO *et al.* 2001), providing the strongest support for a role of these SNBPs in chromatin remodeling.

Phylogenetic analyses by Ausio and colleagues (EIRIN-LOPEZ AND AUSIO 2009) provide strong evidence for the evolutionary derivation of the mammalian TPs and PRMs from the linker histone H1. These investigators proposed a vertical model of diversification with changes occurring in a stepwise manner (#22). An H1 precursor with its characteristic winged-helix fold domain (WHD) is proposed to accumulate clusters of lysine residues in its N- and C- tails, resulting in a derivative called Protamine-like (PL). PL is viewed as an intermediate form because in some species it subsequently lost the WHD but retained the C terminus and became arginine rich. This smaller more basic derivative is referred to as Protamine (P) (EIRIN-LOPEZ AND AUSIO 2009). The H1 hypothesis for the origin of SNBPs has been a unifying evolutionary

perspective. It has been valuable for classifying SNBPs of diverse organisms into three categories: H1-like (H-type), PL-type, or P-type.

Drosophila was not included in the SNBP phylogenetic studies of Ausio *et al*. (EIRIN-LOPEZ AND AUSIO 2009). However, when the first SNBPs were molecularly identified among the products of genes expressing male specific transcripts (msts) in *Drosophila melanogaster*, the MST77F protein was described as H1-like (RUSSELL AND KAISER 1993). Next, MST35Ba and MST35Bb proteins, products of a gene duplication, were described as protamine-like and named dProtamine A and dProtamine B, nomenclature that has persisted to this day (JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005).

An important discovery was made by Rathke *et al.* (RATHKE *et al.* 2007) when they identified a *Drosophila* protein named Transition Protein Like 94D (TPL94D) whose temporal profile was reminiscent of mammalian TPs. They noted that it contained a domain characteristic of High Mobility Group Box proteins. HMG box proteins are eukaryote-specific DNA-binding proteins that contain three alpha-helices predicted to fold in an L-shaped structure (reviewed in (MALARKEY AND CHURCHILL 2012). Some of these proteins such as the vertebrate SRY bind DNA in a sequence-specific manner and act as transcription factors (HARLEY *et al.* 2003), while others such as HMGB1 bind in a sequence-independent manner and are proposed to act as architectural proteins facilitating multiple chromatin-dependent processes (WU *et al.* 2018). The HMG box was subsequently identified in dProtamine A and dProtamine B by Dorus and Karr (DORUS *et al.* 2008) and in MST77F by Rathke (RATHKE *et al.* 2010). Pivotal studies by Garner *et al*. (GARTNER *et al.* 2015) and Doyen *et al.* (DOYEN *et al.* 2015) showed that the number of male-enriched HMG-box containing proteins also included four previously uncharacterized proteins. Doyen *et al.* (DOYEN *et al.* 2015) further showed that the HMG box domain of

Mst35Ba/b, Mst77F, and Tpl94D is typically 50 amino acids, smaller than the 75-amino acid canonical HMG box domain, but it retains three alpha-helices. Their analysis is consistent with the proposal that an HMG precursor protein gained male germ line expression and lost a portion of the HMG box.

Overall, these previous studies noted that the Drosophila SNBPs known to date had an evolutionary origin distinct from H1-type SNBPs and the true protamines. We propose that these important findings justify a new nomenclature with proteins officially designated as MST-HMG box family members. We retain the original conventions that recognize the male specific transcription (MST) and distinctions among members using the gene's cytogenetic localization, or originally identified mutant phenotype. Using this convention, the dProtamineA and dProtamineB proteins are hereafter referred to as MST-HMG35Ba and MST-HMG35Bb and their genes as *Mst-Hmg35Ba* and *Mst-Hmg35Ba*. Protamine-like 99C is renamed MST-HMG99C for the protein and *Mst-Hmg99C* for the gene. Thus far, functional studies have been reported for four *Mst-Hmg* genes and only *Mst-Hmg77F* has a strong effect on fertility when mutant or deleted in males (JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005; KIMURA AND LOPPIN 2016).

Here, we expand the *Drosophila* Mst-Hmg gene family by identifying four additional members. We previously identified one gene, *ddbt* (hereafter renamed *Mst-Hmg ddbt*), based on its mutant phenotype, a paternal effect on chromosome maintenance in the early embryo (YAMAKI *et al.* 2016). Three additional members were identified based on sequence similarity. Our characterization of the MST-HMG66A protein reveals a dynamic localization behavior that differs from those known for other MST-HMG box proteins. We further show that MST-HMG66A and MST-HMG35B function together to ensure that mature sperm are released from

the testis. Our results reveal that SNBPs can function surprisingly late in spermiogenesis to support the interaction of sperm nuclei with the somatic head cyst cell. Overall, the evolutionary and functional characterizations of MST-HMG box protein family members provide new perspectives on the acquisition of protein functions necessary to meet rapidly evolving demands of *Drosophila* sperm formation and function.

Materials and Methods

Identification and comparative analyses of putative MST-HMG genes and proteins

We used *Drosophila melanogaster* modENCODE RNA-seq data available from FlyBase (http://flybase.org) to identify transcripts reported to be expressed in adult males but not females, and in the testis. From this list of 1755 identified genes, we sought those similar in structure to *Mst-Hmg ddbt* so we defined the subset encoding predicted proteins 250 or fewer amino acids in length, enriched in K, R, or H residues, and containing the NFLR motif. This identified the predicted CG14835 protein which we named MST-HMG66A. Two additional proteins, MST-HMG49B and MST-HMG61C were identified in a BLASTp search of the *D. melanogaster* genome using amino acids 84-127 from MST-HMG35Ba. For the interspecific comparisons, sequences were downloaded from FlyBase or National Center for Biotechnology Information (http://ncbi.nlm.nih.gov/nucleotide). BLASTp was used to search for predicted MST-HMG orthologous proteins (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Clustal Omega was used for amino acid sequence alignments (SIEVERS *et al.* 2011). MST-HMG66A secondary structure was predicted using PSIPRED v3.3 (BUCHAN *et al.* 2013). Tertiary structure was predicted using Phyre2 (KELLEY *et al.* 2015) and I-TASSER (ROY *et al.* 2010) both programs yielded essentially the same results.

We used PAML4 (YANG 2007) for $\underline{d}_N/d_S$ analysis of *Mst-Hmg35Ba*, *Mst-Hmg66A*, *Mst-Hmg77F*, and *Mst-Hmg99C*. Each test used sequences from *D. melanogaster*, *D. simulans*, and *D. sechellia*. The number of codons aligned and used to calculate $\underline{d}_N/d_S$ using the CODEML program were: 145 (*Mst-Hmg35Ba*), 142 (*Mst-Hmg66A*), 208 (*Mst-Hmg77F*), and 200 (*Mst-Hmg99C*). Three different tests were performed for neutral vs. positive selection: M1 vs. M2, M7 vs. M8, and M8a vs. M8. Different initial values of $\underline{d}_N/d_S$ were used to ensure convergence, $-2\Delta lnL$ values were calculated, and p values were determined by a chi-square test. Bayes empirical Bayes was used to determine which residues are likely under positive selection.

Drosophila melanogaster strains

Strains were obtained from the Bloomington *Drosophila* Stock Center (BDSC) except strains carrying deficiencies for several genes including *Mst-Hmg35Ba* and *Mst-Hmg35Bb* in 35B6-7 were kindly provided by C. Rathke and R. Renkawitz-Pohl. The deficiencies were previously denoted *protΔ1* and *protΔ2* (RATHKE *et al.* 2010) but we denoted them here as *Df(2L)Rathke1* and *Df(2L)Rathke2*. We obtained the BDSC w^{1118} stock, which is the parental strain of w^{1118}*; Mi{ET1}CG14835*^{MB10076} (BELLEN *et al.* 2011), and found that it had only 50% of the fertility compared to a wildtype strains even when males were crossed to wildtype females. To remove potential background effects, we outcrossed it and a stock containing the truncation allele derived from w^{1118}*; Mi{ET1}CG14835*^{MB10076} to a healthy *y w* strain. Even after four outcross generations to replace all but the chromosome 3 of interest, fertility of the outcrossed control strain was still lower than desired (70%). We therefore used egg hatch rates of progeny as a more reliable comparison of male fertility of control and experimental genotypes (Table 3.2).

We made *egfp-Mst-Hmg66A* and *mCherry-Mst-Hmg66A* fusion genes starting with a PCR amplified genomic region from the y^1; cn^1 bw^1 sp^1 strain. The region contained 1.3 kb upstream of the predicted *Mst-Hmg66A* transcription start site and 600 bp downstream of the poly-adenylation site. The *egfp* or *mCherr*y coding sequences along with a linker sequence encoding YSDLEL were obtained from Clontech vectors and inserted into beginning of the *Mst-Hmg66A* open reading frame. Transgenic lines were created by PhiC31-mediated integration into the *attP40* genomic site as described by Groth *et al.* (2004).

Functional tests and protein localization assays

As a measure of male fertility, we determined the hatch rate of eggs laid by females of the wildtype *Sevelin* strain mated to males with control and experimental genotypes. Males were collected within a day of eclosion (1 d PE) and separated from females for 3 days to allow sperm accumulation and storage before mated to virgin females that were 3-4 d PE. At minimum of 300 laid eggs were collected for each test and the number of hatched larvae was counted after 24 hours.

We tested whether decreased dosage of *Mst-Hmg66A* in males caused paternal chromosome loss in embryos using three different crosses. We set up 35 crosses of single males that had two or zero doses of *Mst-Hmg66A* to 3 females with the following genotypes: *C(2)EN*, *C(3)EN,* or *y w sn; C(4)EN*, *ci* ey^R. The presence of viable progeny from females with the *C(2)EN* or *C(3)EN* compound chromosome was scored as an indicator of loss of the paternal chromosome 2 or 3, respectively since three copies of either chromosome is lethal. From the cross with *C(4)EN ci ey* females, we scored progeny for mosaicism for *ci* or *ey* which results from post-fertilization paternal chromosome 4 loss.

To assay stages of spermatogenesis and fertilization, we prepared reproductive tissues and newly fertilized eggs for microscopy as described by Wilson *et al.* (2006) and Yamaki *et al*. (2016) but with caution to avoid tissue flattening. For spermiogenesis, we used spermatid nuclear morphology as assessed by DNA staining with 4', 6-Diamidino-2-Phenylindole, Dihydrochloride (DAPI) to determine developmental stage. Here we introduce a modified nomenclature of spermiogenic stages to allow easy identification based simply on three-dimensional nuclear shape (see Figure 3.7), rather than by relative timing (e.g., as early vs late canoe or elongating vs fully elongated nuclei, terminology used previously). For assays of spermatid or sperm nuclei in whole mount testes, seminal vesicles and seminal receptacles pretreatment with 0.1% Triton X-100 in phosphate buffer for 20 minutes was required to allow even penetration of DAPI.

For protein localization, we used the following primary antibodies: rabbit anti-acetylated histone H4 (Thermo 1:200), rabbit anti-histone H3 (abcam, 1:500), and the secondary goat anti-rabbit DyLight650 antibody (Thermo, 1:200). For F-actin, we used 100nM Alexa594 conjugated phalloidin (ROGAT AND MILLER 2002). Preparations were mounted in Vectashield (Vector Laboratories, Burlingame, CA).

Preparations were viewed using a DeltaVision Elite imaging system with X 100X/ N.A. 1.4 oil immersion objective and 0.2 um sections or a Leica TCS SP5 II confocal microscope with X 40/ N.A. 1.25 or X63/N.A. 1.4 oil immersion objective and 0.42-um optical sections. Images were processed and relative signal intensities measured using ImageJ (SCHINDELIN *et al.* 2012).

Results

Identification of Drosophila genes encoding proteins with the NFLR motif expands the Mst-Hmg box gene family

We previously reported that the *Drosophila melanogaster* Deadbeat (Ddbt) protein is a telomere-enriched SNBP. Spermatids lacking Ddbt lose telomere capping complexes but develop into mature sperm that are fertilization competent. However, the lack of telomere capping complexes results in failure of paternal chromosome maintenance in the early embryo. We identified an 11-amino acid sequence, which we abbreviated the NFLR motif, that is critical for Ddbt function and conserved over 65 million years of dipteran evolution. We also identified the NFLR motif in MST-HMG77F, MST-HMG35Ba and MST-HMG35Bb (YAMAKI *et al.* 2016).

The Ddbt study motivated us to search the publicly available modENCODE RNA-seq dataset for male-enriched transcripts encoding additional NFLR motif-containing proteins. This search yielded CG14835. Neither Ddbt nor CG14835 was identified as MST-HMG box proteins in other studies (DOYEN *et al.* 2015; GARTNER *et al.* 2015), but when we relaxed search parameters, we found that Ddbt, and the predicted CG14835, CG30056, and CG34269 proteins share a 50-amino acid region with sequence similarity to the family's founding members (Figure 3.1). This region has the potential to form three alpha helices. Helix 1, which contains the NFLR motif, and helix 2 are conserved in length. The intervening loop is variable in length and sequence. Consistent with Doyen's report (DOYEN *et al.* 2015), there are four aromatic amino acids whose positions are conserved. Helix 3 is characteristically followed by clusters of basic residues. These shared features provided evidence that Ddbt, CG14835, CG30056, and CG34269 are *bona fide* members of the MST-HMG box protein family and expanded membership to 12.

To recognize the structural relationships, we introduce the consistent nomenclature for *Mst-Hmg* genes and MST-HMG proteins presented in Figures 3.1, A and B.

Doyen *et al.* showed that *Drosophila* species vary in the number of *Mst-Hmg* box genes (DOYEN *et al.* 2015). We extended their analysis to include new family members. As summarized in Figure 3.1B, *Mst-Hmg* box genes present in *Drosophila melanogaster* are detected as single copy genes in most of the species analyzed. The founding member of the gene family may have been the ancestral *Mst-Hmg35Ba,* as single orthologs are found in all examined Drosophila species and other insects, including the distantly related locust, *Locusta migratoria*. If correct, this suggests the *Mst-Hmg* gene is at least 380 million years old (MISOF *et al.* 2014). This is considerably earlier than the estimated birth of *Mst-Hmg ddbt* 65 million years ago (YAMAKI *et al.* 2016) and the origin of genes, such as *Mst-Hmg66A*, that are presently found only in *Drosophila* species. There are examples of recent *Mst-Hmg66A* loss. Syntenic regions of *D. kikkawaii* and *D. willistoni* show no evidence of *Mst-Hmg66A* sequences, but closely related species have the gene (Table S3.1). Expansion of *Mst-Hmg* genes has occurred repeatedly and with dispersion to different regions of the genome. For example, two to four *Mst-Hmg66A* orthologues were found in the *obscura* group. Remarkably, *D. melanogaster* has acquired 18 Y-linked copies of *Mst-Hmg77F* (KRSTICEVIC *et al.* 2015) and two Y-linked copies of *Mst-Hmg35a* (GRAMATES *et al.* 2017).

Taken together, the observations are consistent with the evolutionary origin of the *Mst-Hmg* box genes summarized in Figure 3.1C. A precursor gene that encoded a canonical HMG box protein gained male specific expression (DOYEN *et al.* 2015). Modifications resulted in the NFRL-containing smaller MST-HMG box. This sequence is consistent with apparent absence of sequences encoding the NFRL motif in the 24 existing Drosophila genes encoding canonical

HMG box proteins. Expansion of the NFLR containing *Mst-Hmg* gene gave rise to new family members and this was followed by sequence divergence. Subsequently, individual family members experienced copy number increases or gene loss in specific *Drosophila* lineages.

Evolutionary analysis indicates bipartite structure of MST-HMG66A

We prioritized MST-HMG66A as a newly discovered member of the MST-HMG Box family to gain insight into gene and protein evolution. From a detailed comparison in 24 species (Table S3.1), we found that the gene lacks introns in *D. melanogaster* and closely related species but contains a single intron in all other species examined (Figure 3.2A). This distribution is consistent with intron loss in the lineage that gave rise to the melanogaster group. In the intron-containing genes, Exon 1 encodes the N terminus and nearly all of the MST-HMG box, from its helix 1 and including all but the last few amino acids of helix 3.

Alignments show that MST-HMG66A orthologous proteins have an N-terminal region that is often short, a conserved MST-HMG box that is followed by considerable variation in sequence and length of the C terminal region (Figure 3.2B). Pairwise comparisons between species were especially informative and revealed a bipartite protein structure related to the exon-intron structure of the intron-containing genes. As shown in the comparisons between *D. melanogaster* and *D. virilis*, species separated by an estimated 50 million years, a clear demarcation between conserved MST-HMG box and the immediately adjacent divergent region is located at the junction of *D. virilis* Exons 1 and 2. Comparison between *D. pseudoobscura* and *D. virilis*, two species with intron-containing genes, showed a similar demarcation (Figure 3.2C). In this comparison, Exon 2 encoded regions could not be aligned, suggesting that these *D. pseudoobscura* and *D. virilis* regions are not orthologous. This variation might have resulted

from differences in splice acceptor site choice in the different lineages. Despite diversity in sequence and length of the C terminal region, it contains alpha-helical or unordered secondary structure and is enriched in basic amino acids. Provided the MST-HMG66A orthologous proteins have similar functions in the different species, these general features of the C terminus may be important for DNA interactions and it is apparent that different lengths of the MST-HMG box helix 3 can be tolerated.

The structure of HMG-MST66A resembles that of the MST-HMG Ddbt protein in having a conserved MST-HMG box and variable N and C terminal regions (YAMAKI *et al.* 2016). In MST-HMG Ddbt, some of the rapidly changing residues in the C terminal region display signatures of positively selection. We asked whether regions of Mst-HMG66A and other family members were also under positive selection. For this, we used three *melanogaster* species (*D. melanogaster*, *D. simulans*, and *D. sechellia*) to calculate the ratio of nonsynonymous to synonymous substitutions (*dN*/*dS*). *Mst-Hmg66A* showed statistically significant signatures of positive selection ($p = 0.01$), with 5.8% of codons having a *dN*/*dS* of 16.8 (Figure 3.2D). The comparison of *Mst-Hmg35Ba* orthologs did not show signatures of positive selection whereas *Mst-Hmg77F* and *Mst-Hmg99C* yielded signatures of positive selection (Table 3.1). Overall, all but one of the sites predicted to be under positive selection were found to be outside the MST-HMG box. The results suggest that *Mst-Hmg* box protein members experience different selective pressures, but rapidly evolving residues are important and have been driven by functional advantages instead of merely relaxed selection.

The absence of Mst-Hmg66A has no detectable impact on male fertility or embryogenesis

To assess the function of the *Mst-Hmg66A* gene, we generated a mutant allele by excising a transposon insertion within the gene. Starting with the strain w^{1118}; $Mi\{ET1\}CG14835^{MB10076}$ (BELLEN *et al.* 2011), we generated several excision alleles and showed by sequence analysis that one imprecise excision event created a premature stop codon after amino acid 41 in the MST-HMG box helix 2 (Figure 3.2C). This mutation, designated $Mst\text{-}Hmg66A^{V42X}$, was used in functional studies.

We first tested whether males homozygous for $Mst\text{-}Hmg66A^{V42X}$ showed abnormal sperm formation or function. Cytological analysis of testis squashes showed no discernable defects (data not shown). We also compared nuclear morphology by DAPI staining of mature sperm produced by five control males (n = 1845 sperm) and five $Mst\text{-}Hmg66A^{V42X}$ males (n= 3672 sperm) and did not detect abnormalities (data not shown). To ask if sperm produced by $Mst\text{-}Hmg66A^{V42X}$ males were fertilization competent, we compared hatch rates of eggs produced by females mated to these males to those mated to control males. As shown in Table 3.2, only 70% of the eggs hatched from the control even after outcrossing (see Materials and Methods) and the experimental cross yielded a similar hatch rate. We conclude that absence of the *Mst-Hmg66A* had no discernable effect on male fertility or sperm morphology or function.

We previously showed that MST-HMG Ddbt is required after sperm entry into the egg to ensure paternal chromosome maintenance in the early embryo. We asked whether MST-HMG66A has a similar role using three genetic assays. When females carrying a compound chromosome, either *C(2)EN* or *C(3)EN*, are mated to control males, no offspring are produced due to triploidy for a major autosome. Survivors are produced in the rare instance of compound chromosome detachment or loss of the paternally contributed autosome in the early embryo .

Crosses of females bearing either *C(2)EN* or *C(3)EN* to 35 single control males or 35 homozygous *Mst-Hmg66A*^{V42X} males, yielded no surviving progeny. We also monitored the loss of chromosome 4. For this test, female carrying *C(4), ci ey*. The recessive *ci* and *ey* markers are detected in offspring only with the loss of the *ci*$^{+}$ *ey*$^{+}$ paternal chromosome 4, either in the entire body or in patches in mosaics. We did not observe such offspring among the over 2000 offspring produced from *C(4), ci ey* females crossed to control or homozygous *Mst-Hmg66A*^{V42X} males (20 single male crosses each). These genetic assays indicate that lack of *Mst-Hmg66A* in sperm has no detectable consequence for paternal chromosome inheritance or stability in early embryos.

MST-HMG66A is an SNBP with a temporal profile that differs from that of known spermatid chromatin proteins

To investigate whether MST-HMG66A is expressed during spermiogenesis and retained in mature sperm, we constructed a transgene that includes the gene's promoter and expresses a transcript with *enhanced green fluorescent protein* (*egfp*) coding sequence placed at the 5' end of the *Mst-Hmg66A* open reading frame. As demonstrated in a later section and in Table 3.2, this transgene rescues a genetic interaction induced by *Mst-Hmg66A*^{V42X}, allowing us to use it as a proxy to monitor MST-HMG66A location.

Stages of *Drosophila* spermatogenesis can be distinguished by nuclear morphology as monitored by DAPI staining and the nuclear localization of specific nuclear proteins. For localization studies, we examined stages from a total of 15 cysts from three males. Pre-meiotic cysts contain 16 spermatocytes and post-meiotic cysts contain 64 spermatids. EGFP- MST-HMG66A was first detected in post-meiotic stages. To compare its profile relative to the dynamic behavior of histones and a spermatid transition protein, we characterized expression

during stages of nuclear transformation. We call these stages: sphere, spheroid, droplet, canoe, rocket, and needle stages to reflect the three-dimensional nuclear shape changes (See Materials and Methods, Figure 3.7). EGFP-MST-HMG66A was first detected at the canoe stage when histones, including acetylated H4, remained easily detected (Figure 3.3A) and the transition protein Vrs tagged with GFP first appeared (BINDER *et al.* 2017). As previously reported, histone and Vrs-GFP signals decreased during later stages (Figure 3.3, A and B). In contrast, the intensity of the EGFP-MST-HMG66A signal increased in the rocket stage. As expected histones and Vrs-GFP were not detected in the needle stage. However, EGFP-MST-HMG66A persisted as the spermatid chromatin became highly condensed and is consistently retained as seen in mature sperm in the seminal vesicles (n =100 sperm). These observations verify that MST-HMG66A is *bona fide* SNBP.

To compare the temporal profile of MST-HMG66A to that of MST-HMG35Bb, we localized EGFP and DsRed tagged transgenic versions respectively. Both transgenes expressed the fusion proteins from their endogenous promoters. As noted above, EGFP-MST-HMG66A was first detected in canoe stage spermatids, earlier than MST-HMG35Bb-DsRed which appears at the rocket stage (Figure 3.3C). MST-HMG35Bb-DsRed is consistently easier to detect than EGFP-MST-HMG66A in mature sperm. We asked how long the proteins persisted as sperm nuclei decondense in the egg cytoplasm during fertilization. Surprisingly, MST-HMG35Bb-DsRed signal dissipated earlier than that the EGFP-MST-HMG66A signal (n=10 sperm). These results with the tagged proteins suggest differential regulation of these two SNBPs in their nuclear association during spermiogenesis and their eviction during fertilization.

Dynamic nuclear pore complex localization reveal spatiotemporally distinct behaviors of histones and MST-HMG box proteins

We and others have described distinct temporal profiles of different *Drosophila* spermatid nuclear proteins. In general, histones are carried over from earlier meiotic stages and largely eliminated post-meiotically (JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005); transition proteins enter and exit post-meiotically (RATHKE *et al.* 2007; BINDER *et al.* 2017); and MST-HMGs enter post-meiotically and are retained in mature sperm (JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005; EREN-GHIANI *et al.* 2015; YAMAKI *et al.* 2016). As documented in ultrastructural studies of spermiogenesis (TOKUYASU 1974), another interesting post-meiotic change is the wholescale redistribution then elimination of nuclear pore complexes (NPCs). We used GFP- and RFP-tagged Nup107, a highly-conserved ring component of the NPC (BECK AND HURT 2017), to ask if NPC changes correlated temporally or spatially with nuclear entry and exit of selected histone and MST-HMG proteins.

Consistent with the TEM studies which detected NPCs as interruptions or fenestrations in the nuclear envelop, we detected tagged Nup107 as asymmetrically distributed around spheroid stage spermatid nuclei. Nup107 then localizes along one of the long nuclear axes in droplet stage (Figure S3.2). The Nup107-enriched region becomes the concavity of canoe stage nuclei. The concavity flattens, and at the rocket stage, Nup107 is eliminated at the base of now fully elongated nuclei. This timing is consistent with elimination of the excess nuclear envelop and nucleoplasm (TOKUYASU 1974). Highly condensed needle stage nuclei lacked NPCs (TOKUYASU 1974) and Nup107 signal.

We observe diminution of the histone H4 acetylation signal between the canoe and rocket stages. This change may be due to deacetylation, degradation of this modified histone, or

eviction from the chromatin and exit through the NPC. Interestingly, immunolocalization signal for histone H3 is detectable in the rocket stage. This persistence after the bulk elimination of the NPCs indicates that H3 does not exit through NPCs; rather loss of signal may be due to destruction of the H3 antigenicity or H3 degradation within the nucleus during condensation.

Localization studies of Nup107, DNA and MST-HMG35Bb-DsRed and EGFP-MST-HMG66A provided a higher resolution view of the differences between the two SNBPs. In rocket and needle stage nuclei, DAPI-stained DNA and MST-HMG35Bb-DsRed signals are coincident suggesting MST-HMG35Bb-DsRed is generally associated with DNA (Figure 3.4, (JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005). However, EGFP-MST-HMG66A showed a dynamic change in spatial distribution as spermatids progress. At the canoe stage, DAPI and EGFP-MST-HMG66A signals are coincident. As shown in Figure 3.4, DNA and EGFP-MST-HMG66A are in the sides of the canoe while RFP-Nup107 is located in the canoe concavity. However, as nuclei transition to the rocket stage, EGFP-MST-HMG66A is enriched in the DNA located closest to the NPC. The significance of this asymmetric accumulation of MST-HMG66A on the DNA unclear, but it may reflect differential condensation of chromatin that is visible in TEM sections at the rocket stage (TOKUYASU 1974).

These and other cytological studies document a remarkable array of different spatiotemporal patterns among spermatid nuclear proteins (Figure 3.7) and even among members of the MST-HMG box protein family. Differences may reflect variation in DNA binding activities, or interacting chromosomal proteins, but NPCs and perhaps other features of the nuclear envelop may play active roles.

Absence of Mst-Hmg66A and Mst-HMG35Ba/b significantly reduces total sperm count

Deletion of some *Drosophila Mst-Hmg* genes do not result in detectable phenotypes or cause only subtle effects on sperm morphology or male fertility (RATHKE *et al.* 2010; EREN-GHIANI *et al.* 2015). However, changes in the dosage of two different *Mst-Hmgs* have been shown to induce more severe effects. For example, males hemizygous for *Mst-Hmg77F* (*Df/Mst-Hmg77F*) and males that completely lack *Mst-Hmg35Ba* and *Mst-Hmg35Bb* (*Df Mst-Hmg35Ba/b)* are fertile (RATHKE *et al.* 2010). However, the combined reduction in dosage results in male sterility (KIMURA AND LOPPIN 2016).

Because hemizygosity for *Mst-Hmg66A*^{V42X} did not reduce male fertility, we assayed for genetic interaction with other *Mst-Hmg* genes. As shown on Table 3.2, of the six deficiency genotypes tested, the only combination that reduced male fertility was *Df Mst-Hmg35Ba/b; Df/Mst-Hmg66A*^{V42X}, which we refer to as the triple mutant as it is entirely deleted for three *Mst-Hmg* genes. When control males deleted for two genes, *Df Mst-Hmg35Ba/b*, or just *Df/Mst-Hmg66A*^{V42X} were mated to wildtype females, 93% and 86% respectively of the eggs hatched, but the triple mutant males yielded a 26% hatch rate.

To determine the cause of the reduced fertility of the triple mutant, we compared total sperm count, sperm transfer to females, and sperm morphology relative to those of control males. No significant differences in total sperm count were observed the seminal vesicles of young males (2 days post-eclosion, 2d PE) but significantly fewer sperm were produced by triple mutant males compared to controls, with 25% and 45% reductions in sperm yields 4d PE and 6d PE respectively (Figure 3.5B). Since the triple mutants produced motile sperm, we asked if their sperm were transferred to females upon mating and successfully stored in the female sperm storage organ, the seminal receptacle. We observed that females mated to triple mutant males

have a 55% reduction in seminal receptacle sperm counts compared to the controls (Figure 3.5C).

Interestingly, sperm produced by triple mutants had abnormal nuclear morphology in the seminal vesicle and seminal receptacle, with variable nuclear lengths and shapes, including knobbly nuclei, and frequent aggregations (Figure 3.5, C and D). A low frequency of abnormal nuclear morphologies was previously reported for sperm produced by *Df Mst-Hmg35Ba/b* males, even though these males have high fertility (Table 3.2).

Absence of Mst-Hmg66A and MsT-HMG35Ba/b impacts the actin-based interaction between sperm heads and the head cyst cell

To determine when the nuclear defects first arose in the triple mutant, we compared spermiogenic stages in control and mutant males. Individualization, the process that separates the 64 interconnected spermatids, initiated and progressed through the head region normally as observed by phalloidin staining to detect the position of actin cones along the cyst. Based on the post-nuclear position of actin cones, elongated and highly condensed spermatid nuclei appeared well aligned and were uniform in shape and size (Figure 3.6, A and B, 30 cysts observed from 5 males of each genotype).

Bundles of 64 individualized sperm remain enclosed by two somatic cells, the head and tail cyst cells. As the cyst moves to the basal-most segment of the testis, the head cyst cell (HCC) embeds in the terminal epithelium (TE) and anchors the cyst to the testis wall. The long spermatid tails are packed laterally for streamlining, then undergo an organized coiling process to draw tails close to the TE before sperm are released from the cyst cells and swim into the seminal vesicle (TOKUYASU *et al.* 1972; DESAI *et al.* 2009; DUBEY *et al.* 2016). This process

requires a connection between the HCC and sperm that is strengthened by HCC actin filaments that interdigitate between sperm heads (DESAI *et al.* 2009). To determine if nuclear morphology was normal at these stages, we visualized the organization of the HCC F-actin cap by phalloidin staining (DESAI *et al.* 2009) and spermatid nuclei by DAPI-staining. In the control males, the HCC actin cap surrounded an orderly cluster of spermatid nuclei (n=30 cysts) (Figure 3.6A). In the triple mutant males, 9 cysts were indistinguishable from those of control males. However, the remaining cysts exhibited a heterogeneous array of abnormalities (Figure 3.6B). Eight cysts had a relatively weak HCC F-actin staining and disorganized nuclear bundles. In the remaining cysts, HCC F-actin staining did delineate an actin cap structure and nuclei were disorganized, bent, and often aggregated. Taken together, these results indicate that MST-HMG66A together with MST-HMG35Ba/b play a post-individualization role in maintaining nuclear integrity. We conclude that these MST-HMG proteins are needed to structurally support interactions between the spermatid nuclei and F-actin cap of the HCC which must be stably established and maintained to enable the tail coiling process.

Discussion

The widespread prevalence of sperm-specific chromosomal proteins has long been documented in animals, but only a few model organisms have been used to carefully analyze their functions. Genetic studies have shown that histone H1-derived transition proteins (TPs) and protamines (PRMs) are required for spermatid nuclear condensation in the mouse (YU *et al.* 2000; CHO *et al.* 2001; ZHAO *et al.* 2001). Almost all *Drosophila* SNBPs known today belong to the MST-HMG box family, most of whose members were first recognized because of their male-enriched expression and characteristic amino acid compositions (RUSSELL AND KAISER 1993;

JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005; RATHKE *et al.* 2007; EREN-GHIANI *et al.* 2015). Because several of these proteins show temporal profiles during spermiogenesis that are highly similar to those of TPs or PRMs, *Drosophila* MST-HMG box proteins have been thought to be functionally analogous to mouse TPs or PRMs despite the lack of experimental evidence for such a role. Thus, the addition of MST-HMG Ddbt to this protein family is important because MST-HMG Ddbt was originally identified in a forward genetic screen for male sterility due to paternal effects and its function is independent of spermatid nuclear condensation (YAMAKI *et al.* 2016). The updated MST-HMG box family clearly provides a different perspective on how each member may have evolved and diversified its function. In this study, we propose a new nomenclature to clarify the evolutionarily and functionally distinct nature of *Drosophila* MST-HMG box proteins compared to PRMs. For this reason, names such as "protamines" or "dProtamine" should be refrained from now on to avoid confusion unless true homologs of PRMs were to be discovered in this species.

It is worth mentioning, however, that the MST-HMG box family contains at least 12 distinct members with 20 Y-linked copies of a few members (KRSTICEVIC *et al.* 2015). Interestingly, the loss of MST-HMG77F allows the detection of MST-HMG35Bb by immunostaining in spermatid nuclei undergoing individualization (KIMURA AND LOPPIN 2016). Normally, MST-HMG35Bb is not detectable by immunostaining in highly condensed spermatid nuclei at this stage likely because of the loss of antigen accessibility. Thus, it is possible that spermatid chromatin is not as condensed in the absence of MST-HMG77F, allowing for access to MST-HMG35Bb antigens. Additionally, sperm DNA in the absence of MST-HMG35Ba/b is more susceptible to X-ray-induced DNA damage (RATHKE *et al.* 2010), suggesting that sperm DNA may be less compacted and more accessible by X-ray in the absence of MST-HMG35Ba/b.

These studies suggest that we cannot yet rule out the possibility that different family members contribute to spermatid chromatin condensation to various extents and that the complete deletion of MST-HMG box proteins may actually result in spermatid nuclear condensation defects.

Based on similarities to the canonical HMG box, it is likely that the MST-HMG box is required for DNA binding. It is also possible that the NFLR motif that is unique to MST-HMG box allows for a different level of DNA compaction in spermatid nuclei. At least for MST-HMG77F, an *in vitro* analysis has shown that its MST-HMG box induces the aggregation of DNA oligonucleotides (KOST *et al.* 2015). If all MST-HMG box proteins use the conserved MST-HMG box for DNA binding, then what is the significance of the highly variable regions flanking the MST-HMG box? Thus far, six members of the MST-HMG box protein family have been localized using tagged versions or by immunostaining, revealing at least three distinct behaviors among the family members. First, there is variation in the timing of their nuclear localization during spermiogenesis. Because NPCs are still present when MST-HMG box proteins enter spermatid nuclei, it is likely some features about these proteins themselves that regulate differential timing of their nuclear entry. One hypothesis is therefore that the flanking regions of the MST-HMG box domain harbors nuclear localization signals (NLSs) that interact with different receptor proteins such as importin proteins. At this point, no *Drosophila* spermatid chromatin proteins have been analyzed for their NLSs. Interestingly, a canonical HMG box protein, SRY, is known to contain two NLSs at the beginning and the end of the HMG box. One NLS becomes bound by calmodulin in response to developmentally regulated calcium waves while the other NLS is a string of basic residues that binds to importin beta. These two NLSs are not functionally redundant, and likely regulate the precise timing of SRY's nuclear localization to turn on the transcriptional program for male sex determination (HARLEY *et al.* 2003). Perhaps

a combination of multiple NLSs is used to regulate the precise timing of the nuclear localization of MST-HMG box proteins in a similar fashion.

MST-HMG box proteins also show distinct behaviors with regards to where they localize in spermatid nuclei. The most extreme case is MST-HMG Ddbt, which is localized at spermatid telomeres (YAMAKI *et al.* 2016). MST-HMG Ddbt requires telomere capping complexes (TCCs) for its proper telomeric localization (YAMAKI *et al.* 2016). Another spatially distinct behavior is exhibited by MST-HMG66A. This protein is widely distributed in the spermatid nucleus at the canoe stage, but the protein becomes concentrated on one side of the nucleus that is adjacent to the NPC-enriched nuclear envelope. Given the ability of NPCs to interact with different chromosomal proteins as shown in somatic cells (BECK AND HURT 2017), it is tempting to speculate that MST-HMG66A asymmetric localization is regulated by NPCs. Perhaps the spatially distinct distributions of MST-HMG box proteins are mediated by the variable regions of each MST-HMG box protein that may interact with specific nuclear proteins.

Lastly, MST-HMG box proteins are removed from chromatin at different times either during spermiogenesiss or fertilization. MST-HMG94D becomes evicted during spermiogenesis at the rocket stage (RATHKE *et al.* 2007). This is an interesting stage of nuclear transformation because extensive nuclear volume reduction is initiated with the elimination of NPCs and the majority of nucleoplasm. MST-HMG94D may simply be exported through the NPCs, degraded within the nucleus through the proteasome system, or disposed of along with the nucleoplasm, or a combination of these mechanism. For MST-HMG box proteins that are retained in mature sperm, their loss occurs after sperm entry into the egg cytoplasm. We document variation in the timing of loss of MST-HMG66A as compared to MST-HMG35Bb. MST-HMG35Bb is likely removed by the maternally supplied chaperone, TAP/p32. Interestingly, MST-HMG35Ba, the

paralogous protein of MST-HMG35Bb, requires a completely different set of chaperone proteins (NAP-1, NLP, and Nph) (EMELYANOV *et al.* 2014). It is thus possible that the highly variable regions of MST-HMG box proteins contain specific features that allow interactions with distinct sets of chaperone proteins, which may be the basis for differential behaviors of MST-HMG box proteins during fertilization. This proper removal of SNBPs to re-establish histone-based chromatin is essential because the failure to do so results in failed embryogenesis. The fact that the variable regions also harbor positively selected residues suggests that the remodeling process during fertilization may be ensured to occur in a species-specific manner with maternal chaperones that have likely co-evolved with these rapidly evolving SNBPs.

What are the functional consequences of the spatiotemporally distinct behaviors of MST-HMG box proteins on spermatid nuclear transformation? Our interpretation of *Mst-Hmg66A* interaction with *Mst-Hmg35Ba/b* provides an alternative perspective on how some MST-HMG box proteins function. Four widely distributed family members (MST-HMG-35Ba, MST-HMG35Bb, MST-HMG-99C, and MST-HMG-66A) can be deleted singly without markedly affecting male fertility. Interestingly, genetic interactions have been demonstrated between MST-HMG-35Ba/b and MST-HMG77F (KIMURA AND LOPPIN 2016), between MST-HMG-35Ba/b and MST-HMG-99C (EREN-GHIANI *et al.* 2015), and between MST-HMG-35Ba/b and MST-HMG-66A. In all these cases, male fertility was reduced, yet spermatid nuclear condensation seemed unaffected. In the current study, we revealed a surprisingly late defect in spermatid nuclei lacking MST-HMG-35Ba/b and MST-HMG-66A. The defect appeared in the terminal epithelium (TE) of the testis after the actin machine has moved past and individualized the head region of the bundle of 64 spermatids. Normally, TE is where sperm tail coiling takes place, and previous studies show that the successful interaction between the head cyst cell (HCC) and

sperm heads is required for this process (DESAI *et al.* 2009). The HCC-sperm head interaction is maintained by the interdigitating actin filament provided by the HCC (DESAI *et al.* 2009). We suggest MST-HMG66A in concert with MST-HMG35Ba/b likely provide structural support to maintain the integrity of highly condensed nuclei during the physically demanding process of coiling. In light of our findings, we further provide an alternative interpretation of the *Mst-Hmg77F* deletion phenotype. The absence of MST-HMG77F results in the disorganization of spermatid heads during individualization. Perhaps MST-HMG77F plays a similar structural role during individualization to withstand the force exerted by the actin machine acting on the spermatid head region (NOGUCHI AND MILLER 2003). Whether during individualization or coiling, the demand for withstanding physical forces likely arose with the extremely long tails of *Drosophila* sperm. Because of the absence of nuclear lamina (FABBRETTI *et al.* 2016) and depletion of cytoskeletal filaments (TOKUYASU 1974) in late spermatids and sperm, some MST-HMG box proteins with their DNA-binding ability may have been co-opted for structural support roles.

In conclusion, our interpretation of several *Mst-Hmg box* mutant phenotypes expands the prevailing view of SNBPs as primarily involved in DNA compaction, and we propose that the expansion of the *Drosophila* MST-HMG box family has likely provided opportunities for a general DNA-binding protein to acquire diverse roles such as sperm telomere capping or structural support after spermatid nuclear condensation. Such specific roles have likely been selected for in this animal lineage to meet the ever-changing demands that are imposed by sperm formation and function.

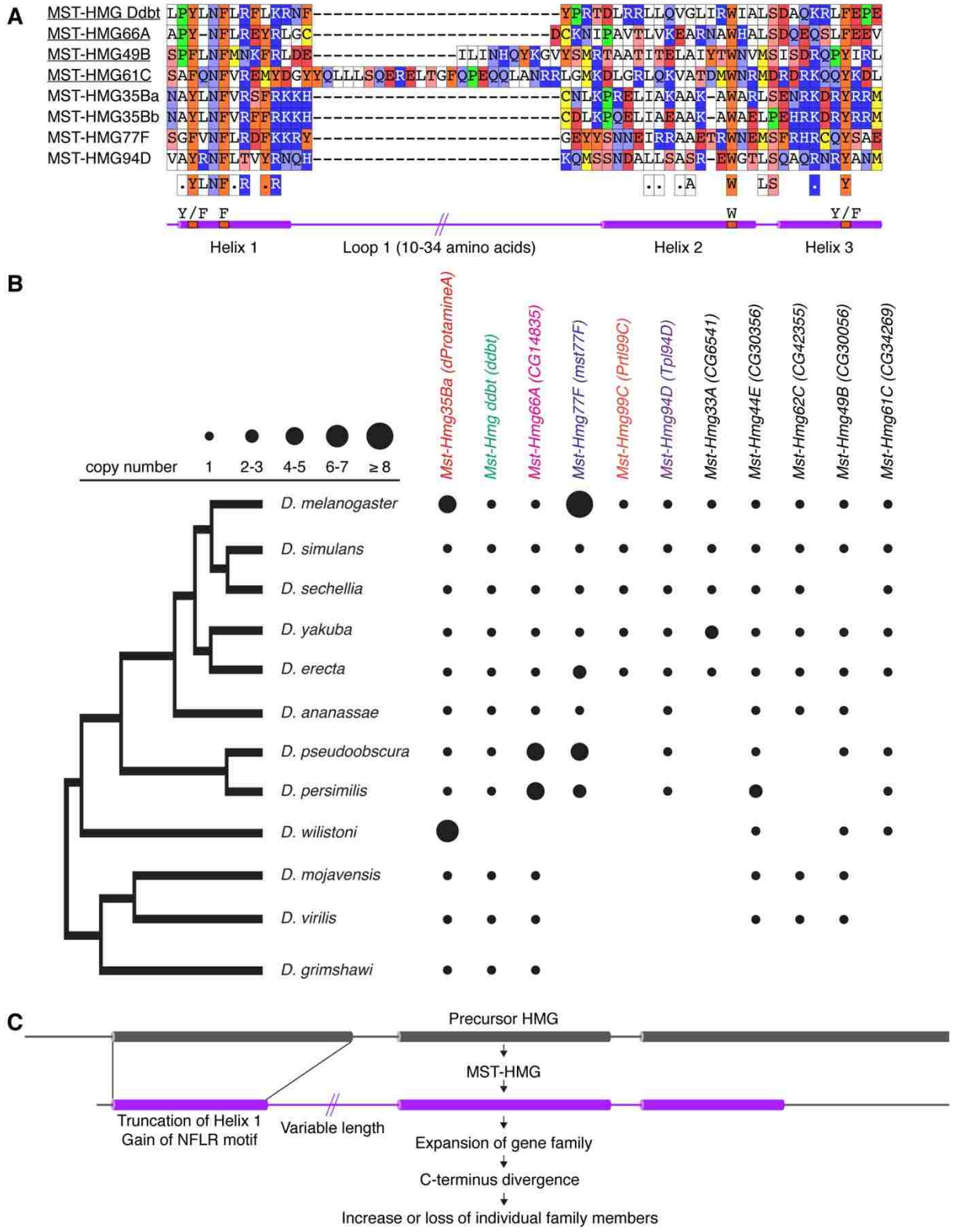


Figure 3.1 The *Drosophila* MST-HMG box protein and *Mst-Hmg* box gene family.
(A) Comparison of the amino acid sequences of four newly identified *Drosophila melanogaster* MST-HMG box proteins (underlined) and four founding members. The NFLR motif in helix 1

and other features are described in the text. (B) Comparison of copy numbers of *Mst-Hmg* genes among *Drosophila* species. Note the proposed nomenclature with corresponding older names in parentheses. All genes show male specific transcription. For six genes (color font), protein localization studies have been performed and confirm expression during spermiogenesis. (C) Proposed evolutionary origin of Mst-Hmg genes.

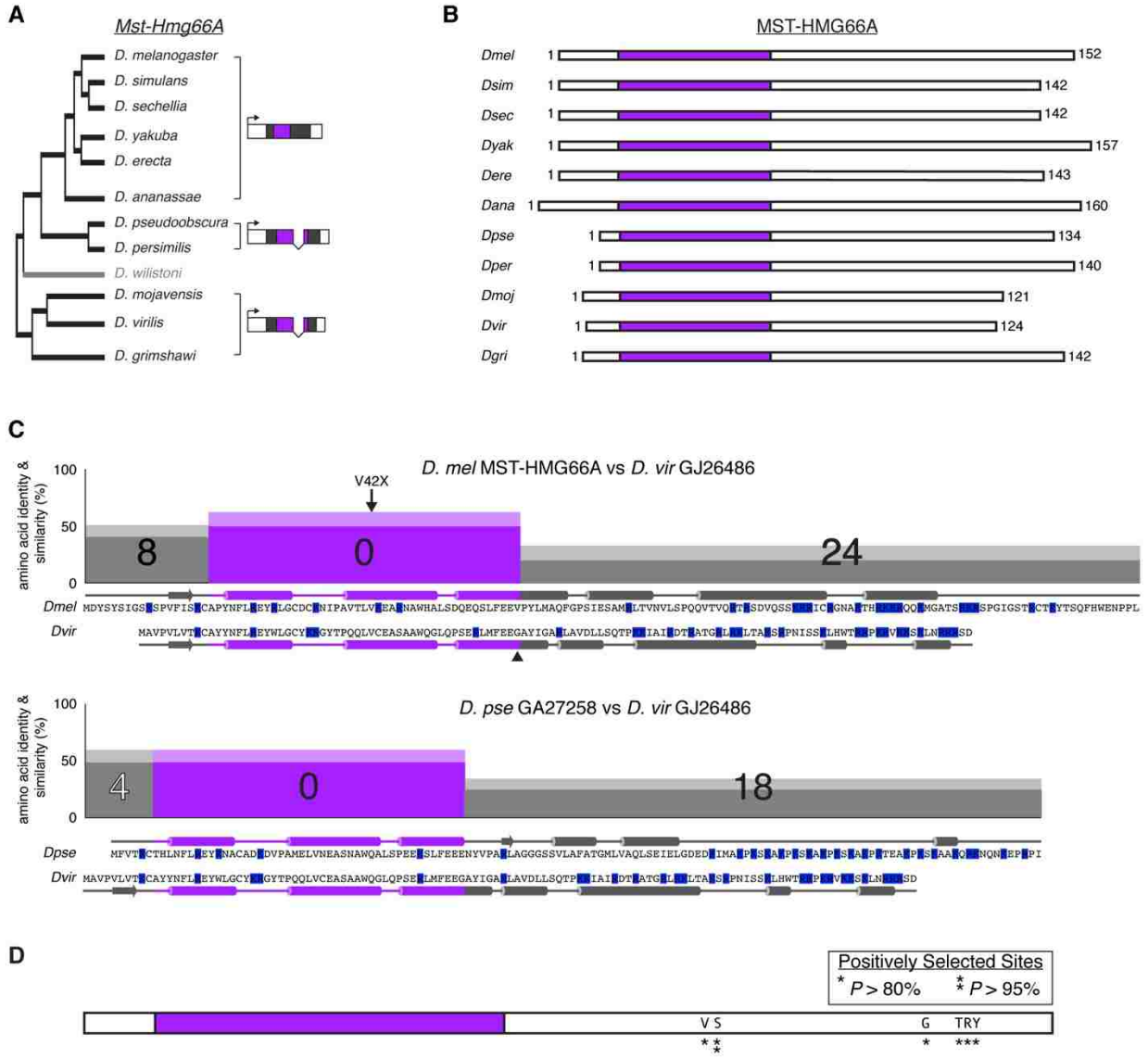


Figure 3.2 Species comparison of orthologous MST-HMG66A sequences.
(A) Gene structure of orthologous *Mst-Hmg66A* genes, showing 5' and 3' untranslated regions (white segments), open reading frame (gray with pink Mst-Hmg box) and intron position in region encoding helix 3 in all but the melanogaster group. (B) Comparisons of MST-HMG66A protein showing location of the MST-HMG box and variable lengths of the flanking regions. (C) Pairwise species comparisons delineating the N terminal, MST-HMG box, and C terminal subregions. In the top diagram, numbers report whether *D. melanogaster* protein has equivalent or more amino acids in each subregion compared to the *D. virilis* protein. Percent amino acid similarity is depicted by box height, inclusive of darker colored (percent identity) and light colored boxes (percent similarity). Shown below the diagram are the protein sequences and alpha-helical regions. Arrow shows the boundary between *D. virilis* exons 1 and 2. Basic residues are noted in blue. In the bottom diagram, the comparison uses the same conventions, except outlined number report the shorter *D. pseudoobscura* N-terminal subregion. (D) Diagram

depicts *D. melanogaster* MST-HMG66A and location of positively selected residues based on analysis of *D. melanogaster, D. simulans,* and *D. sechellia*.

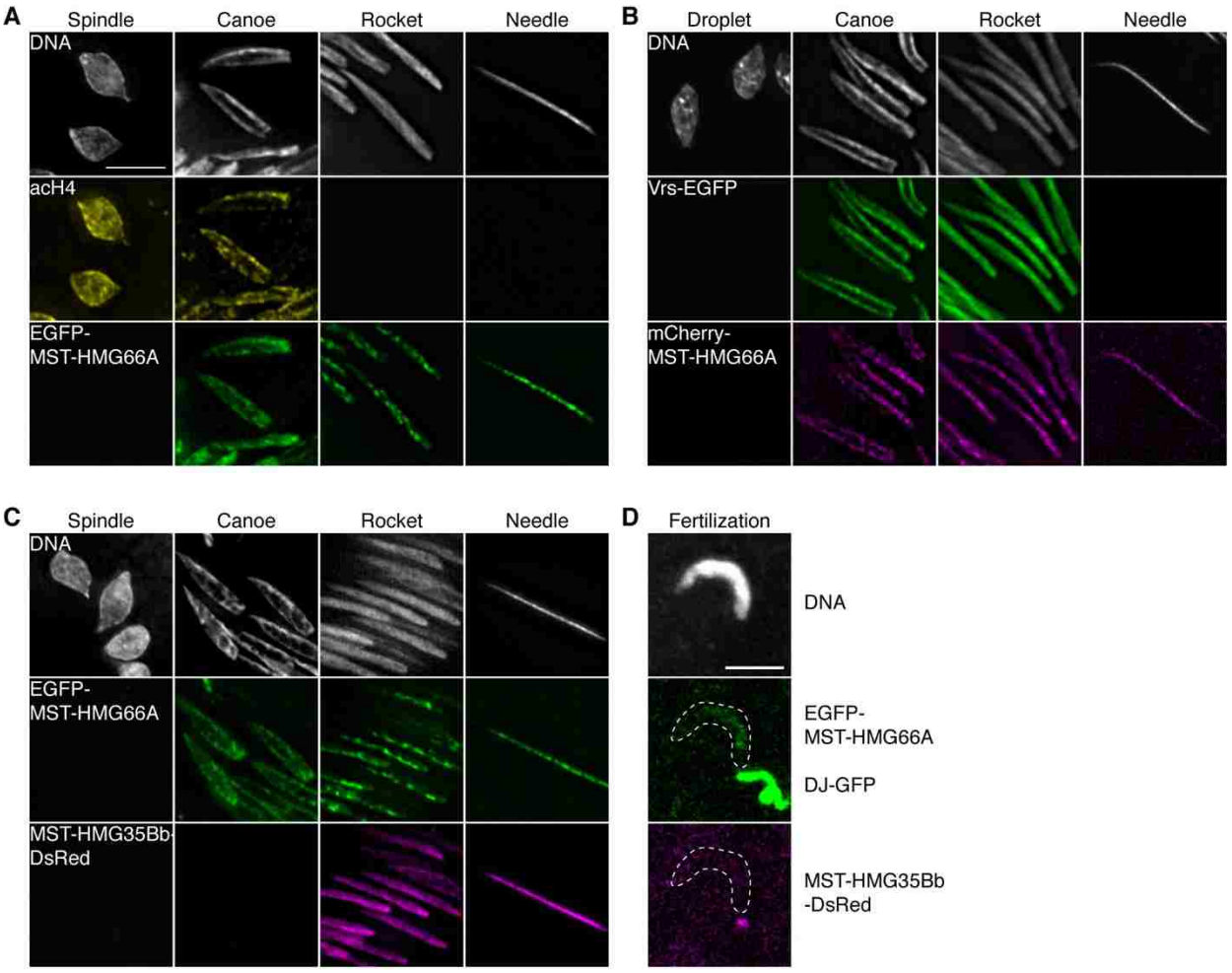


Figure 3.3 Temporal profile of MST-HMG66A during spermiogenesis and fertilization. Spermiogenic stages were distinguished by DNA staining (DAPI), and MST-HMG66A localization was determined using EGFP or mCherry tagged versions relative to timing of nuclear (A) acetylated histone H4 by immunostaining, (B) EGFP-Vrs and (C) MST-HMG35Bb-DsRed. (D) A sperm containing EGFP-MST-HMG66A, MST-HMG35Bb-DsRed, and the tail protein Dj-GFP undergoing nuclear decondensation in the egg cytoplasm. Note that at this stage, MST-HMG66A is still retained but MST-HMG35Bb is not detected. Scale bar corresponds to 5 um.

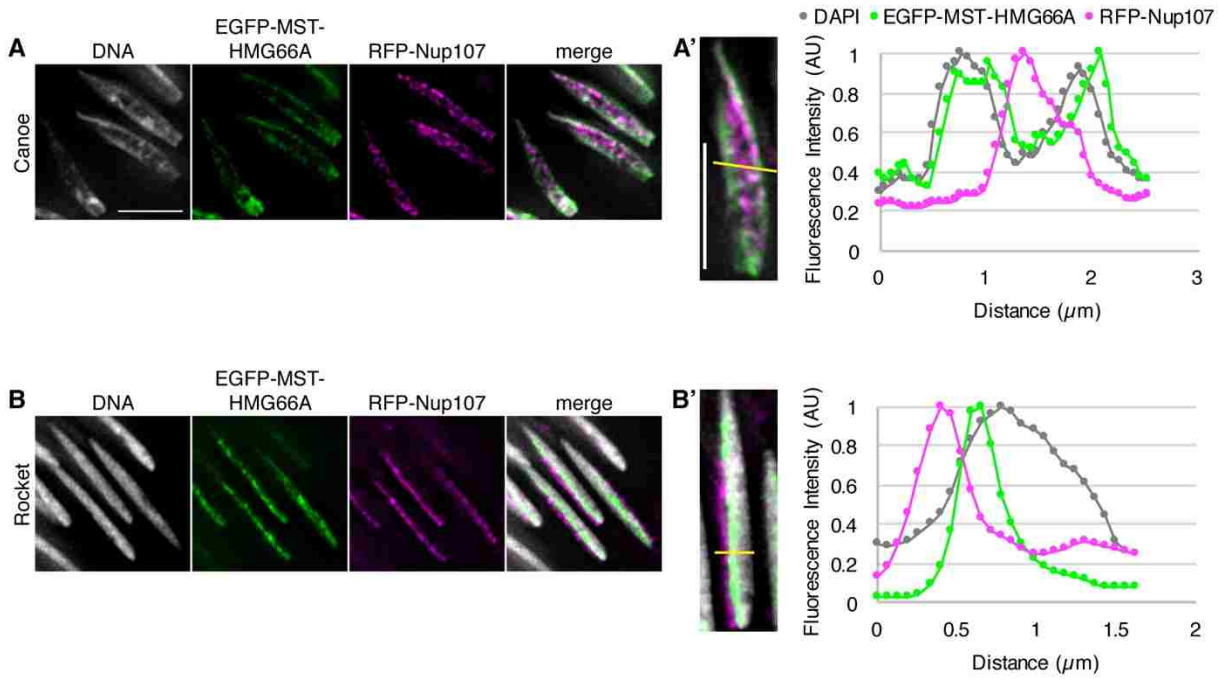


Figure 3.4 Dynamic changes in MST-HMG66A nuclear distribution during spermiogenesis. (A) Localization of EGFP-MST-HMG66A at the canoe stage. (A') Line scan across the nucleus (yellow line) compares signal intensity of DAPI, EGFP-MST-HMG66A and RFP-Nup107. (B) Localization of EGFP-MST-HMG66A at the rocket stage. (B') Similar convention as (A'). Scale bar 5 um.

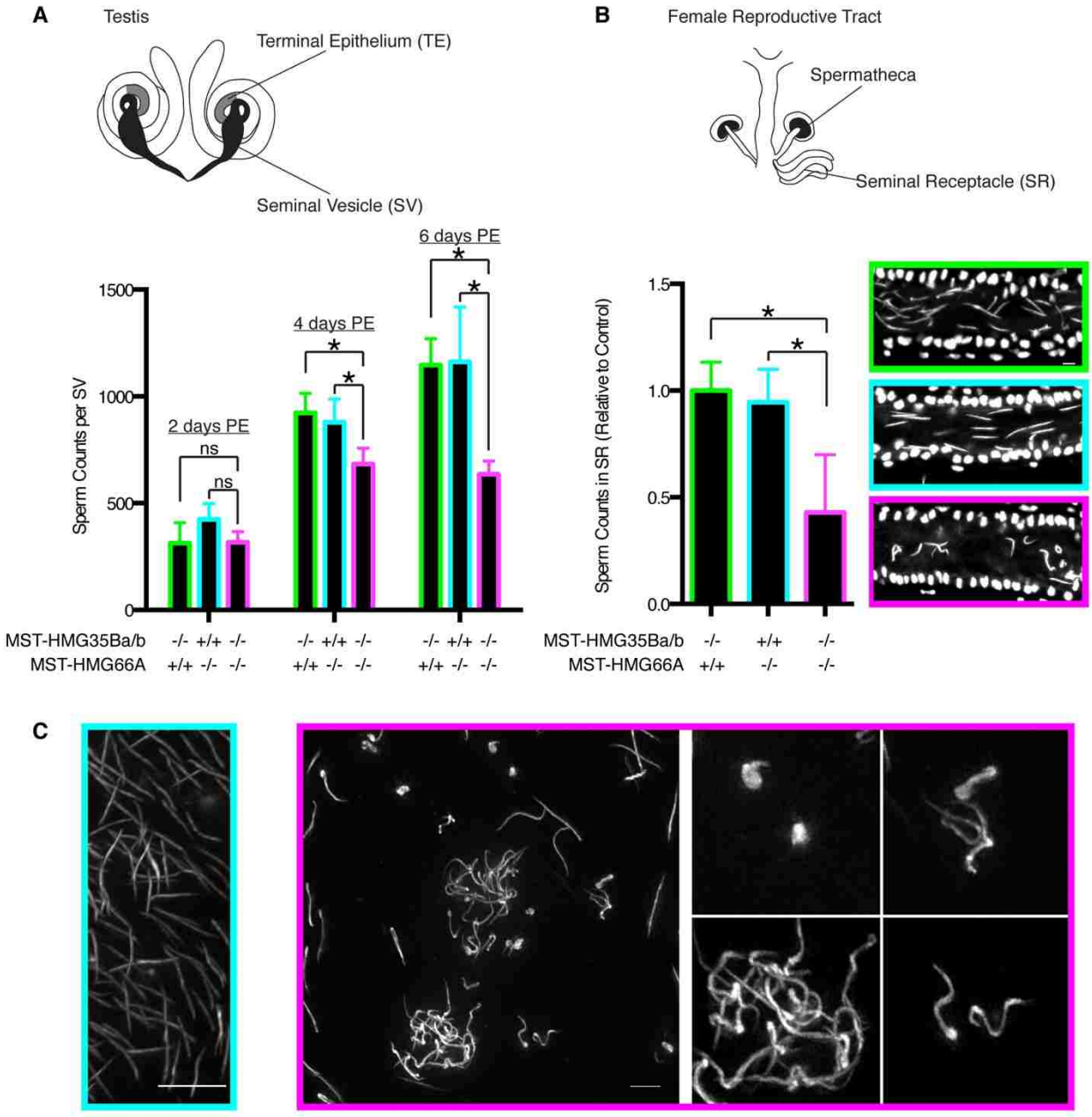


Figure 3.5 Dosage effects of MST-HMG66A and MST-HMG35Ba/b on sperm in the seminal vesicle and seminal receptacle.

Diagram shows the coiled testis tubes with apical tips at the top. The basal segments are lined with terminal epithelia (TE, gray) and connect to seminal vesicles (SV, black). Graph compares the SV sperm counts in males that varied in age (days post-eclosion, PE) and dosage of *Mst-Hmg66A* and *Mst-Hmg35Ba/b* loci. Error bars indicate standard deviation and asterisks indicate significant difference in sperm counts to compared to controls. (B) Diagram of the female reproductive tract showing sperm storage organs, spermathecae and seminal receptacles (SR) (right). Graph compares the SR sperm counts in females mated to similarly aged males that varied in dosage of MST-HMG66A and MST-HMG35Ba/b loci. Right panels show SR segments that were DAPI-stained to show sperm nuclei within the SR epithelial cells. (C)

Morphology of DAPI-stained sperm nuclei in SVs of males that varied in dosage of MST-HMG66A and MST-HMG35Ba/b. Small right panel show higher resolution view of selected sperm or sperm aggregates. Scale bar corresponds to 5 um.

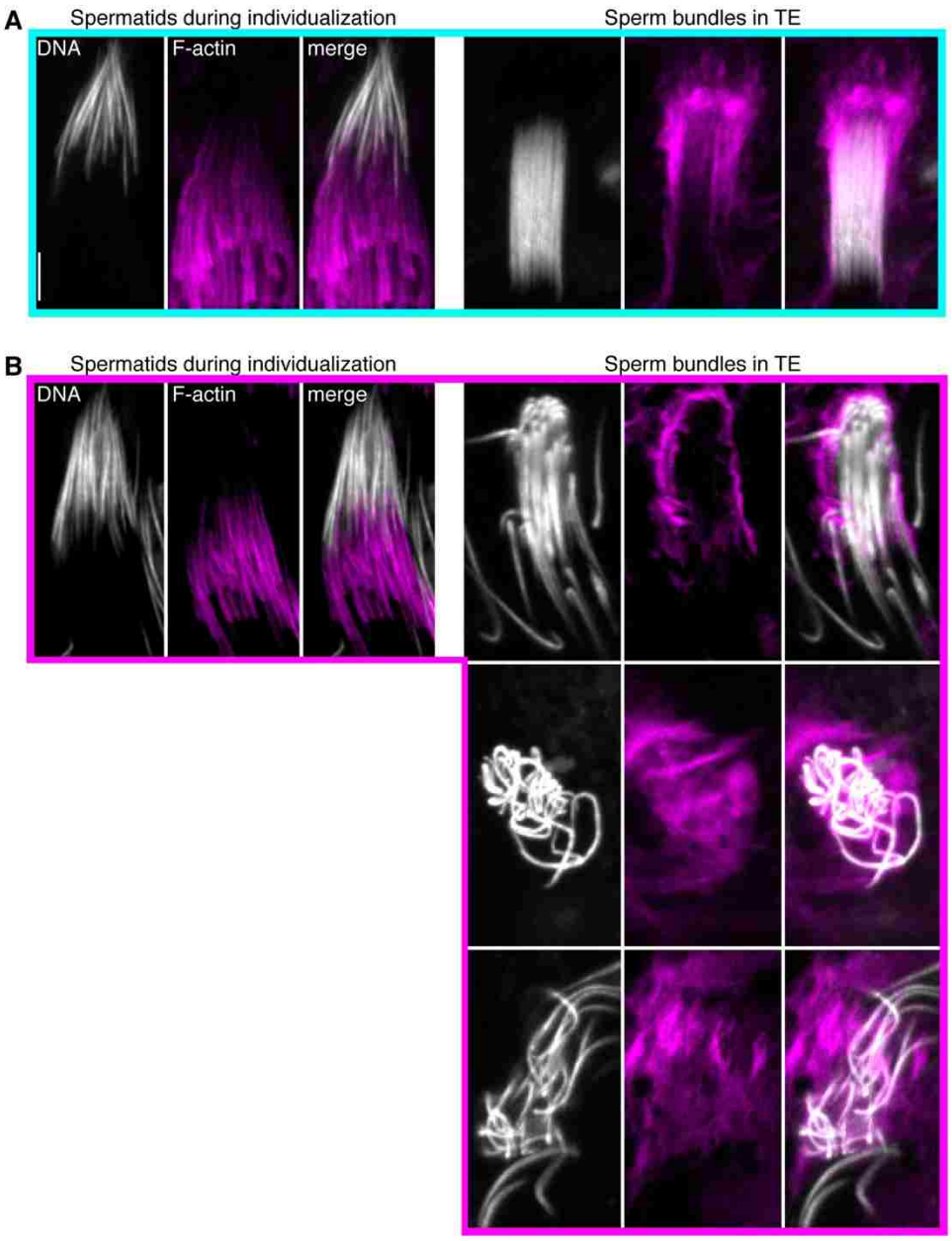
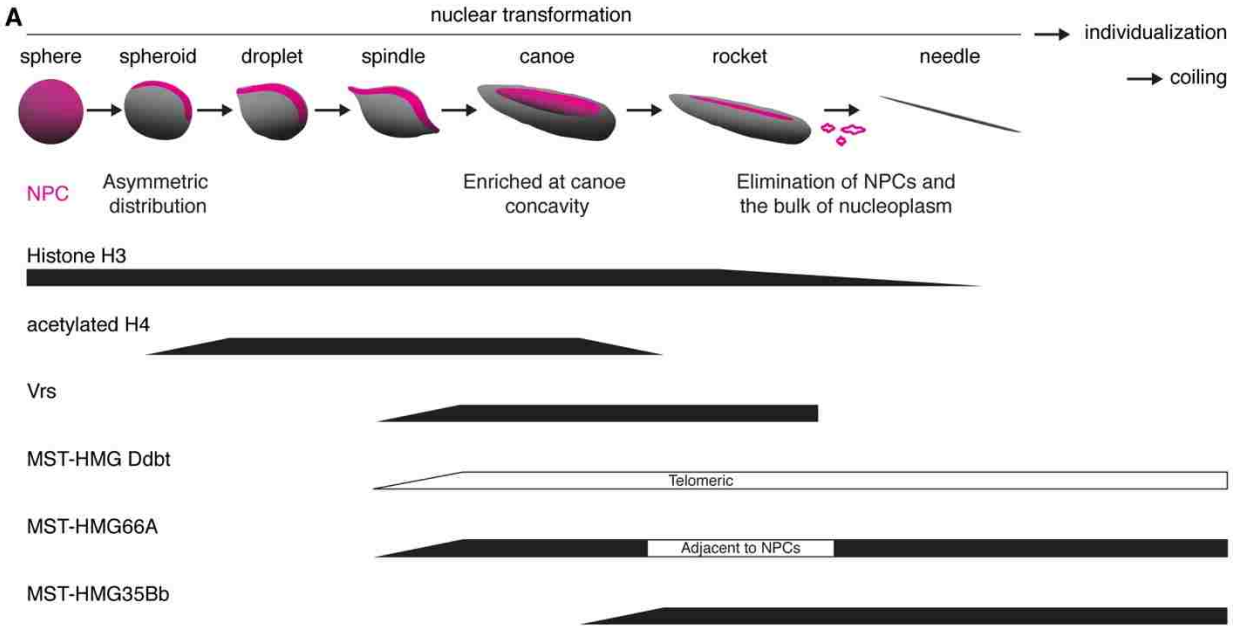
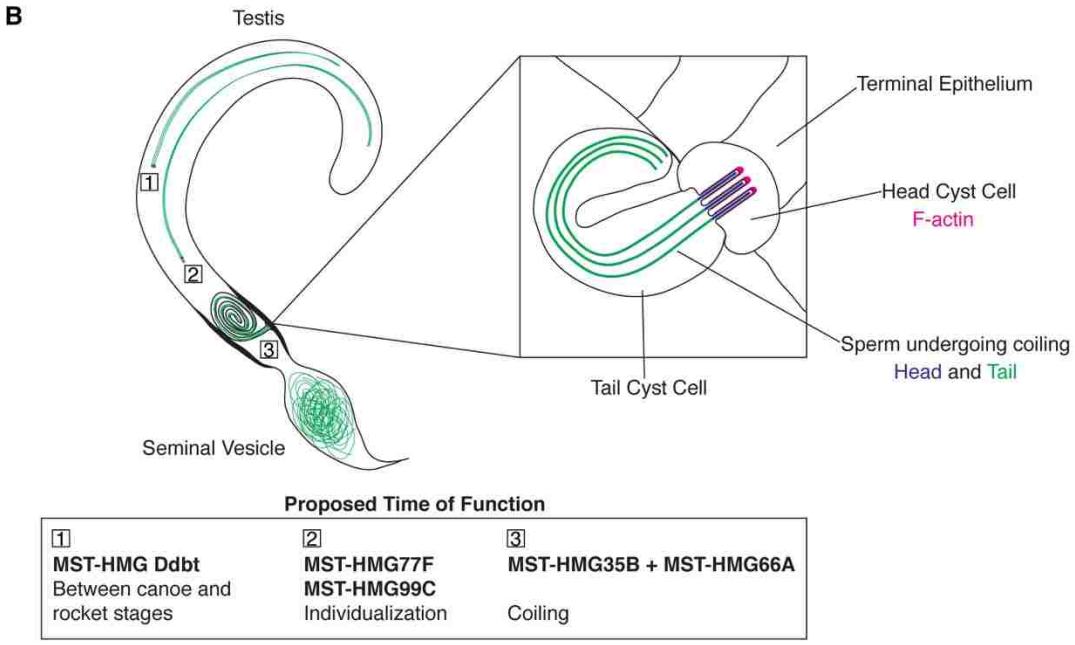


Figure 3.6 Dosage effects of MST-HMG66A and MST-HMG35Ba/b on spermiogenesis during and after individualization.
Spermatids or sperm from males that vary in MST-HMG66A and MST-HMG35Ba/b dosage with color coded genotypes as in Figure 3.5. Nuclei are stained with DAPI and F-actin is visualized using Alexa 594-conjugated phalloidin. (A) In the control, the left panels show a cyst of spermatids during individualization, with actin cones progressed past the head region. The right panels show a cyst embedded in the TE, with the sperm heads bundled properly by the surrounding head cyst cell (HCC) and its well- organized F-actin network. (B) Corresponding stages from males that lack MST-HMG66A and MST-HMG35Ba/b show variable penetrance and expressivity of the abnormal sperm and HCC F-actin organization. Scale bar corresponds to 5 um.





Proposed Time of Function

1	2	3
MST-HMG Ddbt Between canoe and rocket stages	**MST-HMG77F** **MST-HMG99C** Individualization	**MST-HMG35B + MST-HMG66A** Coiling

Figure 3.7 The spatiotemporal dynamics and diverse functions of nuclear proteins during *D. melanogaster* spermiogenesis.
(A) Diagram depict stages of spermatid nuclear transformation, distribution of NPCs, and gradual accumulation and loss of chromosomal proteins. Histone dynamics were originally described Renkawitz –Pohl (). Subnuclear localization is classified as widely distributed (black) or more regionally localized as noted for two MST-HMGs.
(B) Proposed time of function of MST-HMG proteins based on earliest detectable defects resulting from reduced dosage. MST-HMG Ddbt is required to maintain the telomere capping complex just after the canoe stage (1) with consequences for paternal chromosome stability

during embryogenesis. MST-HMG77F and MST-HMG99C are likely required for individualization in the sperm head region (2) while MST-HMG66A and MST-HMG35Ba/b are required together for interaction of sperm with head cyst cell in the testis terminal epithelium (3). Processes (2) and (3) are critical for sperm production.

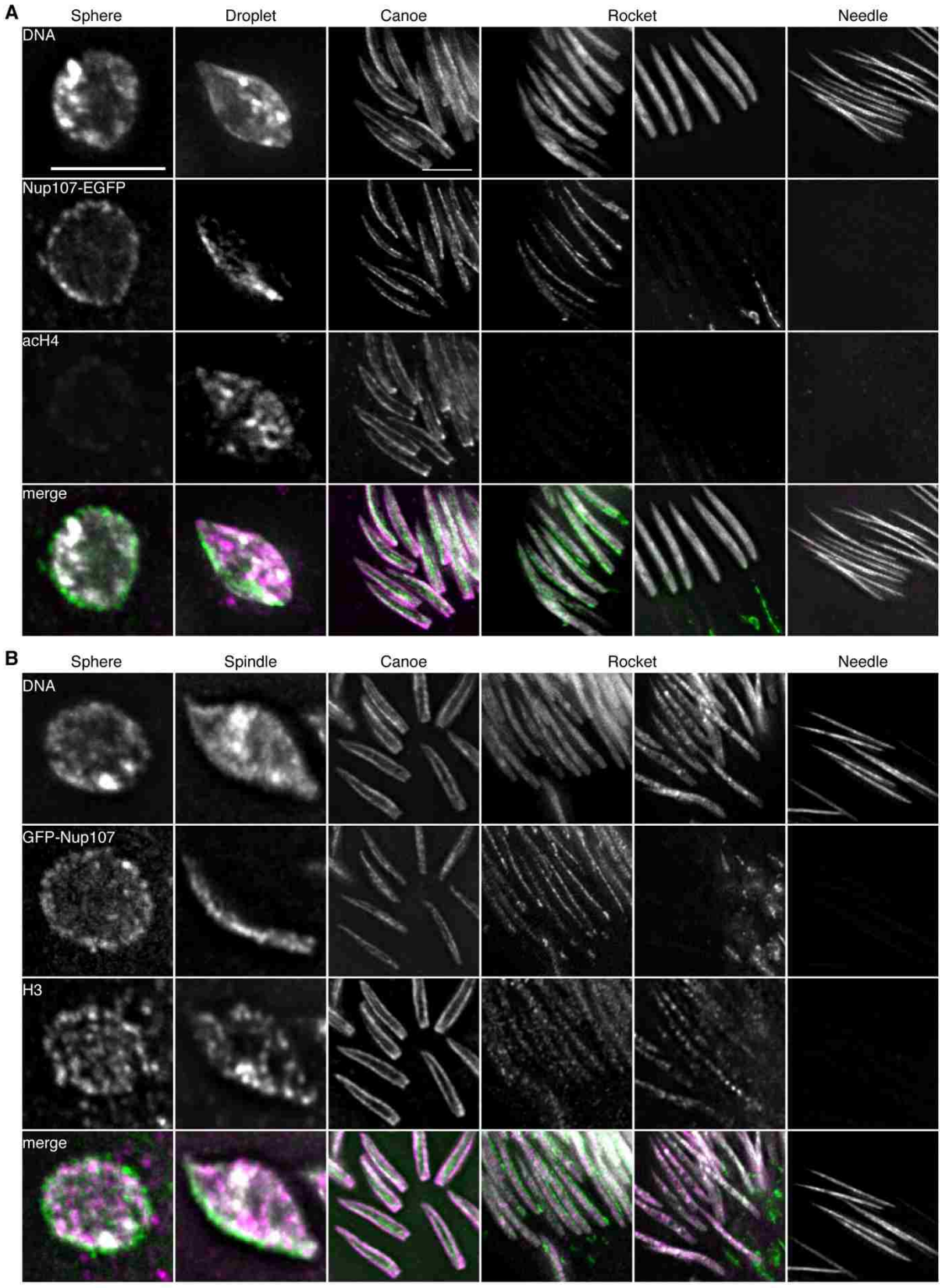

A
Sphere
Droplet
Canoe
Rocket
Needle
DNA
Nup107-EGFP
acH4
merge
B
Sphere
Spindle
Canoe
Rocket
Needle
DNA
GFP-Nup107
H3
merge

Figure S3.1 Dynamic changes in Nup107 distribution relative to histone transitions during spermiogenesis.
Stages during nuclear transformation were distinguished by DNA dye DAPI, Nup 107 was detected by GFP-Nup107, and acetylated histone H4 (A) and histone H3 (B) were detected by immunostaining. Scale bar correspond to 5 um.

Table 3.1 d_N/d_S analysis of *Mst-Hmg box* genes

Models compared	$-2\Delta lnL$	d_N/d_S estimate	% sites[a] with estimated d_N/d_S	Positively selected sites[b]
Mst-Hmg35Ba[c]				
M1 (neutral) vs M2 (selection)	1.64 (p=0.44, d.f. = 2)	N/A	N/A	N/A
M7 (neutral) vs M8 (selection)	1.70 (p=0.43, d.f. = 2)	N/A	N/A	N/A
M8a (neutral with $d_N/d_S = 1$) vs M8 (selection)	1.56 (p=0.21, d.f. = 1)	N/A	N/A	N/A
Mst-Hmg77F[c]				
M1 vs M2	13.9 (p=0.001, d.f. = 2)	82.5	1.4	**142P**
M7 vs M8	14.0 (p=0.0009, d.f. = 2)	82.6	1.4	*110L* **142P** *193E*
M8a vs M8	13.9 (p=0.0001, d.f. = 1)	82.6	1.4	
Mst-Hmg99C[c]				
M1 vs M2	5.48 (p=0.06, d.f. = 2)	N/A	N/A	N/A
M7 vs M8	5.59 (p=0.01, d.f. = 2)	5.57	11.5	*6G 75T 82Y 105S* **108L** *166R 190E*
M8a vs M8	5.48 (p=0.02, d.f. = 1)	5.57	11.5	

This analysis used *D. melanogaster*, *D. simulans*, and *D. sechellia* sequences (Table S1).
[a] Percentage of 145 (*Mst-Hmg35Ba*), 208 (*Mst-Hmg77F*), and 200 (*Mst-Hmg99C*) codons aligned after removal of gaps from species comparisons.
[b] Predicted with P > 90% (bold) or P > 80% (italics).
[c] Positions of the MST-HMG box domain: aa 78-124 (MST-HMG35Ba), aa 45-92 (MST-HMG77F), and aa 20-66 and aa 143-189 (MST-HMG99C).

Table 3.2 Egg hatch rate

Paternal genotype	# eggs	# larvae
yw; Iso3	518	365 (70%)
yw; Mst-Hmg66A^{V42X}	510	389 (76%)
yw; Iso3/Df(3L)BSC459 (*Df Mst-Hmg66A)*	510	382 (75%)
yw; Mst-Hmg66A^{V42X}/Df(3L)BSC459	510	441 (86%)
yw; egfp:Mst-Hmg66A; Mst-Hmg66A^{V42X}/Df(3L)BSC459	510	451 (88%)
w; Df(2L)Rathke1/Df(2L)Rathke2; MKRS/TM2 (*Df Mst-Hmg35Ba/b*)	510	472 (93%)
w; Df(2L)Rathke1/ Df(2L)Rathke2; Mst-Hmg66A^{V42X}/TM2	510	375 (74%)
w; Df(2L)Rathke1/CyO; Mst-Hmg66A^{V42X}/Df(3L)BSC459	510	400 (78%)
w; Df(2L)Rathke2/CyO; Mst-Hmg66A^{V42X}/Df(3L)BSC459	370	302 (81%)
w; Df(2L)Rathke1/ Df(2L)Rathke2; Mst-Hmg66A^{V42X}/Df(3L)BSC459	760	197 (26%)
w; Df(2L)Rathke1/egfp:Mst-Hmg66A Df(2L)Rathke2; Mst-Hmg66A^{V42X}/Df(3L)BSC459	501	324 (65%)
w; Mst-Hmg66A^{V42X} / Df(3L)BSC452 (*Df Mst-Hmg77F*)	500	380 (76%)
w; Df(2L)Rathke1/CyO; +/Df(3L)BSC452	240	160 (67%)
w; Df(2L)Rathke1/CyO; Mst-Hmg66A^{V42X}/Df(3L)BSC452	300	223 (74%)

Table S3.1 *Drosophila* MST-HMG66A orthologous proteins

Species	NCBI Gene ID or Genomic location	Size
D. melanogaster	38832	152
D. simulans	6737140	142
D. sechellia	6611147	142
D. erecta	6544304	143
D. yakuba	6533321	157
D. eugracilis	108102736	149
D. biarmipes	108028450	158
D. suzukii	108005761	194
D. takahashii	108065804	139
D. elegans	108142530	123
D. rhopaloa	108037249	140
D. ficusphila	108086965	130
D. ananassae	6506916	160
D. bipectianta	108119404	160
D. pseudoobscura	6897354	134
D. pseudoobscura	6897351	87
D. pseudoobscura	6901637	157
D. pseudoobscura	6901636	145
D. persimilis	6588024	140
D. persimilis	6588099	208
D. persimilis	6599328	155
D. miranda	108154475	141
D. miranda	108157180	96
D. obscura	111072714	143
D. obscura	111081425	118
D. virilis	26531256	121
D. mojavensis	26528214	124
D. arizonae	108612766	123
D. albomicans	scaffold_66278: 18,427..18,948	173
D. busckii	108600261	130
D. grimshawi	scaffold_15110: 2,810,787..2,811,503	142

Chapter 4: *Drosophila melanogaster* Sperm Head-Enriched Proteome

Abstract

Mass spectrometry analysis of specific cell types provides insights into how they perform cell type-specific functions. Previous proteomic studies using *Drosophila* sperm discovered 1108 proteins but failed to detect most of the known head-specific proteins such as those found in the nucleus or acrosome. This is most likely due to the extremely long tail of *Drosophila* sperm, which is 180 times longer than the head. To identify sperm head-enriched proteins, here we first describe a method to isolate sperm heads from long tails by mechanically disrupting the cells into fragments and isolating the heads by density gradient centrifugation. Using this method, sperm heads were enriched by 180-fold. The sperm head-enriched sample was then analyzed by mass spectrometry to investigate the protein compositions of the *Drosophila* sperm head. We discovered a total of 388 proteins that are likely enriched, if not exclusively found, in the sperm head. Among these proteins, we identified a new member of the MST-HMG box family, a group of sperm-specific chromatin proteins in *Drosophila.* Surprisingly, unlike all other members of this family, this family member is a small protein that is highly acidic due to the gain of acidic clusters in the N-terminal domain and the loss of basic amino acid clusters in the C-terminal domain. The addition of this protein to the *Mst-Hmg box* gene family reinforces the idea that the C-terminal domains of MST-HMG box proteins are highly divergent.

Introduction

While sperm of different animals come in distinct shapes and sizes, in most species, they can be divided into two morphologically and functionally distinct regions: tail and head. The consensus view is that the flagellar axoneme in the sperm tail provides the cell with the ability to migrate towards eggs. Axonemal structure and movement has been extensively studied using various model organisms. Because axonemal structures are well conserved among eukaryotes, general principles learned in one organism can be translated into other systems. Studies of actual sperm cells as well as the flagellum of single cell organisms such as *Chlamydomonas* have contributed to the identification of key axonemal substructures that are required for proper beating of sperm tails (INABA 2007).

The head can be further divided into four distinct sub-components: the nucleus, a membrane-bound organelle called the acrosome, the centriole, and the plasma membrane enclosing these structures. The widely accepted view of how the sperm head functions is as follows; acrosomal exocytosis promotes the degradation of the protective extracellular matrix surrounding the egg and presents fusogenic proteins onto the head plasma membrane, the head plasma membrane physically interacts and fuses with the egg plasma membrane, and the nucleus and the centriole become incorporated into the egg cytoplasm to initiate the first division of the zygote (HIROHASHI *et al.* 2008; GEORGADAKI *et al.* 2016). However, only a few selected groups of animals, namely mouse and aquatic species, have been used to study sperm head functions extensively (HIROHASHI *et al.* 2008), and thus the extent to which these functions can be generalized actually remains an open question.

As a group of genes, those involved in reproduction are rapidly evolving (SWANSON AND VACQUIER 2002). An example of such rapid evolution can be observed in the nucleus. At the

morphological level, sperm nuclei in many animals exhibit a similar characteristic of extreme condensation. Inside such highly condensed nuclei, sperm DNA interacts with a unique set of proteins called sperm nuclear basic proteins (SNBPs). SNBPs are comprised of evolutionarily and biochemically distinct proteins, including histones, histone variants, and sperm-specific nuclear proteins such as protamines (EIRIN-LOPEZ AND AUSIO 2009). Biochemical analyses have shown that there is a tremendous variation among species in the types of chromosomal proteins that constitute SNBPs. For example, mammalian sperm nuclei contain protamines, which have originated from histone H1 (EIRIN-LOPEZ AND AUSIO 2009; CHAMPROUX *et al.* 2016), while *Drosophila* sperm nuclei contain a group of male-enriched proteins that are derived from a high-mobility group (HMG) box protein (i.e. MST-HMG box proteins) (JAYARAMAIAH RAJA AND RENKAWITZ-POHL 2005; RATHKE *et al.* 2007; DOYEN *et al.* 2015; EREN-GHIANI *et al.* 2015; YAMAKI *et al.* 2016). These features inevitably contribute to the relative difficulty of identifying sperm head proteins using other organisms' proteomic data, calling for a more direct approach using the materials obtained from the species of interest.

Previous *Drosophila* sperm proteomic studies used whole sperm extracts and identified 1108 proteins (DORUS *et al.* 2006; WASBROUGH *et al.* 2010). However, these studies failed to pick up almost all known head-specific proteins that are present in the nucleus or acrosomal membrane. This is most likely because the *Drosophila* sperm has a tail that is 180 times longer than the head, so protein extracts prepared from whole sperm are inevitably predominated by those found in the tail. Here in this study, we first report a method that we developed to isolate sperm heads from charlatanically long tails of *Drosophila* sperm. With this method, we were able to enrich sperm heads 180-fold, which was assessed cytologically. Based on the mass spectrometry analysis of the sperm head-enriched sample, we here report *Drosophila*

melanogaster Sperm Head Enriched Proteome (DmSHEP), with a list of 388 proteins that are likely enriched, if not exclusively found, in sperm heads. One of the proteins is predicted to be an HMG-box protein whose expression is enriched in males. Interestingly, this protein is more similar to known members of the MST-HMG box family than canonical HMG box proteins. The mass spectrometry-based identification of sperm head-enriched proteins serves as the platform for the candidate gene-based approach to test systematically how sperm head proteins contribute to sperm formation or function.

Materials and Methods

Fly strains and care

To distinguish sperm heads from tails by epifluorescence microscopy, we used two fluorescently tagged fusion proteins, Dj-GFP (a tail marker) and HMG-MST35Bb-DsRed (a head marker), that were expressed under their endogenous promoters. *ST2* flies expressing these fluorescently tagged proteins were made by introducing the transgenes onto chromosome 2 by recombination. The flies were maintained in bottles, with each containing 200-300 adult flies. Males were collected within a day of eclosion and kept separated from females for 4-5 days in vials before sperm isolation (50 males per vial).

Sperm isolation and whacking

A pair of testes with seminal vesicles were isolated from each male that had stored sperm for 4-5 days. One pair of forceps was used to hold down the seminal vesicle while the other pair was used to puncture a hole in the seminal vesicle, which caused sperm to flow out of the vesicle. The sperm were then pulled slowly and carefully out of the vesicle completely. Sperm

isolated from 500 males this way were transferred to 250µL 1X EBR containing 0.1g of Lysis Matrix B (MP Biomedicals). The tube was shaken in the Bead Mill instrument at the speed of 30Hz for 1 minute, and this was repeated four more times. The tube was briefly spun, then the content was filtered through a 50 µm cell strainer to remove the beads and large tissue clumps.

Sucrose density gradient ultracentrifugation

Sucrose density gradient was set up for sperm head enrichment. 500 males' worth of whacked sperm were first mixed with 90% sucrose solution to obtain the final concentration of 70%. Sucrose gradient consisting of 5 steps (0.5mL 90%, 1mL 80%, 2mL 70%, 0.7mL 65%, and 0.3mL 45%) was created in 5mL ultracentrifuge tubes. The tubes were spun at 100,000 x g for 4 hours at 10 degrees. 1mL of 65% was collected from the top of the tube, and a fresh 0.7mL 65% sucrose and 0.3mL 45% sucrose were overlaid. The second spin was done at 100,000 x g for 12 hours at 10 degrees. After this spin, the 80% fraction contained mostly heads while the 65/70% fractions contained mostly tails.

To remove sucrose from the head-enriched fraction, we used a GHP Nanosep 0.45-µm filtration device (Pall Corporation). The head enriched fraction was repeatedly applied to the spin column until its volume was reduced to 200 µL. Then, fresh 1X EBR was added and spun. This was repeated until the concentration of sucrose became less than 0.1%. The heads were collected in the final volume of 500 µL 1X EBR.

Protein extraction and quantification

Enriched sperm heads were pelleted and consolidated into one tube, containing 250 µL 0.1% *RapiGest* (Waters) in 50mM ammonium bicarbonate. We sonicated the sample using the

following parameters: a power of 4.5 for 10 seconds, with 30 repetitions. After each sonication step, the sample was transferred onto ice for 1 minute to prevent the overheating of the sample. BCA assay was performed to measure protein concentration as per manufacture's protocol (Pierce).

SDS-Polyacrylamide Gel Electrophoresis

To visualize the complexity of whole sperm and that of sperm heads, equal amounts of total protein (500 ng) were separated by SDS-PAGE (4% stacking and 12% resolving). The gel was stained with the Pierce Silver Stain Kit as per manufacture's protocol (Pierce). The developed gel was photographed using the Gel Doc™ EZ System (Bio-Rad).

Mass spectrometry sample preparation, data acquisition, and data analysis

We prepared 20 µg of extracted protein in 0.1% RapiGest. DTT was added to the protein sample at a final concentration of 5mM. The sample was incubated at 60 ° C for 30 minutes. After cooling down to room temperature, IAA was added to the sample at a final concentration of 15mM. After keeping the sample in dark at room temperature for 30 minutes, trypsin was added at a ratio of 1:50 enzyme:protein. After 4-hour shaking at 37 ° C, HCl was added to a final concentration of 200mM to acidify RapiGest. This was done at 37 ° C for 45 minutes. The sample at 14,000 x g at 4 °C for 10 minutes, and the supernatant was transferred to a fresh Eppendorf tube. 30 µg digested protein in acidified RapiGest was applied through a Waters (Cat#186000782) Oasis MCX (Mixed-Mode Cation Exchange) column to remove salts and neutrals. 1 µg of peptides were fractionated in an HPLC, then eluted into a Velos Pro mass spectrometer (ThermoFisher, San Jose, CA). The MS/MS data were searched using COMET

(ENG *et al.* 2013) with dynamic modification searches of 15.994915 methionine and a static modification of 57.021464 Cysteine against a fasta database containing all the protein sequences from FlyBase plus contaminant proteins.

Results

Preparation of sperm head-enriched sample

Drosophila sperm are extremely long with the tail portion making up 99% of the entire cell. To analyze protein compositions of the sperm head specifically, we first developed a method to physically disrupt the sperm cells into small fragments so that we could isolate the sperm heads from tail fragments (Figure 4.1A). Bead beating sperm with a small amount of 0.1 mm silica beads reproducibly yielded sperm fragments of various size, including sperm heads (Figure 4.1B).

Sperm heads and tails are compositionally distinct, leading us to test whether we could isolate heads from tails by sucrose density gradient ultracentrifugation. We prepared sucrose step gradients (45%, 65%, 70%, 80%, and 90%), and placed the fragmented sperm in the 70% fraction. After 4-hour spin at 100,000 x g, sperm tail fragments were found to be enriched at the boundary between 45% and 65% fractions while the heads equilibrated in the 80% sucrose solution. However, even after the first spin, the head-enriched fraction still contained many tail fragments, so we applied this fraction through another ultracentrifugation using the same gradient setup. After the second spin, the 80% fraction contained a significantly enriched amount of sperm heads relative to the tail fragments (Figure 4.1B).

To estimate the efficiency of sperm head recovery after ultracentrifugation, we spotted 1 μL of the concentrated head sample on a cellulose paper and counted the number of sperm nuclei

visualized by MST-HMG35Bb-DsRed. From one batch corresponding to 1,000 males' worth of sperm, we routinely recovered at least 1,000,000 sperm heads. Assuming that each male provides ~4,000 sperm, our recovery efficiency was only ~25%. The loss could be anywhere during the process, but one possibility is that sperm whacking may have been incomplete such that heads were still attached to tails. Also, sperm heads may have been aggregated during ultracentrifugation, and the spot assay using only 1 μL may have led to the underestimation of the actual recovery because 1 μL was only 0.4% of the recovered heads.

Sperm head enrichment assessed by SDS-PAGE

The counting of the number of heads compared to tail fragments per field of view indicated that the heads were enriched by 180-fold. To assess this enrichment of sperm heads molecularly, we compared protein extracts prepared from whole sperm and those from head-enriched samples. We separated 500 ng of total protein by SDS-PAGE, which were visualized with silver stain. The resulting gel revealed two distinct banding patterns between the whole sperm and head-enriched samples (Figure 4.1C). Noticeably, the head-enriched samples showed much fewer bands, indicating that the composition of *Drosophila* sperm head is much less complex than that of the entire cell.

Mass spectrometry-based identification of sperm head proteins

We combined several rounds of sperm head preparations to obtain a total of ~20 μg protein in a RapiGest-based lysis buffer. We chose RapiGest as a detergent because of its compatibility with mass spectrometry analysis as opposed to incompatible ones such as SDS or Triton X-100. The extracted proteins were digested with trypsin, and the resulting peptides were

analyzed by tandem mass spectrometry in the data-dependent acquisition mode in two technical replicates.

Mass spectrometry analysis deduces protein identities by sequencing peptides derived from protease treatment, and sometimes it is impossible to deduce the exact identities because multiple proteins can give rise to identical peptides. We therefore report in this study "protein groups" that our analysis yielded, and only refer to specific protein identities when they were identified unambiguously. The two runs yielded 556 and 564 distinct protein groups. 498 protein groups were identified in both runs. Combining the two runs, we discovered a total of 588 protein groups. These results were compared to the 1108 proteins that had been reported previously in the two *Drosophila* sperm proteome studies (DmSP-I and DmSP-II). Of the 588 proteins, 282 were previously undiscovered proteins (Table 4.1).

Of the previously undiscovered proteins, we manually removed from the list the ones that we predicted to be cytoskeletal or mitochondrial proteins (i.e. tail contaminants) based on their predicted molecular functions or subcellular localizations. 18 proteins were predicted to have cytoskeletal roles and 21 were predicted to localize in mitochondria, reducing the number on the list to 243 proteins (Table 4.1).

Although the majority of the previously reported sperm proteome is likely contributed by tail proteins, at least one of these proteins (MST-HMG35B) is a known sperm nuclear-specific protein. MST-HMG35B was identified in DmSP-I in which one unique peptide (QGPVTNNAYLNFVR) was identified in all three replicate experiments. In our analysis, we also identified MST-HMG35B, but with more unique peptide fragments derived from the same protein (QGPVTNNAYLNFVR, KHCDLKPQELIAEAAK, GLTEMCNHPK, AWAELPEHR). Therefore, we looked for proteins for which we identified more unique peptide fragments in this

study than the previous two studies. Because DmSP-I and DmSP-II were each performed in triplicates, we prioritized this manual search by focusing on the proteins that had been only identified in one of the triplicates. Of the 1108 proteins, 152 (DmSP-I) and 448 (DmSP-II) proteins were only identified in one of the triplicate experiments. We identified 81 of these proteins in our analysis, and 36 of them yielded more unique peptides than the previous studies. After removing 8 proteins because of their predicted cytoskeletal or mitochondrial roles, we designate the rest of the proteins to be included in the head-enriched proteome.

Compared to the previously reported DmSP-I and II, we identified more proteins that are known to be present in the sperm nucleus (Table 4.2). However, because the DDA analysis failed to discover any known acrosomal proteins (Table 4.2), we postulated that this was due to the relatively low abundance of these proteins in the sperm head. Therefore, we analyzed the head-enriched sample by mass spectrometry in the data-independent acquisition (DIA) mode. Unlike DDA, DIA theoretically surveys all peptides in a given window of mass-to-charge ratios (AEBERSOLD AND MANN 2016), enabling the detection and sequencing of low-abundance peptides. Of the 5699 peptides that we analyzed in DIA, there was indeed a peptide that uniquely derived from a known acrosomal protein, Sneaky (Table 4.2). The list of the 706 proteins were filtered in a similar way to the DDA analysis to obtain 117 proteins that are likely in the sperm head. Combining DDA and DIA results, we here report a list of 388 proteins that we predict to be enriched, if not exclusively found, in *Drosophila* sperm heads. This sperm head-enriched proteome is hereafter referred to as *Drosophila melanogaster* Sperm Head Enriched Proteome (DmSHEP).

Categorization of Drosophila sperm head proteome

To gain insights into the composition of DmSHEP, we first created two categories to sort the proteins based on their predicted subcellular localizations: nucleus and membrane. To predict nuclear localization, we used known or predicted molecular structures or functions of the proteins. In this way, we designated 33 proteins to the nuclear category (Table 4.3). Notably, known MST-HMG box proteins were identified (MST-HMG35Ba/b, MST-HMG77F, MST-HMG99C, and MST-HMG77Y), three of which were previously undiscovered by mass spectrometry. Several protein groups that correspond to Histones H2A, H2B, H3, and H4 were identified, but there was no detection of H1 or centromere-specific H3 variant. In addition to SNBPs, we identified several transcription factors, general chromatin architectural proteins, and chromatin remodeling enzymes. Surprisingly, we also discovered Vrs, which was previously reported by cytology to appear in spermatid nuclei but to be absent in mature sperm (BINDER *et al.* 2017), suggesting that the amount of Vrs present in mature sperm may be below the detection limit by fluorescence microscopy.

In addition to the 33 proteins, we found 14 previously uncharacterized proteins that are small (less than 400 amino acids) and highly basic (pI > 10), a general characteristic of known *Drosophila* SNBPs. Because *Drosophila* SNBPs are also predicted to form multiple alpha helices, we looked among these small basic proteins for ones with the capability to form alpha helices. We found three such proteins (CG14658, CG30039, and CG31245). CG14658 and CG30039 have no functions or predicted domains associated with them but are predicted to form several alpha helices and have clusters of basic amino acids. CG31245 has two copies of a domain known as the testicular haploid expressed repeat (THER). THER-containing proteins are

found in round spermatid nuclei of mouse sperm, suggesting that CG31245 has the potential to function in sperm nuclei of *Drosophila*.

We used the Phobius program to predict the presence of signal sequence and transmembrane domains. We discovered a total of 57 proteins with predicted signal sequence or at least one transmembrane domain (Table 4.3), an indication that these proteins function in a membrane compartment whether plasma membrane or acrosomal membrane. Most of these membrane proteins have no predicted molecular functions. The most common molecular features predicted were enzymatic activities (10), transporters (8), and calcium binding (5). A known acrosomal membrane protein Sneaky was successfully identified in the DIA analysis, were not identified, suggesting that either acrosomal membrane proteins are much less abundant or harder to extract than plasma membrane proteins.

While the majority of the rest of the proteins cannot be predicted in terms of their locations or functions, we see various proteins involved in enzymatic processes, including ubiquitin tags as well as several subunits of the proteasome (Table 4.3). We also see five proteins likely functioning at centrioles, including two known spermatid centriolar proteins (Spag4 and Yuri).

Taken together, we conclude that our approach successfully identified known sperm head-specific proteins that previous studies failed to detect. However, our analysis still failed to detect several proteins that are known to be present in the nucleus (MST-HMG Ddbt, K81, and Cid) or the Misfire protein in the acrosome, suggesting that either almost complete removal of tails may be needed to increase the chance of detecting peptides derived from low-abundance head-specific proteins or differential extraction may be needed to analyze different compartments separately.

Rapid evolution of DmSHEP proteins among species

In *Drosophila melanogaster,* 15063 proteins are predicted to be expressed, 54% of which are conserved in humans. In contrast, DmSP-I and DmSP-II are significantly enriched with conserved proteins, with 63% of them having human versions ($p<0.0001$). Interestingly, only 42% of the proteins reported in DmSHEP have human orthologs, and this under-enrichment is statistically significant compared to 54% conservation observed at the whole-fly level. We asked whether the rest of the proteins show any signatures of rapid evolution under positive selection. To assess for positive selection, we used the ratios of non-synonymous to synonymous substitutions (d_N/d_S) as an indicator. A previous study calculated the d_N/d_S for 10,758 protein-coding genes between *D. melanogaster* and *D. simulans*, and the genome-wide average d_N/d_S was reported to be 0.15 (STANLEY AND KULATHINAL 2016). Taking advantage of this already available dataset, we obtained the d_N/d_S for 98 genes that apparently lack human orthologs. The average d_N/d_S among this set of genes was computed to be 0.23, which is slightly higher than the genome-wide average. Found among the top of the list were *vrs*, *Mst-Hmg77F,* and *Mst-Hmg99C,* all of which encode proteins that enter spermatid nuclei during nuclear transformation. Also found above the whole-genome average were previously mentioned potential SNBP-encoding genes, CG30039 ($d_N/d_S = 0.63$) and CG31245 ($d_N/d_S = 0.29$), and membrane protein-encoding genes, CG14841 ($d_N/d_S = 0.97$) and CG32450 ($d_N/d_S = 0.55$). Thus, the combination of molecular evolutionary analysis with detailed sequence analysis will aid in prioritizing functional characterizations of proteins that have no orthologs in other model organisms and whose functions have yet to be analyzed.

Discovery of a new member of the MST-HMG box protein family

Most *Drosophila* SNBPs identified to date contain a conserved domain called the Male-Specific Transcript High-Mobility Group (MST-HMG) box (DOYEN *et al.* 2015). HMG box proteins are found in all eukaryotic branches and have various functions from chromosome architecture to transcription regulation (MALARKEY AND CHURCHILL 2012). HMG box is made up of three alpha helices that fold into a characteristic L shape. MST-HMG box is most notably distinct from canonical HMG box because the first helix is shorter and contains a characteristic amino acid sequence, NFLR. In addition, as Doyen et al described, there are four aromatic amino acids whose positions are conserved within the MST-HMG box (DOYEN *et al.* 2015).

Of the nuclear proteins reported in DmSHEP, we discovered one HMG box protein, CG31010, and thus chose it for further in-depth comparative analysis. To see if the HMG box of CG31010 is the canonical one or the MST-HMG box, we aligned the predicted HMG box region of CG31010 against the MST-HMG box domain of several MST-HMG box proteins and the canonical HMG box domain of *Drosophila melanogaster* HMG-box proteins. Interestingly, CG31010's HMG box was computed to be more similar to MST-HMG box than the canonical HMG box (Figure 4.2A), suggesting that CG31010 is a new member of the MST-HMG box protein family. Following the nomenclature as described in our previous study, *CG31010* has been renamed *Mst-Hmg100A.*

The *Mst-Hmg100A* mRNA, for which there is cDNA evidence, encodes the protein containing 254 amino acids (Figure 4.2B). Surprisingly, MST-HMG100A is highly acidic compared to all the other MST-HMG box proteins identified to date, which are all highly basic. In all the MST-HMG box proteins identified to date, C-terminal regions exhibit tremendous variation in sequence and length but contain clusters of basic residues. In contrast to this, the

regions flanking the MST-HMG box of MST-HMG100A are enriched in acidic residues (N-terminal domain) and lack basic residue clusters (C terminal domain). The *Mst-Hmg100A* gene and its orthologs are present as a single copy gene in all the 12 *Drosophila* species. The orthologs are also found in the Tephritidae family species (e.g. *Bactrocera dorsalis* and *Rhagoletis zephyria*) (Figure 4.2C), indicating that the gene arose 65 million years ago. Protein lengths are highly variable among species because of variable lengths of the N-terminal domains. Taken together, we conclude that *Mst-Hmg100A* likely evolved from an *Mst-Hmg* box gene in the fly lineages specifically and that the flanking regions of the MST-HMG box have undergone significant changes to transform the ancestral protein into a highly acidic one, the only member of the MST-HMG box family with such drastic changes in amino acid compositions.

Discussion

In this study, we developed a method to isolate sperm heads from tails. By performing mass spectrometry analysis of the head-enriched peptides in the DDA mode, we discovered 271 proteins that are likely present in the *Drosophila* sperm head. We successfully identified known SNBPs such as MST-HMG77F and MST-HMG99C that previous mass spectrometry studies had failed to discover. This indicates that sperm head proteins were significantly enriched enough for peptide identification in DDA, which is inherently biased towards abundant peptides for sequencing (AEBERSOLD AND MANN 2016). Furthermore, by performing the less biased approach of DIA, we identified additional 117 head-enriched proteins, including two known acrosomal membrane proteins. The combined list of 388 proteins constitute the first report of DmSHEP.

The majority of *Drosophila* SNBP-encoding genes identified to date form the *Mst-Hmg box* gene family. In our previous study, we described a likely scenario as to how this gene family

originated from a canonical HMG box protein-encoding gene. Briefly, the precursor gene acquired a male-enriched expression and underwent changes that resulted in the truncation of the HMG box. The resulting box is characteristically shorter in the first helix. This helix is also of great importance in identification of MST-HMG box proteins because it contains the NFLR motif, which is not found in any of the 24 canonical HMG box proteins in *Drosophila melanogaster.* Importantly, the gene family is highly dynamic in *Drosophila* lineages, with increased copy number or loss of individual family members. Our proteomic analysis identified a new family member, MST-HMG100A, that shows unique features. Unlike all the other members, this protein lacks basic residue clusters and is highly acidic. This surprising finding expands the *Mst-Hmg box* gene family to the 13 distinct members. This also provides support to our previously proposed model of MST-HMG box protein evolution with dynamic and recurring C terminal divergence among family members.

DmSHEP also provides the first direct evidence in support of previous studies suggesting the presence of functional Y-linked copies of the *Drosophila melanogaster Mst-Hmg77F* gene (*Mst-Hmg77Y*) (KRSTICEVIC *et al.* 2015). These studies combined DNA and mRNA sequencing to reveal that there are 18 copies of *Mst-Hmg77Y* in *Drosophila melanogaster* (KRSTICEVIC et al. 2015). When combining the autosomal and Y-liked copies, the contribution of the Y-linked copies is estimated to be ~20% at the mRNA level (KRSTICEVIC *et al.* 2015). In our analysis, we discovered one unique peptide that belongs to the MST-HMG77Y protein group. Comparing the spectra counts between MST-HMG77F and MST-HMG77Y, we see significant enrichment of the autosomal version over the Y-liked copies (35 counts vs. 2 counts). Our analysis is not considered quantitative, but it is consistent with the previous analysis comparing the levels of mRNA contributed by the autosomal gene compared to the Y-linked copies.

The overall composition of the *Drosophila* sperm nucleus cannot be conclusively determined from this study because our analysis failed to discover some known head-specific proteins. Yet, the proteins we discovered suggest an interesting parallel between insect and mammalian sperm nuclei. In addition to sperm-specific nuclear proteins, we identified histones, transcription factors, and chromatin architectural proteins on our list, all of which have also been observed in mammalian sperm (CASTILLO *et al.* 2014). Retained histones in mature sperm are enriched at specific loci such developmentally important genes (e.g. *Hox* gene cluster) (HAMMOUD *et al.* 2009). The presence of histones in *Drosophila* sperm *per se* does not imply their analogous roles in potentially regulating gene expression in the early embryo. Thus, it will be of great interest to map these proteins molecularly in sperm nuclei by using methods such as ChIP-seq. The presence of transcription factors in the sperm nucleus also potentially impacts gene regulation in the early embryo. However, because these proteins can also have spermatogenic roles, it is difficult to functionally dissect and separate their spermatogenic and post-fertilization roles. In *Drosophila*, a protein degradation tool, deGradFP, has been developed to target GFP-tagged proteins for degradation (CAUSSINUS *et al.* 2011). This tool has been used to eliminate the centromeric histone variant post-meiotically to study its impact on paternal chromosome segregation in the early embryo (RAYCHAUDHURI *et al.* 2012). Therefore, deGradFP may be applied to analyze the consequences on gene regulation and development of the removal of transcription factors that are normally retained in sperm nuclei.

The nuclear proteomes of human and mouse sperm indicate a remarkably heterogenous mixture of proteins in the sperm nucleus considering how little activity there is in the nucleus of this cell. In particular, the human and mouse sperm nuclei have been reported to contain hundreds of proteins (CASTILLO *et al.* 2014). In contrast, DmSHEP only reports a few dozen

nuclear proteins. However, this does not necessarily mean that the *Drosophila* sperm nucleus is much less complex than that of mammalian sperm and may simply be due to the difficulty of extracting nuclear proteins. Previous proteomic research using cell culture and larval samples identified ~60% of the predicted *Drosophila* proteome (BRUNNER *et al.* 2007). Interestingly, these authors noted that small proteins and highly basic proteins are more under-represented. To identify more of these proteins, they used techniques such as gel-filtration to enrich for small proteins and free-flow electrophoresis to enrich for basic proteins. Therefore, using a method that is more tailored towards the extraction of specific protein types (e.g. SNBPs, nucleoplasm proteins, etc.) should unveil the actual complexity of the *Drosophila* sperm nucleus.

In conclusion, DmSHEP has made an important step towards the understanding of the complexity of the *Drosophila* sperm head. The suspiring discovery of MST-HMG100A, a highly acidic member of the MST-HMG box family, shows a much more heterogenous nature and dynamic evolution of this protein family. It also demonstrates the power of the proteomics-based identification of proteins that would otherwise be difficult, if not impossible, to discover simply based on comparative sequence analysis. Also, there is accumulating evidence that sperm head proteins are not only used for spermatogenesis and fertilization, but a subset of them are required for the successful initiation of embryogenesis. Thus, the comprehensive identification and functional characterization of the sperm head proteome will shed light on how sperm proteins may impact the earliest stages of embryonic development.

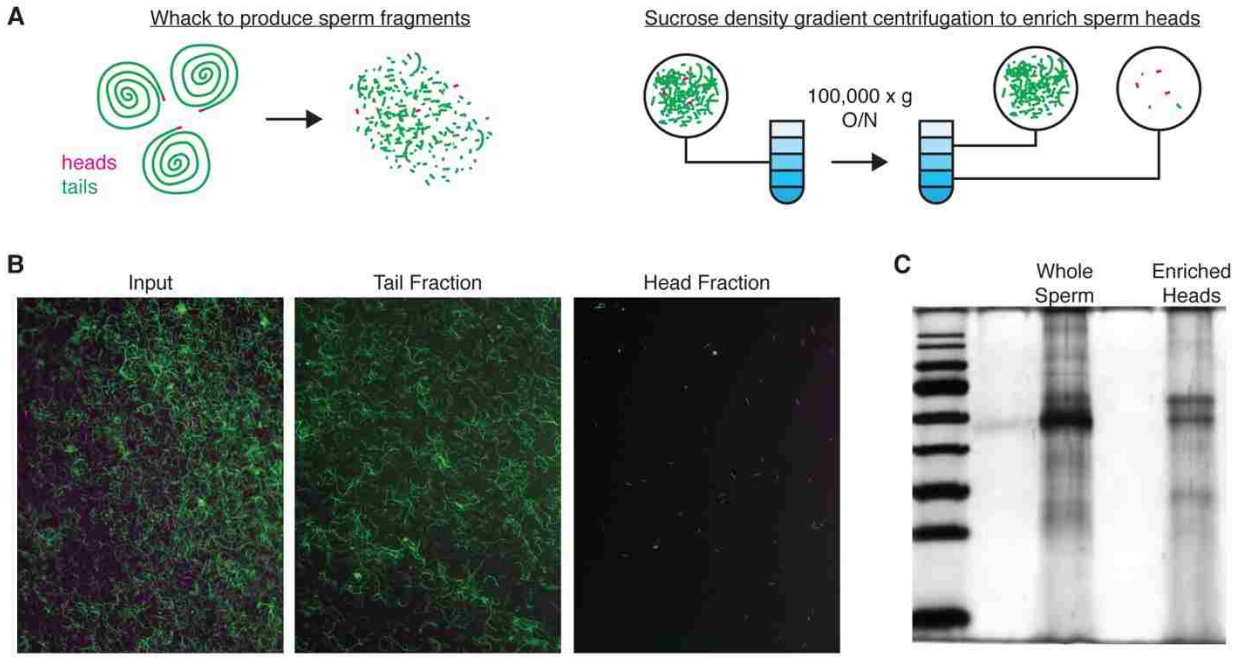


Figure 4.1 Identification of *Drosophila* sperm head-enriched proteins.
(A) Flow diagram showing sperm head enrichment protocol. (B) Images show sperm heads (magenta) and tails (green) before and after sucrose gradient ultracentrifugation. (C) Silver stained gel showing whole sperm extracts compared to sperm head-enriched extracts. 500 ng of total protein was loaded in each lane.

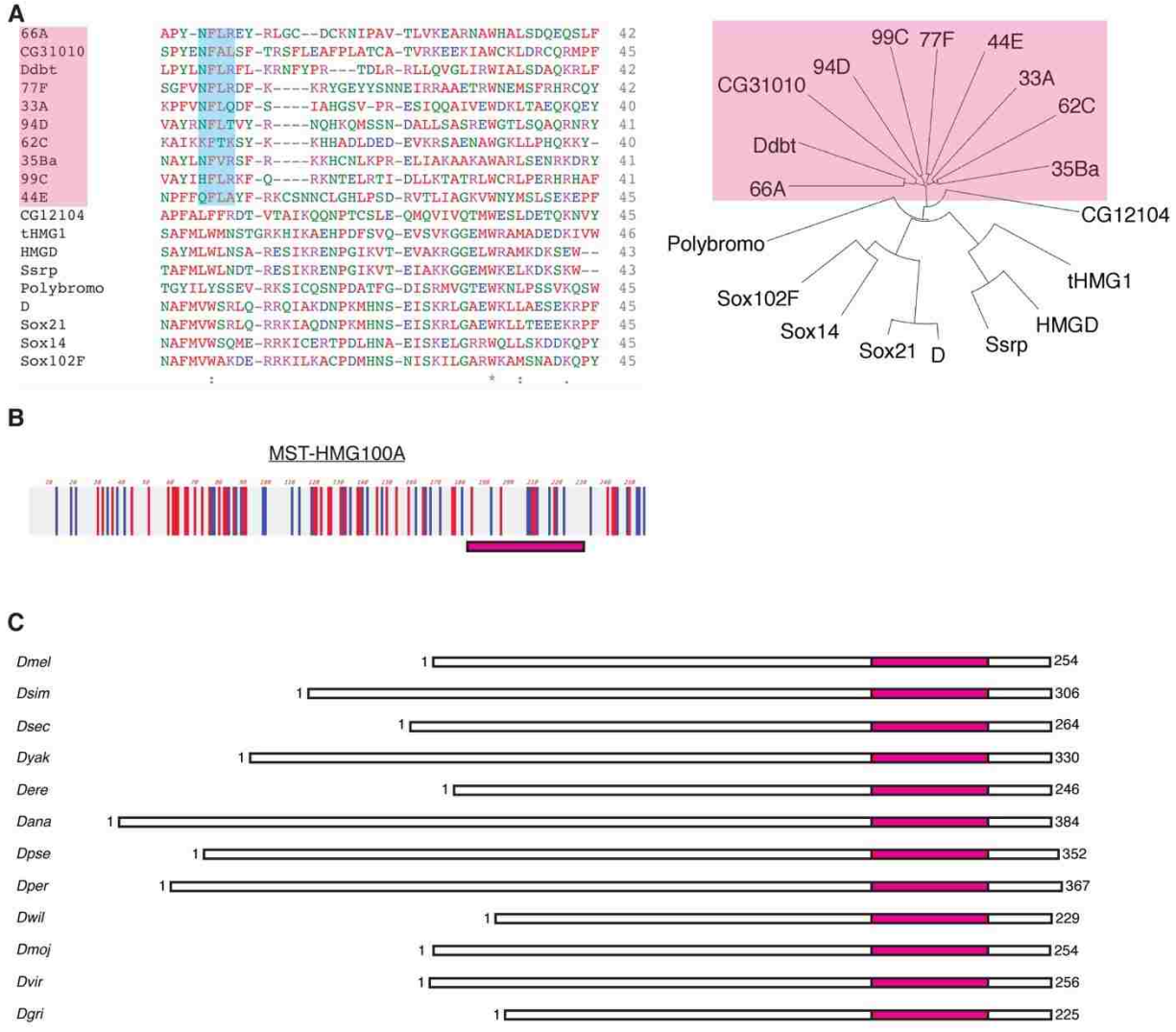


Figure 4.2 Discovery of MST-HMG100A (CG31010).
(A) Alignment of the predicted HMG box domain of CG31010 against the MST-HMG box of all members of MST-HMG box family (highlighted in pink) and the HMG box of several *Drosophila* HMG box proteins. The tree on the right is based on the alignment on the left. NFLR motif of MST-HMG box proteins is highlighted in blue. Note CG31010 falls into the same group as the MST-HMG box family. (B) Predicted MST-HMG100A showing acidic amino acids (red) and basic amino acids (blue). The MST-HMG box is indicated by the pink box. (C) MST-HMG100A compared to its orthologous proteins in other *Drosophila* species.

Table 4.1 Summary of mass spectrometry results

		Tail contaminants		
	previously unreported proteins	cytoskeletal proteins[a]	mitochondrial proteins[a]	proteins likely enriched in sperm head
DDA				
A total of 588 proteins	282	18	21	243
Previously reported but more enriched[b] compared to DmSP-I or II	N/A	2	6	28
DIA				
A total of 706 proteins	164	11	36	117

DmSHEP = 388 proteins

[a]based on predicted subcellular localizations or molecular functions
[b]see the text for what qualifies as "more enriched"

Table 4.2 Discovery of known head-specific proteins by mass spectrometry

	Nucleus (8 proteins)	Acrosome (2 proteins)	Plasma membrane (1 protein)	Centriole (4 proteins)	Enzyme (1 protein)
DmSP-I and II					
1108 proteins	1	0	0	0	0
DDA					
588 proteins	3	0	0	1	0
DIA					
706 proteins	3	1	0	1	0

Table 4.3 Protein categorization based on predicted structure or function

	Nucleus	Membrane	Centriole	Enzymes[a]	Unknown
271 proteins from DDA	24	55	4	52	136
117 proteins from DIA	9	At least 2[b]	1	19	Up to 86

[a]nuclear, membrane, and centriolar enzymes are excluded

[b]including two acrosomal proteins

Literature Cited

AEBERSOLD, R., AND M. MANN, 2016 Mass-spectrometric exploration of proteome structure and function. Nature 537**:** 347-355.

AMANN, R. P., 2008 The cycle of the seminiferous epithelium in humans: a need to revisit? J Androl 29**:** 469-487.

AMARAL, A., J. CASTILLO, J. M. ESTANYOL, J. L. BALLESCA, J. RAMALHO-SANTOS *et al.*, 2013 Human sperm tail proteome suggests new endogenous metabolic pathways. Mol Cell Proteomics 12**:** 330-342.

ARPANAHI, A., M. BRINKWORTH, D. ILES, S. A. KRAWETZ, A. PARADOWSKA *et al.*, 2009 Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. Genome Res 19**:** 1338-1349.

BAKER, J., W. E. THEURKAUF AND G. SCHUBIGER, 1993 Dynamic changes in microtubule configuration correlate with nuclear migration in the preblastoderm Drosophila embryo. J Cell Biol 122**:** 113-121.

BALHORN, R., 2007 The protamine family of sperm nuclear proteins. Genome Biol 8**:** 227.

BARRAL, S., Y. MOROZUMI, H. TANAKA, E. MONTELLIER, J. GOVIN *et al.*, 2017 Histone Variant H2A.L.2 Guides Transition Protein-Dependent Protamine Assembly in Male Germ Cells. Mol Cell 66**:** 89-101 e108.

BECK, M., AND E. HURT, 2017 The nuclear pore complex: understanding its function through structural insight. Nat Rev Mol Cell Biol 18**:** 73-89.

BELLEN, H. J., R. W. LEVIS, Y. HE, J. W. CARLSON, M. EVANS-HOLM *et al.*, 2011 The Drosophila gene disruption project: progress using transposons with distinctive site specificities. Genetics 188**:** 731-743.

BINDER, A. M., B. T. WAKIMOTO, C. DAVIS, J. CHMIELEWSKI AND J. E. TOMKIEL DEAN, 2017 Versager is expressed at the histone-to-protamine transition during spermiogenesis and is required for embryonic chromosome transmission in Drosophila melanogaster. Research Journal of Developmental Biology 4**:** 1.

BLACK, J. A., AND G. H. DIXON, 1967 Evolution of protamine: a further example of partial gene duplication. Nature 216**:** 152-154.

BLOCH, D. P., 1969 A catalog of sperm histones. Genetics 61**:** Suppl:93-111.

BRODSKY, M. H., B. T. WEINERT, G. TSANG, Y. S. RONG, N. M. MCGINNIS *et al.*, 2004 Drosophila melanogaster MNK/Chk2 and p53 regulate multiple DNA repair and apoptotic pathways following DNA damage. Mol Cell Biol 24**:** 1219-1231.

BRUNK, K., B. VERNAY, E. GRIFFITH, N. L. REYNOLDS, D. STRUTT *et al.*, 2007 Microcephalin coordinates mitosis in the syncytial Drosophila embryo. J Cell Sci 120**:** 3578-3588.

BRUNNER, E., C. H. AHRENS, S. MOHANTY, H. BAETSCHMANN, S. LOEVENICH *et al.*, 2007 A high-quality catalog of the Drosophila melanogaster proteome. Nat Biotechnol 25**:** 576-583.

BUCHAN, D. W., F. MINNECI, T. C. NUGENT, K. BRYSON AND D. T. JONES, 2013 Scalable web services for the PSIPRED Protein Analysis Workbench. Nucleic Acids Res 41**:** W349-357.

CARRELL, D. T., AND S. S. HAMMOUD, 2010 The human sperm epigenome and its potential role in embryonic development. Mol Hum Reprod 16**:** 37-47.

CASTILLO, J., A. AMARAL AND R. OLIVA, 2014 Sperm nuclear proteome and its epigenetic potential. Andrology 2**:** 326-338.

CAUSSINUS, E., O. KANCA AND M. AFFOLTER, 2011 Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol 19**:** 117-121.

CENCI, G., L. CIAPPONI AND M. GATTI, 2005 The mechanism of telomere protection: a comparison between Drosophila and humans. Chromosoma 114**:** 135-145.

CENCI, G., R. B. RAWSON, G. BELLONI, D. H. CASTRILLON, M. TUDOR *et al.*, 1997 UbcD1, a Drosophila ubiquitin-conjugating enzyme required for proper telomere behavior. Genes Dev 11**:** 863-875.

CHAMPROUX, A., J. TORRES-CARREIRA, P. GHARAGOZLOO, J. R. DREVET AND A. KOCER, 2016 Mammalian sperm nuclear organization: resiliencies and vulnerabilities. Basic Clin Androl 26**:** 17.

CHO, C., H. JUNG-HA, W. D. WILLIS, E. H. GOULDING, P. STEIN *et al.*, 2003 Protamine 2 deficiency leads to sperm DNA damage and embryo death in mice. Biol Reprod 69**:** 211-217.

CHO, C., W. D. WILLIS, E. H. GOULDING, H. JUNG-HA, Y. C. CHOI *et al.*, 2001 Haploinsufficiency of protamine-1 or -2 causes infertility in mice. Nat Genet 28**:** 82-86.

COUNCE, S. J., 1973 The causal analysis of insect embryogenesis, pp. 1-156 in *Developmental Systems: Insects*, edited by S. J. Counce and C. H. Waddington. Academic Press, New York.

CROOKS, G. E., G. HON, J. M. CHANDONIA AND S. E. BRENNER, 2004 WebLogo: a sequence logo generator. Genome Res 14**:** 1188-1190.

DESAI, B. S., S. SHIROLIKAR AND K. RAY, 2009 F-actin-based extensions of the head cyst cell adhere to the maturing spermatids to maintain them in a tight bundle and prevent their premature release in Drosophila testis. BMC Biol 7**:** 19.

DORUS, S., S. A. BUSBY, U. GERIKE, J. SHABANOWITZ, D. F. HUNT *et al.*, 2006 Genomic and functional evolution of the Drosophila melanogaster sperm proteome. Nat Genet 38**:** 1440-1445.

DORUS, S., Z. N. FREEMAN, E. R. PARKER, B. D. HEATH AND T. L. KARR, 2008 Recent origins of sperm genes in Drosophila. Mol Biol Evol 25**:** 2157-2166.

DOS SANTOS, G., A. J. SCHROEDER, J. L. GOODMAN, V. B. STRELETS, M. A. CROSBY *et al.*, 2015 FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res 43**:** D690-697.

DOYEN, C. M., G. E. CHALKLEY, O. VOETS, K. BEZSTAROSTI, J. A. DEMMERS *et al.*, 2015 A Testis-Specific Chaperone and the Chromatin Remodeler ISWI Mediate Repackaging of the Paternal Genome. Cell Rep 13**:** 1310-1318.

DROSOPHILA 12 GENOMES, C., A. G. CLARK, M. B. EISEN, D. R. SMITH, C. M. BERGMAN *et al.*, 2007 Evolution of genes and genomes on the Drosophila phylogeny. Nature 450**:** 203-218.

DUBEY, P., S. SHIROLIKAR AND K. RAY, 2016 Localized, Reactive F-Actin Dynamics Prevents Abnormal Somatic Cell Penetration by Mature Spermatids. Dev Cell 38**:** 507-521.

DUBRUILLE, R., AND B. LOPPIN, 2015 Protection of Drosophila chromosome ends through minimal telomere capping. J Cell Sci 128**:** 1969-1981.

DUBRUILLE, R., G. A. ORSI, L. DELABAERE, E. CORTIER, P. COUBLE *et al.*, 2010 Specialization of a Drosophila capping protein essential for the protection of sperm telomeres. Curr Biol 20**:** 2090-2099.

EIRIN-LOPEZ, J. M., AND J. AUSIO, 2009 Origin and evolution of chromosomal sperm proteins. Bioessays 31**:** 1062-1070.

EISMAN, R. C., AND T. C. KAUFMAN, 2013 Probing the boundaries of orthology: the unanticipated rapid evolution of Drosophila centrosomin. Genetics 194**:** 903-926.

EMELYANOV, A. V., J. RABBANI, M. MEHTA, E. VERSHILOVA, M. C. KEOGH *et al.*, 2014 Drosophila TAP/p32 is a core histone chaperone that cooperates with NAP-1, NLP, and nucleophosmin in sperm chromatin remodeling during fertilization. Genes Dev 28**:** 2027-2040.

ENG, J. K., T. A. JAHAN AND M. R. HOOPMANN, 2013 Comet: an open-source MS/MS sequence database search tool. Proteomics 13**:** 22-24.

EREN-GHIANI, Z., C. RATHKE, I. THEOFEL AND R. RENKAWITZ-POHL, 2015 Prtl99C Acts Together with Protamines and Safeguards Male Fertility in Drosophila. Cell Rep 13**:** 2327-2335.

FABBRETTI, F., I. IANNETTI, L. GUGLIELMI, S. PERCONTI, C. EVANGELISTELLA *et al.*, 2016 Confocal Analysis of Nuclear Lamina Behavior during Male Meiosis and Spermatogenesis in Drosophila melanogaster. PLoS One 11**:** e0151231.

FARRELL, J. A., AND P. H. O'FARRELL, 2014 From egg to gastrula: how the cell cycle is remodeled during the Drosophila mid-blastula transition. Annu Rev Genet 48**:** 269-294.

FOE, V. E., G. M. ODELL AND B. A. EDGAR, 1993 Mitosis and morphogenesis in the Drosophila embryo: Point and counterpoint, pp. 149-300 in *The Development of Drosophila melanogaster*, edited by M. Bate and A. M. Arias. Cold Spring Harbor Laboratory Press, New York.

FOGARTY, P., S. D. CAMPBELL, R. ABU-SHUMAYS, B. S. PHALLE, K. R. YU *et al.*, 1997 The Drosophila grapes gene is related to checkpoint gene chk1/rad27 and is required for late syncytial division fidelity. Curr Biol 7**:** 418-426.

FUYAMA, Y., 1984 Gynogenesis in Drosophila melanogaster. Jpn. J. Genet 59**:** 91–96.

FUYAMA, Y., R. W. HARDY AND D. L. LINDSLEY, 1998 Parthenogenesis in *Drosophila melanogaster*. The mechanisms of gynogenesis induced by *ms(3)K81* sperm. Jpn J Genet 63**:** 553-554.

GAO, G., X. BI, J. CHEN, D. SRIKANTA AND Y. S. RONG, 2009 Mre11-Rad50-Nbs complex is required to cap telomeres during Drosophila embryogenesis. Proc Natl Acad Sci U S A 106**:** 10728-10733.

GAO, G., Y. CHENG, N. WESOLOWSKA AND Y. S. RONG, 2011 Paternal imprint essential for the inheritance of telomere identity in Drosophila. Proc Natl Acad Sci U S A 108**:** 4932-4937.

GAO, G., J. C. WALSER, M. L. BEAUCHER, P. MORCIANO, N. WESOLOWSKA *et al.*, 2010 HipHop interacts with HOAP and HP1 to protect Drosophila telomeres in a sequence-independent manner. EMBO J 29**:** 819-829.

GARTNER, S. M., S. ROTHENBUSCH, M. K. BUXA, I. THEOFEL, R. RENKAWITZ *et al.*, 2015 The HMG-box-containing proteins tHMG-1 and tHMG-2 interact during the histone-to-protamine transition in Drosophila spermatogenesis. Eur J Cell Biol 94**:** 46-59.

GEORGADAKI, K., N. KHOURY, D. A. SPANDIDOS AND V. ZOUMPOURLIS, 2016 The molecular basis of fertilization (Review). Int J Mol Med 38**:** 979-986.

GRAMATES, L. S., S. J. MARYGOLD, G. D. SANTOS, J. M. URBANO, G. ANTONAZZO *et al.*, 2017 FlyBase at 25: looking to the future. Nucleic Acids Res 45**:** D663-D671.

GRAVELEY, B. R., A. N. BROOKS, J. W. CARLSON, M. O. DUFF, J. M. LANDOLIN *et al.*, 2011 The developmental transcriptome of Drosophila melanogaster. Nature 471**:** 473-479.

GROTH, A. C., M. FISH, R. NUSSE AND M. P. CALOS, 2004 Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. Genetics 166**:** 1775-1782.

HAMMOUD, S. S., D. A. NIX, H. ZHANG, J. PURWAR, D. T. CARRELL *et al.*, 2009 Distinctive chromatin in human sperm packages genes for embryo development. Nature 460**:** 473-478.

HARLEY, V. R., S. LAYFIELD, C. L. MITCHELL, J. K. FORWOOD, A. P. JOHN *et al.*, 2003 Defective importin beta recognition and nuclear import of the sex-determining factor SRY are associated with XY sex-reversing mutations. Proc Natl Acad Sci U S A 100**:** 7045-7050.

HARTWELL, L. H., AND T. A. WEINERT, 1989 Checkpoints: controls that ensure the order of cell cycle events. Science 246**:** 629-634.

HIROHASHI, N., N. KAMEI, H. KUBO, H. SAWADA, M. MATSUMOTO *et al.*, 2008 Egg and sperm recognition systems during fertilization. Dev Growth Differ 50 Suppl 1**:** S221-238.

IAMPIETRO, C., J. BERGALET, X. WANG, N. A. CODY, A. CHIN *et al.*, 2014 Developmentally regulated elimination of damaged nuclei involves a Chk2-dependent mechanism of mRNA nuclear retention. Dev Cell 29**:** 468-481.

INABA, K., 2007 Molecular basis of sperm flagellar axonemes: structural and evolutionary aspects. Ann N Y Acad Sci 1101**:** 506-526.

JANKOWSKI, J. M., J. C. STATES AND G. H. DIXON, 1986 Evidence of sequences resembling avian retrovirus long terminal repeats flanking the trout protamine gene. J Mol Evol 23**:** 1-10.

JAYARAMAIAH RAJA, S., AND R. RENKAWITZ-POHL, 2005 Replacement by Drosophila melanogaster protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus. Mol Cell Biol 25**:** 6165-6177.

KAUL, Z., A. J. CESARE, L. I. HUSCHTSCHA, A. A. NEUMANN AND R. R. REDDEL, 2011 Five dysfunctional telomeres predict onset of senescence in human cells. EMBO Rep 13**:** 52-59.

KE, M. T., S. FUJIMOTO AND T. IMAI, 2013 SeeDB: a simple and morphology-preserving optical clearing agent for neuronal circuit reconstruction. Nat Neurosci 16**:** 1154-1161.

KELLEY, L. A., S. MEZULIS, C. M. YATES, M. N. WASS AND M. J. STERNBERG, 2015 The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10**:** 845-858.

KIERSZENBAUM, A. L., AND L. L. TRES, 1975 Structural and transcriptional features of the mouse spermatid genome. J Cell Biol 65**:** 258-270.

KIMURA, S., AND B. LOPPIN, 2016 The Drosophila chromosomal protein Mst77F is processed to generate an essential component of mature sperm chromatin. Open Biol 6.

KLATTENHOFF, C., H. XI, C. LI, S. LEE, J. XU *et al.*, 2009 The Drosophila HP1 homolog Rhino is required for transposon silencing and piRNA production by dual-strand clusters. Cell 138**:** 1137-1149.

KOST, N., S. KAISER, Y. OSTWAL, D. RIEDEL, A. STUTZER *et al.*, 2015 Multimerization of Drosophila sperm protein Mst77F causes a unique condensed chromatin structure. Nucleic Acids Res 43**:** 3033-3045.

KRSTICEVIC, F. J., C. G. SCHRAGO AND A. B. CARVALHO, 2015 Long-Read Single Molecule Sequencing to Resolve Tandem Gene Copies: The Mst77Y Region on the Drosophila melanogaster Y Chromosome. G3 (Bethesda) 5**:** 1145-1150.

LAKE, C. M., J. K. HOLSCLAW, S. P. BELLENDIR, J. SEKELSKY AND R. S. HAWLEY, 2013 The development of a monoclonal antibody recognizing the Drosophila melanogaster phosphorylated histone H2A variant (gamma-H2AV). G3 (Bethesda) 3**:** 1539-1543.

LANGLEY, C. H., M. CREPEAU, C. CARDENO, R. CORBETT-DETIG AND K. STEVENS, 2011 Circumventing heterozygosity: sequencing the amplified genome of a single haploid Drosophila melanogaster embryo. Genetics 188**:** 239-246.

LEWIS, J. D., N. SAPERAS, Y. SONG, M. J. ZAMORA, M. CHIVA *et al.*, 2004 Histone H1 and the origin of protamines. Proc Natl Acad Sci U S A 101**:** 4148-4152.

LINDSLEY, D. L., AND K. T. TOKUYASU, 1980 Spermatogenesis, pp. 225-294 in *The Genetics and Biology of Drosophila*, edited by M. Ashburner and T. R. F. Wright. Academic Press, New York.

LOPPIN, B., R. DUBRUILLE AND B. HORARD, 2015 The intimate genetics of Drosophila fertilization. Open Biol 5.

MALARKEY, C. S., AND M. E. CHURCHILL, 2012 The high mobility group box: the ultimate utility player of a cell. Trends Biochem Sci 37**:** 553-562.

MAMPELL, K., 1945 Analysis of a Mutator. Genetics 30**:** 496-505.

MASON, J. M., T. A. RANDALL AND R. CAPKOVA FRYDRYCHOVA, 2016 Telomerase lost? Chromosoma 125**:** 65-73.

MIESCHER, F., 1874 Das Protamin, eine neue organische Base aus den Samenfäden des Rheinlachses. Berichte der deutschen chemischen Gesellschaft 7**:** 376-379.

MISOF, B., S. LIU, K. MEUSEMANN, R. S. PETERS, A. DONATH *et al.*, 2014 Phylogenomics resolves the timing and pattern of insect evolution. Science 346**:** 763-767.

MUSARO, M., L. CIAPPONI, B. FASULO, M. GATTI AND G. CENCI, 2008 Unprotected Drosophila melanogaster telomeres activate the spindle assembly checkpoint. Nat Genet 40**:** 362-366.

NOGUCHI, T., AND K. G. MILLER, 2003 A role for actin dynamics in individualization during spermatogenesis in Drosophila melanogaster. Development 130**:** 1805-1816.

O'FARRELL, P. H., J. STUMPFF AND T. T. SU, 2004 Embryonic cleavage cycles: how is a mouse like a fly? Curr Biol 14**:** R35-45.

O'SULLIVAN, R. J., AND J. KARLSEDER, 2010 Telomeres: protecting chromosomes against genome instability. Nat Rev Mol Cell Biol 11**:** 171-181.

PALMER, D. K., K. O'DAY AND R. L. MARGOLIS, 1990 The centromere specific histone CENP-A is selectively retained in discrete foci in mammalian sperm nuclei. Chromosoma 100**:** 32-36.

PFEIFFER, B. D., A. JENETT, A. S. HAMMONDS, T. T. NGO, S. MISRA *et al.*, 2008 Tools for neuroanatomy and neurogenetics in Drosophila. Proc Natl Acad Sci U S A 105**:** 9715-9720.

PIMPINELLI, S., 2006 Drosophila telomeres, pp. 433-459 in *Telomeres*, edited by T. L. de Lange, V. Lundbald and E. H. Blackburn. Cold Spring Harbor Laboratory Press, New York.

RAFFA, G. D., L. CIAPPONI, G. CENCI AND M. GATTI, 2011 Terminin: a protein complex that mediates epigenetic maintenance of Drosophila telomeres. Nucleus 2**:** 383-391.

RATHKE, C., W. M. BAARENDS, S. AWE AND R. RENKAWITZ-POHL, 2014 Chromatin dynamics during spermiogenesis. Biochim Biophys Acta 1839**:** 155-168.

RATHKE, C., W. M. BAARENDS, S. JAYARAMAIAH-RAJA, M. BARTKUHN, R. RENKAWITZ *et al.*, 2007 Transition from a nucleosome-based to a protamine-based chromatin configuration during spermiogenesis in Drosophila. J Cell Sci 120**:** 1689-1700.

RATHKE, C., B. BARCKMANN, S. BURKHARD, S. JAYARAMAIAH-RAJA, J. ROOTE *et al.*, 2010 Distinct functions of Mst77F and protamines in nuclear shaping and chromatin condensation during Drosophila spermiogenesis. Eur J Cell Biol 89**:** 326-338.

RAYCHAUDHURI, N., R. DUBRUILLE, G. A. ORSI, H. C. BAGHERI, B. LOPPIN *et al.*, 2012 Transgenerational propagation and quantitative maintenance of paternal centromeres depends on Cid/Cenp-A presence in Drosophila sperm. PLoS Biol 10**:** e1001434.

RICKMYRE, J. L., S. DASGUPTA, D. L. OOI, J. KEEL, E. LEE *et al.*, 2007 The Drosophila homolog of MCPH1, a human microcephaly gene, is required for genomic stability in the early embryo. J Cell Sci 120**:** 3565-3577.

ROGAT, A. D., AND K. G. MILLER, 2002 A role for myosin VI in actin dynamics at sites of membrane remodeling during Drosophila spermatogenesis. J Cell Sci 115**:** 4855-4865.

RONG, Y. S., 2008 Telomere capping in Drosophila: dealing with chromosome ends that most resemble DNA breaks. Chromosoma 117**:** 235-242.

ROTHWELL, W. F., AND W. SULLIVAN, 2000 Fluorescent analysis of Drosophila embryos, pp. 141-157 in *Drosophila protocols*, edited by W. Sullivan, M. Ashburner and R. S. Hawley. Cold Spring Harbor Laboratory Press, New York.

ROY, A., A. KUCUKURAL AND Y. ZHANG, 2010 I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5**:** 725-738.

ROYOU, A., M. E. GAGOU, R. KARESS AND W. SULLIVAN, 2010 BubR1- and Polo-coated DNA tethers facilitate poleward segregation of acentric chromatids. Cell 140**:** 235-245.

RUSSELL, S. R., AND K. KAISER, 1993 Drosophila melanogaster male germ line-specific transcripts with autosomal and Y-linked genes. Genetics 134**:** 293-308.

SCHINDELIN, J., I. ARGANDA-CARRERAS, E. FRISE, V. KAYNIG, M. LONGAIR *et al.*, 2012 Fiji: an open-source platform for biological-image analysis. Nat Methods 9**:** 676-682.

SCRIBA, M. E. L., 1964 1964 Beeinflussung der frühen Embryonalentwicklung von *Drosophila melanogaster* durch Chromosomenaberrationen. Zool Jb Anat 81**:** 435-490.

SHIMIZU, Y., K. MITA, M. TAMURA, K. ONITAKE AND M. YAMASHITA, 2000 Requirement of protamine for maintaining nuclear condensation of medaka (Oryzias latipes) spermatozoa shed into water but not for promoting nuclear condensation during spermatogenesis. Int J Dev Biol 44**:** 195-199.

SIBON, O. C., V. A. STEVENSON AND W. E. THEURKAUF, 1997 DNA-replication checkpoint control at the Drosophila midblastula transition. Nature 388**:** 93-97.

SIEVERS, F., A. WILM, D. DINEEN, T. J. GIBSON, K. KARPLUS *et al.*, 2011 Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7**:** 539.

STANLEY, C. E., JR., AND R. J. KULATHINAL, 2016 flyDIVaS: A Comparative Genomics Resource for Drosophila Divergence and Selection. G3 (Bethesda) 6**:** 2355-2363.

STEWART, J. A., M. F. CHAIKEN, F. WANG AND C. M. PRICE, 2012 Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. Mutat Res 730**:** 12-19.

STURTEVANT, A. H., 1923 Inheritance of Direction of Coilling in Limnaea. Science 58**:** 269-270.

SUBIRANA, J. A., C. COZCOLLUELA, J. PALAU AND M. UNZETA, 1973 Protamines and other basic proteins from spermatozoa of molluscs. Biochim Biophys Acta 317**:** 364-379.

SWANSON, W. J., AND V. D. VACQUIER, 2002 The rapid evolution of reproductive proteins. Nat Rev Genet 3**:** 137-144.

SWANSON, W. J., A. WONG, M. F. WOLFNER AND C. F. AQUADRO, 2004 Evolutionary expressed sequence tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. Genetics 168**:** 1457-1465.

TAKADA, S., A. KELKAR AND W. E. THEURKAUF, 2003 Drosophila checkpoint kinase 2 couples centrosome function and spindle assembly to genomic integrity. Cell 113**:** 87-99.

TATES, A. D., 1971 *Cytodifferentiation during spermatogenesis in Drosophila melanogaster : an electron microscopic study*. Drukkerij J. H. Pasmans, The Hague, The Netherlands.

TIRMARCHE, S., S. KIMURA, L. SAPEY-TRIOMPHE, W. SULLIVAN, F. LANDMANN *et al.*, 2014 Drosophila protamine-like Mst35Ba and Mst35Bb are required for proper sperm nuclear morphology but are dispensable for male fertility. G3 (Bethesda) 4**:** 2241-2245.

TOKUYASU, K. T., 1974 Dynamics of spermiogenesis in Drosophila melanogaster. IV. Nuclear transformation. J Ultrastruct Res 48**:** 284-303.

TOKUYASU, K. T., W. J. PEACOCK AND R. W. HARDY, 1972 Dynamics of spermiogenesis in Drosophila melanogaster. II. Coiling process. Z Zellforsch Mikrosk Anat 127**:** 492-525.

VAN DE WERKEN, C., G. W. VAN DER HEIJDEN, C. ELEVELD, M. TEEUWSSEN, M. ALBERT *et al.*, 2014 Paternal heterochromatin formation in human embryos is H3K9/HP1 directed and primed by sperm-derived histone modifications. Nat Commun 5**:** 5868.

WAKIMOTO, B. T., AND M. G. HEARN, 1990 The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of Drosophila melanogaster. Genetics 125**:** 141-154.

WAKIMOTO, B. T., D. L. LINDSLEY AND C. HERRERA, 2004 Toward a comprehensive genetic analysis of male fertility in Drosophila melanogaster. Genetics 167**:** 207-216.

WASBROUGH, E. R., S. DORUS, S. HESTER, J. HOWARD-MURKIN, K. LILLEY *et al.*, 2010 The Drosophila melanogaster sperm proteome-II (DmSP-II). J Proteomics 73**:** 2171-2185.

WESOLOWSKA, N., F. L. AMARIEI AND Y. S. RONG, 2013 Clustering and protein dynamics of Drosophila melanogaster telomeres. Genetics 195**:** 381-391.

WIEGMANN, B. M., M. D. TRAUTWEIN, I. S. WINKLER, N. B. BARR, J. W. KIM *et al.*, 2011 Episodic radiations in the fly tree of life. Proc Natl Acad Sci U S A 108**:** 5690-5695.

WILSON, K. L., K. R. FITCH, B. T. BAFUS AND B. T. WAKIMOTO, 2006 Sperm plasma membrane breakdown during Drosophila fertilization requires sneaky, an acrosomal membrane protein. Development 133**:** 4871-4879.

WU, S. F., H. ZHANG AND B. R. CAIRNS, 2011 Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res 21**:** 578-589.

WU, X. J., Y. Y. CHEN, C. C. GONG AND D. S. PEI, 2018 The role of high-mobility group protein box 1 in lung cancer. J Cell Biochem.

YAMAKI, T., G. K. YASUDA AND B. T. WAKIMOTO, 2016 The Deadbeat Paternal Effect of Uncapped Sperm Telomeres on Cell Cycle Progression and Chromosome Behavior in Drosophila melanogaster. Genetics 203**:** 799-816.

YANG, P., W. WU AND T. S. MACFARLAN, 2015 Maternal histone variants and their chaperones promote paternal genome activation and boost somatic cell reprogramming. Bioessays 37**:** 52-59.

YANG, Z., 2007 PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24**:** 1586-1591.

YASUDA, G. K., G. SCHUBIGER AND B. T. WAKIMOTO, 1995 Genetic characterization of ms (3) K81, a paternal effect gene of Drosophila melanogaster. Genetics 140**:** 219-229.

YU, Y. E., Y. ZHANG, E. UNNI, C. R. SHIRLEY, J. M. DENG *et al.*, 2000 Abnormal spermatogenesis and reduced fertility in transition nuclear protein 1-deficient mice. Proc Natl Acad Sci U S A 97**:** 4683-4688.

ZHAO, M., C. R. SHIRLEY, Y. E. YU, B. MOHAPATRA, Y. ZHANG *et al.*, 2001 Targeted disruption of the transition protein 2 gene affects sperm chromatin structure and reduces fertility in mice. Mol Cell Biol 21**:** 7243-7255.