npg

## REVIEW

# Duplicated proteasome subunit genes in *Drosophila* and their roles in spermatogenesis

JM Belote and L Zhong[1]
*Department of Biology, Syracuse University, Syracuse, NY, USA*

The proteasome is a large, multisubunit complex that acts as the cell's 'protein-degrading machine' in the ubiquitin-mediated proteolytic pathway for regulated protein turnover. Although proteasomes are usually thought of as being homogeneous structures, recent studies have revealed their more dynamic and heterogeneous nature. For example, in a number of plant and animal species, multiple isoforms of several proteasome subunits, encoded by paralogous genes, have been discovered, and in some cases, these alternative isoforms have been shown to be functionally distinct from their conventional counterparts. A particularly striking example of this phenomenon is seen in *Drosophila melanogaster*, where 12 of the 33 subunits that make up the 26S proteasome holoenzyme are represented in the genome by multiple paralogous genes. Remarkably, in every case, the 'extra' genes are expressed in a testis-specific manner. Here, we describe the extent and nature of these testis-specific gene duplications and discuss their functional significance, and speculate on why this situation might have evolved.
*Heredity* (2009) **103**, 23–31; doi:10.1038/hdy.2009.23; published online 11 March 2009

## Introduction

The major pathway for regulated proteolysis in eukaryotes is the ubiquitin–proteasome system, UPS (Glickman and Ciechanover, 2002). By controlling the rapid and irreversible turnover of key regulatory proteins, the UPS has important roles in a wide variety of biological processes, including cell-cycle progression, transcriptional regulation, signal transduction, cell fate determination and metabolic regulation. The UPS also has a 'housekeeping' function, ridding cells of abnormal or misfolded polypeptides that can form harmful aggregates (Goldberg, 2003). In mammals, the UPS has an additional role in the antiviral immune response, generating antigenic peptides for presentation to the cell surface by major histocompatibility complex (MHC) class I molecules (Kloetzel, 2004).

In the UPS, specific proteins are targeted for destruction by the covalent attachment of multiple copies of the highly conserved, 76-amino-acid protein, ubiquitin (Ub). The Ubs are sequentially added through a cycle of reactions catalyzed by the E1 (Ub-activating), E2 (Ub-conjugating) and E3 (Ub-ligase) enzymes, and the multi-Ub tag serves as a recognition signal for a large, multisubunit proteolytic complex called the 26S proteasome. Although Ub tagging usually destines a protein for destruction, the activity of deubiquitinating enzymes can provide a mechanism to control the rate of a protein's turnover, by removing the Ub tag before the substrate reaches the proteolytic machinery (Kraut *et al.*, 2007).

It is not completely understood how proteins are selected for ubiquitination, although some important signal sequences have been identified on particular short-lived proteins (for example, PEST sequences, the destruction box, the F-box, the N-terminal amino acid). In some cases, these degradation signals are activated by phosphorylation, thereby triggering the rapid turnover of the affected protein.

## The proteasome: a protein-degrading machine

The 26S proteasome is an extraordinary molecular machine made up of more than 30 different subunits, and having a molecular mass of over 2400 kDa (Figure 1). The core of this complex is a hollow cylindrical particle called the 20S proteasome. At each end of the 20S core is a 'cap' complex, the 19S regulator, that has the role of gatekeeper, capturing Ub-tagged proteins and facilitating their entry into the core's degradation chamber. Once inside, the doomed polypeptide is cut into peptide remnants, which diffuse out, possibly through side vents. X-ray crystallography studies have shown that the 20S proteasome is made up of four stacked rings (Unno *et al.*, 2002; Gröll *et al.*, 1997). The outer rings comprise seven evolutionarily related α-type subunits and the inner rings are each made up of seven homologous β-type subunits, with three of the β-type subunits conferring the peptide bond hydrolyzing activities.

The 20S proteasome is sealed off at each end by the N-terminal tails of a subset of the α subunits (Gröll *et al.*,
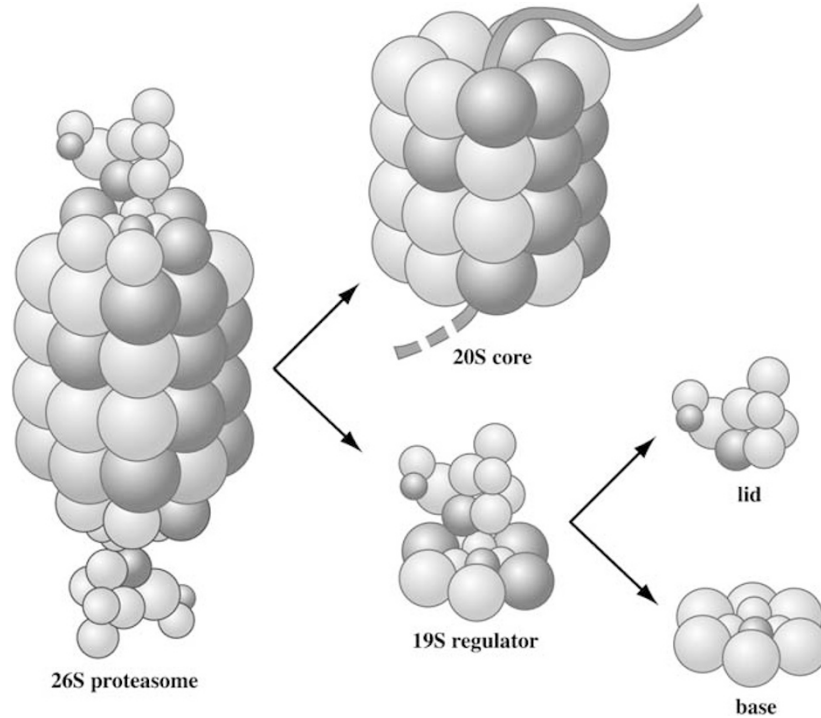
**Figure 1** Subunit structure of the 26S proteasome. The subunit organization of the 20S core has been solved by X-ray crystallography of yeast and bovine proteasomes (Gröll *et al.*, 1997; Unno *et al.*, 2002). A model of subunit arrangement of the 19S lid has been proposed by Sharon *et al.* (2006) based on mass spectrometry results. Darkly shaded subunits have testis-specific isoforms (see Table 1 for details).

2000; Unno *et al.*, 2002). Thus, for a protein substrate to enter the proteolytic chamber the 'gate' must be opened, a job usually performed by the 19S cap. This large (890 kDa) complex can be separated into two parts, the base and the lid. The base consists of six ATPase subunits of the AAA family of proteins, and four non-ATPase subunits. It is thought that the AAA-ATPase subunits are important for unfolding the protein substrates (Braun *et al.*, 1999), opening the gated channel (Köhler *et al.*, 2001) and translocating unfolded polypeptides into the central core (Navon and Goldberg, 2001). The lid is an asymmetrical complex that sits atop the base, oriented in opposite directions at the two ends of the 26S proteasome (Voges *et al.*, 1999; Sharon *et al.*, 2006). It consists of at least eight different subunits, most of which have no known molecular function.

## Proteasomes can exist in multiple structural forms

Although the proteasome has conventionally been thought of as a relatively homogeneous structure, evidence accumulated over the past few years has underscored its more dynamic and heterogeneous nature (Glickman and Raveh, 2005). This structural heterogeneity is manifested in a number of ways. First, several of the subunits of both the core and the 19S regulatory cap can undergo posttranslational modifications, including phosphorylation (Mason *et al.*, 1998), *N*-acetylation (Kimura *et al.*, 2000) and O-linked glycosylation (Sümegi *et al.*, 2003). The functional significance of these modifications is poorly understood, although it has been speculated that such alterations might have regulatory significance.

A second way in which proteasomes can exhibit structural heterogeneity is through association of the 20S core particle with alternative regulators that substitute for the usual 19S caps (Rechsteiner and Hill, 2005). One well-studied alternative regulator is called REG (also known as 11S or PA28). There are two types of REG complexes. One, found primarily in the cytoplasm, is a hetero-heptameric ring of two homologous subunits REGα and REGβ (Li and Rechsteiner, 2001), and the other, confined to the nucleus, is a homo-heptamer of the related protein REGγ. In both cases, the REG complex binds to the α-rings of 20S core, replacing the usual 19S cap and opening the gate in an ATP-independent manner (Whitby *et al.*, 2000).

A third type of proteasome structural heterogeneity that exists in some organisms involves the regulated expression of alternative proteasome subunit isoforms that are encoded by paralogous genes (Yang *et al.*, 2003; Kloetzel, 2004; Belote and Zhong, 2005). For example, in mammals there is one well-documented case in which specialized 20S proteasomes containing alternative subunits occur: during the antiviral immune response, γ-interferon induces the synthesis of three new β-type subunits, β1i, β2i and β5i, that replace the catalytic β1, β2 and β5 subunits, respectively. These reconfigured 'immunoproteasomes' have altered proteolytic properties that are more efficient at producing peptide antigens for MHC class I-mediated antigen presentation (Kloetzel, 2004).

Perhaps the most conspicuous example of this phenomenon is seen in *Drosophila melanogaster*, where 12 of the 33 subunits that make up the 26S proteasome have

two, or in some cases three, alternative isoforms that are encoded by paralogous genes (Yuan *et al.*, 1996; Ma *et al.*, 2002; Zhong and Belote, 2007). In every case, one form of each subunit is widely expressed at all developmental stages and in all tissues (these will be referred to as 'conventional' proteasome subunits), whereas all of the additional isoforms are testis specific. This represents an unprecedented example of developmental regulation of alternative proteasome subunit expression, and it suggests that there might be a specialization of proteasome function during spermatogenesis. The large number of proteasome subunit isoform genes, and the collective shift of the expression patterns between the conventional subunit genes and their testis-specific counterparts (see below) suggests that there might be a testis-specific proteasome that is dynamically assembled during *Drosophila* spermatogenesis. If so, what is its functional significance? What specific role(s) does this testis-specific proteasome have in spermatogenesis? Why has this peculiar situation evolved?

## Testis-specific proteasome subunit genes

The remarkable extent of gene duplications encoding testis-specific isoforms of proteasome subunits is shown in Table 1. In some cases (see below), the testis-specific expression patterns have been examined in detail, using βGal- or green fluorescent protein (GFP)-tagged proteasome subunit reporter transgenes. For the rest, testis-specific expression of the 'extra' genes has been inferred from the observations that (1) northern blot analyses demonstrated that they are expressed in adult males, but not in females (Ma *et al.*, 2002), and (2) examination of expressed sequence tag (EST) databases revealed that they are all preferentially represented by ESTs from testis cDNA libraries, whereas their conventional counterparts are highly represented in a variety of cDNA libraries (Boutanaev *et al.*, 2002; JM Belote and L Zhong, unpublished). As a first step to understanding the functional role(s) of the presumed testis-specific proteasome, the expression and subcellular localization of a subset of these proteins has been investigated in detail.

**Table 1** 26S proteasome subunit genes of *D. melanogaster*

| Gene subunit (type) | Name[a] | Map position[b] | Testis-specific isoform | Gene name[a] | Map position | Amino-acid identity (%)[c] | Reporter transgene[d] |
|---|---|---|---|---|---|---|---|
| α1 (core) | CG18495 | 43E (2R) | — | — | — | — | — |
| α2 (core) | CG5266 | 87B (3R) | — | — | — | — | — |
| α3 (core) | CG9327 | 57B (2R) | α3T | CG1736 | 100B (3R) | 58 | GFP |
| α4 (core) | CG3422 | 14B (X) | α4T1 | CG17268 | 92F (3R) | 74 | βGal |
|  |  |  | α4T2 | CG4569 | 60D (2R) | 54 | βGal |
| α5 (core) | CG10938 | 54C (2R) | — | — | — | — | — |
| α6 (core) | CG4904 | 31C (2L) | α6T | CG5648 | 34B (2L) | 64 | GFP |
| α7 (core) | CG1519 | 46B (2R) | — | — | — | — | — |
| β1 (core) | CG8392 | 52E (2R) | — | — | — | — | — |
| β2 (core) | CG3329 | 71A (3L) | β2R1 | CG18341 | 5B (X) | 62 | — |
|  |  |  | β2R2 | CG12161 | 82F (3R) | 35 | — |
| β3 (core) | CG11981 | 85C (3R) | — | — | — | — | — |
| β4 (core) | CG17331 | 36A (2L) | β4R1 | CG17301 | 23A (2L) | 43 | — |
|  |  |  | β4R2 | CG17302 | 23A (2L) | 55 | — |
| β5 (core) | CG12323 | 47B (2R) | β5R1 | CG9868 | 59D (2R) | 53 | GFP |
|  |  |  | β5R2 | CG31742 | 36E (2L) | 47 | GFP |
| β6 (core) | CG4097 | 73B (3L) | — | — | — | — | — |
| β7 (core) | CG12000 | 83A (3R) | — | — | — | — | — |
| Rpt1 (base) | CG1341 | 43E (2R) | — | — | — | — | — |
| Rpt2 (base) | CG5289 | 95C (3R) | — | — | — | — | — |
| Rpt3 (base) | CG16916 | 10B (X) | Rpt3R | CG9475 | 85E (3R) | 77 | GFP |
| Rpt4 (base) | CG3455 | 5E (X) | Rpt4R | CG7257 | 68E (3L) | 82 | GFP |
| Rpt5 (base) | CG10370 | 95B (3R) | — | — | — | — | — |
| Rpt6 (base) | CG1489 | 19F (X) | Rpt6R | CG2241 | 99F (3R) | 88 | — |
| Rpn1 (base) | CG7762 | 76D (3L) | — | — | — | — | — |
| Rpn2 (base) | CG11888 | 99A (3R) | — | — | — | — | — |
| Rpn3 (lid) | CG10484 | 37B (2L) | — | — | — | — | — |
| Rpn5 (lid) | CG1100 | 83C (3R) | — | — | — | — | — |
| Rpn6 (lid) | CG10149 | 51C (2R) | — | — | — | — | — |
| Rpn7 (lid) | CG5378 | 94B (3R) | — | — | — | — | — |
| Rpn8 (lid) | CG3416 | 60C (2R) | — | — | — | — | — |
| Rpn9 (lid) | CG10230 | 95B (3R) | — | — | — | — | — |
| Rpn10 (base) | CG7619 | 78E (3L) | — | — | — | — | — |
| Rpn11 (lid) | CG18174 | 25C (2L) | — | — | — | — | — |
| Rpn12 (lid) | CG4157 | 73A (3L) | Rpn12R | CG11552 | 71A (3L) | 49 | — |
| Rpn13 (base) | CG13349 | 50C (2R) | Rpn13R | CG6789 | 4F (X) | 32 | — |
| Uch37 (lid) | CG3431 | 67B (3L) | Uch37R | CG1950 | 11A (X) | 56 | — |

[a]CG identifier from annotated database (FlyBase, http://flybase.bio.indiana.edu/).
[b]Salivary gland chromosome cytological map position (with chromosome arm).
[c]Amino-acid identity compared to the conventional subunit sequence.
[d]Reporter fusion transgene used to determine testis-specific expression.
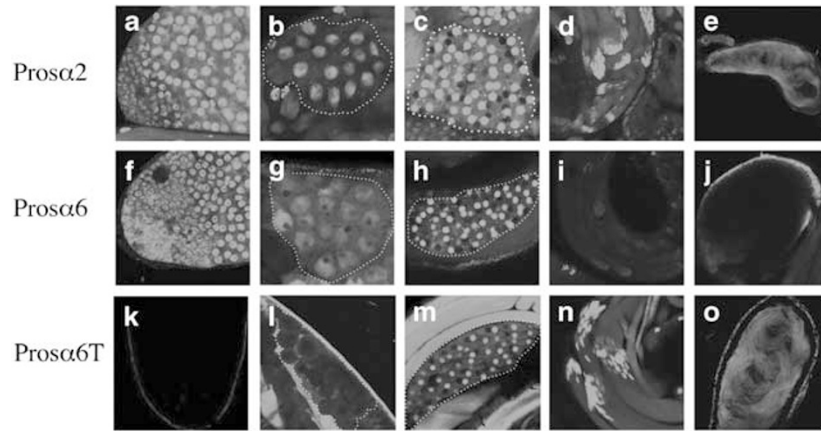
**Figure 2** Spermatogenic expression profiles of the conventional subunit α2 (**a–e**), the conventional subunit with a testis-specific isoform, α6 (**f–j**) and the testis-specific subunit α6T (**k–o**), as seen by green fluorescent proteins (GFP)-tagged transgenes. (**a, f, k**) = apical tips of testes showing spermatogonial cells; (**b, g, l**) = 16-cell primary spermatocyte cysts; (**c, h, m**) = 64-cell spermatid cysts (bright spots are nuclei, dark spots are nebenkerns); (**d, i, n**) = elongated spermatid bundles (bright areas are bundles of spermatid nuclei); (**e, j, o**) = seminal vesicles containing mature sperm (most of the GFP fluorescence is in the sperm tails, although nuclei are also labeled).

To explain the expression patterns, a brief description of *Drosophila* spermatogenesis is helpful (Fuller, 1993). Stem cells located at the tip of the coiled testis divide to produce spermatogonial cells that undergo four cycles of mitosis, each yielding a cyst of 16 interconnected primary spermatocytes surrounded by 2 somatic cyst cells. Following a growth phase, the primary spermatocytes undergo meiosis, resulting in a cyst of 64 spherical, haploid spermatids, each containing a prominent round nucleus and a phase-dark structure, the nebenkern, consisting of fused mitochondria. In the later stages of sperm development, known as spermiogenesis, the spermatids undergo tremendous elongation. The nuclei condense and experience a dramatic shape change, becoming long and thin. The spermatids then undergo individualization, during which the syncytial spermatid bundle is resolved into 64 separate sperm cells. The individualization complex (IC), a cytoskeletal-membranous complex containing a cluster of structures called actin cones, accomplishes this step. During individualization, the IC moves down the spermatid bundle, eliminating all cytoplasmic bridges and pushing the excess cytoplasm and organelles into a 'cystic bulge' that eventually becomes the waste bag, which is ultimately reabsorbed. As the IC progresses down the bundle, each spermatid becomes sheathed in its own plasma membrane. In the final stage of spermiogenesis, the sperm coil up, the cyst ruptures and the mature sperm move into the seminal vesicle for storage.

The most detailed analyses of when and where the various proteasome subunit isoforms are expressed have been done with the α3T, α6T and Rpt3R subunits, and their conventional counterparts, α3, α6 and Rpt3. Figure 2 illustrates the spermatogenic expression profile for the α6/α6T pair, as well as that of α2, a conventional subunit with no testis-specific isoform. The α2 subunit, similar to all conventional subunits, is expressed somatically throughout the fly, and in the testis it is present in the cytoplasm and nuclei in all spermatogenic stages (Figures 2a–e). For α6, a conventional subunit with a testis-specific isoform, expression occurs early in spermatogenesis and remains prominent up through the 64-cell spermatid stage, but its abundance then drops and it is not detected in mature sperm stored in the seminal vesicle (Figures 2f–j). In contrast, the testis-specific α6T subunit is not seen premeiotically, but it begins to appear during meiosis, becomes conspicuous in the postmeiotic stages and is seen in mature sperm (Figures 2k–o). Similar expression patterns are seen for the α3/α3T subunits, except that α3T is preferentially localized to the nuclei of late-stage spermatids and is not seen distributed along the sperm tails of mature sperm. A slightly different pattern is seen for Rpt3/Rpt3R. In this case, Rpt3R first appears in the 16-cell primary spermatocyte stage (that is, premeiotically) and is prominent in nuclei and cytoplasm of spermatid bundles throughout the sperm elongation stage. However, it disappears following individualization and is not apparent in mature sperm.

Although each of the testis-specific subunits that have been examined using GFP-transgene reporters (for example, α3T, α6T, β5R1, β5R2, Rpt3R and Rpt4R) show slightly different patterns of expression, they have in common the following features: (1) none are expressed somatically or during the early gonial stages of spermatogenesis, (2) all are prominently expressed in the cytoplasm and nuclei of 64-cell spermatids and elongating stage cysts and (3) all are found associated with the IC as numerous, punctate 'speckles' trailing the actin cones as they move down the spermatid bundle during individualization (Figure 3). That these speckles are proteasome-related structures, and not merely artifactual aggregates of GFP-tagged proteins, has been shown by the fact that immunostaining testes with anti-proteasome antibodies labels the same speckles (Zhong and Belote, 2007). The subcellular nature of these structures is unknown.

## Functional significance of testis-specific proteasome genes

One critical question concerns whether these testis-specific proteasome subunits have any functional significance. That is, do they have an essential role in spermatogenesis, or are they merely functionally redundant copies
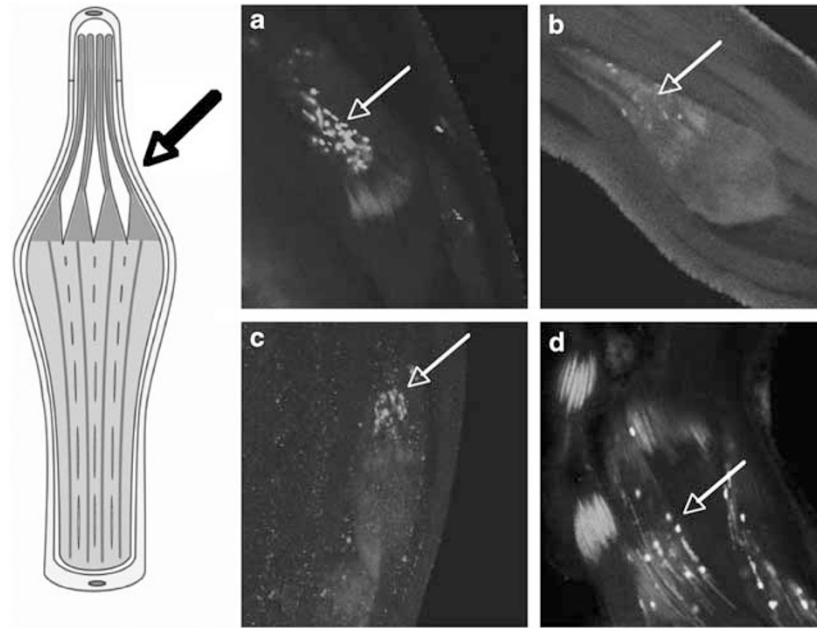
**Figure 3** Proteasomal 'speckles' associated with the individualization complex (IC). On the left is a cartoon of the IC composed of actin cones (orange) within the cystic bulge, formed as the IC moves down the spermatid bundle. Sperm nuclei (blue) are at the top. For clarity, only 4 of the 64 spermatids are shown. (**a**) Speckles of α3T-GFP trailing the actin cones (red); (**b**) speckles of α6T-GFP trailing the actin cones; (**c**) immunostaining wild-type testis with anti-proteasome antibody (green); (**d**) immunostaining α6T-GFP transgenic testis with anti-proteasome antibody showing colocalization (yellow) of the α6T-GFP and proteasome signals.

of the conventional subunit genes? To date, this question has been addressed only with respect to α6T, Rpt3R and Rpt4R, using a mutational approach. For α6T and Rpt3R, knockout mutants have been generated (*Prosα6T¹* and *Rpt3R⁶¹⁵*, respectively) and found to exhibit a male-sterile phenotype, demonstrating that, at least for these cases, the testis-specific proteasome subunit isoforms are necessary for normal spermatogenesis (Li, 2006; Zhong and Belote, 2007). Consistent with the observed expression patterns, there is no detectable effect of these mutants on any stage before and including the spermatid elongation stage. However, in both cases, sperm individualization is disrupted, and mature sperm fail to be produced.

The effect on individualization is understandable given the association of the testis-specific proteasome subunits with proteasomal 'speckles' seen trailing the actin cones of the IC as it moves down the spermatid bundle. Interestingly, the unusual pattern of proteasomal 'speckles' is strikingly similar to that reported for the apoptotic protein DRONC (Huh *et al.*, 2004). Could this similarity be meaningful? It is known that during the process of spermatogenesis, the apoptotic pathway is used in a spatially restricted manner to mediate the cytoplasmic breakdown within the cystic bulge during the individualization process. One current model is that DRONC undergoes activation following the degradation of its inhibitory protein partner, DIAP, and that DRONC then activates the effector caspase drICE to trigger apoptosis within the IC. If one role of the testis-specific proteasome is to carry out the destruction of DIAP, then in the *Prosα6T¹* mutant, DRONC might fail to be activated due to a compromised UPS, with the subsequent failure of drICE to be activated as well. Consistent with this model, it is found that in the mutant, unlike the case in wild type, active DRONC fails to accumulate in
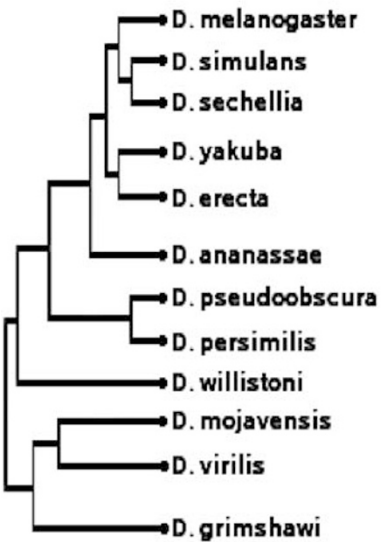
the cystic bulge in the mutant, and active drICE is significantly reduced (Zhong and Belote, 2007).

In addition to this proposed specific role of the UPS during individualization, it is likely that testis-specific proteasomes have a more general role, as well. For example, as the IC moves down the spermatid bundle, there is a need for the cytoplasmic contents to be disassembled to facilitate their removal into the 'waste bag'. During this process, a wide range of protein substrates are likely to be targeted for destruction, and proteasome-mediated proteolysis would be expected to be unusually active in or near the IC.

In contrast to these results with α6T and Rpt3R mutants, the *Rpt4R⁹* null mutant is viable and fertile in both sexes, demonstrating that this gene does not have an essential role in spermatogenesis. It is possible that, in this case, the function of Rpt4R can be replaced by that of its paralog, Rpt4. This presumed redundancy of function between the conventional subunit and its testis-specific paralog is not unreasonable, because it has been shown that in the case of the α6/α6T subunits, ectopically expressing the conventional subunit in testes can rescue the *Prosα6T¹* mutant phenotype (Zhong and Belote, 2007). These results suggest that the testis-specific proteasomes might not be qualitatively that much different in their function from that of the conventional somatic proteasomes, although it is possible that they have a subtly distinct function, or that they are more efficient at mediating rapid protein degradation during spermiogenesis.
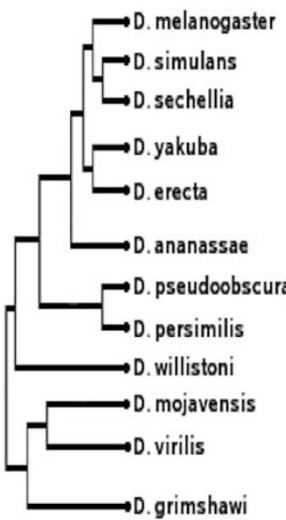
## Proteasome gene duplications are widespread in the *Drosophila* genus

The phenomenon of testis-specific proteasome subunit gene duplication is not limited to *D. melanogaster*, as seen

**Table 2** Numbers of paralogous 20S proteasome subunit genes in various *Drosophila* species

| Species | α1 | α2 | α3 | α4 | α5 | α6 | α7 | β1 | β2 | β3 | β4 | β5 | β6 | β7 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D. melanogaster | 1 | 1 | 2 | 3 | 1 | 2 | 1 | 1 | 3 | 1 | 3 | 3 | 1 | 1 |
| D. simulans | 1 | 1 | 2 | 3 | 1 | 2 | 1 | 1 | 3 | 1 | 3 | 3 | 1 | 1 |
| D. sechellia | 1 | 1 | 2 | 3 | 1 | 2 | 1 | 1 | 3 | 1 | 3 | 3 | 1 | 1 |
| D. yakuba | 1 | 1 | 2 | 3 | 1 | 2 | 1 | 1 | 3 | 1 | 3 | 3 | 1 | 1 |
| D. erecta | 1 | 1 | 2 | 3 | 1 | 2 | 1 | 1 | 3 | 1 | 3 | 3 | 1 | 1 |
| D. ananassae | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 3 | 1 | 3 | 4 | 1 | 2 |
| D. pseudoobscura | 1 | 2 | 2 | 3 | 1 | 3 | 1 | 3 | 3 | 3 | 3 | 3 | 1 | 3 |
| D. persimilis | 1 | 2 | 2 | 3 | 1 | 4 | 1 | 3 | 3 | 3 | 3 | 3 | 2 | 3 |
| D. willistoni | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 2 | 5 | 2 |
| D. mojavensis | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 2 | 1 | 2 |
| D. virilis | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 3 | 2 | 2 | 3 | 3 | 2 |
| D. grimshawi | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 3 | 1 | 2 | 3 | 3 | 2 |

See Supplementary Table S1 for details and gene identifiers.

**Table 3** Numbers of paralogous 19S cap proteasome subunit genes in various *Drosophila* species

| Species | RPT1 | RPT2 | RPT3 | RPT4 | RPT5 | RPT6 | RPN1 | RPN2 | RPN3 | RPN5 | RPN6 | RPN7 | RPN8 | RPN9 | RPN10 | RPN11 | RPN12 | RPN13 | UCHP37 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D. melanogaster | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| D. simulans | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| D. sechellia | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 |
| D. yakuba | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 |
| D. erecta | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| D. ananassae | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| D. pseudoobscura | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | 2 |
| D. persimilis | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 |
| D. willistoni | 3 | 1 | 2 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 2 |
| D. mojavensis | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| D. virilis | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| D. grimshawi | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |

See Supplementary Table S2 for details and gene identifiers.

by the existence of orthologs of many of these genes in other *Drosophila* species (Tables 2 and 3, see also Supplementary Tables S1 and S2). Although the testis-specific expression of most of these 'extra' genes has not been confirmed, in one case where it has been directly tested (that is, the α4/α4T subunits of *D. virilis*) this was found to be the case (Belote *et al.*, 1998).

The comparative study of proteasome gene duplications was greatly facilitated by the recent sequencing of the genomes of 11 other *Drosophila* species (Drosophila 12 Genomes Consortium, 2007). Several interesting features emerge from this analysis. First, when all species are considered, the great majority of proteasome subunit genes have undergone duplication in at least one lineage. This is especially true for the 20S proteasome subunits, where all but α1 and α5 have paralogous genes in one or more species. Second, the proteasome gene makeup among the various species examined appears extraordinarily

dynamic, with gene duplications appearing early in some lineages, and then undergoing further gene duplication or gene loss in later lineages. As a result, the number of paralogous genes often varies from one in some species to three, or more, in others (Tables 2 and 3). Although some cases of apparent gene loss might be explained by incomplete sequence information or annotation errors, it is unlikely that all of these examples of gene loss are artifactual. There does seem to be a much higher degree of duplication and loss in genes corresponding to subunits of the 20S proteasome as compared to those encoding subunits of the 19S cap, although the reasons for this are unclear. Despite the widespread occurrence of proteasome gene duplication in *Drosophila*, this does not seem to be a common feature of other insect genomes. For example, sequence analyses of *Anopheles gambiae*, *Aedes aegypti*, *Bombyx mori* and *Tribolium castaneum* shows that there are only a few sporadic examples of proteasome gene duplications, and in these cases there is no information on whether any exhibit testis-specific expression.

## Why have *Drosophila* proteasome genes undergone such rampant gene duplication?

Gene duplication is a common evolutionary mechanism for the creation of new genes with specialized functions (Lynch, 2002). The generation of testis-specific genes by such a mechanism might indicate that some features of spermatogenesis require an increased amount of some particular gene product, or that there is a need for a version of the protein with an altered activity better adapted to its specific role in spermatogenic cells. The occurrence of testis-specific isoforms of various proteins generated by gene duplication events is not uncommon in *Drosophila* (for example, casein kinase 2 (Kalmykova *et al.*, 2002), mitochondrial translocase of the outer membrane (Hwa *et al.*, 2004), TAFs (Hiller *et al.*, 2004), OXPHOS genes (Tripoli *et al.*, 2005), myosin VI light chain (Frank *et al.*, 2006), glycogen synthase kinase-3 (Kalamegham *et al.,* 2007), eIF4G (Baker and Fuller, 2007; Franklin-Dumont *et al.*, 2007), see also Mikhaylova *et al.*, 2008)), and is also documented for some mammalian genes (for example, phosphoglycerate kinase (McCarrey and Thomas, 1987), cytochrome *c* (Hake *et al.*, 1994), ornithine decarboxylase antizyme (Ivanov *et al.*, 2000), TAF(II)250 (Wang and Page, 2002)). However, the degree to which the proteasome subunit genes have undergone such rapid proliferation in the *Drosophila* genome to produce testis-specific genes is remarkable.

There are several possible mechanisms that could explain the enriched testis-specific duplication of proteasome genes. One explanation is that these genes are duplicated to allow increased abundance of proteasome particles in spermatids to accommodate the demands of spermatogenesis. *Drosophila* sperm are extremely long, and this is a trait that is rapidly evolving (Pitnick *et al.*, 1995). Given the important role that the UPS appears to have in sperm individualization, and in the cytoplasmic structural breakdown accompanying the elimination of cellular contents, it is likely that the extreme length of *Drosophila* spermatid bundles would pose a daunting challenge. Because transcription of virtually all genes ceases after the primary spermatocyte stage, proteasome

subunit abundance might be a limiting factor for the successful completion this process, and thus there might be strong selective pressure to increase the level of proteasome gene expression in spermatocytes. Proteasome subunit gene duplication is one way to immediately increase the level of proteasome gene expression, and once it happens one of the two paralogous genes would be free to evolve a testis-specific expression pattern.

A logical consequence of this hypothesis is that the duplicate genes would retain most or all of the functions of their parental genes. The observations that over-expressing the conventional α6 subunit in testes rescues the *Prosα6T¹* mutant, and that the *Rpt4R⁹* null mutant is viable and fertile, support this idea. On the other hand, it cannot be ruled out that the duplicate gene might also evolve altered coding sequences to produce proteasomes that are better adapted to accomplish the rapid and expensive breakdown that occurs during the latter stages of sperm development. In fact, ectopically expressing the testis-specific α3T subunit in the soma results in lethality, indicating that at least for this subunit the testis-specific paralog is not merely a redundant copy of its conventional counterpart (Ma *et al.*, 2002).

A related hypothesis that has been put forward to explain the high occurrence of gene duplication is based on the observation that in many species, including *Drosophila*, the X chromosome undergoes precocious inactivation in the male germ line (Betrán *et al.*, 2004). Under this hypothesis, X-linked genes that undergo duplication onto an autosome (by retroposition for example), thereby escape early inactivation during spermatogenesis, and, if they have a role in spermatogenesis, are therefore favored by natural selection. While some of the proteasome genes that have undergone duplication fit this X to autosome mobilization scenario (for example, α4, *Rpt3*, *Rpt4*, and *Rpt6*), most of the proteasome gene duplications represent autosome to autosome, or even autosome to X, duplication events. Thus, this explanation cannot fully explain why proteasome genes have undergone duplication to testis-specific forms.

One feature of proteasome subunits that might also promote the tendency for them to be represented in the genome by paralogous genes is that they function as parts of a large multisubunit complex. Normally, immediately following a gene duplication event, paralogous genes are likely to be functionally redundant. Thus, the selective constraints on one copy should be relaxed because deleterious mutations are hidden from natural selection by the functioning of the other gene copy. In such a case, a degenerative mutation that occurs in one copy is usually ignored by natural selection and, as a result, that copy quickly evolves into a nonfunctional pseudo-gene and is lost. In the case of a gene duplication involving a component of a multisubunit complex, however, a degenerative mutation in one copy might not be ignored by natural selection because it might result in the altered subunit exhibiting a dominant negative effect (Gibson and Spring, 1998). In fact, dominant negative missense mutations have been documented for *Drosophila* proteasome genes (Saville and Belote, 1993; Smyth and Belote, 1999). It has been proposed that gene duplications that encode components of a multisubunit complex are more likely to be

maintained in the genome as functional gene copies because degenerative mutations in one copy are more likely to be removed by purifying selection (Gibson and Spring, 1998). This property would allow functional gene duplicates to persist longer and thereby increase the likelihood that the paralogous genes would evolve specialized subfunctions. It is interesting that another system of large multisubunit complexes (the mitochondrial oxidative phosphorylation, OXPHOS, machinery) is characterized by a strikingly similar high level of gene duplication and testis-specific specialization (Tripoli *et al.*, 2005). In that case, of 78 nuclear genes in *D. melanogaster* encoding subunits of 5 large OXPHOS complexes, 20 have paralogs (either duplications or triplications), and most, if not all, of the duplicated genes show testis-biased expression (Tripoli *et al.*, 2005).

Whatever the driving force, it is clear that the genes encoding components of *Drosophila* proteasomes have undergone an unusual degree of gene duplication to produce testis-specific isoforms for a significant fraction of subunits. It would be interesting to know if proteasomes in spermatogenic cells have altered activities or substrate preferences. Future studies, using biochemical approaches, may be able to address this question, although the small size of *Drosophila* testes poses a challenge.

## Acknowledgements

## References

Baker CC, Fuller MT (2007). Translational control of meiotic cell cycle progression and spermatid differentiation in male germ cells by a novel eIF4G homolog. *Development* **134**: 2863–2869.

Belote JM, Miller M, Smyth KA (1998). Evolutionary conservation of a testes-specific proteasome subunit gene in *Drosophila*. *Gene* **215**: 93–100.

Belote JM, Zhong L (2005). Proteasome gene duplications in mammals, flies and plants. *Recent Res Dev Genes Genomes* **1**: 107–129.

Betrán E, Emerson JJ, Kaessmann H, Long M (2004). Sex chromosomes and male functions: where do new genes go? *Cell Cycle* **3**: 873–875.

Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI (2002). Large clusters of co-expressed genes in the *Drosophila* genome. *Nature* **420**: 666–669.

Braun BC, Glickman M, Kraft R, Dahlmann B, Kloetzel PM, Finley D *et al.* (1999). The base of the proteasome regulatory particle exhibits chaperone-like activity. *Nat Cell Biol* **1**: 221–226.

*Drosophila* 12 Genomes Consortium (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.

Frank DJ, Martin SR, Gruender BN, Lee YS, Simonette RA, Bayley PM *et al.* (2006). Androcam is a tissue-specific light chain for myosin VI in the *Drosophila* testis. *J Biol Chem* **281**: 24728–24736.

Franklin-Dumont TM, Chatterjee C, Wasserman SA, Dinardo S (2007). A novel eIF4G homolog, off-schedule, couples translational control to meiosis and differentiation in *Drosophila* spermatocytes. *Development* **134**: 2851–2861.

Fuller MT (1993). Spermatogenesis. In: Bate M, Martinez Arias A (eds). *The Development of Drosophila melanogaster*, vol 1. Cold Spring Harbor Laboratory Press: New York, pp 71–147.

Gibson TJ, Spring J (1998). Genetic redundancy in vertebrates: polyploidy and persistence of genes encoding multidomain proteins. *Trends Genet* **14**: 46–49.

Glickman MH, Ciechanover A (2002). The ubiquitin-proteasome pathway: destruction for the sake of construction. *Physiol Rev* **82**: 373–428.

Glickman MH, Raveh D (2005). Proteasome plasticity. *FEBS Lett* **579**: 3214–3223.

Goldberg AL (2003). Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**: 895–899.

Gröll M, Bajorek M, Köhler A, Moroder L, Rubin DM, Huber R *et al.* (2000). A gated channel into the proteasome core particle. *Nat Struct Biol* **7**: 1062–1067.

Gröll M, Ditzel L, Löwe J, Stock D, Botchler M, Bartunik H *et al.* (1997). Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **386**: 463–471.

Hake LE, Kuemmerle N, Hecht NB, Kozak CA (1994). The genes encoding the somatic and testis-specific isotypes of the mouse cytochrome *c* genes map to paralogous regions of chromosomes 6 and 2. *Genomics* **20**: 503–505.

Hiller M, Chen X, Pringle MJ, Suchorolski M, Sancak Y, Viswanathan S *et al.* (2004). Testis-specific TAF homologs collaborate to control a tissue-specific transcription program. *Development* **131**: 5297–5308.

Huh JR, Vernooy SY, Yu H, Yan N, Shi Y, Guo M *et al.* (2004). Multiple apoptotic caspase cascades are required in non-apoptotic roles for *Drosophila* spermatid individualization. *PLoS Biol* **2**: 43–53.

hwa JJ, Zhu AJ, Hiller MA, Kon CY, Fuller MT, Santel A (2004). Germ-line specific variants of components of the mitochondrial outer membrane import machinery in *Drosophila*. *FEBS Lett* **572**: 141–146.

Ivanov IP, Rohrwasser A, Terreros DA, Gesteland RF, Atkins JF (2000). Discovery of a spermatogenesis stage-specific ornithine decarboxylase antizyme: antizyme 3. *Proc Natl Acad Sci USA* **97**: 4808–4813.

Kalamegham R, Sturgill D, Siegfried E, Oliver B (2007). Drosophila mojoless, a retroposed GSK-3, has functionally diverged to acquire an essential role in male fertility. *Mol Biol Evol* **24**: 732–742.

Kalmykova AI, Shevelyov YY, Polesskaya OO, Dobritsa AA, Evstafieva AG, Boldyreff B *et al.* (2002). CK2(beta)tes gene encodes a testis-specific isoform of the regulatory subunit of casein kinase 2 in *Drosophila melanogaster*. *Eur J Biochem* **269**: 1418–1427.

Kimura Y, Takaoka M, Tanaka S, Sassa H, Tanaka K, Polevoda B *et al.* (2000). N(α)-acetlylation and proteolytic activity of the yeast 20S proteasome. *J Biol Chem* **275**: 4635–4639.

Kloetzel PM (2004). The proteasome and MHC class I antigen processing. *Biochim Biophys Acta* **1695**: 225–233.

Köhler A, Cascio P, Leggett DS, Woo KM, Goldberg AL, Finley D (2001). The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Mol Cell* **7**: 1143–1152.

Kraut DA, Prakash S, Matouschek A (2007). To degrade or release: ubiquitin-chain remodeling. *Trends Cell Biol* **17**: 419–421.

Li J, Rechsteiner M (2001). Molecular dissection of the 11S REG (PA28) proteasome activators. *Biochimie* **83**: 373–383.

Li X (2006). Functional Study of the Testis-specific Proteasome Subunits, Rpt3R and Rpt4R, of *Drosophila melanogaster*. MSc Thesis, Syracuse University.

Lynch M (2002). Genomics. Gene duplication and evolution. *Science* **297**: 945–947.

Ma J, Katz E, Belote JM (2002). Expression of proteasome subunit isoforms during spermatogenesis in *Drosophila melanogaster*. *Insect Mol Biol* **11**: 627–639.

Mason GG, Murray RZ, Pappin D, Rivett AJ (1998). Phosphorylation of ATPase subunits of the 26S proteasome. *FEBS Lett* **430**: 269–274.

McCarrey JR, Thomas K (1987). Human testis-specific *PGK* gene lacks introns and possesses characteristics of a processed gene. *Nature* **326**: 501–505.

Mikhaylova LM, Nguyen K, Nurminsky DI (2008). Analysis of the *Drosophila melanogaster* testes transcriptome revelas coordinate regulation of paralogous genes. *Genetics* **179**: 305–315.

Navon A, Goldberg AL (2001). Proteins are unfolded on the surface of the ATPase ring before transport into the proteasome. *Mol Cell* **8**: 1339–1349.

Pitnick S, Spicer GS, Markow TA (1995). How long is a giant sperm? *Nature* **375**: 109.

Rechsteiner M, Hill CP (2005). Mobilizing the proteolytic machine: cell biological roles of proteasome activators and inhibitors. *Trends Cell Biol* **15**: 27–33.

Saville KJ, Belote JM (1993). Identification of an essential gene, *l(3)73Ai*, with a dominant temperature-sensitive lethal allele, encoding a *Drosophila* proteasome subunit. *Proc Natl Acad Sci USA* **90**: 8842–8846.

Sharon M, Taverner T, Ambroggio XI, Deshaies RJ, Robinson CV (2006). Structural organization of the 19S proteasome lid: Insights from MS of intact complexes. *PLoS Biol* **4**: 11314–11322.

Smyth KA, Belote JM (1999). The dominant temperature-sensitive lethal mutant *DTS7* of *Drosophila melanogaster* encodes an altered 20S proteasome ß-type subunit. *Genetics* **151**: 211–220.

Sümegi M, Hunyadi-Gulyas E, Medzihradszky KF, Udvardy A (2003). 26S proteasome subunits are O-linked *N*-acetylglucosamine-modified in *Drosophila melanogaster*. *Biochem Biophys Res Commun* **312**: 1284–1289.

Tripoli G, D'Elia D, Barsanti P, Caggese C (2005). Comparison of the oxidative phosphorylation (OXPHOS) nuclear genes in the genomes of *Drosophila melanogaster*, *Drosophila pseudoobscura* and *Anopheles gambiae*. *Genome Biol* **6**: R11.

Unno M, Mizushima T, Morimoto Y, Tomisugi Y, Tanaka K, Yasuoka N et al. (2002). The structure of the mammalian 20S proteasome at 2.75 A resolution. *Structure* **10**: 609–618.

Voges D, Zwickl P, Baumeister W (1999). The 26S proteasome: a molecular machine designed for controlled proteolysis. *Annu Rev Biochem* **68**: 1015–1068.

Wang PJ, Page DC (2002). Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. *Hum Mol Genet* **11**: 2341–2346.

Whitby FG, Masters EI, Kramer L, Knowlton JR, Yao Y, Wang CC et al. (2000). Structural basis for the activation of 20S proteasomes by 11S regulators. *Nature* **408**: 115–120.

Yang P, Fu H, Walker J, Papa CM, Smalle J, Ju YM et al. (2003). Purification of the *Arabidopsis* 26S proteasome: Biochemical and molecular analyses revealed the presence of multiple isoforms. *J Biol Chem* **279**: 3410–3413.

Yuan X, Miller M, Belote JM (1996). Duplicated proteasome subunit genes in *Drosophila melanogaster* encoding testes-specific isoforms. *Genetics* **144**: 147–157.

Zhong L, Belote JM (2007). The testis-specific proteasome subunit Prosα6 T of *Drosophila melanogaster* is required for individualization and nuclear maturation during spermatogenesis. *Development* **134**: 3517–3525.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)