Comparative genomic analysis of novel conserved peptide upstream open reading frames in Drosophila melanogaster and other dipteran species
Celine A Hayden and Giovanni Bosco*

Address: Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ 85721, USA
Email: Celine A Hayden - chayden@email.arizona.edu; Giovanni Bosco* - gbosco@email.arizona.edu
* Corresponding author

Abstract

Background: Upstream open reading frames (uORFs) are elements found in the 5'-region of an mRNA transcript, capable of regulating protein production of the largest, or major ORF (mORF), and impacting organismal development and growth in fungi, plants, and animals. In Drosophila, approximately 40% of transcripts contain upstream start codons (uAUGs) but there is little evidence that these are translated and affect their associated mORF.

Results: Analyzing 19,389 Drosophila melanogaster transcript annotations and 666,153 dipteran EST sequences we have identified 44 putative conserved peptide uORFs (CPuORFs) in Drosophila melanogaster that show evidence of negative selection, and therefore are likely to be translated. Transcripts with CPuORFs constitute approximately 0.3% of the total number of transcripts, a similar frequency to the Arabidopsis genome, and have a mean length of 70 amino acids, much larger than the mean length of plant CPuORFs (40 amino acids). There is a statistically significant clustering of CPuORFs at cytological band 57 (p = 10^-5), a phenomenon that has never been described for uORFs. Based on GO term and Interpro domain analyses, genes in the uORF dataset show a higher frequency of ORFs implicated in mitochondrial import than the genome-wide frequency (p < 0.01) as well as methyltransferases (p < 0.02).

Conclusion: Based on these data, it is clear that Drosophila contain putative CPuORFs at frequencies similar to those found in plants. They are distinguished, however, by the type of mORF they tend to associate with, Drosophila CPuORFs preferentially occurring in transcripts encoding mitochondrial proteins and methyltransferases. This provides a basis for the study of CPuORFs and their putative regulatory role in mitochondrial function and disease.
levels in the cell [9]. For example, the yeast General Control Nondepressible 4 (GCN4) transcript contains multiple uORFs that differentially regulate the protein level of a transcription factor-encoding mORF under starvation and non-starvation conditions. In turn, the protein produced from the mORF, the GCN4 protein, is essential to the transcriptional activation of some 40 genes involved in amino acid biosynthesis [6]. Because uORFs influence the levels of mORF protein, it is not surprising that disruption of the uAUG can lead to human disease such as thrombocythemia [10], a disease which is thought to arise as a result of increased mORF protein product, thrombopoietin (TPO). In addition, uAUGs occur in transcripts coding for oncogenes more frequently than other mammalian transcripts [11]. Indeed, oncogenes Mdm2 [12], her-2 [13], MYEOV [14], Bcl-2 [15], and SCL [16], all contain uORFs that affect the level of oncoproteins produced.

Potentially thousands of genes are regulated via uORFs, but there are no demonstrated examples of uORFs affecting mORF protein production in Drosophila or other insect species. Several uORF-containing genes have been well studied in fungi, plants, and mammals [17] and genome-wide searches of conserved uORFs have been conducted using fungal, mammalian and plant transcripts [4,18-21]. Given the examples found in other eukaryotic species, it is plausible that uORFs fill a regulatory role in the arthropod lineage as well.

There is some evidence that regulatory uORFs may occur in insect species. Firstly, a Drosophila gene coding for a putative mannosyl transferase contains a uORF-mORF pair that seems to be evolutionarily conserved in insects [19]. Secondly, there are several examples of Drosophila dicistronic transcripts in which the first open reading frame could be regulatory to the second [22-24]. However, polycistronic transcripts do not all code for putative uORFs; many transcripts defined as polycistronic are initially transcribed as pre-mRNA with two or more ORFs, but are subsequently processed into separate monocistronic transcripts [25]. For this reason, we prefer to use the terminology ‘uORF’ to refer to an ORF (a) which is upstream of a mORF on a single mature mRNA, and (b) which is itself translated as a polypeptide distinct from protein translated from a mORF. In addition, polycistronic transcripts that are not processed into separate mRNA molecules are at times part of this uORF/mORF classification. The computational identification of dicistronic transcripts by Misra et al [22] resulted in the reannotation of 31 gene models, some of which may contain conserved uORF-mORF pairs. However, their search was limited to polycistronic transcripts with ORFs greater than 50 a.a., and it is known that uORF peptides as short as 6 a.a. can regulate mORF translation in mammals [26].

Their analysis also discarded overlapping ORFs, some of which are important for the regulation of mORFs [27].

To identify transcripts with uORFs that are likely to be translated, we took a comparative genomics approach using D. melanogaster transcript annotations, Anopheles gambiae transcript annotations, and dipteran expressed sequence tags (ESTs). Using this approach, we determined the prevalence, diversity, and genomic clustering of CpuORFs under negative selection in dipteran genomes and compared these findings to those reported for the plant lineage.

**Results and Discussion**

**Identification of conserved peptide uORFs in D. melanogaster**

To determine the prevalence of uORFs most likely to be translated, Drosophila melanogaster release 4.3 transcript sequences (19,389) were used to identify the largest, or major, ORF (mORF). Of these, 13,746 contain unique Flybase gene numbers, 5,851 of which contain one or more AUGs upstream of the mORF. This suggests that 43% of Drosophila mORF proteins could be affected in their expression level by translated uORFs. Our calculated percentage is slightly lower than previously reported Drosophila uAUG frequencies [2], but this discrepancy can be explained by the smaller dataset used in the previous study.

Putative dipteran homologs were found by comparing D. melanogaster mORFs to 666,153 NCBI ESTs using tBLASTn. Many of the EST sequences contained truncated uORF and mORF sequences, therefore the search was limited to species that diverged from D. melanogaster more than 15 Mya (non-melanogaster group species; AAA: 12 Drosophila Genomes Website) [28,29], to increase detection of negative selection acting on short protein sequences. For each pair of homologs, global alignment of uORFs identified candidate CpuORFs and $K_u/K_s$ ratios were used to further verify evolutionary conservation of the uORF amino acid sequence. In addition, Flybase transcript annotations were used to discard any genes in which the putative CpuORF was fused to the mORF in any given transcript splice variant.

$K_u/K_s$ ratios < 1 indicate that a sequence is under negative selection, $K_u/K_s$ ratios close to 1 imply that the sequence is undergoing drift, and $K_u/K_s$ ratios > 1 suggest that the sequence is under positive selection. We found a total of 44 CpuORFs with a $K_u/K_s$ ratio significantly less than one (Table 1; Additional File 1). Importantly, our $K_u/K_s$ ratio analysis distinguishes between high-scoring amino acid alignments that reflect conservation of nucleotide sequences versus alignments that reflect true evolutionary
Table 1: \(K_a/K_s\) values of uORF and associated mORFs correlated to most distantly related organism containing uORF-mORF association in an EST

| CG identifier | uORF  | mORF  | Most distantly related organism (NCBI accession #), most closely shared taxonomic classification with D. melanogaster\(^a\) |
|---------------|-------|-------|---------------------------------------------------------------------------------------------------------------------------------|
| CG18624       | 0.11***** | 0.06***** | Boomic\(^b\) (CV448373), Arthropoda                                                                                               |
| CG12664       | 0.26*** | 0.15*** | Drovir (EB568517), Drosophilida                                                                                                    |
| CG12788/CG17767 | 0.32**** | 0.28**** | Angam (CD747020), Diptera                                                                                                           |
| CG33713/CG33714 | 0.06**** | 0.13**** | Carmoe (DV250045), Pancrustacea                                                                                                   |
| CG3240        | 0.11**** | 0.11**** | Dromoj (EB613491), Drosophilida                                                                                                    |
| CG9960/CG9958 | 0.02**** | 0.07**** | Dapmog (DY937460), Pancrustacea                                                                                                   |
| CG31917       | 0.00**** | 0.09**** | Bonmor (DY230769), Endopterygota                                                                                                |
| CG31919/CG33995 | 0.10**** | 0.31*  | Glomor (DV616490), Schizophora                                                                                                     |
| CG18042       | 0.01**** | 0.29*   | Bonmor (AU003981), Endopterygota                                                                                                |
| CG7400        | 0.10**  | 0.06*** | Dropse (DR124033), Sophophora                                                                                                     |
| CG16974       | 0.00**  | 0.14**** | Dropse (DR133486), Sophophora                                                                                                     |
| CG4824        | 0.04*** | 0.17*** | Dropse (DR131819), Sophophora                                                                                                     |
| CG17325       | 0.08**** | 0.07**** | Drogri (EB611588), Drosophilida                                                                                                    |
| CG10570       | 0.28**  | 0.19**** | Drogri (EB601583), Drosophilida                                                                                                    |
| CG11508       | 0.13**** | 0.54**** | Glomor (DV620389), Schizophora                                                                                                     |
| CG8026        | 0.31**** | 0.04**** | Drogri (EB598775), Drosophilida                                                                                                    |
| CG17759 (uORF2) | 0.33*  | 0.02**** | Dromoj (EB608824), Drosophilida                                                                                                    |
| CG33671/CG33672 | 0.07**** | 0.14**** | Apimel (DB747777), Endopterygota                                                                                                   |
| CG6191        | 0.13**  | 0.05**** | Drogri (EB625487), Drosophilida                                                                                                    |
| CG30100       | 0.00**  | 0.09**** | Ixosca (DN974785), Arthropoda                                                                                                     |
| CG17725       | 0.00**  | 0.06**** | Droll (EB488086), Sophophora                                                                                                      |
| CG5469        | 0.10**** | 0.07**** | Aedoeg (EB099927), Diptera                                                                                                           |
| CG33786/CG33785 | 0.03**** | 0.16**** | Bonmor (EB992822), Endopterygota                                                                                                   |
| CG9865 (uORF1) | 0.12**** | 0.30**** | Droll (EB454746), Sophophora                                                                                                       |
| CG9865 (uORF2) | 0.07**** | 0.30**** | Aedoeg (DV278474), Diptera                                                                                                          |
| CG9865 (uORF3) | 0.04**** | 0.30**** | Acypis (CV847404), Neoptera                                                                                                         |
| CG9878        | 0.30**** | 0.04**** | Ixosca (AF483733), Arthropoda                                                                                                      |
| CG30290       | 0.00**** | 0.05**** | Carmoe (DY308116), Pancrustacea                                                                                                   |
| CG12016       | 0.12**** | 0.12**** | Acypis (CN762015), Neoptera                                                                                                         |
| CG32573       | 0.42**** | 0.19**** | Droll (EB501531), Sophophora                                                                                                       |
| CG11989       | 0.04**** | 0.01**** | Myzper (EE261505), Neoptera                                                                                                         |
| CG7869        | 0.09**** | 0.12**** | Dromoj (EB608881), Drosophilida                                                                                                    |
| CG7628        | 0.10**** | 0.03**** | Glomor (DV613491), Schizophora                                                                                                     |
| CG9666        | 0.29**** | 0.03**** | Artfra (BQ605225), Pancrustacea                                                                                                   |
| CG2128        | 0.24**** | 0.00**** | Bonmor (BY914486), Endopterygota                                                                                                   |
| CG9288        | 0.16**** | 0.17**** | Aedoeg (DV427990), Diptera                                                                                                          |
| CG9924        | 0.08**  | 0.12**** | Drobir (EB563704), Drosophilida                                                                                                    |
| CG31241       | 0.23**** | 0.00**** | Dromoj (EB603524), Drosophilida                                                                                                    |
| CG31178       | 0.31**** | 0.33**** | Drobir (EB560303), Drosophilida                                                                                                    |
| CG7071/CG34131 | 0.08**** | 0.20**** | Lutlon (AM099995), Diptera                                                                                                          |
| CG10238       | 0.29**** | 0.12**** | Taegut (DV959401), Coelomata                                                                                                       |
| CG5116        | 0.15**** | 0.16**** | Droll (EB489685), Sophophora                                                                                                       |
| CG14550       | 0.13*   | 0.21**** | Bonmor (CK362143), Endopterygota                                                                                                   |
| CG7950        | 0.35**** | 0.04**** | Myzper (EE263186), Neoptera                                                                                                         |

\(^a\) D. melanogaster taxonomic classification as described by NCBI

\(^b\) Abbreviations: Boomic, Boophilus microplus; Drovir, Drosophila virilis; Angam, Anopheles gambiæ; Carmoe, Carcinus maenas; Dromoj, Drosophila majovensis; Dapmog, Daphnia magna; Bonmor, Bambyx mori; Glomor, Glossina morsitans; Dropse, Drosophilida pseudobaculata; Drovir, Drosophila virilis; Drogri, Drosophilida grimshawi; Apimel, Apis mellifera; Ixosca, Ixodes scapularis; Droll, Drosophilida willistoni; Aedoeg, Aedes aegypti; Acypis, Acrhythosiphon pisum; Myzper, Myzus persicae; Artfra, Artemia franciscana; Lutlon, Lutzomyia longipalpis; Toepyg, Toeniopygia guttata

\(* p-value < 0.05; H_0: K_a/K_s = 1, H_A: K_a/K_s < 1\)

\(** p-value < 0.01\)

\(*** p-value < 0.001\)

\(**** p-value < 0.0001\)
conservation of the amino acid sequence, and therefore are good indicators of translation.

Another indicator of translation is start codon context. Based on nucleotide frequencies of sequences surrounding mORFs, it is predicted that the *Drosophila* optimal consensus sequence is CAAaAUACg [2,30], but no functional experiments have been conducted in insects to validate the strength of this initiation context. Therefore, although the predominant CpSuORF start context (AAaAUACa) seems to be weaker than the predominant mORF context, it remains to be determined whether ribosomes initiate efficiently at the uORF AUG. It is also quite likely that initiation of some CpSuORFs is dependent upon cellular conditions, as has been shown in various genes [6,31], leading to regulation of mORF protein levels.

A number of uORF-mORF pairs were used as positive controls for the modified uORF-Finder program. In a previous study, CG9865 was shown to contain a putative uORF-mORF pair that has been conserved among distantly related insect species [19]. This gene was identified by our analysis, therefore validating our approach. *Drosophila Tat-like (DTL)*, a gene containing a uORF with amino acid similarity in *D. melanogaster* and *D. pseudoobscura* [24] was also found by the uORF-Finder program. A third gene identified by our analysis, CG10238, is a bicistronic transcript encoding the small and large subunit of Molybdopterin synthase 2 (MOCS2) [23]. It is well conserved across distantly related eukaryotic species (see Additional File 2). In addition, 5 of the 31 dicistronic genes described by Misra et al [22] were shown to contain CpSuORFs (Table 2; denoted by Misra and colleagues as CG33071ORFA-CG33071ORFB, Tim9b-CG12788, CG33009ORFA-CG33009ORFB, CG33005ORFA-CG33005ORFB, and snapin-CG9960, but subsequently renamed CG33713-CG33714, CG12788-CG17767, CG33671-CG33672, CG33786-CG33785, and CG9960-CG9958, respectively). Many of the dicistronic transcripts identified by Misra et al [22] are transcripts with ORF pairs that are not well conserved among the *Drosophila* species. For example, the mei217-mei218uAUG is not conserved in any of the 11 other sequenced *Drosophila* genomes (UCSC *D. melanogaster* genome browser) [32], therefore it is not surprising that a number of the dicistronic genes were not identified by the uORF-Finder program. Additionally, it is likely that neither the *D. melanogaster* annotations nor the dipteran ESTs are representative of the complete transcript population within each species due to the incomplete annotation of 5' transcription start sites [33], and incomplete coverage of the genomes by ESTs.

Initially, 41 genes and 43 uORFs showed evidence of mild to strong purifying selection (*Ks*/Ka ratio significantly < 1), and an additional gene with one uORF was detected during subsequent duplication analysis (see below). The proportion of genes in the *Drosophila* genome showing evidence of CpSuORFs is approximately 0.3% (42 genes out of 14,040 genes), which is similar to the frequency predicted for the Arabidopsis genome (0.4–0.5%) [19]. The present study likely underestimates the prevalence of CpSuORFs due to incomplete EST resources and potentially misannotated 5' regions in *D. melanogaster*.

Consistent with calculated *Ks*/Ka values, the majority of CpSuORFs with a low *Ks*/Ka ratio are present in lineages beyond the Drosophilidae (Table 1) and therefore have been conserved more than 40 My (Assembly/Alignment/Annotation of 12 Drosophila species) [28,29]. Those uORFs that exhibit a low *Ks*/Ka ratio but are only found within Drosophila species may represent uORFs that have recently emerged within the Drosophila lineage but are nonetheless under mild to strong selection pressures.

**Insect CpSuORFs are longer in average length than plant CpSuORFs**

Two studies have shown that the length of a uORF can influence the ability of a ribosome to reinitiate scanning and translation initiation at a mORF [34,35]. The plant and mammalian cell systems used in these studies show that reinitiation at a downstream AUG is generally more efficient in the presence of shorter uORFs, and in plant protoplasts reinitiation drops sharply in constructs containing uORFs longer than 34 amino acids. Both studies were carried out using viral components, and as such it is not clear whether these observations extend to mRNAs in a native eukaryotic cellular environment. Nonetheless, uORF length could play an important role in the regulation of mORFs, therefore we analyzed Drosophila CpSuORFs in terms of their amino acid lengths. Initial characterization of the 44 putative CpSuORFs under negative selection reveals a wide distribution of lengths, ranging from 15 to 179 amino acids (Table 2, Figure 1A).

To date, most, if not all, functionally characterized uORFs are smaller than 100 amino acids, but more than one fourth (12/44) of *D. melanogaster* CpSuORFs are above this size. In general, Drosophila CpSuORFs seem to be larger than those found in plants. While 83% of Arabidopsis CpSuORFs are between 21 and 60 amino acids in length (mean of 40 amino acids ± 16 standard deviation; Figure 1B), the Drosophila uORF length distribution peaks between 41 and 80 amino acids (mean of 76 amino acids ± 44; Figure 1A). These plant and insect datasets were not generated by comparing species with the same evolutionary distance but a more convincing comparison can be made by analyzing uORFs that have been conserved over more than 200 My: between Arabidopsis and rice, and between Drosophila and non-Brachycera lineages (e.g. *Anopheles*). The Arabidopsis distribution peak remains...
essentially unchanged under these restrictions (mean of 39 amino acids ± 13), whereas the distribution of Drosophila uORFs peaks at an even greater length, 81–100 amino acids (mean of 92 amino acids ± 29; Figure 1C). Longer uORF lengths in Drosophila may reflect a need for stronger suppression of mORF translation than in plants, consistent with the observations of the above-mentioned cell culture studies. Alternatively, insect cells may exhibit more efficient reinitiation resulting in a requirement for longer uORFs to attenuate mORF translation.

| Flybase transcript identifier and uORF number (FBtrXXXXX_#) | CG identifier | Cytological gene location | uORF length (a.a.) |
|-------------------------------------------------------------|---------------|---------------------------|---------------------|
| FBtr0071140_1                                               | CG18624       | 7C2-7C2                   | 54                  |
| FBtr0071349_3                                               | CG12664       | 8C11-8C13                 | 41                  |
| FBtr0074767_3                                               | CG12788/C17767| 18D3-18D7                 | 117                 |
| FBtr0077227_1                                               | CG33713/C33714| 19F4-19F4                 | 90                  |
| FBtr0077747_1                                               | CG3240        | 23A1-23A1                 | 179                 |
| FBtr0077737_2                                               | CG9960/C9958b | 23A3-23A3                 | 134                 |
| FBtr0079037_2                                               | CG31917       | 25C1-25C1                 | 73                  |
| FBtr0079006_1                                               | CG31919/C33995b| 25C1-25C1                | 44                  |
| FBtr0079695_3                                               | CG18042       | 29D4-29D5                 | 85                  |
| FBtr0080133_1                                               | CG7400        | 31F4-31F5                 | 20                  |
| FBtr0080489_1                                               | CG16974       | 34A8-34A8                 | 21                  |
| FBtr0080803_5                                               | CG4824        | 35E2-35E2                 | 44                  |
| FBtr0081102_1                                               | CG17325       | 37A4-37A5                 | 48                  |
| FBtr008122_5                                               | CG10570       | 37A4                      | 50                  |
| FBtr0088187_5                                               | CG11508       | 44B3-44B3                 | 150                 |
| FBtr0088610_3                                               | CG8026        | 45B3-45B3                 | 48                  |
| FBtr0087829_3                                               | CG17759       | 49B8-49B9                 | 31                  |
| FBtr0091650_2                                               | CG33671/C33672b| 49B10-49B10               | 86                  |
| FBtr0087678_3                                               | CG6191        | 50B3-50B4                 | 21                  |
| FBtr0087140_1                                               | CG30100       | 53B1-53B1                 | 70                  |
| FBtr0086701_1                                               | CG17725       | 55D3-55D3                 | 27                  |
| FBtr0086654_7                                               | CG5469        | 55E5-55E5                 | 121                 |
| FBtr0091786_1                                               | CG33786/C33785b| 57A8-57A9                 | 108                 |
| FBtr0071680_7                                               | CG9865 (uORF1) | 57F7-57F7                 | 65                  |
| FBtr0071680_5                                               | CG9865 (uORF2) | 57F7-57F7                 | 84                  |
| FBtr0071680_4                                               | CG9865 (uORF3) | 57F7-57F7                 | 76                  |
| FBtr0071676_1                                               | CG9878        | 57F8-57F8                 | 65                  |
| FBtr0071672_1                                               | CG30290       | 57F8-57F9                 | 94                  |
| FBtr0073063_4                                               | CG12016       | 63D1-63D1                 | 81                  |
| FBtr0074315_3                                               | CG32573       | 14F5-14F5                 | 109                 |
| FBtr0076348_2                                               | CG11989       | 67D2-67D2                 | 50                  |
| FBtr0076203_3                                               | CG7869        | 68A4-68A4                 | 68                  |
| FBtr0076213_1                                               | CG7628        | 68A7-68A8                 | 18                  |
| FBtr0074991_5                                               | CG9666       | 76A3-76A3                 | 129                 |
| FBtr0078767_1                                               | CG2128        | 83A4-83A4                 | 38                  |
| FBtr0082829_3                                               | CG9288        | 87F13-87F13               | 80                  |
| FBtr0082871_2                                               | CG9924        | 88A3-88A4                 | 25                  |
| FBtr0083570_4                                               | CG31241       | 90F11-90F11               | 178                 |
| FBtr0084138_3                                               | CG31178       | 93F14-93F14               | 40                  |
| FBtr0084211_1                                               | CG7071/C34131b| 94A6-94A6                 | 157                 |
| FBtr0084782_2                                               | CG10238       | 96C1-96C1                 | 90                  |
| FBtr0084877_1                                               | CG5116        | 96E2-96E2                 | 15                  |
| FBtr0084974_2                                               | CG14550       | 96F10-96F10               | 111                 |
| FBtr0085563_1                                               | CG7950        | 99D3-99D3                 | 111                 |

* Gene with multiple CPuORFs in the same 5'UTR

b Different gene identifiers annotated as producing the same transcript; the first CG identifier predicts the translation of the mORF and the second CG identifier predicts the translation of the uORF.

Physical mapping reveals clustering of CPuORFs independently of gene duplication In insect and mammalian genomes, clusters of closely related genes can sometimes occur, such as the Hox gene clusters [36]. To determine whether genes with uORFs...
Figure 1
Conserved peptide uORF length distribution. A. A total of 44 CPuORFs identified in Drosophila melanogaster. B. CPuORFs in Arabidopsis thaliana as described by Hayden and Jorgensen [19]. C. CPuORFs conserved between D. melanogaster and non-Brachycera species.
cluster in certain parts of the genome, the 44 uORFs were placed on the *D. melanogaster* cytological map (Table 1) and compared to a random distribution (Methods). uORF frequencies were not statistically different from a randomly generated dataset except for a cluster of 6 uORFs residing on band 57 (p-value = 10^{-5}), five of which fall on a much smaller segment of the chromosome, band 57F. Upon closer examination, some of these uORFs may have arisen as a result of tandem duplications; one uORF found in the CG30290 transcript as well as two uORFs found in the CG9865 transcript (uORF1 and uORF3) all contain twin CX_9C motifs. Interestingly, the observed clustering is not dependent upon the putative duplication events of CX_9C motif-containing uORFs. Eliminating the duplication bias by collapsing CX_9C-containing uORFs to one representative, clustering is still statistically significant, with 4 uORFs on cytological band 57 (p-value = 0.004) and 3 uORFs on band 57F (p-value = 0.0002). Therefore, the data suggest that there is a preponderance of both clustering and duplicate retention of uORFs on band 57. Clustering at this region could be an indicator of chromatin interactions at this site that could mediate CPuORF regulation.

The twin CX_9C motif is an integral part of coiled-coil helix, coiled-coil helix (CHCH) domains, a domain previously implicated in uORF-mORF associations in group 8 plant uORFs [19]. In fact, the group 8-like *Drosophila* uORF member described in the plant study is uORF3 of CG9865. It is interesting to note that the plant group 8 uORF has consistently lost its duplicate copy during both recent and ancient polyploidy events whereas the *Drosophila* group 8 putative homologue may be retaining its twin CX_9C motifs. Interestingly, the observed clustering is not dependent upon the putative duplication events of CX_9C motif-containing uORFs. Eliminating the duplication bias by collapsing CX_9C-containing uORFs to one representative, clustering is still statistically significant, with 4 uORFs on cytological band 57 (p-value = 0.004) and 3 uORFs on band 57F (p-value = 0.0002). Therefore, the data suggest that there is a preponderance of both clustering and duplicate retention of uORFs on band 57. Clustering at this region could be an indicator of chromatin interactions at this site that could mediate CPuORF regulation.

**CPuORF-mORF pair duplicate retention is low within Drosophila melanogaster**

To determine whether there has been retention of uORF-mORF pair duplicates within the *Drosophila* genome itself, the 41 mORFs with strongly conserved uORFs were compared to the *D. melanogaster* transcriptome. A single gene, CG17325 showed evidence of a duplicate copy, CG10570, in which the uORF-mORF pair is conserved (See Additional File 3). CG10570 was not detected by our program due to the short length of its mORF (< 100 amino acids), therefore this gene was added to our list of CPuORF pairs following our duplication analysis (Tables 1 and 2). CG17325 and CG10570 reside adjacent to one another on chromosome 2, band 37A4-A5, and are transcribed on opposite strands away from one another. The close proximity of the genes suggests a segmental duplication gave rise to the two genes, both of which are conserved throughout the *Drosophila* lineage and exhibit a K_i/K_s ratio < 0.28 (Table 1). This duplication presumably occurred more than 40 Mya since both loci are present in *D. melanogaster*, *D. grimshawi*, and *D. virilis*. Unlike the extensive uORF-mORF duplication retention history of the Arabidopsis genome, CG17325 and CG10570 were the only example of gene duplicate retention in *Drosophila*.

**GO term and protein domain analysis suggest a link between CPuORFs and both mitochondrial proteins and methyltransferases**

Further differences between plant and insect CPuORFs were observed following gene ontology (GO) term analysis. GO term frequencies in the *D. melanogaster* genome were compared to frequencies in the insect uORF dataset to look for overrepresentation of terms. P-values were determined using the Bonferroni correction method, a method that accounts for multiple comparisons and calculates a conservative p-value. Also, the recent tandem duplicate (see above) was not included in the analysis to eliminate bias from recent duplication events. Because GO terms have been assigned to all ORFs found in bicistronic transcripts, GO terms were extracted for both uORF and mORF gene identifiers, designated hereafter as the uORF dataset (41 mORFs and 7 uORFs). This analysis differs from previous analyses in plants; it not only identifies 1) classes of mORF proteins that tend to associate with CPuORFs, but it also identifies 2) ORFs that preferentially associate with other ORFs on a single transcript. In plants, a large proportion of CPuORFs associate with mORFs encoding transcription factors, however this trend was not observed in insects. Instead, mORF proteins showing evidence of N-methyltransferase activity (GO term for CG9666 and CG9960 mORFs; Table 3) tend to associate with CPuORFs (p = 0.02). This methyltransferase activity may act on DNA or RNA, since both types of Interpro domains are overrepresented in these two genes.

Additionally, overrepresentation of GO term ‘protein import into the mitochondrial inner membrane’ is driven by two proteins in the Drosophila uORF dataset, CG9878 (Translocase of inner membrane 10, Tim10) and CG17767 (Tim9b), which contain the Interpro Zn-finger Tim10/DDP-type domain (p = 0.01). Unlike the overrepresented methyltransferase domain, the Tim10/DDP-type domain is not limited to the mORFs, but appears in either the uORF or mORF, demonstrating that these ORFs show a preference for associating with other ORFs in a transcript. Specifically, Tim10 is encoded by the mORF of its transcript while Tim9b is encoded by the uORF. This does not imply that Tim9b does not act as a regulatory uORF, however. Tim9b may act both as a chaperone in the intermembrane space, as well as a regulatory element controlling the translation of its associated mORF.
In support of a model in which mitochondrial proteins preferentially associate with other ORFs on a single transcript, a further connection to the mitochondrial inner membrane is found when examining other genes in the uORF dataset. The CG8026 mORF encodes a putative mitochondrial folate transport protein [37,38] (Table 4). Interestingly, this trend may extend to the mammalian lineage, exemplified by the human Uncoupling protein 2 (UCP2) mORF, a putative inner mitochondrial membrane transporter. The UCP2 mORF is not only associated with what appears to be a CPuORF, but it is regulated by its uORF in a glutamine-dependent manner [39]. B-cell lymphoma 2 (BCL-2) is another mammalian oncogene that produces a protein from its mORF, BCL-2, which is localized to mitochondria [40] and is associated with a functional uORF [15].

Other Drosophila genes also have potential links to the mitochondrion, such as CG18624, a putative NADH dehydrogenase that is predicted to act in mitochondrial electron transport (Table 4). Also, uORF1 of CG9865 is a putative homolog of p8Mature T-Cell Proliferation 1 (p8MTCP1), an ORF that is transcribed on the same mRNA as p13MTCP1, is targeted to mitochondria [41], and may play a role in oncogenesis [42,43]. CG9865 uORF1 has a twin CX8C motif, as do p8MTCP1 and other proteins targeted to mitochondria, namely yeast proteins Mitochondrial Ribosomal Protein 10 (Mrp10p) [44], Cytochrome Oxidase 19 (Cox19p) [45], Cytochrome Oxidase 17 (Cox17p) [46], and Mitochondrial intermembrane space Import and Assembly 40 (Mia40p) [47]. In humans, the twin CX8C motif found in Mia40p is required for import and stable accumulation of Mia40 in the intermembrane space [48]. Several genes in the uORF dataset contain ORFs with CX8C motifs, such as uORFs 1 and 3 of CG9865, the uORFs of CG30290 and CG9288, and the mORF of CG7950 (See Additional File 2). These open reading frames could be interacting with other ORFs on the same transcript to target them to the mitochondria or to form a stabilizing protein complex.

It is possible that these ORF associations are vestiges of ancient prokaryotic operons that originated in the mitochondrion and were transferred to the nuclear genome over time. This hypothesis runs counter to the prevailing thought that mitochondrial proteins involved in transport are generally of eukayotic origin [49]. Regardless of their origin, nuclear ORFs coding for mitochondrial proteins may maintain an association with other ORFs on a single transcript over long periods of evolutionary time for several reasons. Both ORFs may be co-regulated at the transcriptional level and be required at similar times in development, thus providing more efficient transcription of DNA. Alternatively, the uORF may be regulating expression of the mORF with important biological consequences. These possibilities are not mutually exclusive and further experimentation will be required to determine whether this energy-producing organelle is influenced by the translational regulation of uORF-mORF pairs on single transcripts.

Interestingly, the trend in animal mitochondrial ORFs was not observed in plants. Instead, plant uORFs tend to associate with mORFs encoding transcription factors [19]. Perhaps these unique characteristics reflect fundamental differences in the two eukaryotic lineages. Despite their differences, plants and animals both seem to contain uORF-mORF pairs involved in a wide range of biochemical and regulatory pathways (Table 4). There is some evidence in the literature that transcripts with uORFs can occur in similar biochemical pathways, such as genes affecting the polyamine biochemical pathway [50], but this is the exception rather than the rule and no additional examples have been born out by our analyses. To facilitate future studies of these elements, all CPuORF annotations will be submitted to Flybase.

**Conclusion**

The identification and characterization of putative CPuORFs has established a knowledge base from which many hypotheses have been generated and can now be...
Table 4: Predicted function and biological processes of uORF-mORF pairs in Drosophila

| CG identifier | Gene name synonymsa | Inferred functionb | Inferred biological processb | Supporting evidence |
|---------------|---------------------|-------------------|-----------------------------|---------------------|
| CG18624       | Putative NADH dehydrogenase | Mitochondrial electron transport | Pfam domain; GO term designation |
| CG12664       | ld14, fend⁹         | Unknown           | Neuromuscular development   | [61, 62]            |
| CG12788/CG17767⁹ | Tim9b (uORF)        | Mitochondrial inner membrane translocase subunit (uORF) | Transport across mitochondrial inner membrane (uORF) | Interpro domain |
| CG33713/CG33714⁹ |     | Acyl-CoA binding (mORF) RNA binding (uORF) | Unknown | Interpro domain |
| CG3240        | Rad1⁹               | Putative 3'->5' exonuclease activity | DNA repair | [63, 64] |
| CG9960/CG9958⁹ | snapin (uORF)       | Putative methyltransferase (mORF) Putative Biogenesis of lysosome-related Organelles Complex-I-like (BLOC-I-like) subunit (uORF) | Biogenesis of lysosome-related organelles (eg, melanosomes and platelet dense granules; uORF) | [65] (uORF) Interpro domain (mORF) |
| CG31917       | TFBS (uORF)         | Putative TFIIH subunit (uORF) | Transcription and DNA repair (uORF) | [66, 67]; Interpro domain |
| CG31919/CG33995⁹ |     | Ankyrin repeat, protein-protein interactions | Target of transcription factor Glial cells missing (Gcm), involved in neuronal development and function | Interpro domain; [68] |
| CG18042       | lmgb                | Putative component of Anaphase Promoting Complex (uORF) | Mitosis; Neural development (unclear whether it is the uORF, mORF or both) | [69, 70]; Flybase personal communication FBrf0125046; [71]; NCBI Conserved Domain Search |
| CG7400        | Fatp⁹               | Putative very-long-chain fatty acyl-CoA synthetase | Fatty acid metabolism | [72] |
| CG16974       | Member of LIG superfamilyp | Leucine-rich repeat and Immunoglobulin domain-containing protein | Unknown | [73, 74] |
| CG4824        | Bic⁹               | RNA binding protein | Anterior-Posterior patterning | [75–77] |
| CG17325       | Unknown             | Unknown           | Unknown                      | [75] |
| CG10570       | Unknown             | Unknown           | Transcription of splicing factors | [78] |
| CG11508       | DmSNAP50, DmPBP49⁹ | Subunit of an snRNA transcriptional activator protein | Mitochondrial folate transport | [37, 38] |
| CG8026        | Mitochondrial carrier protein | Photoreceptor signal transduction; Axonal guidance | Unknown | [79–81] |
| CG17759⁹ (uORF2) | Galpha49B, Gq12 ⁹ | G-protein subunit | Isoprenoid production (mORF) | [82, 83] |
| CG33671/CG33672⁹ |     | Mevalonate kinase (mORF); BoLA-like protein, putative nucleic acid binding protein (uORF) | Unknown | [84–86] |
| CG6191        | Unknown             | Unknown           | Unknown                      | GO term designation |
| CG30100       | Pepck⁹              | Translation release factor | Translation termination | [84–86] |
| CG17725       | Putative phosphoenolpyruvate carboxykinase | Gluconeogenesis; Starvation; Glyceroneogenesis | Unknown | [87]; FBrf0189302 |
| CG5469        | Gint3⁹              | Ubiquitin regulatory X domain (UBX), putative RNA binding | Unknown | Interpro domain |
| CG33786/CG33785⁹ |     | Unknown           | Translation (mORF) | Interpro domain |
| CG9865⁹ (uORF1) | Putative mannosyl transferase | Unknown | Interpro domain |
| CG9865⁹ (uORF2) | Putative mannosyl transferase | Unknown | Interpro domain |
| CG9865⁹ (uORF3) | Putative mannosyl transferase | Unknown | Interpro domain |
| CG9878        | Tim10⁹              | Putative inner mitochondrial membrane translocase | Protein transport across mitochondrial membrane | [88] |
| CG30290       | Putative flavoprotein enzyme | Unknown | Interpro domain |
| CG12016       | Unknown             | Unknown           | Unknown                      | Interpro domain |
| CG32573       | Unknown             | Unknown           | Unknown                      | Interpro domain |
tested. CPuORFs in dipterans show similarities to their plant counterparts in terms of their prevalence within the genome and diversity of sequence, but differ in their greater average length, their genome clustering, and their preferential association with methyltransferases. In addition, the present analysis has shown a significant correlation between mitochondrially-targeted proteins and transcripts containing uORFs, an observation that could lead to important discoveries impacting our understanding of human disease. Given the wealth of genetic tools available in Drosophila, this model system is ideally suited to the basic understanding of uORF-containing transcripts and post-transcriptional regulation.

Methods
Identification of conserved peptide uORFs

Drosophila melanogaster transcript sequences, release 4.3 (19,389 sequences) were downloaded from Flybase [51]. Anopheles gambiae transcript sequences, build 3.4 (14,127 sequences) were downloaded from Ensembl [52], and dipteran expressed sequence tags (ESTs) (666,153) were downloaded from NCBI [53] December 15, 2006. Because the melanogaster group members (includes D. simulans, D. yakuba, D. erecta, and D. ananassae) diverged from D. melanogaster relatively recently [28,29], their transcript sequences are of limited use in detecting strong negative selection over short sequence lengths due to the accumulation of few synonymous and non-synonymous substitutions. Therefore these species were excluded from this first comparison, as were D. melanogaster ESTs.

Comparative analysis of D. melanogaster and A. gambiae sequences was performed using uORF-Finder [19], a program that identifies the longest open reading frame of a transcript in the first species (defined as the mORF), finds the putative homolog in the second species, and aligns all open reading frames upstream of these homologs to identify putatively conserved uORFs. uORF-Finder was designed to compare full-length cDNA sequences from two species, therefore to accommodate a D. melanogaster full-length transcript-to-dipteran EST comparison, the program was modified and putative homologs in the ESTs were identified using the first 100 amino acids of the D. melanogaster mORFs. uORF size was also limited to 200 amino acids (no additional uORFs were found when uORF size was limited to 300 a.a.).

The presence of putative CPuORFs was established in at least three different species by either extracting the first 100 amino acids of the D. melanogaster mORF sequence and searching the NCBI EST database using tBLASTn for putative homologs with conserved uORF sequences, or by scanning the UCSC D. melanogaster genome browser and inspecting other Drosophila genomes for conservation of

| uORF Name | mORF Name | Putative Function | Biological Process |
|-----------|-----------|------------------|--------------------|
| CG11989   | ArdlA    | Putative N-Acetyltransferase catalytic subunit | Unknown |
| CG7869    | SuURb    | DNA binding      | Unknown |
| CG7628    | Hdacbc   | Histone deacetylase | Unknown |
| CG9666    | Putative methyltransferase | Wing development; Chromatin remodeling |
| CG2128    | CG1298   | CG31241          | CG34131           |
| CG9288    | CG9924   | Rdxb             | Unknown |
| CG31241   | DTLbc    | Putative RNA methylase | Late larval development |
| CG31178   | Unknown   | Unknown           | Unknown |
| CG10238   | CG14550  | CG5116           | CG7975            |
| CG11989   | CG9924   | CG31241          | CG34131           |

* refers to mORF unless otherwise noted

**References:**
[1] Drosophila melanogaster transcript sequences, release 4.3 (19,389 sequences) were downloaded from Flybase [51].
[2] Anopheles gambiae transcript sequences, build 3.4 (14,127 sequences) were downloaded from Ensembl [52].
[3] Dipteran expressed sequence tags (ESTs) (666,153) were downloaded from NCBI [53] December 15, 2006.
[4] The presence of putative CPuORFs was established in at least three different species by either extracting the first 100 amino acids of the D. melanogaster mORF sequence and searching the NCBI EST database using tBLASTn for putative homologs with conserved uORF sequences, or by scanning the UCSC D. melanogaster genome browser and inspecting other Drosophila genomes for conservation of...
uORF start and stop codons [32]. Any putative uORF sequences that showed evidence of in-frame fusion with the mORF on the UCSC browser (in an alternative splice form, for example) were not included in the final list of CPuORF-containing transcripts.

**Calculation of \( K_a/K_s \)**

The \( K_a/K_s \) ratio was determined using pairwise_kaks.PLS (version 1.7) [54] and is derived from the highest scoring BLAST homolog in the *D. melanogaster*-dipteran high scoring pairs. Both the approximate method (option -kaks yn00) and the maximum likelihood method (-kaks codeml) were used. Only the approximate method calculation is reported in Table 1 due to the typically short evolutionary distance between the organisms found in the highest scoring BLAST pairs. The Nei-Gojobori p-distance and alignment of putatively duplicated D. melanogaster genes, an NIH R01 grant to GB (GM069462).

**References**

1. Cavener DR, Cavener BA: Translation start sites and mRNA leaders. In An Atlas of Drosophila genes Edited by: Maroni G. New York, Oxford University Press; 1993:359-377.
2. Rogozin IB, Kochetov AV, Kondrashov FA, Koonin EV, Milanesi L: Presence of ATG triplets in 5’ untranslated regions of
eukaryotic cDNAs correlate with a 'weak' context of the start codon. Bioinformatics 2001, 17:890-900.

3. Churbanov A, Rogozin IB, Babenko VN, Rzhetsky A, Koonin EV: Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes. Nucleic Acids Res 2005, 33:5512-5520.

4. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Bastaftome M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Szczaszczocziak C, Farman M, Butler J, Purcell S, Harris S, Braun GS, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Penalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Kochi K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA, Birren BW: Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae. Nature 2005, 438:1105-1115.

5. Kawaguchi R, Bailey-Serres J: Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated region of Mrna sequence features that contribute to translational regulation in Arabidopsis. Nucleic Acids Res 2005, 33:955-965.

6. Hinnebusch AG: Translational regulation of yeast GCN4. A window on factors that control initiation-tRNA binding to the ribosome. J Biol Chem 1997, 272:21661-21664.

7. Law GL, Raney A, Heusner C, Morris DR: Polymamine regulation of ribosome pausing at the upstream open reading frame of S-adenosylmethionine decarboxylase. J Biol Chem 2001, 276:38036-38043.

8. Hensley C, Elliott KA, Franceschetti M, Mayer MJ, Illingworth C, Michael AJ: A dual upstream open reading frame-based autoregulatory circuit controlling polyamine-responsive translation. J Biol Chem 2005, 280:39229-39237.

9. Wiese A, Etzinger N, Wobbes B, Smeekens S: A conserved upstream open reading frame mediates sucrose-induced repression of translation. Plant Cell 2004, 16:1717-1729.

10. Ghiardi N, Wiestner A, Vorien K, Ohsaka A, Skoda RC: Hereditary thrombocythaemia in a Japanese family is caused by a novel point mutation in the thrombopoietin gene. Br J Haematol 2009, 144:285-292.

11. Kozak M: An analysis of 5'-noncencing sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 1987, 15:8125-8148.

12. Jin X, Turcott E, Englehardt S, Mize GJ, Morris DR: The two upstream open reading frames of oncogene mdm2 have different translational regulatory properties. J Biol Chem 1999, 274:25716-25721.

13. Child SJ, Miller MK, Geballe AP: Cell type-dependent and independent control of HER-2/neu translation. Int J Biochem Cell Biol 2002, 34:201-213.

14. Alves de Almeida R, Heuser T, Blaschke R, Bartram CR, Janssen JW: Control of MYEOV protein synthesis by upstream open reading frames. J Biol Chem 2006, 281:695-704.

15. Pratt MA, White D, Kushwaha N, Tibbo E, Niu MY: Cytoplasmic mutant p53 induces Bcl-2 expression in estrogen receptor-positive breast cancer cells. Apoptosis 2007, 12:657-669.

16. Calhoun CF, Muller C, Martin R, Kroml G, Pietsch H, Hwang T, Leutz A: Translational control of SCL isoform expression in hematopoietic lineage choice. Genes Dev 2003, 17:959-964.

17. Geballe AP, Sachs MS: Translational control by upstream open reading frames. In Translational control of gene expression edited by: Sonenberg N, Hershey JWB and Mathews MB. Cold Spring Harbor, New York, CSHL Press: 2000:595-614.

18. Crowe ML, Wang XQ, Rothnagel JA: Evidence for conservation and selection of upstream open reading frames suggests probable encoding of bioactive peptides. BMC Genomics 2006, 7:16.

19. Hayden CA, Jorgensen RA: Identification of novel conserved peptide uORF homology groups in Arabidopsis and rice reveal ancient eukaryotic origin of select groups and preferential association with transcription factor-encoding genes. BMC Biol 2007, 5:32.

20. Neufsey DE, Galagan JE: "Dual Modes of Natural Selection on Upstream Open Reading Frames". Mol Biol Evol 2007, 24(4):1744-51. Epub 2007 May 9.

21. Zhang Z, Dietrich FS: Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated regions (UTR) of genes in Saccharomyces cerevisiae. Curr Genet 2005, 48:77-87.
44. Jin C, Myers AM, Tzagoloff A: Cloning and characterization of MRP10, a yeast gene coding for a mitochondrial ribosomal protein. Biochim. Biophys. Acta 1997, 1353:228-234.

45. Nobrega MP, Bandeira SC, Beers J, Tzagoloff A: Characterization of COX19, a widely distributed gene required for expression of mitochondrial cytochrome oxidase. J Biol Chem 2002, 277:40206-40211.

46. Myers J, Glaser DM, Tzagoloff A: Purification, characterization, and localization of yeast Cox17p, a mitochondrial copper shuttle. J Biol Chem 1997, 272:33191-33196.

47. Mesecke N, Terzisika N, Kozany C, Baumann F, Neupert W, Hell K, Herrmann JM: A disulfide relay system in the intermembrane space of mitochondria that mediates protein import. Cell 2005, 121:1059-1069.

48. Hofmann S, Rothbauer U, Muhlenbein N, Baiker K, Hell K, Bauer MF: Functional and mutational characterization of human MIA40 acting during import into the mitochondrial intermembrane space of Drosophila. J Biol Chem 2005, 280:517-528.

49. Gray MW, Burger G, Lang BF: The origin and early evolution of mitochondria. Genome Biol 2001, 2:REVIEW0118.

50. Chang KS, Lee SH, Hwang SB, Park KY: Interpro domain database. [ftp://ftp.ebi.ac.uk/pub/databases/interpro/entry.list]

51. Ensembl Anopheles resources. [http://flybase.org]

52. Pairwise KaKis program. [http://flybase.org/static_pages/down loads/builddata7.html]

53. NCBI. [http://flybase.org]

54. GO term association files. [http://flybase.org/static_pages/down loads/builddata7.html]

55. Biomart. [http://flybase.org/static_pages/down loads/builddata7.html]

56. Drosophila GO term files. [http://flybase.org/static_pages/down loads/builddata7.html]

57. Drosophila GO term files. [http://flybase.org/static_pages/down loads/builddata7.html]

58. Drosophila GO term files. [http://flybase.org/static_pages/down loads/builddata7.html]

59. Umemiya T, Takasu E, Takeichi M, Aigaki T, Nose A: Cloning and characterization of a putative SCF ubiquitin ligase complex in Drosophila. Proc Nat Acad Sci U S A 1998, 95:7532-7537.

60. Tchurikov NA, Kretova OV, Chernov BK, Golova YB, Zhimulev IF, Zykov IA: The Orct gene of Drosophila melanogaster codes for a putative organic cation transporter with six or 12 transmembrane domains. Gene 1997, 201:69-74.

61. Kraut R, Menon K, Zinn K: A gain-of-function screen for genes controlling motor axon guidance and synaptogenesis in Drosophila. J Neurobiol 2001, 47:407-420.

62. Obi Y, Sato M, Ojioka M, Inouye S: Enzymatic and genetic characterization of firefly luciferase and Drosophila CG6178 as a fatty acyl-CoA synthetase. Biosci Biotechnol Biochem 2005, 69:819-828.

63. Vogel C, Teichmann SA, Chothia C: The immunoglobulin superfamily in Drosophila melanogaster and Caenorhabditis elegans and the evolution of complexity. Development 2003, 130:6317-6328.

64. MacLaren CM, Evans TA, Alvarado D, Duffy JB: Comparative analysis of the Kellken molecule-related ligands of the LIG superfamily. Dev Genes Evol 2004, 214:360-366.

65. Mohler J, Wieschaus EF: Dominant maternal-effect mutations of Drosophila melanogaster causing the production of double-abdomen embryos. Genetics 1986, 112:803-822.

66. Drapkin R, Reardon JT, Ansari A, Huang JC, Zawel L, Ahn K, Sancar A: Mammalian Rad9, Hus1 and Rad17 participate in sliding clamp and DNA replication. J Biol Chem 1997, 272:13891-13897.

67. Li C, Harding GA, Parise J, McNamara-Schroeder KJ, Stumpf WE: Architectural arrangement of cloned proximal sequence element protein binding subunits on Drosophila U1 and U6 snRNA gene promoters. Mol Cell Biol 2004, 24:1897-1906.

68. Lee YJ, Shah S, Suzuki F, Zars T, O’Day PM, Hyde DR: The Drosophila dgg gene encodes a G alpha protein that mediates phototransduction. Neuron 1994, 13:1143-1157.

69. Goldring ML, Herman-Borgmeyer I, Grenningloh G, Zopf D: A genetic screen for suppressors of Drosophila NSF2 neuromuscular disease. Genetics 2003, 163:835-843.

70. Lander ES, Stylker GM, Altshuler D, Pollack P, Ecker J, Nusbaum C, Lills A, Zody M, Kallali M, Heath S: Systematic identification of novel protein domain families associated with nuclear functions. Genome Res 2002, 12:47-56.

71. Webb CT, Gorman MA, Lazarou M, Ryan MT, Gulbis JM: Crystal structure of the mitochondrial chaperone TIM9. Nature 2000, 408:769-773.

72. Ranish JA, Hahn S, Lu Y, Yi EC, Li XJ, Eng J, Aebersold R: Identification of a complex of TIM-23 and two Hsp70 homologs in mammalian mitochondria. J Biol Chem 1999, 274:25769-25776.

73. Takahashi A, Nakamura K, Iwatsuki K, Tanaka A, Inoue Y, Nishida E, Kuroda K, Hirose Y, Inoue T, Shimizu Y: Suppression of food intake and growth by amino acids in Drosophila: the role of pum, a fat body expressed gene with homology to vertebrate glycine cleavage system. Development 1999, 126:5275-5284.

74. Okamura T, Shimizu H, Nagaoka A, Ueda R, Ishii S: ATF-2 regulates fat metabolism in Drosophila. Mol Biol Cell 2007, 18:1519-1529.

75. Doerks T, Copley RR, Schultz J, Ponting CP, Bork P: Comparative analysis of novel protein domain families associated with nuclear functions. Genome Res 2002, 12:47-56.

76. Williams BC, Garrett-Engele CM, Li Z, Williams GV, Rosenman ED, Goldberg ML: Two putative actintransferases, san and deco, are required for establishing sister chromatid cohesion in Drosophila. Curr Biol 2005, 15:2033-2038.

77. Belyaeva ES, Zhimulev IF, Volkova EA, Alekseyenko AA, Moshkin YM, Koryakov DE: Su(NUR)ES: a gene suppressing DNA underreplication in intercalary and pericentric heterochromatin of Drosophila melanogaster polytene chromosomes. Proc Natl Acad Sci U S A 1998, 95:7532-7537.

78. Tchurikov NA, Kretova OV, Chernov BK, Golova YB, Zhimulev IF, Zykov IA: SuUR protein binds to the boundary regions separating founder domains in Drosophila melanogaster. J Biol Chem 2004, 279:1705-1710.
92. Marenda DR, Zraly CB, Dingwall AK: The Drosophila Brahma (SWI/SNF) chromatin remodeling complex exhibits cell-type specific activation and repression functions. Dev Biol 2004, 267:279-293.

93. Pandey R, Muller A, Napoli CA, Selinger DA, Pikaard CS, Richards EJ, Bender J, Mount DW, Jorgensen RA: Analysis of histone acetyltransferase and histone deacetylase families of Arabidopsis thaliana suggests functional diversification of chromatin modification among multicellular eukaryotes. Nucleic Acids Res 2002, 30:5036-5055.

94. Kent D, Bush EW, Hooper JE: Roadkill attenuates Hedgehog responses through degradation of Cubitus interruptus. Development 2006, 133:2001-2010.

95. Nybakken K, Vokes SA, Lin TY, McMahon AP, Perrimon N: A genome-wide RNA interference screen in Drosophila melanogaster cells for new components of the Hh signaling pathway. Nat Genet 2005, 37:1323-1332.

96. Reiss J, Dorche C, Stallmeyer B, Mendel RR, Cohen N, Zabot MT: Human molybdopterin synthase gene: genomic structure and mutations in molybdenum cofactor deficiency type B. Am J Hum Genet 1999, 64:706-711.