DOI: 10.1534/genetics.109.104844

Conservation of Domain Structure in a Fast-Evolving Heterochromatic SUUR Protein in Drosophilids

Anna A. Yurlova, Igor V. Makunin, Tatyana D. Kolesnikova, Olga V. Posukh, Elena S. Belyaeva and Igor F. Zhimulev[1]

Department of Molecular and Cellular Biology, Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

Manuscript received May 8, 2009
Accepted for publication June 28, 2009

ABSTRACT

Different genomic regions replicate at a distinct time during S-phase. The *SuUR* mutation alters replication timing and the polytenization level of intercalary and pericentric heterochromatin in *Drosophila melanogaster* salivary gland polytene chromosomes. We analyzed *SuUR* in different insects, identified conserved regions in the protein, substituted conserved amino acid residues, and studied effects of the mutations on SUUR function. *SuUR* orthologs were identified in all sequenced drosophilids, and a highly divergent ortholog was found in the mosquito genome. We demonstrated that SUUR evolves at very high rate comparable with that of Transformer. Remarkably, upstream ORF within 5′ UTR of the gene is more conserved than SUUR in drosophilids, but it is absent in the mosquito. The domain structure and charge of SUUR are maintained in drosophilids despite the high divergence of the proteins. The N-terminal part of SUUR with similarity to the SNF2/SWI2 proteins displays the highest level of conservation. Mutation of two conserved amino acid residues in this region impairs binding of SUUR to polytene chromosomes and reduces the ability of the protein to cause DNA underreplication. The least conserved middle part of SUUR interacting with HP1 retains positively and negatively charged clusters and nuclear localization signals. The C terminus contains interlacing conserved and variable motifs. Our results suggest that SUUR domains evolve with different rates and patterns but maintain their features.

IT is well established that replication timing in the S-phase generally correlates with the preceding transcriptional activity of the chromatin domain (Schubeler *et al.* 2002; MacAlpine *et al.* 2004; Donaldson 2005). As a rule, late replication is observed in transcriptionally silent and condensed chromosome regions, mostly composed of pericentric heterochromatin (PH). Late-replicating regions in euchromatin are represented by 100- to 200-kb chromatin domains (MacAlpine *et al.* 2004; White *et al.* 2004), which are often denoted as foci of late replication in the interphase nuclei (Berezney *et al.* 2000).

One of the peculiar features that advances *Drosophila melanogaster* as a model for studying late replication is its giant larval salivary gland polytene chromosomes that enable easy and precise identification of late-replicating regions. There are ~240 regions showing late replication apart from the PH in *D. melanogaster* polytene chromosomes. These regions are scattered over the euchromatic chromosome arms and also display characteristic features of heterochromatin, such as dense packaging and low transcription level (Zhimulev *et al.* 2003a). Many late-replicating regions are underreplicated. Underreplication results from the early start of the G-phase before the S-phase is actually complete; hence many late-replicating chromosome sequences fail to complete replication by the end of each endocycle (Gall *et al.* 1971; Smith and Orr-Weaver 1991; Lilly and Spradling 1996). Morphologically, underreplication in these regions appears as "weak spots" or chromosome breaks on polytene chromosome squashes, which serve as a convenient cytological marker of late replication and underreplication. These regions are collectively referred to as intercalary heterochromatin (IH). Many of them are known to be bound by repressive Pc-G protein complexes and are mainly composed of deeply silenced genes (Zhimulev and Belyaeva 2003; Zhimulev *et al.* 2003a; Belyakin *et al.* 2005).

Underreplication is also known to be significantly affected by a product of the *SuUR* gene. This gene encodes a protein that is specifically associated with PH and IH (Makunin *et al.* 2002; Zhimulev *et al.* 2003b; Pindyurin *et al.* 2007). The only known mutation of the gene, $SuUR^{ES}$, is caused by an ~6-kb insertion into the last exon (Makunin *et al.* 2002). $SuUR^{ES}$ larvae show altered replication timing in late-replicating regions. Namely, replication in these regions completes sooner than in the wild-type strain, so the polytenization level in

Supporting information is available online at http://www.genetics.org/cgi/content/full/genetics.109.104844/DC1.

[1]*Corresponding author:* Department of Molecular and Cellular Biology, Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Lavrentiev Ave. 8, 630090 Novosibirsk, Russia. E-mail: zhimulev@bionet.nsc.ru

IH is restored to that of the euchromatin. This is also accompanied by an increase in the degree of polytenization of many sequences in PH and by the concomitant structuring of the chromocenter (Belyaeva *et al.* 1998; Moshkin *et al.* 2001; Zhimulev *et al.* 2003a). Conversely, an increase in the *SuUR* gene copy number enhances the underreplication in IH regions (Zhimulev *et al.* 2003a; Belyakin *et al.* 2005). Ectopic expression of SUUR in follicular cells suppresses the amplification of chorion gene clusters (Volkova *et al.* 2003). Finally, strong SUUR overexpression in third instar larval salivary glands leads to structural changes ("swellings") in chromosome morphology of PH and IH regions (Zhimulev *et al.* 2003c).

SuUR gene has four exons and a very short promoter region devoid of recognizable regulatory elements in addition to two presumptive E2F-binding sites (Makunin *et al.* 2002). Recently, an upstream open reading frame (uORF) was identified in the 5′ UTR of *SuUR* (Hayden and Bosco 2008). The gene encodes a 962-aa protein without any homologs reported in protein databases (Makunin *et al.* 2002). Nevertheless, the N terminus of the protein shows moderate similarity to the ATPase/helicase domain of chromatin-remodeling proteins from the SWI2/SNF2 group. ATP-dependent chromatin-remodeling factors are known to serve as molecular motors that alter the accessibility of DNA in chromatin, thereby regulating many aspects of transcription and replication (Havas *et al.* 2001). While the strongest similarity between SUUR and SNF2/SWI2 proteins is observed within Walker A and Walker B motifs involved in ATP binding and hydrolysis (Walker *et al.* 1982), the SUUR sequence differs significantly from the canonical motifs (Makunin *et al.* 2002). It is unknown whether SUUR could bind and hydrolyze ATP, but the fragment containing the first 360 amino acid residues shows a dominant-negative effect and displaces endogenous SUUR from polytene chromosomes (Kolesnikova *et al.* 2005).

Previously, we demonstrated that the C-terminal fragment $SUUR_{495–962}$ controls underreplication, although it is unable to induce structural changes in chromatin when overexpressed. On the contrary, the N-terminal fragments $SUUR_{1–599}$ and $SUUR_{1–779}$ had no effect on endoreplication, but were able to bind PH and IH regions and to induce formation of chromosome swellings such as the full-length SUUR (Kolesnikova *et al.* 2005). Here we demonstrate that SUUR is present in other Drosophila species and that it affects the break formation in salivary gland polytene chromosomes in *D. simulans*. Comparative analysis of SUUR in 11 Drosophila species showed that the protein belongs to a group of fast-evolving genes although its domain organization is conserved in Drosophila. We introduced targeted point mutations in two conserved regions within N- and C-terminal parts of SUUR and analyzed how these substitutions affect the protein function. We showed that point mutations in the N-terminal region of SUUR abolish its specific binding to the late-replicating regions of polytene chromosomes and decrease the ability of the protein to suppress polytenization in these regions. We also performed a more precise functional mapping of the C-terminal region of SUUR, which was known to cause underreplication.

MATERIALS AND METHODS

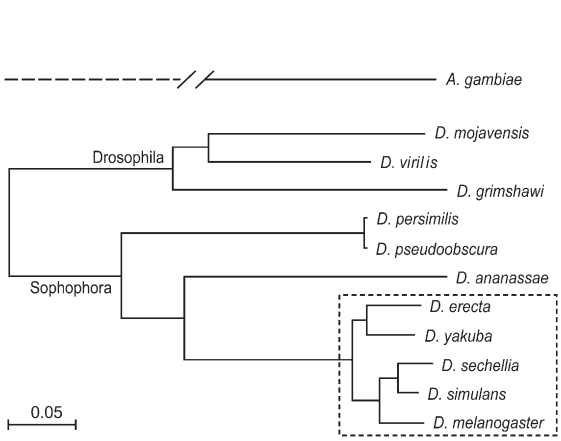
Drosophila stocks and genetics: Fly stocks were kept on standard Drosophila cornmeal medium at 25°. The following stocks carrying GAL4 drivers were used: *da*-GAL4 for ubiquitous expression (Wodarz *et al.* 1995), *Sgs3*-GAL4 for expression in salivary glands starting from the mid-third instar (PS1–PS11) (Cherbas *et al.* 2003), *AB1*-GAL4 for expression in salivary glands from early embryogenesis (Drysdale *et al.* 2005), *arm*-GAL4 for weak variegated expression in salivary glands (Kolesnikova *et al.* 2005), and *C323*-GAL4 for expression in follicle cells (Manseau *et al.* 1997). The *w; $SuUR^{ES}$* stock was described in Belyaeva *et al.* (1998). Oregon-R was used as a wild-type stock. We used *D. erecta* and *D. virilis* from the laboratory stock collection and *D. ananassae* (strain 14024-0371.13) from the Tucson Drosophila Stock Center.

Molecular procedures: All molecular procedures were performed as described in Sambrook and Russell (2001). DNA-modifying enzymes were purchased from New England Biolabs. Genomic DNA from *D. erecta* was amplified by PCR using *SuUR*-specific primers, and sequencing of PCR products was done on the ABI PRISM 310 Genetic Analyzer (Applied Biosystems) at the DNA Sequencing Center of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia (http://sequest.niboch.nsc.ru). RNA from ovaries was isolated using Trizol (Gibco BRL). A RT–PCR kit (Promega) was used according to the manufacturer's recommendations. Primer sequences used for the RT–PCR and amplification of *SuUR* genomic sequences in different Drosophila species are available upon request.

Constructs for transformations: Transgenic constructs are described in the supporting information (File S1). Plasmids were co-injected with pUChsΔ2-3wc (pTURBO) transposase helper plasmid into *y^1 w^{67}* embryos, and several independent transgenic lines were obtained for each construct (Rubin and Spradling 1982).

Immunostaining of polytene chromosomes: Indirect immunofluorescent analysis of polytene chromosomes was performed as described in Poux *et al.* (2001). We used E45 antibodies raised against the middle part of SUUR (Makunin *et al.* 2002) and antibodies against hemagglutinin tag (HA) provided by V. Pirrotta. The E45 antibodies and HA antiserum were used at a 1:50 and 1:10 dilutions, respectively. For immunostaining of *D. simulans* polytene chromsomes and analysis of chromosome binding of $SUUR_{Nmut}$ protein, the double-squash approach was used when a positive control was present on the same slide with the investigated polytenes.

Data analysis: We used the BLAT program (Kent 2002; http://genome.ucsc.edu) to map the *SuUR* orthologs in the genomic sequences available. Multiple protein alignments were constructed using ClustalW (Thompson *et al.* 1994; http://www.ebi.ac.uk/clustalw). K-Estimator 6.1v software (Comeron 1999; http://en.bio-soft.net/format/KEstimator.html) was used to calculate the number of synonymous (K_s) and nonsynonymous (K_a) substitutions. The SAPS program was used for the statistical analysis of protein sequences (Brendel *et al.* 1992; http://www.isrec.isb-sib.ch/software/SAPS_form.html). The phylogenetic tree was built in MEGA4



	Size, aa	Charge	Weight, kDa
A. gambiae	732	-11	81,0
D. mojavensis	1035	+39	114,8
D. virilis	1024	+45	114,3
D. grimshawi	1014	+43	112,6
D. persimilis	989	+38	110,7
D. pseudoobscura	986	+41	110,4
D. ananassae	959	+34	107,6
D. erecta	962	+46	107,5
D. yakuba	965	+40	107,5
D. sechellia	957	+42	107,1
D. simulans	954	+44	106,9
D. melanogaster	962	+45	107,6

Figure 1.—Phylogenetic tree based on SUUR proteins and the main properties of each protein. The tree was built in MEGA4 software using protein alignment with the following settings: neighbor-joining method, complete deletion of gaps, Poisson correction, and uniform rates among sites. Species from the *melanogaster* subgroup are boxed with a dashed line.

(Kumar *et al.* 2008). Identification of protein motifs and structure predictions were performed using MotifScan (http://myhits.isb-sib.ch/cgi-bin/motif_scan) and Predict Protein (Rost *et al.* 2004; http://www.predictprotein.org).

RESULTS

Identification of SUUR protein in insects: Southern blot hybridization of *SuUR* cDNA with genomic DNAs from various Drosophila species produced signals in species from the *melanogaster* subgroup only (data not shown). Among these, *D. erecta* was one of the most distant species from *D. melanogaster* (Figure 1). We amplified and sequenced the genomic DNA from the *SuUR* locus in *D. erecta* (GenBank accession no. AJ539550). The exon–intron structure of the *SuUR* gene in *D. erecta* was confirmed by comparison of the genomic sequence and the sequence of *SuUR* cDNA fragment obtained from *D. erecta* total ovarian RNA by RT–PCR. Splice sites are conserved between *D. melanogaster* and *D. erecta*.

We also used predicted SUUR sequences from nine recently sequenced Drosophila species for which genomic sequences are available at the UCSC Genome Browser website (Figure 1). We noted that the annotations of the *SuUR* gene produced by some annotation projects differ significantly from the exon–intron structure of the gene in *D. melanogaster* in five of nine species: *D. simulans*, *D. yakuba*, *D. ananassae*, *D. persimilis*, and *D. virilis*. The differences include the prediction of an additional exon in the 5′-end of the gene, which merged the uORF with the main ORF, and the prediction of additional introns and lack thereof, notably by the Genescan annotations (Burge and Karlin 1997). We determined the exon–intron structure of *SuUR* in *D. yakuba*, *D. ananassae*, and *D. virilis* by sequencing RT–PCR products obtained from total fly RNA. Sequences of PCR products confirmed discrepancies in Genescan annotation; therefore, we used our version of *SuUR* annotation. For *D. simulans* and *D. persimilis*, we transferred annotation of *SuUR* from the closely related species *D. melanogaster* and *D. pseudoobscura*, respectively. Sequences of *SuUR* ORFs used in this study are given in Figure S1. These data confirmed the integrity of the conserved uORF predicted in *SuUR* 5′ UTR for all analyzed species (Hayden and Bosco 2008). Our review of *SuUR* annotation in Drosophila species demonstrates that the computer gene annotations should be used with great care.

The *SuUR* gene is not annotated outside of Drosophila. However, the BLAST search identified a weak similarity (~25% identities, *E*-value $7e^{-6}$) with ENSANGP00000027713.1 protein from *Anopheles gambiae* (contemporary gene name AGAP005819; coordinates: chr2L:21,832,968–21,835,239; AgamP3 genome assembly). The similarity was limited to the N-terminal region of SUUR (aa 51–276). The rest of the protein sequence in *A. gambiae* is highly diverged, making comparison of the full-size proteins impossible. In contrast to the Drosophila SUUR, the ENSANGP00000027713.1 protein has no negatively or positively charged regions in the middle part of the protein. There is no apparent uORF upstream of the ENSANGP00000027713.1 main ORF. In *D. melanogaster*, the *CG6310* gene is located downstream of *SuUR*. Similarly, the mosquito *CG6310* homolog ENSANGT00000010378.2 (contemporary name AGAP005820; chr2L:21,835,318–21,836,802; AgamP3 genome assembly) is located downstream of ENSANGP00000027713.1, indicating that the latter indeed represents a highly diverged version of Drosophila *SuUR*. As Drosophila SUUR, the mosquito protein contains noncanonical Walker A and Walker B motif sites: it has substitution in GKT sequence from the putative nucleotide-binding loop and in DExH box. The predicted mosquito protein is smaller and has a negative charge while Drosophila SUUR has a positive total charge (Figure 1). We were unable to identify SUUR orthologs in other sequenced nondipteran insect species.

SUUR contributes to the formation of chromosome breaks in *D. simulans*: In many Drosophila species, the salivary gland polytene chromosomes display specific chromosome breaks and constrictions due to underreplication (Zhimulev 1998), suggesting that SUUR contributes to underreplication in these species. Immunostaining of *D. simulans* salivary gland polytene chro-

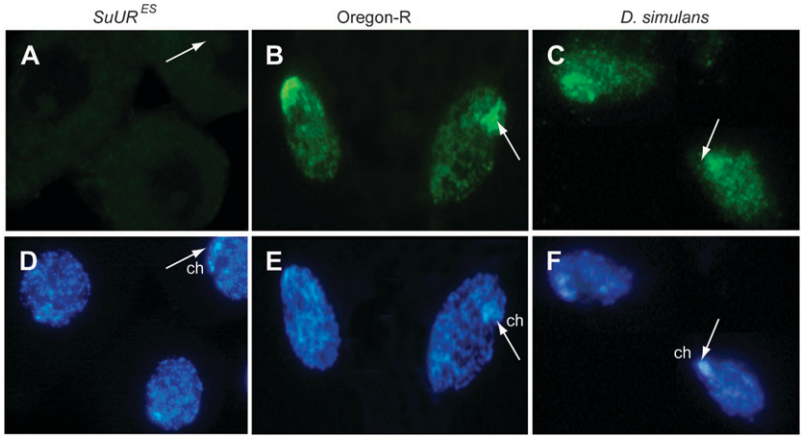


Figure 2.—Immunodetection of SUUR in follicular cells with E45 antibodies. (A) Negative control: no staining is observed in follicle cells of $SuUR^{ES}$. (B) Positive control: strong staining is observed in the nucleus and especially the chromocenter (ch, indicated by arrows) in follicle cells of *D. melanogaster* (Oregon-R). (C) In *D. simulans*, the antibodies produce staining similar to that observed in Oregon-R. (D–F) Hoechst staining.

mosomes with anti-SUUR antibodies does not produce any pronounced pattern. A chromocenter-specific signal could be detected only in rare nuclei. Immunofluorescent analysis of SUUR localization on polytene chromosomes of *D. melanogaster* × *D. simulans* hybrid larvae shows absence of staining in ~90% of nuclei, with ~10% of nuclei demonstrating a staining pattern characteristic of *D. melanogaster*. Notwithstanding, we did observe strong SUUR staining in follicle-cell nuclei preparations of whole-mount ovaries of *D. simulans* (Figure 2C). In both wild-type *D. melanogaster* and *D. simulans* the antibodies produce staining throughout the nucleus with a strong signal immediately adjacent to the chromocenter while virtually no staining is observed in the $SuUR^{ES}$ mutant (Figure 2).

To prove that SUUR does contribute to underreplication in *D. simulans*, we crossed *AB1*-GAL4>UAS-$SuUR_{1–458}$ *D. melanogaster* females with *D. simulans* males. The *AB1*-GAL4>UAS-$SuUR_{1–458}$ transgenic combination provides expression of the N-terminal half of SUUR in salivary glands from an early developmental stage. Expression of this fragment ($SUUR_{1–458}$) under early *AB1*-GAL4 driver has a dominant-negative effect, and results in the complete disappearance of weak spots from the polytene chromosomes, similar to the *SuUR* mutant phenotype (Kolesnikova *et al.* 2005). Consistently, overexpression of $SUUR_{1–458}$ leads to the disappearance of weak spots on both homologs in *D. melanogaster* × *D. simulans* hybrid progeny (Figure S2). It needs to be pointed out that the chromosomes of both *D. simulans* and *D. melanogaster* (Oregon-R) × *D. simulans* hybrids demonstrate weak spots in the same regions as *D. melanogaster*. This result argues in favor of a common mechanism of weak spot formation in both species and that the SUUR protein has a key role in this process.

SUUR orthologs display high levels of substitutions in different Drosophila species: Comparison of SUUR orthologs from Drosophila species revealed high numbers of amino acid substitutions, insertions, and deletions, even in closely related species (Table 1, Figure S3). Strikingly, the level of amino acid conservation is much higher within uORF, which encompasses 68 residues in *D. melanogaster*, than is observed for SUUR main ORF (Table 1). We calculated the numbers of synonymous (K_s) and nonsynonymous (K_a) substitutions per site for the species from the *melanogaster* subgroup, using K-Estimator software (Comeron 1999) (Table 2). We excluded distantly related species from this analysis because of the ambiguity in alignment, especially in the middle part of the protein (see below). The number of nonsynonymous substitutions per site in the *SuUR* gene between *D. melanogaster* and *D. yakuba* is 0.052. This is very similar to the K_a value characteristic for the fast-evolving genes in Drosophila (Schmid and Tautz 1997). The size and charge of the SUUR protein are retained in the course of evolution despite the high substitution rate (Figure 1). The secondary structure predictions even in very distant species, such as *D. melanogaster* and *D. grimshawi*, turned out to be mostly identical. Numerous helices and extended sheets were predicted in the N-terminal part while the rest of the protein was less structured (Figure S4).

The phylogenetic tree created for available SUUR proteins is fairly consistent with the tree obtained in genomewide analysis (Stark *et al.* 2007); *e.g.*, *D. yakuba* and *D. erecta* are grouped together, and the *D. pseudoobscura* branch is shorter (Figure 1). While the Drosophila subgenus branch appeared somewhat longer in the SUUR tree in comparison to the whole-genome tree, it could be just a consequence of a rooting problem: unfortunately, the sequence of the *D. willistoni* genome was not available at the UCSC Genome Browser website at the time of our analysis, and use of SUUR sequence from this species could affect the position of the tree root, and hence could affect the length of the Drosophila subgenus branch.

Distribution of substitutions across the protein is nonuniform (Figure 3A, Figure S3). The N-terminal region of SUUR is the most conserved part of the protein. It has a relatively low level of substitutions, no insertions, and no deletions even in distantly related Drosophila species. The middle part of SUUR (*D.*

TABLE 1

Pairwise comparison of amino acid identities for SUUR protein and uORF

	mel	sim	sec	yak	ere	ana	pse	per	vir	moj	gri
mel		93.1	91.9	89.3	87.7	65.4	60.0	59.9	48.7	46.4	44.7
sim	100		95.2	88.8	87.6	66.2	61.0	60.7	49.3	46.8	45.6
sec	98.5	98.5		87.5	86.2	65.0	60.1	59.6	48.8	46.8	45.3
yak	100	100	98.5		90.7	66.2	61.0	60.8	48.9	46.7	45.8
ere	92.9	92.9	91.5	92.9		66.0	61.1	60.8	49.2	45.8	45.4
ana	76.4	76.4	76.4	76.4	70.4		59.6	59.5	48.4	45.2	46.1
pse	66.1	66.1	64.7	66.1	60.5	69.1		99.0	50.7	48.3	49.3
per	66.1	66.1	64.7	66.1	60.5	69.1	100		50.9	48.3	49.3
vir	63.2	63.2	61.7	63.2	57.7	63.0	70.7	70.7		81.2	63.8
moj	57.3	57.3	57.3	57.3	52.1	59.0	70.1	70.1	64.5		58.3
gri	57.3	57.3	57.3	57.3	52.1	60.0	63.6	63.6	82.5	79.6	

Identity for SUUR protein is shown above the diagonal, and identity for uORF from *SuUR* transcript is shown below the diagonal. mel, *D. melanogaster*; sim, *D. simulans*; sec, *D. sechellia*; yak, *D. yakuba*; ere, *D. erecta*; ana, *D. ananassae*; pse, *D. pseudoobscura*; per, *D. persimilis*; vir, *D. virilis*; moj, *D. mojavensis*; gri, *D. grimshawi*.

melanogaster residues 280–581) shows the lowest level of amino acid identity across species (Table 3). Two distantly related species from the subgenus Drosophila, *D. mojavensis* and *D. virilis*, have long insertions in this region of SUUR (Figure S3). Despite an extremely high level of primary sequence divergence, the negatively and positively charged regions located in the middle part (Figure 3B) consistently maintained their properties in other species (Table 3). For example, the SAPS program (http://www.isrec.isb-sib.ch/software/SAPS_form.html) predicts statistically significant spacing between positively charged residues on the sides of a negatively charged region of SUUR in *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. pseudoobscura*, and *D. persimilis*. In *D. mojavensis*, *D. virilis*, and *D. grimshawi*, this negatively charged region is interrupted by a single positively charged residue (Figure S3). In *D. ananassae*, both negatively and positively charged clusters are smaller (Table 3). Intriguingly, these regions display a very similar total charge in different SUUR orthologs, although the vast majority of the charged residues *per se* are not conserved (Figure S3).

The middle part of SUUR largely coincides with the region (aa 339–671) known to interact with another heterochromatic protein, HP1, in the yeast two-hybrid assay (Pindyurin *et al.* 2008). Surprisingly, this part undergoes very rapid evolution. Even the sequence that displays similarity to the HP1-interacting motif (LRVSL, aa 429–433; Pindyurin *et al.* 2008) diverged significantly in Drosophila species (Figure S3), with only two species, *D. yakuba* and *D. erecta*, containing this motif unaltered.

A search for the known protein motifs in SUUR using PredictProtein and MotifScan identified type I or type II nuclear localization signals (NLS) in the middle part of the protein in all species, except for *D. mojavensis* in which no bipartite NLS was found. In addition to NLS, a motif homologous to the AT hook was present in the middle of the protein in five species, from *D. melanogaster* to *D. ananassae*. No known motifs were identified within the C-terminal part of SUUR, although this region encompassed alternating stretches of conserved and nonconserved sequences (Figure S3).

Targeted mutagenesis of SUUR and the effect of mutations on endoreplication: On the basis of the protein alignment, we substituted conserved amino acid residues in two regions of SUUR protein. In the N-terminal part, we introduced L57R/G58R substitutions ($SUUR_{Nmut}$) within a conserved region with similarity to the Walker A motif of ATPase/helicase domain (Figure 3, B and C). It has previously been established that ectopic expression of truncated SUUR protein containing amino acids 1–779 ($SUUR_{1–779}$) does not suppress endoreplication in salivary glands (Figure 3D), indicating that the protein domain (or its crucial part) involved in the suppression of endoreplication is located downstream of the nonsense mutation in $SUUR_{1–779}$. Therefore, we substituted two invariably conserved amino acid residues within this region, F816S and F817D, to obtain $SUUR_{Cmut}$. Mutated ORFs were cloned into the pUAST vector (Brand and Perrimon 1993), and several independent transformants were generated for each construct.

TABLE 2

Synonymous (above diagonal) and nonsynonymous (below diagonal) substitutions per site in *SuUR* coding region

	mel	sim	sec	yak	ere
mel		0.146	0.175	0.324	0.289
sim	0.031		0.086	0.295	0.285
sec	0.038	0.020		0.324	0.301
yak	0.052	0.055	0.061		0.220
ere	0.058	0.056	0.066	0.042	

mel, *D. melanogaster*; sim, *D. simulans*; sec, *D. sechellia*; yak, *D. yakuba*; ere, *D. erecta*.

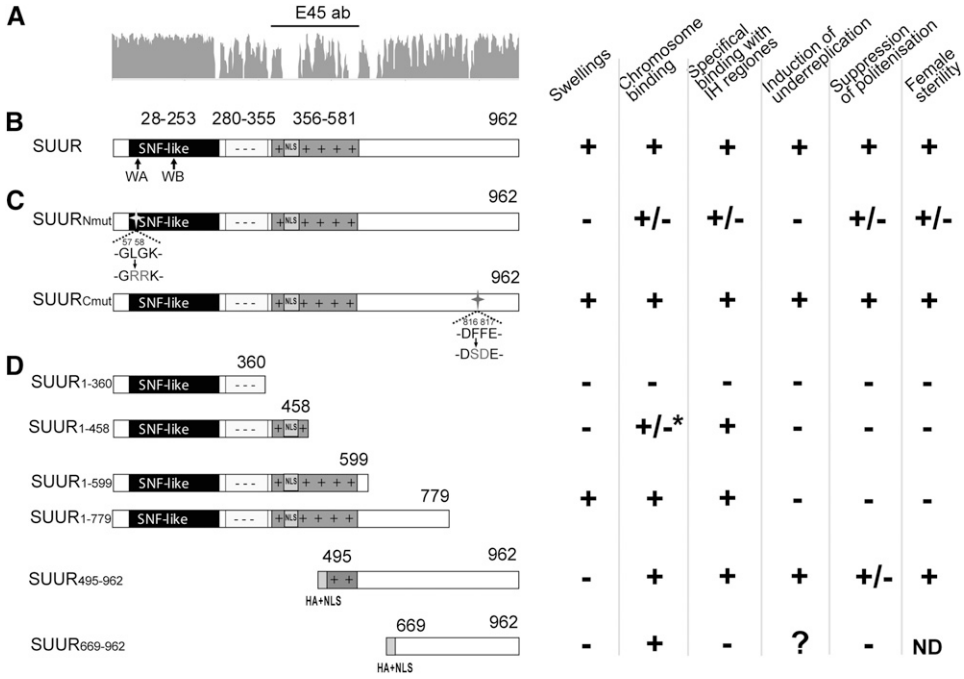


FIGURE 3.—Organization and features of different SUUR isoforms. (A) Conservation plot, based on the SUUR sequences from 11 Drosophila species, and the protein region used to generate E45 antibodies (E45 ab). (B) Domain organization of SUUR protein. The SNF-like domain is solid, the negatively charged region is marked with - -, and the positively charged amino acid region is darkly shaded with + +. NLS, nuclear localization signal. Arrows show the position of regions with similarity to Walker A (WA) and Walker B (WB) motifs of SNF2/SWI2 proteins. (C) Point mutations introduced within the N- and C-terminal part of SUUR are indicated by arrows. (D) Truncated fragments of SUUR are described by KOLESNIKOVA *et al.* (2005) except for $SUUR_{669–962}$. HA+NLS, hemagglutinin tag and nuclear localization signal sequence. An asterisk indicates that granules in intercalary and pericentric heterochromatin were observed.

We examined the effects of ectopically expressed mutated proteins in the UAS-GAL4 system (BRAND and PERRIMON 1993). Permanent strong expression of SUUR in salivary glands under the *AB1*-GAL4 driver suppresses endoreplication and results in miniature salivary glands (VOLKOVA *et al.* 2003). Expression of $SUUR_{Nmut}$ under the control of *AB1*-GAL4 causes only partial suppression of endoreplication. The nuclei of salivary glands from *AB1*-GAL4>UAS-$SuUR_{Nmut}$ larvae are larger than those with ectopic expression of the full-length SUUR from *AB1*-GAL4>UAS-*SuUR* larvae, although they are smaller than Oregon-R salivary gland nuclei (Figure 4).

Ectopic expression of UAS-*SuUR* in follicular cells under the control of the *C323*-GAL4 driver suppresses amplification of chorion genes and results in complete female sterility (VOLKOVA *et al.* 2003). When $SUUR_{Nmut}$ was ectopically expressed under the *C323*-GAL4 driver, we observed weak suppression of the female sterile phenotype (20 crosses were set for 10 independent

TABLE 3

Conservation of charge for middle part

	Negative charge region			Positive charge region		
Species	Position	Identity	Charge	Position	Identity	Charge
mel	280–355		−17	356–581		+39
sim	280–350	86.8	−16	351–573	85.3	+37
sec	280–350	85.5	−16	351–576	84.0	+35
yak	280–356	80.5	−18	357–583	78.0	+35
ere	280–356	79.2	−18	357–580	75.8	+41
ana	280–351	56.5	−13	352–580	39.6	+29
pse	280–359	46.9	−18	360–609	40.9	+37
per	280–359	46.3	−18	360–612	40.0	+36
vir	280–371	30.4	−16	372–659	24.3	+43
moj	280–369	26.9	−17	370–676	24.8	+43
gri	280–371	31.5	−16	372–664	21.4	+44

Amino acid identity was calculated between *D. melanogaster* and other species. mel, *D. melanogaster*; sim, *D. simulans*; sec, *D. sechellia*; yak, *D. yakuba*; ere, *D. erecta*; ana, *D. ananassae*; pse, *D. pseudoobscura*; per, *D. persimilis*; vir, *D. virilis*; moj, *D. mojavensis*; gri, *D. grimshawi.*

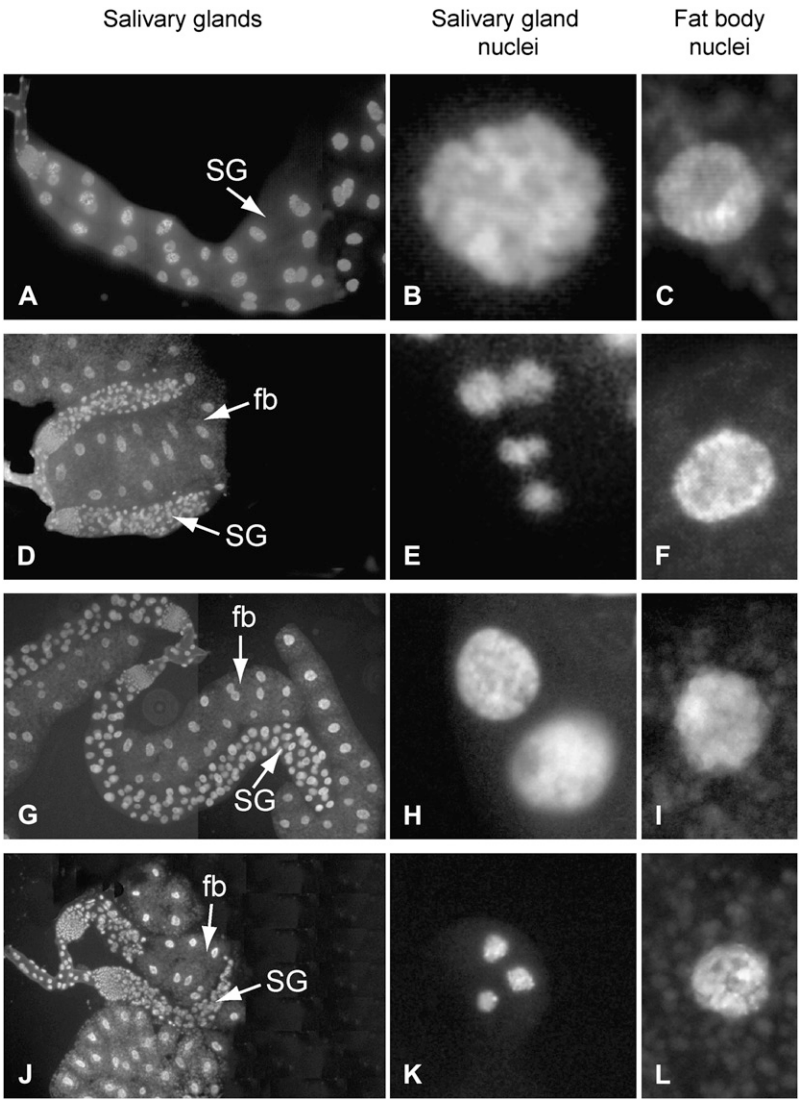


FIGURE 4.—Effects of point mutations in SUUR on its ability to suppress endoreplication in salivary glands upon overexpression from early development. Left column, salivary gland; middle column, salivary gland nuclei; right column, fat body nuclei photographed under the same magnification. (A–C) Oregon-R. (D) Ectopic expression of SUUR protein in salivary glands under the *AB1*-GAL4 driver suppresses polytenization, which leads to the formation of miniature salivary gland. The size of the salivary gland nuclei in *AB1*-GAL4>UAS-*SuUR* larvae is dramatically reduced (E) as compared to the fat body nuclei (F) and wild-type salivary gland nuclei (B). (G) Ectopic expression of $SUUR_{Nmut}$ with the *AB1*-GAL4 driver results in partial suppression of polytenization. Salivary gland nuclei in *AB1*-GAL4>UAS-$SuUR_{Nmut}$ larvae (H) are similar in size to those of the fat body (I), only slightly smaller than in Oregon-R (B), but significantly larger than in *AB1*-GAL4>UAS-*SuUR* larvae (E). (J–L) Effects of ectopic expression of $SUUR_{Cmut}$ on polytenization in the salivary gland are indistinguishable from those of the full-length SUUR. Salivary gland (SG) nuclei and fat body (fb) nuclei were stained with Hoechst.

transgenic stocks, and in 2 crosses from different stocks single escapers were observed).

Contrary to our expectations, substitutions in the C-terminal part of the SUUR had no detectable effects on the protein. Overexpression of $SUUR_{Cmut}$ under the *AB1*-GAL4 driver results in miniature salivary glands similar in size to those observed upon ectopic overexpression of full-length SUUR in *AB1*-GAL4>UAS-*SuUR* larvae (Figure 4). Ectopic expression of $SUUR_{Cmut}$ under *C323*-GAL4 resulted in complete female sterility similar to the ectopic expression of full-length SUUR.

Mutation in N terminus impairs the protein's ability to associate with chromosomes and alters the chromatin structure: In wild-type polytene chromosomes, SUUR is detected in late-replicating regions. When UAS-*SuUR* is expressed under the control of the weak mosaic *arm*-GAL4 driver, 20% of salivary gland nuclei demonstrate weak spots in the IH regions and an immunostaining pattern similar to those of wild-type SUUR (KOLESNIKOVA *et al.* 2005). It is a convenient system for the expression of the protein at the level similar to that of the wild type. When $SUUR_{Nmut}$ is expressed under the control of the *arm*-GAL4 driver, no protein is detected in PH or IH or elsewhere on the *arm*-GAL4>UAS-$SuUR_{Nmut}$; $SuUR^{ES}$ chromosomes except for a weak signal in nucleolus (Figure S5), and no weak spots were observed (data not shown). To test whether $SUUR_{Nmut}$ is capable of any chromosome binding, we employed a strong salivary-gland-specific *Sgs3*-GAL4 driver, which is active in mid-third instars when most of the replication in the salivary gland has ceased. *Sgs3*-GAL4>UAS-*SuUR*; $SuUR^{ES}$ larvae chromosomes display distinct binding signals in all bands and the chromocenter (Figure 5A). In contrast, in *Sgs3*-GAL4>UAS-$SuUR_{Nmut}$; $SuUR^{ES}$ larvae, the immunolocalization signal for $SUUR_{Nmut}$ is weak and dim (Figure 5B). These results suggest that the introduced substitutions within the N-terminal regions of SUUR dramatically decrease the binding of the protein to chromosomes.

Ectopic expression of UAS-*SuUR* under *Sgs3*-GAL4 induces swellings in IH regions (ZHIMULEV *et al.* 2003c). In contrast, no changes in chromosome

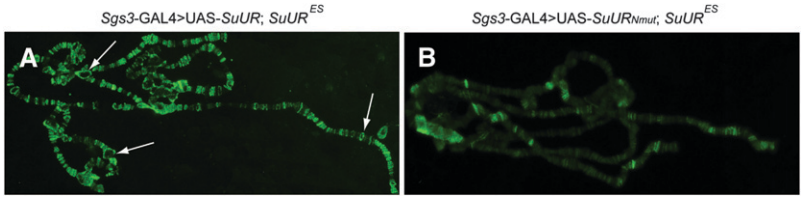


FIGURE 5.—$SUUR_{Nmut}$ displays reduced binding to chromosomes when overexpressed under the control of the *Sgs3*-GAL4 driver. Chromosomes from larvae expressing full-length SUUR and $SUUR_{Nmut}$ were photographed using identical exposure times. (A) On *Sgs3*-GAL4>UAS-*SuUR; SuUR*^{ES} polytene chromosomes, SUUR is found in almost all bands. Swellings are indicated by arrows. (B) On *Sgs3*-GAL4>UAS-$SuUR_{Nmut}$; *SuUR*^{ES} polytene chromosomes, binding of mutant protein is drastically reduced.

morphology are observed when $SUUR_{Nmut}$ is expressed with the same *Sgs3*-GAL4 driver (data not shown).

The C-terminal part of SUUR binds to polytene chromosomes: Mutation of conserved amino acid residues F816S and F817D in the C-terminal region has no pronounced effect on the ability of the SUUR protein to cause underreplication. Earlier we showed that the C-terminal SUUR fragment $SUUR_{495–962}$ suppresses endoreplication while ectopic expression of the SUUR fragment $SUUR_{1–779}$ lacking residues 780–962 does not (KOLESNIKOVA *et al.* 2005). We decided to test the overexpression effects of the smaller conserved C-terminal region of SUUR (aa 669–962). We cloned the fragment of the *SuUR* ORF that contained the last 293 codons fused to the HA tag and NLS into the pUAST vector (see File S1, Figure 3D) to obtain the UAS-$SuUR_{669–962}$ construct (hereafter, $SUUR_{669–962}$).

When $SUUR_{669–962}$ was expressed from the onset of development under control of the *AB1*-GAL4 driver, the size of the salivary glands remained unaffected. However, the analysis of polytene chromosomes from *AB1*-GAL4>UAS-$SuUR_{669–962}$ larvae revealed general disorganization of polytene chromosomes (Figure 6A). In contrast to the wild-type chromosomes where ectopic fibers typically link IH regions, in *AB1*-GAL4>UAS-$SuUR_{669–962}$ chromosomes we observed numerous ectopic fibers that were formed along the chromosome arms.

Ectopic expression of $SUUR_{669-962}$ under *arm*-GAL4 resulted in a range of uniformly staining chromosomes of varying intensities (data not shown), as detected with anti-HA antibodies. The antibodies do not stain polytene chromosomes of the wild-type strain (data not shown). Also, it has been shown elsewhere that neither HA tag nor NLS bind polytene chromosomes on their own (JAQUET *et al.* 2002), suggesting that the observed localization pattern reflects a property of the $SUUR_{669–962}$ fragment. Overexpression of $SUUR_{669–962}$ with *Sgs3*-GAL4 driver results in extremely strong nonspecific binding of SUUR to the chromosomes, regardless of the banding pattern (Figure 6E). Notably, under these conditions polytene chromosome morphology remained unchanged (Figure 6B), even though the chromosomes appeared totally covered by the $SUUR_{669–962}$ (Figure 6E). Endogenous SUUR was found to be specifically associated with its typical chromosomal sites (Figure 6F), indicating that the C-terminal part of the protein does not have a dominant-negative effect. Our results indicate that, while the C-terminal part of the protein is essential for under-

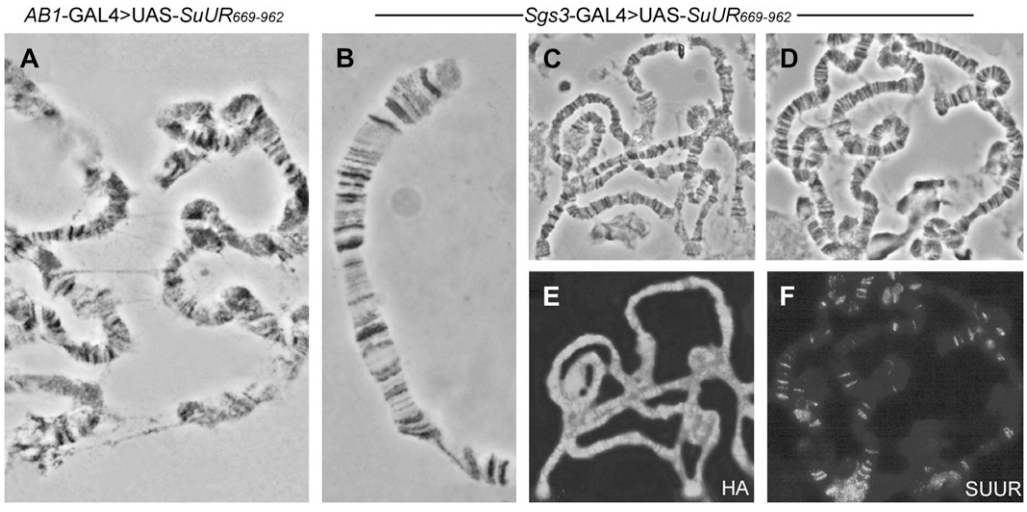


FIGURE 6.—Chromosome binding of $SUUR_{669–962}$ upon ectopic expression and its effects on chromatin structure. (A) Ectopic expression from early development in *AB1*-GAL4>UAS-$SuUR_{669–962}$ larvae results in the formation of ectopic fibers. (B) Polytene chromosomes from *Sgs3*-GAL4>UAS-$SuUR_{669-962}$ larvae resemble wild type in that they show no foam or swellings. (C and E) In *Sgs3*-GAL4>UAS-$SuUR_{669-962}$ larvae, the chromosomes are nonspecifically covered with the ectopically expressed truncated protein, as detected with anti-HA antibodies. (D and F) Endogenous SUUR in *Sgs3*-GAL4>UAS-$SuUR_{669–962}$ larvae can be detected at its typical chromosomal sites with E45 antibodies. (A–D) Phase contrast. (E and F) Immunostaining.

replication, it does not significantly suppress endoreplication on its own.

DISCUSSION

SUUR orthologs are present in the genomes of 11 Drosophila species. Notably, all these species display chromosome breaks and constrictions marking local DNA underreplication in salivary gland polytene chromosomes (Zhimulev 1998). We observed a dominant-negative effect of $SUUR_{1–458}$ overexpression in the salivary glands of hybrid *D. melanogaster* × *D. simulans* larvae, which manifested as a disappearance of chromosome breaks. This indirectly supports the idea that *D. simulans* SUUR protein is functional and affects late replication in heterochromatic regions in a way similar to that of *D. melanogaster* SUUR.

The potential SUUR ortholog in the mosquito lacks functionally important protein domains such as positively and negatively charged clusters, so its function in the mosquito remains in question. A BLAST search of the *A. gambiae* protein at NCBI resulted in high-confidence hits to predicted proteins in the yellow fever mosquito *Aedes aegypti* (*E*-value e^{-25} and e^{-22}) and in the southern house mosquito *Culex quinquefasciatus* (*E*-value e^{-22}). Interestingly, a BLAST search of SUUR from some Drosophila and Anopheles species detected mammalian ERCC6 protein, which is important in transcription-coupled excision repair (Troelstra *et al.* 1992) as the best hit. Mutations in ERCC6 lead to Cockayne syndrome (Mallery *et al.* 1998; Laugel *et al.* 2008). The similarity is restricted to the N-terminal part of the protein, and *E*-values range from e^{-11} for the search with the *D. mojavensis* protein to e^{-6} with *A. gambiae*.

The K_a/K_s ratio for SUUR orthologs varies from 0.16 to 0.23 within species from the *melanogaster* subgroup. The K_a/K_s ratio between *D. melanogaster* and *D. yakuba* orthologs is 0.16, suggesting that about one-half of the 1850 Drosophila-specific proteins evolve under stronger selection pressure than SUUR, judging from recent genomewide analysis of the Drosophila proteome (Zhang *et al.* 2007). The same study demonstrated that proteins with orthologs in distant species tend to evolve under stronger selection pressure than Drosophila-specific proteins. SUUR protein has a high substitution rate similar to the fast-evolving genes in Drosophila (Schmid and Tautz 1997). The evolution rate of the *SuUR* gene is comparable to that of *transformer* (*tra*), a gene involved in the primary somatic sex-determination pathway (O'Neil and Belote 1992). Specifically, the amino acid identity level in TRA and SUUR in *D. melanogaster* and *D. simulans* is, respectively, 92.4% and 93.1% and that in *D. melanogaster* and *D. erecta* is, respectively, 87.0% and 87.7% (O'Neil and Belote 1992). In other sequenced insect species, such as *Bombyx mori* and *Apis mellifera*, no *SuUR* orthologs were identified. On one hand, this is very typical for the fast-evolving genes: there is no true ortholog of *tra* in *A. mellifera*; however, a distant *tra* homolog, *csd*, is present (Beye *et al.* 2003; Cho *et al.* 2006). On the other hand, *A. mellifera* was not reported to have polyploid tissues. Possibly, the SUUR ortholog is absent in honeybees not because of the rapid evolution of the gene, but because of the fact that a mechanism involving SUUR does not exist in this species.

Conserved uORF was identified within 5′ UTR of *SuUR* (Hayden and Bosco 2008). Both the uORF and the main ORF in *SuUR* are maintained in drosophilids, and high divergence of SUUR in mosquito coincides with the lack of uORF in this species. This observation further supports the possibility that *SuUR* expression could be controlled via uORF. Although no examples of uORFs affecting downstream ORF protein production in Drosophila or other insect species have been demonstrated to date, 44 conserved uORFs were recently identified in *D. melanogaster* (Hayden and Bosco 2008).

Domain organization of SUUR is conserved in all Drosophila species analyzed, and different domains display different rates of amino acid substitutions. The most conserved region of the protein is found at its N terminus where it coincides with the region possessing similarity to the SNF2/SWI2 domain. When this region is absent in SUUR, the protein can no longer display some of its prominent effects in the overexpression system: $SUUR_{495–962}$ (Figure 3D) fails to induce swellings, weakly induces underreplication, and only mildly suppresses polytenization (Kolesnikova *et al.* 2005). Even though the Walker A and Walker B motifs within this region are noncanonical, they are invariably conserved in all species analyzed (Figure S3). Substitution of two amino acid residues in the ATPase-like region, L57R/G58R, attenuates SUUR ectopic expression phenotypes due to its decreased binding to IH regions (this is best seen when we expressed $SUUR_{Nmut}$ under the control of *arm*-GAL4). To some extent $SUUR_{Nmut}$ functionally resembles $SUUR_{495–962}$ much more than SUUR, because it fails to form chromosome swellings and only partially suppresses polytenization. Interestingly, the $SUUR_{1–360}$ fragment does not bind to polytene chromosomes (Kolesnikova *et al.* 2005), but the substitutions in this region decrease the binding of the full-length protein.

The least conserved middle region of the protein encompasses negatively and positively charged clusters, and it is important for the specific binding of SUUR with chromosomes (Kolesnikova *et al.* 2005). The middle part of SUUR maintains charged regions as well as their net charge across all Drosophila species, despite the deletions, insertions, and high rate of charged amino acid substitutions. Ratios of K_a/K_s for this part of the protein are ~0.5 for species from the *melanogaster* subgroup (Table S1), indicating that even this part of the protein apparently is under mild negative selection. A possible HP1-interacting motif has been described

within the middle part of SUUR (Pindyurin *et al.* 2008), but it shows no conservation beyond the *melanogaster* subgroup.

The C terminus of SUUR is moderately conserved, but no characterized protein domains were identified in this region. Comparison of two C-terminal fragments, $SUUR_{495–962}$ and $SUUR_{669–962}$, further supports our conclusion that the middle part of SUUR (specifically, the positively charged region) is indispensable for the specific chromosome binding and suppression of underreplication (Figure 3). Expression of $SUUR_{669–962}$ does not induce underreplication, although we speculate that ectopic fibers formed along the chromosome arms upon early overexpression of $SUUR_{669–962}$ might result from very weak nonspecific underreplication. However, substitutions of the conserved aromatic residues 816 and 817 within its C terminus do not disrupt the ability of SUUR to induce underreplication.

Thus, even though SUUR falls into a class of fast-evolving genes, the protein has some highly conserved regions and maintains its domain structure in drosophilids. Substitution of two conserved amino acids or a truncation of an N-terminal half of the protein could modify the overexpression phenotypes equally well. Presence of an uORF in the *SuUR* gene that is far more conserved than the SUUR protein opens exciting possibilities for studying gene regulation.

The authors thank Andrey Gorchakov for critical reading of the manuscript and stimulating discussion and Vincenzo Pirrotta for providing pHA plasmid and the antibodies. This work was supported by the following Russian Foundation for Basic Research (RFBR) grants: RFBR 08-04-00521-a and RFBR 08-04-01105-a.

LITERATURE CITED

Belyaeva, E. S., I. F. Zhimulev, E. I. Volkova, A. A. Alekseyenko, Y. M. Moshkin *et al.*, 1998 *Su(UR)ES*: a gene suppressing DNA underreplication in intercalary and pericentric heterochromatin of *Drosophila melanogaster* polytene chromosomes. Proc. Natl. Acad. Sci. USA **95:** 7532–7537.

Belyakin, S. N., G. K. Christophides, A. A. Alekseyenko, E. V. Kriventseva, E. S. Belyaeva *et al.*, 2005 Genomic analysis of Drosophila chromosome underreplication reveals a link between replication control and transcriptional territories. Proc. Natl. Acad. Sci. USA **102:** 8269–8274.

Berezney, R., D. D. Dubey and J. A. Huberman, 2000 Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. Chromosoma **108:** 471–484.

Beye, M., M. Hasselmann, M. K. Fondrk, R. E. Page and S. W. Omholt, 2003 The gene *csd* is the primary signal for sexual development in the honeybee and encodes an SR-type protein. Cell **114:** 419–429.

Brand, A. H., and N. Perrimon, 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development **118:** 401–415.

Brendel, V., P. Bucher, I. R. Nourbakhsh, B. E. Blaisdell and S. Karlin, 1992 Methods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci.USA **89:** 2002–2006.

Burge, C., and S. Karlin, 1997 Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. **268:** 78–94.

Cherbas, L., X. Hu, I. Zhimulev, E. Belyaeva and P. Cherbas, 2003 EcR isoforms in *Drosophila*: testing tissue-specific requirements by targeted blockade and rescue. Development **130:** 271–284.

Cho, S., Z. Y. Huang, D. R. Green, D. R. Smith and J. Zhang, 2006 Evolution of the complementary sex-determination gene of honey bees: balancing selection and trans-species polymorphisms. Genome Res. **16:** 1366–1375.

Comeron, J. M., 1999 K-estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals. Bioinformatics **15:** 763–764.

Donaldson, A. D., 2005 Shaping time: chromatin structure and the DNA replication programme. Trends Genet. **21:** 444–449.

Drysdale, R. A., M. A. Crosby, W. Gelbart, K. Campbell, D. Emmert *et al.*, 2005 FlyBase: genes and gene models. Nucleic Acids Res. **33:** D390–D395.

Gall, J. G., E. H. Cohen and M. L. Polan, 1971 Reptitive DNA sequences in Drosophila. Chromosoma **33:** 319–344.

Havas, K., I. Whitehouse and T. Owen-Hughes, 2001 ATP-dependent chromatin remodeling activities. Cell. Mol. Life Sci. **58:** 673–682.

Hayden, C. A., and G. Bosco, 2008 Comparative genomic analysis of novel conserved peptide upstream open reading frames in *Drosophila melanogaster* and other dipteran species. BMC Genomics **9:** 61.

Jaquet, Y., M. Delattre, A. Spierer and P. Spierer, 2002 Functional dissection of the *Drosophila* modifier of variegation *Su(var)3–7*. Development **129:** 3975–3982.

Kent, W. J., 2002 BLAT: the BLAST-like alignment tool. Genome Res. **12:** 656–664.

Kolesnikova, T. D., I. V. Makunin, E. I. Volkova, V. Pirrotta, E. S. Belyaeva *et al.*, 2005 Functional dissection of the Suppressor of UnderReplication protein of *Drosophila melanogaster*: identification of domains influencing chromosome binding and DNA replication. Genetica **124:** 187–200.

Kumar, S., J. Dudley, M. Nei and K. Tamura, 2008 MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinformatics **9:** 299–306.

Laugel, V., C. Dalloz, A. Stary, V. Cormier-Daire, I. Desguerre *et al.*, 2008 Deletion of 5′ sequences of the *CSB* gene provides insight into the pathophysiology of Cockayne syndrome. Eur. J. Hum. Genet. **16:** 320–327.

Lilly, M. A., and A. C. Spradling, 1996 The Drosophila endocycle is controlled by cyclin E and lacks a checkpoint ensuring S-phase completion. Genes Dev. **10:** 2514–2526.

MacAlpine, D. M., H. K. Rodriguez and S. P. Bell, 2004 Coordination of replication and transcription along a *Drosophila* chromosome. Genes Dev. **18:** 3094–3105.

Makunin, I. V., E. I. Volkova, E. S. Belyaeva, E. N. Nabirochkina, V. Pirrotta *et al.*, 2002 The Drosophila Suppressor of Underreplication protein binds to late-replicating regions of polytene chromosomes. Genetics **160:** 1023–1034.

Mallery, D. L., B. Tanganelli, S. Colella, H. Steingrimsdottir, A. J. van Gool *et al.*, 1998 Molecular analysis of mutations in the *CSB* (*ERCC6*) gene in patients with Cockayne syndrome. Am. J. Hum. Genet. **62:** 77–85.

Manseau, L., A. Baradaran, D. Brower, A. Budhu, F. Elefant *et al.*, 1997 GAL4 enhancer traps expressed in the embryo, larval brain, imaginal discs, and ovary of Drosophila. Dev. Dyn. **209:** 310–322.

Moshkin, Y. M., A. A. Alekseyenko, V. F. Semeshin, A. Spierer, P. Spierer *et al.*, 2001 The *bithorax* complex of *Drosophila melanogaster*: underreplication and morphology in polytene chromosomes. Proc. Natl. Acad. Sci. USA **98:** 570–574.

O'Neil, M. T., and J. M. Belote, 1992 Interspecific comparison of the *transformer* gene of Drosophila reveals an unusually high degree of evolutionary divergence. Genetics **131:** 113–128.

Pindyurin, A. V., C. Moorman, E. de Wit, S. N. Belyakin, E. S. Belyaeva *et al.*, 2007 SUUR joins separate subsets of PcG, HP1 and B-type lamin targets in *Drosophila*. J. Cell Sci. **120:** 2344–2351.

Pindyurin, A. V., L. V. Boldyreva, V. V. Shloma, T. D. Kolesnikova, G. V. Pokholkova *et al.*, 2008 Interaction between the *Drosophila* heterochromatin proteins SUUR and HP1. J. Cell Sci. **121:** 1693–1703.

Poux, S., D. McCabe and V. Pirrotta, 2001 Recruitment of components of Polycomb group chromatin complexes in *Drosophila*. Development **128:** 75–85.

Rost, B., G. Yachdav and J. Liu, 2004 The PredictProtein server. Nucleic Acids Res. **32:** W321–326.

Rubin, G. M., and A. C. Spradling, 1982 Genetic transformation of Drosophila with transposable element vectors. Science **218:** 348–353.

Sambrook, J., and D. V. Russell, 2001 *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schmid, K. J., and D. Tautz, 1997 A screen for fast evolving genes from *Drosophila.* Proc. Natl. Acad. Sci. USA **94:** 9746–9750.

Schubeler, D., D. Scalzo, C. Kooperberg, B. van Steensel, J. Delrow *et al.*, 2002 Genome-wide DNA replication profile for *Drosophila melanogaster*: a link between transcription and replication timing. Nat. Genet. **32:** 438–442.

Smith, A. V., and T. L. Orr-Weaver, 1991 The regulation of the cell cycle during *Drosophila* embryogenesis: the transition to polyteny. Development **112:** 997–1008.

Stark, A., M.F. Lin, P. Kheradpour, J.S. Pedersen, L. Parts *et al.*, 2007 Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. Nature **450:** 219–232.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Troelstra, C., A. van Gool, J. de Wit, W. Vermeulen, D. Bootsma *et al.*, 1992 ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell **71:** 939–953.

Volkova, E. I., A. A. Yurlova, T. D. Kolesnikova, I. V. Makunin and I. F. Zhimulev, 2003 Ectopic expression of the *Suppressor of Underreplication* gene inhibits endocycles but not the mitotic cell cycle in *Drosophila melanogaster.* Mol. Genet. Genomics **270:** 387–393.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay, 1982 Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. **1:** 945–951.

White, E. J., O. Emanuelsson, D. Scalzo, T. Royce, S. Kosak *et al.*, 2004 DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states. Proc. Natl. Acad. Sci. USA **101:** 17771–17776.

Wodarz, A., U. Hinz, M. Engelbert and E. Knust, 1995 Expression of crumbs confers apical character on plasma membrane domains of ectodermal epithelia of Drosophila. Cell **82:** 67–76.

Zhang, G., H. Wang, J. Shi, X. Wang, H. Zheng *et al.*, 2007 Identification and characterization of insect-specific proteins by genome data analysis. BMC Genomics **8:** 93.

Zhimulev, I. F., 1998 Polytene chromosomes, heterochromatin, and position effect variegation. Adv. Genet. **37:** 1–566.

Zhimulev, I. F., and E. S. Belyaeva, 2003 Intercalary heterochromatin and genetic silencing. BioEssays **25:** 1040–1051.

Zhimulev, I. F., E. S. Belyaeva, I. V. Makunin, V. Pirrotta, E. I. Volkova *et al.*, 2003a Influence of the *SuUR* gene on intercalary heterochromatin in *Drosophila melanogaster* polytene chromosomes. Chromosoma **111:** 377–398.

Zhimulev, I. F., E. S. Belyaeva, I. V. Makunin, V. Pirrotta, V. F. Semeshin *et al.*, 2003b Intercalary heterochromatin in *Drosophila melanogaster* polytene chromosomes and the problem of genetic silencing. Genetica **117:** 259–270.

Zhimulev, I. F., E. S. Belyaeva, V. F. Semeshin, V. V. Shloma, I. V. Makunin *et al.*, 2003c Overexpression of the *SuUR* gene induces reversible modifications at pericentric, telomeric and intercalary heterochromatin of *Drosophila melanogaster* polytene chromosomes. J. Cell Sci. **116:** 169–176.

Communicating editor: J. A. Birchler

GENETICS

Supporting Information
http://www.genetics.org/cgi/content/full/genetics.109.104844/DC1

Conservation of Domain Structure in a Fast-Evolving Heterochromatic SUUR Protein in Drosophilids

Anna A. Yurlova, Igor V. Makunin, Tatyana D. Kolesnikova, Olga V. Posukh, Elena S. Belyaeva and Igor F. Zhimulev

DOI: 10.1534/genetics.109.104844

FILE S1

Constructs for transformation

Targeted substitutions were introduced by method described by Intine and Nazar (INTINE and NAZAR 1998) as suggested by For replacement of Leu57 and Gly58 to arginines in N-terminal part of SUUR we made PCR with the following combinations of primers: standard T3 – Ig2 and Ig1 – T7 using DNA of clone 31 (MAKUNIN *et al.* 2002) as a template. Substitutions in the primer sequences are underlined. The PCR products were mixed together and used for PCR with standard T3 – T7 primers to obtain the fragment with the substitutions. On next step we replaced *Pst*1-*Hin*dIII fragment in clone f40 (MAKUNIN *et al.* 2002). The insertion was excised with NotI-KpnI and cloned into pUAST (BRAND and PERRIMON 1993) resulting in construct $SuUR_{Nmut}$.

For replacement of Phe816 and Phe817 for Ser and Asp, respectively we made PCR with Ig8F – A8(Smut)R and A7(Smut)F – St1A using DNA of f40 plasmid. PCR fragments were combined and amplified with Ig8F and St1A primers. The 137 bp *Sph*I-*Spe*I fragment containing the substitutions was excised from the PCR product and used for replacement of original *Sph*I-*Spe*I fragment in f40 clone. The insertion was cloned into pUAST as described above resulting in construct $SuUR_{Cmut}$.

For expression of the C-terminal part of SUUR (amino acids 669-962) cDNA clone was digested with *Bam*HI and sticky ends were filled with dNTPs by Klenov. The 1,4 kb *Bam*HI-*Xho*I fragment was released by digestion with *Xho*I and cloned in pHA plasmid (N. Hulo and V. Pirrotta, unpublished, see below) between NaeI and XhoI sites. Resulting plasmid was digested with *Xba*I and sticky ends were filled with dNTPs by Klenov. The insert was excised by digestion with *Xho*I and ligated into pUAST Drosophila transformation vector (BRAND and PERRIMON 1993) between EcoRI (blunted with dNTPs by Klenov) and *Xho*I sites.

The sequences of the final plasmids were verified.

LITERATURE CITED

BRAND, A. H., and N. PERRIMON, 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development **118:** 401-415.

INTINE, R. V., and R. N. NAZAR, 1998 PCR-mediated mutagenesis in sequences recalcitrant to homogeneous amplification. Biotechniques **25:** 364-366.

MAKUNIN, I. V., E. I. VOLKOVA, E. S. BELYAEVA, E. N. NABIROCHKINA, V. PIRROTTA *et al.*, 2002 The Drosophila Suppressor of Underreplication protein binds to late-replicating regions of polytene chromosomes. Genetics **160:** 1023-1034.

Primers used for directed mutagenesis:

Ig8F (5'-GTGGTGTCTATTCCGCTT-3'),

A8(Smut)R (5'-CGTTATCTCA**TC**G**G**AGTCATCTTC-3'),

A7(Smut)F (5'-GAAGATGACT**C**C**GA**TGAGATAACG-3'),

St1A (5'-CTGCTGCACTTGTGGATGAA-3'),

Ig1 (5'- GAGTGGCC**G**T**A**GAAAGGTCGCT-3'),

Ig2 (5'- AGCGACCTTTC**T**A**C**GGCCACTC-3')

Sequence of insertion in pHA (TAGS) plasmid provided by N. Hulo and V. Pirrotta (unpublished). The tags are cloned in modified pUC19 plasmid. Tag epitopes (HA Kreis and HA) and nuclear localization signal (NLS) are shown above the sequence. Termination codons are indicated by asterisks. Sites for restriction enzymes are signed below the sequence.

```
                        HA Kreis                >  <      NLS
          M  Q  D  L  P  G  N  D  N  S  T  A  P  P  T  K  K
TCTAGACAACATGCAGGACCTGCCAGGCAACGACAACAGCACCGCCCCCCCGACGAAGAA  60
XbaI

       >     <           HA          >
 R  V  K  L  Y  P  Y  D  V  P  D  Y  A                    *
GCGCGTCAAGCTTTACCCCTACGACGTGCCCGACTACGCCGGCCCGGGATATCCTCGAGT 120
                                      NaeI SmaI EcoRV XhoI

   *   *
AATTGATTGAGTCGAC
          SalI
```

TABLE S1

Synonimous (above diagonal) and non-synonimous (below diagonal) substitutions per site in the middle part of *SuUR* region encoding aa 280-581 in *D. melanogaster*.

	mel	sim	sec	yak	ere
mel		0,111	0,131	0,312	0,259
sim	0,060		0,049	0,263	0,227
sec	0,073	0,043		0,300	0,257
yak	0,111	0,112	0,125		0,212
ere	0,119	0,113	0,133	0,080	

mel – *D. melanogaster*, sim – *D. simulans*, sec – *D. sechellia*, yak – *D. yakuba*, ere – *D. erecta*.

***>D.melanogaster SuUR* ORF**

ATGTATCACTTTGTATCCGAGCAAACGCCGGAGGTGCGGCTCACAGATGAGGCACTGGTCACGAGCCACGTGACGCAGTACCTCAA
GAGCTTTCAACTGGATGCCGTCCGCTTCGTCTACGATCGTTTGGCTAAGCGGGAATTCTGCATCCTGAATGATGAGAGTGGCCTTG
GAAAGGTCGCTACGGTGGCGGCACTTCTAAGTGCCCTACCACCCGCCAAAAAAACACTTGTGGTGCTCCAGAACGATGAGCAACTG
CTCACCGGTTGGCGATTCCATTTGGACACTCTCACTGACCTCCAGGTGTACATCATTCAGGGGGTGCAAGATACCACCGACTCCCC
GCACAGCGTTTACCTGGCGAAATGGAGCCAACTACGGAGCATTGGAGATCTCAGTCGCCTCAAGTTCGACTACATTATGGTGGATA
ATCGGGGGCATTCGCTGAACAACAGCTTTTGCACTTCCATGCTGCTGAAGCAATTCGAGGGACGGGTAAACGTTCTTATCTCCAGT
GTTGACGTTACGTCCGACGTTAGGTTGCTCTACAATGTTCTACGGTTGGGCGGCCGACTGGAGCATCAGTACAAAAGCTTTGCGAG
CTTCGATCGAAAATTCCATCTGCCAGATCCGAAGGAGGTATTCAGTAAGCGCATAGATCTCGAGGAATACTACAAGCAGCGGGGAT
TCCTCAGCGAGTACATCAAGGACTTTCGTCTGCGTCGTTTTCGTCATCAATTCGACAAGTCACTGCCACTTGTCGCACCGGAGCAA
TACAAACACAATCTGAATCTCTGGTTGGCCTCAAAGAACAGCCAGAGCACAATAAGCGGCTCAGATGTGTGCTCTACAATCGCTTC
CATAGACAACAATCCGGCTCAGCAAAATAAAACTGGGCTCTTTGAAGAAACGGATAGGTTATCCGAGCACAGCGTGGATGATGTGG
CAATGTCACCGCTGATATTCGAGTATTCCGAGTCTGACGATGAACCCCTAACAGTAGAACCTGATGCAGATCAAAACCCGGTATTA
GTTGTATCCAGCGATGATTGTGAGATTGTAACACCTCCAAGTACTCCGCAAAATCGAACGCCCTCATTAAATGAATCACCAAGGAC
GAAATCAAAGAAAAAATTTAGCAAAAAAACGTCACCGCGCAAAAAAGCTGATCTCACCGATTCCGAAGAGGATGATGAAGCGACCG
ATAATATGCCGCCAAGGAAGAGAACTCGTGCAGCCACCGTTCATCTGACACCTAAGACAAGGAGACTCAATGTTCGAATTTTAAGG
GTTTCCCTCGATACTCTGAGCACACCACCACCGTCCAGAACCACTACTGCGATCGTAACGCCCAAAACGGAGCCCACAGCTAGAAG
AAAAAATCTGAAAAAGCGCACGGTGAGCCCAGTGGATGTGGGTCGTCCAGCCACAAGGGGAATGCAGCGCCTAACACGATCCGCTG
AAACAAAAATAAACAGTAAGTACTTGAAACATCGCTCATTGGATGATGTCAAGCGTAGCTTTCCAAGGAGAGTCAAGTTGGAGGGT
AACCAAACGCCCAGAAGTAGCAAACAAATAGTGAAGCAGGAGCCTAAGTCTAAAGTCGGGCAGGAAAAGAAACAGAAGACGGTGGA
TGTACCTGCTCAAGGAACTGCGAAACGAAAGCCTGGACGTCCGAGAAAATGCCAAACCAAGACTGAAGATTTGGGGAAGACAAAGA
CTAAGCCGAATTCAAAGCACTTGCCACCTACTCCTCAAGTTCTCAGTGGTAGCTCGTTGTCCTCGGAGTACATGCAATGTGCTCAG
CGAATACCAGACAATTTGGATGCCATTGAGTCGCCCGCTTTTCGTGTGCCCTTTACACCTCAACAAACTCCAATGCTTCTAACCTT
GCCGTCTACACACAATTTGCTCAACGATTCTGAGGTGGTGTCTATTCCGCTTTACAAAGATCCAGTTGAGACAGTGGTGATCAATA
GCTCGCACGATGAGTGTTCTCCGCAGGATCCATCGCAGAGTCGTCGCACCAAGGCTTTGAAAAGAAAACGCAAACCCGTGACGTCC
GTCAATTCAAGTTTCGGCGGCGGATTGGGTCTACCACCGGCCAAGAGGAGTGCCAACAAATCTCCAGATCTATTTAGCATATCCTC
GGAACACTCGCAGATTCCGCTGGCGCAACCAAGACCCTCATCGCCCTTCGAAGGCTTTAAGATCTTTGGCTCTGAGGTGAAGCAGT
TTCAGCAGCAGCTAGCGAAAGTAAATATATCGGTGCCAAAAAAGAAGCGAGATCGCTCGTGTTTGGATATCCTGGAGCAAATGTTC
GAGCCACGACAACAGCAGTCGGCTAAAACTAGTCCCAAAGTCTTGCCAACCTTGCCACTGACCCAGAAGGATGATGCAGAGTCTAC
CATTACGCAGAGAAGAAGAACCCTGCTGGAAGATGACTTCTTTGAGATAACGAACAATGGCCAATTTGGCAGTCGCATGCGATTGA
ACTCCTCTGGCGAAGTATCGCCCGTGCAACCGGATCAACAGTCCGTCAGGCCGTCGCAGGCCAATAAAATCACCAATTATTTGATA
GGTTCTGGAATTACGCAGGAGCGAACGCAACCTTCGAATGGTAATCGCAATTCCATTCTTGCATCGCTGCGCAAGTCACCCAAATC
TCCCAAACAGGGTGCTAAATCCACTCAGGCCACGAAACTAACACGCTGGTTTGGCTCCGTCTTCGGCGGTGGAGCCTCACAGACCA
GCTCCGTGGAATCGGTTAGTGCACCAAGTACGCCGGTGAATTCATCCACAAGTGCAGCAGCTTGTCAAACGCGATCGGCGAGAAGT
GGTGGAGCATCAGGACCGACCAAAAGAAAGCGATTGGAACTGTTCAAGTGA

***>D.simulans SuUR* ORF**

ATGTATCACTTTGTATCCGAGCAAACGCCGGAGTTGCGGCTCTCGGAGGAGACACTGGTCACAAGCCACGTGACGCAGTACCTCAA
GAGCTTCCAACTGGATGCTGTCCGTTTCGTCTACGATCGTCTGTCCAAACGGGAATACTGCATCCTTAATGATGAGAGTGGCCTTG
GAAAGGTCGCTACGGTGGCGGCACTTCTAAGTGCCCTACCACCCGCCAAAAAAACACTTGTGGTGCTTCAGAACGATGAGCAACTG
CTCACCGGTTGGCGATTCCATTTGGACACTCTCACTGACCTCCAGGTGTACATCATTCAGGGGGTGCAAGATACCACCGAATCCCC
GCACAGCGTTTACCTGGCAAAATGGAGCCAACTACGGAGCATTGGAGATCTCAGTCGCCTCAAGTTCGACTACATTATGGTGGATA
ATCGAGGGCACTCGCTGAACAACAGCTTTTGCACTTCCATGCTGCTGAAGCAATTCGAGGGACGGGTGAACGTTCTTATCTCCAGT
GTTGACATTACGTCCGACGTTAGGTTGCTCTACAATGTTCTACGGTTGGGCGGCCGACTGGAGCATCAGTACAAAAGCTTTGCGAG
CTTCGATCGTAAATTCCACCTGCCAGATCCGAAGGAGGTATTCAGCAAGCGCATAGATCTCGAGGAATACTACAAGCAGCGGGGAT
TCCTCAGCGAGTACATCAAAGACTTTCGTCTGCGTCGTTTCCGTCATCAATTCGACAAGTTGCTGCCTCTTGTCGCACCGGAGCAA
TACAAACACAATCTGAATCTCTGGTTGGCCTCAAAAAACAGCCAGAGCACAATAAGCGGCTCAGAGGTTTGCTCTACAATCGCTTC
CATAGAAAACAATCCGGCTCAGCAAAATAAAACTGGGCTCTTTGAAGAATCGGACAGTGTAGATGATGTGGCAATGTCACCGCTGA
TATTCGAGTCCTCCGAATCTGACGATGAACCCCTAACAGTAGAACCTGGTGCAGATCAAAACCCGGTATTAGTTGTATCCAGCGAT
GATTGTGAGATTGTAACACCTCCAAATACTCCGCCAAATCGAACGCCCTCATTAAATGAATCGCCTAAGACGAAATCAAAGAAAAA
ATATAACAAGCAAACGTCGCCGCGCAAAAAAGCTGATCTCACCGATTCCGAAGAGGATGATGAAGCGATGCCGCCAAGGAAGAGAA
CTCGTGCGGCCACCGTTCATCTAACACCCAAGACAAGGAGACTCAATGTTCGAATTTTAAGGTGTTCCCTAGATAATCTAAGCACG
CCACCACCGTCTAGAACCACTACTGCGATCGTAACACCGAAAACAGAGCCCTCAGTTAGAAGAAAAAATGTGAAAAAGCGCACGGT
GAGCCCAGTGGATGTGGGTCGTCCAGCCACAAGGGGAATGCAGCGCCTAACACGATCCGCTGAAACAAAACTAAACAGCAAGTACT
TGAAACACCGCACATTGGATGATGTCAAGCGTAACTTTCCAAGGAGAGTCAAGATTGAGGGGACCCAAACGCCCAGGAGTAGCAAA
CGAAAAGTGAAGCAGGAGCCTAAGTCTAAAGCCGAGCAGGAAAAGAAACAGAAGACGGTGGACAAACCTGCTCAAGAAACTCCGAA
ACGAAAGCCTGGACGTCCGAAAAAATGCAAAACTATGACTGAAGATTTGGGAAAGACAACGACTAAGCCGGATTCAAACCCCTTGC
CATCTACTCCTCAAGTTCTCAGTGGCAGCTCCTTGTCGTCGGAGTACATGCAATGTGCTCAGCGAATACCAGACAATTTGGATGCC
ATTGAGTCGCCTGCTTTCCGTGTGCCCTTCACACCTCAACAAACTCCAATGCTTCTAACGCTGCCGTCTACTCATAACTTGCTCAA
CGACTCCGAGGTGGTGGCTATTCCGCTTCACAAAGATCAAGTTGAGACAGTGGTGATCAACAGCTCCCACGATGAGAGTTCGCCGC
AGGATCCATCGCAGAGTCGTCGCACCAAGGCTTTGAAAAGAAAACGCAAACCCGTGACGTCCGTTAATTCAAGTTTTGGCGGAGGA
TTGGGTCTACCACCGGCCAAGAGGAGTGCCAACAAATCTCCAGATCTGTTTAGCATATCCTCGGATCACTCCCAGATTCCGCTGGC
GCAACCGCGACCCTCGTCGCCCTTCGAAGGCTTTAAGATCTTTGGCTCTGAGGTTAAGCAGTTTCAGCAGCAACACGCGAATGTAA
ATATATCAGTGCCAAAAAAGAAGCGAGATCGCTCGTGCTTGGATATCCTGGAGCAAATGTTCGAGCCACGACAACAGCAGTCGACG
AAAACTAGTCCCAAAGTCTTGCCAACCTTGCCACTGACCCAGAAGGATGATGCAACGCCTACGATTACGCAGAGGAGAAGAACCCT
GCTGGAAGATGACTTCTTCGAGATAACGAACAATGGCCAGTTTGGCAGTCGCATGCGGTTGAACTCCTCTGGCGAGGTCTCGCCCG
TGCAACCGGATCAACCGTCCGTCAGGCCGTCGCAGGCCAATAAAATCACCAATTACTTGATAGGTTCCGGAATTACCCAGGAGAGA

ACGCAACCCTCCAATGGTAATCGCAACTCCATTGTCGCATCGCTGCGCAAGTCACCCAAATCTCCGAGACAGGCTGCTAAATCCAC
GCAGGCCACGAAGCTAACACGCTGGTTTGGCTCGGTCTTCGGCGGAGTAGGCTCACAGACCAGCTCCGTGGAGTCGGTCAGTGCAC
CAAGTACGCCGGTTAATTCATCCACAAGTGCAGCTGCGTGTCAAACGCGATCGGCGAGAAGTGGTGGAGCATCAGGACCGACCAAA
AGAAAGCGACTGGAACTGTTCAAGTGA

>*D.sechellia SuUR* ORF

ATGTATCACTTTGTATCCGAGCAAACGCCGGAGTTGCGGCTCTCGGAGGAGGCACTAGTCACAGGCCACGTGACGCAGTACCTCAA
GAGCTTCCAACTGGATGCTGTCCGTTTCGTATACGATCGTCTGTCCAAGCATGAATTCTGCATCCTGAATGATGAGAGTGGCCTAG
GAAAAGTCGCCACGGTGGCGGCACTTCTAAGTGCCCTGCCACCCGCCAAAAAAACACTTGTCGTGCTCCAGAACGATGAGCAACTG
CTCACCGGTTGGCGATTTCATTTGGACACTCTCACTGACCTCCAGGTGTACATCATTCAGGGGGTGAAAGATACCACCGACTCCCC
GCACAGCGTTTACCTGGCGAAATGGAGCCAACTACGGAGCATTGGAGATCTCAGTCGCCTCAAGTTCGACTACATTATGGTGGATA
ATCGGGGGCACTCGCTGAACAACAACTTTTGCACTTCCATGCTGCTGAAGCAATTCGAGGGACGGGTAAACGTTCTTATCTCCAGT
GTTGACATTACGTCTGACGTAAGGTTGCTCTACAATGTTCTACGTTTGGGCGGCCGATTGGAGCATCAGTACAAAAGCTTTGCGAG
CTTCGATCGCAAATTCCACCTGCCAGATCCGAAGGAAGTATTCAGCAAGCGTATAGATCTCGAGGAATACTACAAGCAGCGGGGAT
TCCTCAGCGAGTACATCAAGGACTTTCGTCTGCGTCGTTTCCGCCATCAATTCGACAAGTCGCTGCCTCTTGTCGCACCGGAGCAA
TACAAACACAATCTGAATCTCTGGTTGGCCTCAAAAAACAGCCAGAGCACAATAAGCGGCTCAGAGGTTTGCTCTACAATCGCTTC
CATAGAACACAATCCGGCTCAGCAAAATAAAACTGGTCTATTTGAAGAATCGGACAGTGTAGATGATGTGGCAATGTCACCGCTGA
TATTCGAGTCCTCCGAATCTGACGATGAACCCCTAACAGTAGAACCTGGTGCAGATCAAAACCCGGTATTAGTTGTATCCAGCGAT
GATTGTGAGATTGTAACACCTCCAAATACTCCGCCAAATCGAACGCCGTCATTAAATGAATCGCCAAGGACGAAATCAAAGAAACA
ATATAGCAAGAAAATGTCGCCGCGCAAAAAAGCTGATCTCACCGATTCCGAAGAGGATGATAAAGCGATGGATAATATGCCGCCAA
AAAAGAGAACTCGTGCGGCCACCGTTCATCTAACACCCAAGACAAGGAGACTCAATGTTCGAATTTTGAGGTGTTCCCTAGATAAT
CTAAGCACCCCACCACCGTCTAGAACCACTACTGCGATCGTAACACCGAAAACAGAGCCCTCAGCTAGAAGAAAAAATCTGAAAAA
GCGCACGGTGAGCCCAGTGGATGTGGGTCGTCCAGCCACAAGGGGAATGCAGCGCCTAACACGATCCGCTGAAACAAAACTAAACA
GCAAGTACTTGAAACACCGCGCATTGGATGATGTCAAGCGTAGCGCTCCAAGGAGAGTCAAGATTGGATGTACCCAAAAGCTCAGG
AGTAGCAAACGAATAGTGAAGCAGGAGCCTATGTCTAAAGCCGAGCAGGAAAACAAAGAGAAGACGGTGAACGAACCTGCTCAAGA
AACTCCGAAACGAAAGCCTGGACGTCCGAGAAAATGCAAAACTATGACTGAAGATTTGGGGAAGACAACGACTAAGCCGGATTCAA
ACTCCTTGCCATCTACTCCTCAAGTTCTCAGTGGCAGCTCCTTGTCCTGGGAGTACATGCAATGTGCTCAGCGAATACCAGACAAT
TTGGACGCCATTGAGTCGCCTGCTTTCCGTGTGCCCTTTACACCTCAACAAACTCCAGTGCTTCTAACGCTGCCGTCTACTCACAA
CTTGCTCAACGACTCTGAGGTGGTGGCAATTCCGCTTCACAAAGAGCAAGTTGAGACAGTGGTGATCAACAGCTCCCACGATGAGA
GTTCGCCGCAGGATCCATCGCAGAGTCGTCGCACCAAGGCTTTGAAAAGAAAACGCAAACCAGTGACGTCTGTCAATTCAAGTTTT
GGCGGGGGATTGGGTCTACCACCGGCCAAGAGGAGTGCCAACAAATCTCCAGATCTGTTTAGCATATCCTCGGATCACTCCCAGAT
TCCGCTGGCGCAACCGCGACCTTCGTCGCCCTTCGAAGGCTTTAAGATCTTTGGCTCTGAGGTGAAGCAGTTTCAGCAGCAACACG
CGAATGTAAATATATCAGTGCCAAAAAAGAAACGAGATCGCTCGTGTTTGGATATACTAGAGCAAATGTTCGAGCCACGACAACAG
CAGTCGACGAAAACTAGTCCCAAAGTCTTGCCAACCTTGCCACTGACCCAGAAGGATGATGCAACGCCCACCATAACGCAGAGGAG
AAGAACCCTGCTGGAAGATGACTTTTTCGAGATAACGAACAATGGCCAATTTGGCAGTCGCATGCGGTTGAATTCCTCTGGCGAGG
TCTCGCCCGTGCAACCGGATCAGCCGTCCGTCAGGTCGTCGCAGGCCAATAAAATCACCAATTACTTGATAGGTTCCGGAATTACC
CAGGAGCGAACGCAGCCCTCCAATGGTAATCGCAACTCCATTGTCGCATCGCTGCGCAAGTCACCCAAATCTCCGAGACAGGCTGC
TAAATCCACGCAGGCCACGAAGCTCACACGCTGGTTTGGCTCGGTCTTCGGCGGAGTAATCTCACAGACCAGCTCCGTAGAATCGG
TCAGTGCACCAAGTACGCCGGTCAATTCATCCACAAGTGCAGCTGCGTGTCAAACGCGATCGGCGAGAAGTGGTGGAGCATCAGGA
CCGACCAAAAGAAAGCGACTGGAACTGTTCAAGTGA

>*D.yakuba SuUR* ORF

ATGTATCACTTTGCATCCGAACAAACACCGGAGTTGCGGCTTTCGGCAGAGGCACTGGTCACGAGCCACGTGACGCAGTACCTCAA
GGGCTTCCAGCTGGATGGAGTGCGCTTCGTCTACGATCGTCTGGCCAGGCGCGAATTCTGCATCTTGAATGATGAGAGTGGGCTTG
GAAAAGTCGCCACGGTGGCGGCGCTACTCAGTGCCCTGCCTCCCGCCAAGAAAACACTTGTGGTGCTGCAGAACGATGAGCAACTG
CTCACCGGTTGGCGGTTCCATTTGGACACTCTCACTGACCTTCAGGTGTACATCATTCAAGGGGTGCACGACACCACAGACTCACC
GCACAGCGTTTACCTGGCGAAATGGAGCCAGCTACGGAGCATTGGAGATCTCAGTCGCCTCAAGTTCGACTACATTATGGTGGACA
ATCGGGGGCACTCGCTGAACAACAGCTTTTGCACTTCCATGCTGCTGAAGCAATTCGAGGGACGCGTAAACGTTCTTATCTCCAGT
GTTGACATTACGTCCGACGTGAAGTTGCTATACAATGTTTTACGGTTGGGCGGCCGACTGGAGCATCAGTACAAAAGCTTTGCAAG
CTTTGATCGCAAATTCCACCTGCCAGATCCGAAGGAGGTATTCAGCAAGCGCATAGATCTCGAGGAATACTACAAGCAGCGGGGAT
TCCTCAGCGAGTATATCAAGGACTTTCGTCTGCGTCGTTTTCGCCATCAATTCGACAAGACGCTGCCCTTTGTCGCACCGGAGCAA
TATAAGCACAATCTGAACCTCTGGTTGTCCTCCAAAAACAGCCAGAGCACACTAAGTGGCTCAGAGGTGTGCTCAACAGTCGCTTC
CATCGACAACAATCCGGCTCAGCAAAATGAAGCAGTGCTCGTCGAGGAATCGGATAAGTTATCCGAGCATAGCGTAGATGATGTTG
TGGCAATGTCACCGCTGATATTCGAGTCCTCCGAATCTGACGATGAACCCATAACAGTGGAAGCTCCAGCAAGTGAAAACCCGGTA
TTAGTTGTATCCAGCGATGATTGTGAGATTGTGACTCCTCCAAATACACCGCCGAATCGAATGCCTATATTAAGTGAATCGCCAAG
GACGAAATCAAAGAAAAAATTTACCAAGAAATCGTCGCCGAGCAAAAAAGCTGATCTCACCGATTCCGAAGAGGATGATGAAGCGT
TGACAATATCGCCAAGAAAGAGAACTCGTGCGGCTACCGTTCATTTTACACCCAAGACAAGGAGACTCAATGTGCGAATTCTAAGG
GTCTCTCTAGATACTCTGAGCACACCACCTCCCTCTGGAGCCACCACTGCCATCGTAACGCCCAAGACAGAGCCCTCAGCTAGAAG
AAAATATCAGAAAAAGCGGCCGGTGAGCCCAATGGATGTGGGTCGTCCAACTACAAGAGGAATGCAGCGCTTGACACGATCCGCTG
AATCAAAAATAAACAGCAAGTACTTGAAGCACCACGCACTGGATGATGCCAAGCGTAGCTTGCCCAGGAGAATCAAGGCTGAGGGG
AACCAAACGCCCAAATCTAGCAAACGAATAGTGAAGCAGGAGCCTAAGGCAAGGCCCAAGCCCGAGCAGGAAAAGGAACCAAAGAC
TGTGGACAAACCTGCACAAGAAACTCCGAAACGAAAACCTGGACGCCCGAGAAAATCAAAAACGGTGACTGAAACTTTAGAAAAGT
CAAAGACTAAGCCGGATTCAAAACCCATGCCACCAACTCCTCAAGTTCTCAGTGGAAGCTCCTTGTCCTCGGAGTACATGCAATGT
GCTCAGCGTATACCAGACAATTTGGATGCCATTGAGTCACCTGCCTTCCGAGTACCCTTCACGCCCCAACAAACTCCTATGCTTCT
AACCTTGCCGTCTACTCACAACTTGCTCAACGACTCTGAGGTCGTGGCTATTCCACTTTCCAAAGATCCCGTTGAAACGGTGGTGA
TCAACAGCTCCCACGATGAGAGTTCGCCGCAGGATCCATCGCAGAGTCGGCGCACCAAGGCGTTAAAAAGAAAACGCAAACCCGAG

GCGCCTGTCAATTCAAGTTTCGGCGGTGGATTGGGTCTGCCACCGGCTAAGAGGAGTGCCAACAAATCTCCAGATCTGTTTAGCAT
ATCCTCGGATCTCTCGCAGATTCCTCTGGCACAGCCACGACCCTCATCGCCCTTCGAAGGCTTTAAGATCTTTGGCTCTGAGGTGA
AGCAGTTTCAGCAGCAGCACGCGAAAGTAAATACACCAGCGCCGAAAAAGAAGCGAGATCGCTCGTGTTTGGATATCCTGGAGCAA
ATGTTCGAGCCACGGGAGCAGCAGGAGTCGGCTAAAACTAGTCCCAAAGTCTTGCCCACCTTGCCACTGACCCAGAAAGATGATGC
AGCGCCTACCATTACGCAGCGAAGAAGAACCCTCCTGGAGGATGACTTCTTTGAGATAACGAACAACGGCCAGTTTGGCAGTCGCA
TGCGGTTGAACTCCTCCGGAGAGGTTTCGCCCGTGCAACAGGATCAGCAGTCCGTCAGGCCGTCGCAGGCCAATAAAATCACTAAA
TACTTGATAGGTTCTGGAATTACCCAGGAGAAGACACAGCCTTCTAATGGCAATCGCAACTCCATCGTCGCATCGCTGCGCAAGTC
CCCAAAATCGCCGAAACAGAGTGCAAAAACCACTCAAGCCACGAAGCTAACGCGCTGGTTTGGCTCCGTCTTCGGAGGAGGAGCCT
CACAGACCAGCTCCGTGGAGTCGGTCAGCGCACCGAGTACCCCGGTTAATCCTTCCACGAGTGCAGCAGCGTGTCAAACGCGATCG
GCGCGAAGTGGTGGAGCCTCGGGACCCAGCAAAAGAAAGCGATTGGAACTGTTCAAGTAA

***>D.erecta SuUR* ORF**

ATGTATCACTTCGTATCCGAACAAACGCCGGAGTTGCGGCTCTCAGCGGAGACACTGGTCACGAGCCACGTGACGCAGTACCTCAA
GAGCTTCCAGCTGGATGCAGTGCGCTTCGTCTACGATCGGCTGGCAAAGCGCGAGTTCTGCATCTTAAATGATGAGAGTGGTCTTG
GAAAAGTCGCCACGGTGGCGGCACTACTAAGTGCGCTTCCACCCGCCAAGAAAACACTTGTGGTTCTCCAGAACGATGAGCAACTG
CTCACCGGTTGGCGATTCCATTTGGACACTCTCACTGACCTCCAAGTGTACATCATTCAAGGGGTGCACGACACCACAGACTCGCC
GCACAGCGTTTACCTGGCGAAATGGAGCCAACTACGGAGCATTGGAGATCTCAGTCGCCTCAAGTTCGACTACATTATGGTGGATA
ATCGGGGGCACTCGCTGAACAACAGCTTTTGTACTTCCATGCTGCTGAAGCAATTCGAGGGACGGGTAAACGTTCTTATCTCTAGT
GTTGACATTACGTCCGACGTGAAGTTGCTCTACAATGTTTTACGACTGGGCGACCGACTAGAGCATCAGTACAAAAGCTTTTCAAG
GTTCGATCGCAAATTCCACCTGCCAGATCCAAAGGAGGTATTCAGCAAGCGCATAGATCTCGAGGAATACTACAAGCAGCGCGGAT
TCCTCAGCGAGTATATCAAGGACTTTCGTCTGCGTCGTTTCCGTCATCAATTCGACAAGACCCTGCCACTTGTCGCACCGGAGCAA
TACAAGCACAATCTGAATCTCTGGTTGGCCTCTAAAAACAGCCAGAGCACATTAAGTGGCTCAGAGGTGTGCTCTACAGTCGCTTC
CATAGAAAACAATCCGCCTCAGCAAAATGAAACTGTGCTCGTTGAAGAATCGGATAGGATATCCGAGCATAGCGTAGATGATGTTG
TGGCAATGTCACCCCTGATATTTGAGTCCTCCGAATCTGACGATGAACCCATAACCGTAGATCCCGTAGCAAATGAAAACCCGGTA
TTAGTTGTATCCAGCGATGATTGTGAGATTGTGACTCCTCCAAATACACCGCCGAATCGAACGCCCTTAGTAAATAAATCGCCAAG
GACGAAATCAAAGAAGAAATGTAGCAAGAAACCGTCACCGTGCAAAGAAGCAGATCTCACCGATTCCGAAAAGGATGATGAAGGGT
TGACAATGCCGCCAAGAAAGAGTACTCGTGCGGCTACCGTTCATTTTACACCCAAGACAAGAAGACTTAATGTGCGACTTTTAAGG
GTTTCCCTAGATGCTCTAAGCACACCAGCCCCGTCTGGAGCCACTACTGCGATCATAACGCCCAAAACAGAGCCCTCAGCTAGAAG
AAAAAATCGGACAAAACAGCCAATGGATGTGGGTCGCCCAGCTACAAGAGGAATGCAGCGCCTAACACGATCCGCTGAATCAAAAA
TAAACAGCAAGTACTTGAAGCACCACGTACTTGATGATGCCAAGCGTAACTTTCCAAGAAGAATCAAGGCTGAGGGGAACCAAACA
CCCAAAACTAGCAAACGAATAGTGAAGCAGGAGTCCAAGGCAAAGGCTAAACCCGAGCAGAAAAAGAAAATAAAGACCGTGGACAA
GCCTGCTCAAGAAACTCCGAAACGAAAACCTGGACGTCCGAGAAAGTGCAAAACGCTGACGGAAACTTTAGGGAAGTCAAAGACTA
AGCCGAATTCAAAGCCCTTGCCACCTACTCCTCAAGTTCTCAGTGGTAGCTCTTTGTCCTCGGAGTACATGCAATGTGCTCAGCGT
ATCCCAGACGATTTGGATGCCATTGAGTCACCAGCTTTCCGGGTGCCTTTCACGCCCCAACAAACTCCTATGCTTCTAACCTTGCC
GTCTACTCACAATTTGCTTAACGACTCTGAAGTGGTGGCTATTCCGCTTTTCAAAGATCAAGTTGAAACAGTGGTGATCAACAGCT
CCCACGATGAGAGTTCCCCGCAGGATCCATCGCAGAGTCGTCGCACCAAGGCGTTAAAAAGAAAACGCAAACCCGACGCGCCTGTC
AATTCAAGCTTTGGCGGGGGATTGGGTCTACCACAGGCTAAGAGGAGTGCCACCAACAAATCTCCAGATCTGTTTAGCATATCCTC
TGATCTCTCACAGATTCCTCTGGCGCAGCCACGACCCTCATCGCCCTTCGAAGGCTTCAAGATCTTTGGCTCTGAGGTGAAGCAGT
TTCAGCAGCAGCATGCGAAAGTAACTATACCAGCGCCAAAAAAGAAGCGAGATCGCTCGTGTTTGGATATCCTAGAGCAAATGTTC
GAGCCACGACAACAGCAGTCGTCTAAAACTAGTCCCAAAGTCTTGCCTACCTTGCCACTGATCCAGAAAGATGATGCAACAACTAC
CTTTACGCAGAGAAGAGTAACTCTCCTGGAAGATGACTTCTTTGAGATAACGAACAATGGGCAGTTTGGCAGTCGCATGAGGTTGA
ACGCTTCTGGAGAGGTCTCGCCCGTGCAACAGGATCAACAGTCCGTCAGGCCGACGCAGGCCAATAAAATCACAAATTACTTGATT
GGTTCTGGAATTACCCAGGAGCGAACGCAGCCTTCTAATGGTAATCGCAACTCCATCGTCGCATCGCTTCGCAAGTCCCCCAAATC
CCCGAAACATGGTGCAAGAACCACTCAAGCCACGAAGTTAACGCGCTGGTTTGGCTCGGTCTTCGGAGGAGGAGCCTCACAGACCA
GCTCCGTGGAATCGGTTAGCGCACCGAGTACCCCGGTAAATCCTTCCACAAGTGCAGCAGCATGTCAAACGCGAACGGCGAGAAGT
GGTGGAGCCACAGGACCCACCAAAAGGAAGCGATTGGAACTGTTCAAGTAA

***>D.ananassae SuUR* ORF**

ATGTATCACTTTGTGTCCGAACAGATGCCGGAGCTTCGGCTATCGGCGGATGCTCTAGTCACCAGCCATGTGACACAGTACCTGAA
GAGTTTCCAGCTGGAGGCGGTTAAATATCTCTATGAACGCCTGAGTAAGCGGGAATTCTGTATACTAAATGATGAAAGTGGTCTGG
GGAAGCAGGCTACAGTGGCGGCGCTCCTTAGTGCACTGGCTCCAGAGAAGAAAGTCTTGCTAGTGCTTCAGAACGATGACCACCTT
CTTGCCGGCTGGCGATTTCATCTGGACACACTAACCGACCTAGCGGTTTACATCATCCAAGATGTCCAAGACACCACGGAGACACC
GCACAGCGTATATCTTGCCAAGTGGAGCCAGCTGCGTAGCATCGGAGATCTTAGCCGACTGAAGTTCGACTACATTTTCGTGGACA
ATCGCGGTCACTCATTAAACAACAGCTTCTGCACCTCCATGCTGCTAAAGCATTATGAGGGCCGGGTAAACATGGTTATCTCCAGT
GTTGACATTACGTCGGATGTCAAGTTGCTGTTTAATATTTTGCGCTTGGGCGGACGTCTGGAGCATCAGTATCGGAACTTCCAGAG
CTTCAATCGAAAGTTCCATCTTCCAGACCCGAAAGAGGTGTTTGAAAAGTGCGTTGATCTCGAGGAATACTACAAGCAGCGGGGAT
TCCTCAGCGACTATATAAAGGACTTTCGGCTTCGCCGTTTTCGCCATCAGTTCGATACGACCCTGCCACTCGTCACAGTCGAGCAA
TACAAGAACAATCTCAACCTGTGGCTGGCATCGCAAAATAGCCAGAGCACGCTTAGCGGCTCCTCAGATGTCTGTTCCACGATCGC
ATCTGTGGAAAACAACCTCCATCGGGCAAACGAAGATTCTGACAAGCTGTCCGAGCACAGTGATGAGATCGTGGCCATGTCGCCGC
TTAGATTCGAATCATCCGACTCTGAGGTTGAGCCCATCATTTTGGAGCCGTCACCTGAAGCCCAAAAAGTCCTTGTGGTCTCCAGC
GACGATTGTGAAATTGTAATTCCCACAAAGCTCTCTCCAAGCGTAACTAATTCTCCGAGACTAACACCAAGAAGGAAGAAGAGAAT
GCCGCAAAAGAAAACATTGCAAATCCCGGACTCGTCGTCATCGGAGGACGAGAAGGATGAAAGTGAAGAGAAGGTGGATAATAATC
CCAAGCCCATAACCCCTCCAAAAAAAAGTTGCTACTCCGAAAAAAGCAATGAAGCTTAACATCCGACTCAAGAGGATGGAAGTAACA
GTAACTCCTCCAAATACTGAGACTGCTCCCCCACTCGCCACCCCGAAAACCGAGCCATCGGTTGGAAGAGTTAGGCGTAAAAGGAG
TAAACAGGGTACTACTAGTCCAGGTGATCGTCCTGCCACTCGCGGCATGCAGCGTTTAACGAGGTCCGCCGAATCCAAGCTCAACA
GCAAATACGTGAAACACCACAGGCTGGAGGATACAAAGCGACACACTCCCAGGCGGATCAAGGTTGAAGCAACTCAAACGCCCACT

ACAAGCAAGCCCAGAGTGAAGCGAGAGGAGAAGGAAAAGCCAACTAAAGAGCCGAGTATAGAAAGTGCACAAAAATCGGAGGCCCC
AACGACAGTAGATACTCCTAAAAGGAAAAGAGGGCGCCAAAGGAAGTGCGAAGAGGAACCTCAGCCAACAAAAGCATCCAACAGTA
AACCAGAAACCAATCATGTAACACCTACTCCTCAAATTTTAAGCGGCAGTTCATTGTCTTCAGAGTATCTGCAATGTGCTCAGCGC
TTCCCTGACAACCTAGACATGTTAGAAGCTCCCACTTTTCGAGTACCTTTAACACCTCAGCAGACCCCAATGATTTTAAGACTCCC
TTCGGCTAACCTGCTGAGCGATTCAGAAGTAATTGATGTTTCGCAAGGAAATCAACCAGCGGAGACTGTGGTCATATCCAGTTCTC
TGGATGAGAGTTCGCAAAACTGCCCTTCGGAGAGCCGGCGCACAAAGGCTTTAAAGCGTAAAAGAAAACCGGAGCCATCGCCGCCA
ACGTCAAGTTTTTCTGTAGGCTTACCGCCTCCGAAGAGGAAATCGAACAAGTCCCCGGATATATTCAGCATTACGTCGGATTTGTC
GCAGATTCCCTTGGCTCAACCGCCCCCCGCATCAACTCCCTTCGAAGGCTTCAAGATCTTTGGCTCTGAGGTGAAACAGTTTCAAC
AGCAACACGCTACCACGGGTGCAGCTCTTTCGAAAAAGAAGAGAGAACGGTCCTGCCTGGATATCCTCGAGCAAATGTTTGAACCT
CGCCAGCGAACGCAGCAGGAGCCATCGAATCCCAAAGTCTTGCCCACCTTGCCGCAGACACAGAAGCAGAGCAATGCACCCACCAT
AACCCAGAGAAGACAAACAATGGTGGAGGAGGACTTCTTCGAGATCACCAACAATGGAGAGTTCGGCAGTCGCATGAGATTGAACT
CTTCTGGCGACGTGTCACCCGTTCAGACGGAGCAACCGAGTGCCAAGCCAACCCAGGCCAACAAAATTACAAAATACCTGATCGGT
TCAGGAATTACGCAGGAGAGAACCCAACCGAATCGTAGTACTTCCTCTGCATCGCAGGCTGTGCGCAAGTCCCCCAAATCCCCAAA
ACAAGGCGGGAAAATGGCTCAGGCAACCAAGCTGACCCGGTGGTTTGGAACGGTTTTCGGAGGGTCTTCGCAGACCAGCTCCGTGG
AGTCAGTCAGTGCTCCCAGTACGCCGGTAGCTCCATCCACTAGTGAAGCGGCTTGCCGAGCGCGATCGTCAAGAAGTGGAGGACCA
TCTGGACCCACCAGAAGAAAGCGGCTAGATCTTTTCAAATGA

> *D.pseudoobscura SuUR* ORF

ATGTATCACTTTGTGTCCGAGCAGACACCGGAACTGCGGCTCTCGGAGAAGGCACTCGTAACCAGCCACGTAACGCAGTATTTAAA
GAGCTTCCAGCTCGATGCGGTGAGATTCCTGTACGAACGCCTCACCAAGGGAGAGTTTTGCATCCTCAACGATGAGAGTGGGCTGG
GGAAGTTGGCCTCCGTGGCGGCCCTGCTCAGTGCTCTGGAGCCCACCAAGAAAACGCTGATTGTGCTGCAAAACGACGAACAGCTA
ATCGCGGGCTGGCAGTTTCACCTAAACACACTCACAGATCTCTCCGTCTACACCATTCAGAGTGTCAAAGACACCACCGAGTCGCC
GCACAGCGTTTATCTGGCCAAATGGAGCAATCTGCGCAGCATCGGGGATCTCAGCAAACTTAAATTCGACTACGTCTTTGTGGACA
ATCGCGGCCATACATTGAACAACCACTTCTGCAACTCGATGCTGGTGAAGCATTTCGAAGGCAAAGTAAACATTGTTATCTCCAGT
GTTGACATAACGTCTGACGTGAAGCTCCTATACAATGTGTTGCGGCTGGGCGGACGCCTAGAGCATCAGCACAGAAACTTTCAGAG
TTTCGATCGCAAGTTCCACCTGCCAGACCCCAAAGAAGTGTTCAGCAAGCGCATAGATCTCGAGGAGTACTACAAACAGCGGGGTT
TTCTAAGTGAATACATAAAGGACTTTCGTCTGCGTCGCTTTCGTCACCAGTTCGATCAGAGCCTGCCCCTCGTGCCGCTTGAGCAA
TACAAGGCGAATCTGAATCTGTGGCTGGCCTCTAAGAACAGTCAGAGCACGCTCAGTGGCTCCTCGGAAGTCTGCTCCACAGTGGC
ATCATCTGTGGAGATCAATCCAGAAGATCCCATAGAAGATCAAGGGAGGGCTGAGCCACTGTCGGAGCATAGCAATGAAGCCGTGA
TCATGCCGCCCTTGCTTTTTGAATCATCCGACTCGGAAGATGAGCTAGTGATGGTGCTCCCATCAGATGCTCCCGTTCTCGAGAAC
AATGCAGTTATGATCGTCTCCAGCGACGATTGCGAGATCATTACGTCCCCAAGCACGCCGCCTACGCGCAAGCAATCAAAGGAGTC
GCCTTCAATTAAACCGAAGAGGAAATACGTCAAACGCCAGCCTCCTGCCAAGCAGCCAGAACTTACCGAATCGGAGCCAGAGGACG
AAGATAGTTCGAGGCAGAAAACGCGTTCAACTCCTGCTGCTGCTGCTGTATCGCCTACCGAAAAGACAAGAAAAGTCAATGTTCGG
CTGAGGAGAATGTCTTTGGCGGCGCAACAGCGCGAGAGCACGCCACCTAATAACGGAGCAGAACCGCCGCTGGTGGCTCCCAAGAC
AGAGCCAACGGCCAAGCGCCGTCAGGTCTTGAGCCCATCATCACCAATAAGACGTCCAGCGACACGCGGAATGCAGCGTCTGACGC
GCTCCGCAGACGCGAAACTGAACAGCAAATACTTGATGCACAGAATCAGATTGGAAGACTGCAGAAAGACCACGCCCAGGAGACCT
AAGGCGGAGTCTAGTCAAACTCCAAGGACAACCACATCGAGGGGGAAGCAAGAGATCGATGTGCCGCAGGAAACGCCCAAGAGGCG
ACCAGGTCGTGGTCGACCAGCCAAAAAAGCAATCGAAGAGAAATCCAACACCAGGACAACACCCAAGGCTGCAGAAGAAACACCCA
AGCGAAAACGAGGACGTCAACCCAACAAAGGACCGCAGGAGAGAGAACCACCCGCTCCCAAATCAAAGGAAACGCCTACCGTAAAG
AACTCGAAGATTAAACCCATGGATCCCACACCCCAAGTGGTGAGCAGCAGCTCTCTGTCCTCTGAATACATGCAGTGCGGGCAAAA
GTTGCCGGACAACGTGGATGTTCTGGGTCTGCCCGCCTTTCGCATACCCTTCACGCCCCAGCCGGCGCCCCTGATGCTACCATCGA
CCTTCACACTGTTCAGCGACTCGGAAGTGCTGTCCGTTCCCCTGACCAGCCCACAGCAGCGGGAGACTGTGGTCATTAGCAGCTCC
CACGATGAGAGCTCGCTGCCGTGCTCATCCACGACGCAGAGCCGCCGCACGAAGGCACTGAAGCGGAAGCGGAAGCAGGAGACACC
GACAAGTGGCTCCTCACACTTCGGATTGCTGCTCTCGCAGCAGCGAAACCCGATTAAATCCCCGGATATCTTTAGCAACTGCTCGG
AAATTTCTCAACTGACGTTGGCGCAGCCGCCACCGAAGCCCTCGTCTCCCTTCGAGTTCGAAGGCTTCAAGATCTTTGGGTCTGAA
GTAAAGCAGTTCCAGCAGCAGCATGCGAAGGCAATCGTTGGACCTGCCAAGAAGAAGCGCGAACGCTCCTGCCTCGACATCCTAGA
GCAGATGTTCGAGCCGCGTCCGCCGGTGACATCCAGCCCCAGGGTCCTACCCAGCCTTCCGAATGGAACACAGTCCTTGCAGCTGG
AGCAGCAGCAGAACACGCCGAAAAGGATAGCCCAACGGCGGCGTTCTTTTGTAGAGGATGACTTCTTTGAGATTACCAACAACGGG
GAGTTTGGCAGCCGCCTGCGCATCAACTCATCGGGTGATGTGTCACCTGTCCAGCAGGAACGCGCTACGCAATCGAACAAGATCAC
CAACTATCTGATTAGCTCGGGGCCAGAGAAGTCGAGCACCAGCAACAATCGCACATCTGCCACGCAAACGTTGCGCAAATCTCCCA
AAGCCGCCAGCAAGTCCACCCAGTCTACGAAGTTGACGCGTTGGTTTGGGGCCGCCTTTGGAGTCTCCTCCCAGAACAGCTCCATA
TCCGAGTCCCTGAGTGCACCCAGCACCCCCATGGCGCCGTCGACGAGCGAGGCAGCTTGTCGGGCACGCTCAGCACGCAGTGCAGG
ACCAACCAAGAGGAAGCGTTTGGATCTGTACAAGTGA

>*D.persimilis SuUR* ORF

ATGTATCACTTTGTGTCCGAGCAGACACCGGAACTGCGGCTCTCAGAGAAGGCACTCGTAACCAGCCACGTAACGCAGTATTTAAA
GAGCTTCCAGCTCGATGCGGTGAGATTCCTGTACGAACGCCTCACCAAGGGAGAGTTTTGCATCCTCAACGATGAGAGTGGACTGG
GGAAGTTGGCCTCCGTGGCGGCCCTGCTCAGTGCTCTGGAGCCCACCAAGAAAACGCTGATTGTCCTGCAAAACGACGAACAGCTC
ATCGCGGGCTGGCAGTTTCACCTAAACACACTCACAGATCTCTCCGTCTACACCATTCAGAGTGTCCAAGACACCACCGAGTCGCC
GCACAGCGTTTATCTGGCCAAATGGAGCAATCTGCGCAGCATCGGGGATCTCAGCAAACTTAAATTCGACTACGTCTTTGTGGACA
ATCGCGGCCATACGTTGAACAACCACTTCTGCAACTCGATGCTGGTGAAGCATTTCGAAGGCAAAGTAAACATTGTTATCTCCAGT
GTTGACATAACGTCTGACGTGAAGCTCCTATACAATGTGTTGCGGCTGGGCGGACGCCTAGAGCATCAGCACAGAAACTTTCAGAG
TTTCGATCGCAAGTTCCACCTGCCAGACCCCAAAGAAGTGTTCAGCAAGCGCATAGATCTCGAGGAGTACTACAAACAGCGGGGTT
TTCTAAGTGAATACATAAAGGACTTTCGTCTGCGTCGCTTTCGTCACCAGTTCGATCAGAGCCTGCCCCTCGTGCCGCTTGAGCAA
TACAAGGCGAATCTGAATCTGTGGCTGGCCTCTAAGAACAGTCAGAGCACGCTCAGTGGCTCCTCGGAAGTCTGCTCCACAGTGGC
ATCATCTGTGGAGATCAATCCAGAAGATCCCATCGAAGATCCAGGGAGGGCTGAGCCACTGTCGGAGCATAGCAATGAAGCCGTGA

TCATGCCGCCCTTGCTTTTTGAATCATCCGACTCGGAAGATGAGCTAGTGATGGTGCTCCCATCAGATGCTCCCGTTCTCGAGAAC
AATGCAGTTATGATCGTCTCCAGCGACGATTGCGAGATCATTACGTCCCCAAGCACGCCACCTACGCGCAAGCAATCAAAGGAGTC
GCCTTCAATTAAACCGAAGAGGAAATACGTCAAACGCCAGCCTCCTGCCAAGCAGCCAGAACTTACCGAATCGGAGCCAGAGGACG
AAGATAGTTCGAGGCAGAAAACGCGTTCAACTCCTGCTGCTGCTGCTGCTGCTGTATCGCCTACCGAAAAGACAAGAAAAGTC
AATGTTCGGCTGAGGAGAATGTCTTTGGCGGCGCAACAGCGTGAGAGCACGCCACCTAATAACGGAGCAGAACCGCCGCTGGTGGC
TCCCAAGACAGAGCCAACGGCCAAGCGCCGTCAGGTCTTGAGCCCATCATCACCAATAAGACGTCCAGCGACACGCGGAATGCAGC
GTCTGACGCGCTCCGCAGACGCGAAACTGAACAGCAAATACTTGATGCACAGAATCAGATTGGAAGACTGCAGAAAGACCACGCCC
AGGAGACCTAAGGCGGAGTCTAGTCAAACTCCAAGGACAACCACATCGAGGGGGAAGCAAGAGATCGATGTGCCGCAGGAAACGCC
CAAGAGGCGACCAGGTCGTGGTCGTCCAGCCAAAAAAGCAATCGAAGAGAAATCCAACACCAGGACAACACCCAAGGCTGAAGAAG
AAACACCCAAGCGAAAACGAGGACGTCAACCCAACAAAGGACCGCAGGAGAGAGAACCACCCGCACCCAAATCAAAGGAAACGCCT
ACCGTAAAGAACTCGAAGATTAAACCCATGGATCCCACACCCCAAGTGGTGAGCAGCAGCTCTCTGTCCTCTGAATACATGCAGTG
CGGGCAAAAGTTGCCGGACAACGTGGATGTTCTGGGTCTGCCCGCCTTTCGCATACCCTTCACGCCCCAGCCGGCGCCCCTGATGC
TACCATCGACCTTCACACTGTTCAGCGACTCGGAAGTGCTGTCTGTTCCCCTGACCAGCCCACAGCAGCGGGAGACTGTGGTCATT
AGCAGCTCCCACGATGAGAGCTCGCTGCCGTGCTCATCCACGACGCAGAGCCGCCGTACGAAGGCACTGAAGAGGAAGCGGAAGCA
GGAGACACCGACAAGTGGCTCCTCACACTTCGGACTGCTGCTCTCGCAGCAGCGAAACCCGATTAAATCCCCGGATATCTTTAGCA
ACTGCTCGGAAATTTCTCAACTGACGTTGGCGCAGCCGCCACCGAAGCCCTCGTCTCCCTTCGAGTTCGAAGGCTTCAAGATCTTT
GGGTCTGAAGTAAAGCAGTTCCAGCAGCAGCATGCGAAGGCAATCGTTGGACCTGCCAAGAAGAAGCGCGAACGCTCCTGCCTCGA
CATCCTAGAGCAGATGTTCGAGCCGCGTCCGCCGGTGACATCCAGCCCCAGGGTCCTACCCAGCCTTCCGAATGGAACACAGCCCT
TGCAGCTGGAGCAGCAGCAGAACACGCCGAAAATGATAGCCCAACGGCGGCGTTCTTTTGTAGAGGATGACTTCTTTGAGATTACC
AACAACGGGGAGTTTGGCAGCCGCCTGCGCATCAACTCATCTGGTGATGTGTCACCTGTCCAGCAGGAACGCGCTACGCAATCGAA
CAAGATCACCAACTATCTGATTAGCTCGGGGCCAGAGAAGTCGAGCACCAGCAACAATCGCACATCTGCCACGCAAACGTTGCGCA
AATCTCCCAAAGCCGTCAGCAAGTCCACGCAGTCTACGAAGTTGACGCGTTGGTTTGGGGCCGCCTTTGGAGTCTCCTCCCAGAAC
AGCTCCATATCCGAGTCCCTGAGTGCACCCAGCACTCCAATGGCGCCGTCGACGAGCGAGGCAGCTTGTCGGGCACGCTCAGCACG
CAGTGCAGGACCAACCAAGAGGAAGCGTTTGGATCTGTACAAGTGA

***>D.mojavensis SuUR* ORF**

ATGTATCACTTTGTCTCAGAGCGAACAGCGGAAGTACGCGTCTCCGAGAACGTGCTGATTCCCAGCAAGACGACGCAGTACTTAAA
GAGCTTTCAACTCGAAGCAGTGCGTTTTCTCTATAAACGGCTGTCCCAACAAGAGTTTTGCATTTTGAATGATGAGAGCGGTTTGG
GCAAAACAGCGTCTGTTGTGGCCCTGTTAAGTGGACTGGGAACGGCCAAGAAGACGCTAGTTGTGTTGCAAAACGATGAGCAGTTG
CTGGCTGGCTGGCAGTTTCACCTGGGAGTGCTCACTGATTTGTCCGTATGCGTTATTAAGGACGTTAATGACACCACAGAGTCGGC
GCATAGCGTTTATCTGAGCAAGTGGAGTGTATTACGCAGCATAGGCGATCTCAGCAAGCTGAATTTTGACTATATCATTGTAGATC
ATCGCGGATACACATTAAATAATAACTTTTGCACCTCCATGCTGCTGAAGCAGTTCGAAGGCAAAGTGAACATTGTGATCTCTAGT
GTGGACATAACGTCCGACGTTAAGTTGCTGTATAAGGTGATGCAATTGGGAGGCTGTTTGGATCATCAATTCAATAGCTTTAGAAG
TTTCGAGAGCAAGTTTCGTCTACCAGATGTTCAGGAGGTGCTAAGCAAGCGTGTCGATCTCGAGGAGTACTATAAGCAGCGGGGTA
TCCTCGGCGAATACATCAAGGATTATCGTCTTCGTCGCTACCGCCATCAGTTCGATTCATGCCTGCCTCTTGTCACGGCCGAGCAA
TACAAGAGGAACGTCTCTCTGTGGCTGGGCGAGAACAATAGTAACAGCACAATAAGTGACTCAACTGGAGGTTCAAAAACATCAGA
AGTTCCTTCCACAGGCGACACCGAAGAAATCTTTGGGTGTCTATTAAGCCTTCAACAGGAGCGACATTTGGCGGAACGCCAAGAGC
CAGATAAAATATCGCTTTCTGAGCAGAGTGATGACGTGATTGCCATGGAGCCTTTGATTTTGGAGTCCGAATCAGAACCTGAGGTT
GAAGCGGATGCCCATGCGAACAAGTGCAGCACTAATGTGGTATTGCTTTCCAGCGATGATTGCGAAATTATTGAGTCGTCAAAATC
GGCGCCAATGACATCGAACTCGCCAGCCAAACGAAAATACACGAAACGTGTACAAGCTCCAAAGTCAGTGGAGCTGACTGAATCGG
AAACGGAAGAGCCGCCTAATACATCGACGATAACAAGGAAAGTACTTAACGTGAGGCTAAGACGAGCTTCCATACAGATCGACCCC
ACAAACTCTGCAAAGGCGTCGGAGCTGCCGACGAAAGCTGTCAAGGCTTCACCTAAGAAATCACCCAATTCTCCGCGTAAGCCGAA
GTTACCTATAAAACCCGAGCCGAAAACTCCGAAACGTGACACATTGTTGCCAGTTACGCCTAAAACTGAGCCAAGGTCAGGACGCT
CTAATGGAAATAAAAACGTTGACGTGGAACCCAGACAGACGCGTGGAATGCAAAGATTAACGCGTTCGGCGGAGTCACGGTCGAAC
AGCAAGTACATGACGCAGCTGAAGGCCCTGGACTATGAGAAGAAAGGGACACCCAGAAGAAAACGTAATTCCAGCACTAATCAAAC
GCCCAAAAGTAGCAAGGTCAAAGAGGAGAAGGCTACTGTGAAGCAAGAGATTGCTGTCACGCCTAGGACATCAAAGACAACGCCAA
TGAAAATGAACCAAAAGGAACCAGAACCCAAAAGGACAACGAGGAGCCGAGCGGAAAAGGAACATTTTTCCAATAAGCTGACCTCC
TTGCCTGTTAATAAAAGAGTCGAAAAGGTCAGTTCTAAAACGATTAAGGCTCCGACCGCTGCCACCTTGAAGCCATCGGAATCTGT
TCCGAATAACGAAAGAAAGGGCAAGGCCAGTCCAAAAGTTGTAACAGCTTCATCAGTTACGCCCGATCTACCGAATTCTCCTCTTG
GCAGTCAAAAAAATGGAAAGGCCAGCTCGAAGCCCGTGACAGCTGCTCCCGCTGCAACACCTGAGATACTAAGCTCTGGTTCCATT
TTGTCTGATAACTCGTATTTGCAATGCGCACAAAAGCTCCCAGAGAACCTCGCAGAACTTGAGGCTTTGGAACTGCCCGAGTTTCG
GGTGCCGCACGCACCGCCGACAACTCAGCTGCTGTTGCCCGTGAACCTGAATCTGTTTAGTGGCTCCGAGGTGGCTGTCGTGCCTA
GTCCAAACCCAAAGCCAGATGTCGTGGTGATCAATAGTTCCCATGATGAGAGCTCGCAGCCCACGACTAAGCCATCGCAATCCCGG
CGCACTCGAGCATTGAAGAGGAAGTCTGTAGCTCCTGCCAAGGAACAACAGTCCCAGCCGACCACCTCCCGATTCGGCCTCATGCT
GGCTCAGCACCGCGGCTCGGTGAACAAATCCCCAGACATCTTCAGCAACTGCTCGGAAACGTCGCAAGTTTCGCTAGCGCAGCCGG
TGCCATTTGAGGGCTTCAAGATATTCGGTTCGGAGGTGAGGCAGTCGCATCAGCAGCTGCATGACAAAACAGGTGTAGGCATGCCG
GCTAAAAGGAAGCGCGAACGCTCCTGCCTGGAACTTTTGGAGAAAATGTTTGAGCCCCATCAGGCTAATCATAACGAGCCGGGCGT
TCAATTGCTGCCGGATCTGCCCAGTCCCAAGCAGCAGCTGCCACAGCGTCGGTTCACCTTGCTGGACGACGACATCTTCGAGATAA
CCAATAATGGGGAATTCGGCAGTCGCCTGCGAGTTGACTCCACAGGCAACGTGTCGCCCGTGCAGACACAGCAGCTGGCACAGCGT
AACAAAATCACAAACTATCTGATCAGCTCTGGTGGCGCTCCGGAGGATACTGGCCAGCGTACCCAATCCAACCAATTGCTGCGCAA
GTCACCCAAACCGGTCAAGTCCACGCAATCTACAAAGCTGACGCGCTGGTTTGGCACGGCCATCTCAGGCGGATCCCAGAGCGCCC
CTAGCACACCTGTGTTGCCTACCGAGGCAGCAGACCGCGCCCGATTGGCGCGCAGCGCGGGATCGAAGAAACGCAAGCGACTAGTG
CTGGACAAGTGA

***>D.virilis SuUR* ORF**

ATGTATCACTTTGTCTCAGAACAAACAGCGGAACTGCGCCTCTCCGAGAATGTGCTTATTTCCAGCCATACGACACAGTACCTGAA
GTGCTTTCAACTGGAAGCTGTGCGTTTTCTCTATGAACGGCTGTCCAAACAAGAATTTTGCATATTTAACGATGAGAGCGGCTTGG
GCAAAAGTGCGCCAGTTGTGGCACTGCTAAATGGCCTGGGAATATCCAAAAAGACATTAATTGTGTTGCAGAACGATGAGCAGTTG
CTGGCTGGCTGGCAGTTTCACTTGGGCGTCCTATCCGACCTGCCTGTGTGCGTTATAAAGGACGTCAATGACAGTACGGATTCGGC
GCATAGCGTTTATCTAAGCAAATGGAGCGTTCTACGCAGCATTGGCGATCTCAGCAAGCTAAACTTTGACTACATCATTGTCGATC
ATCGTGGCTATACATTAAATAACAACTTTTGCACTTCCATGCTTCTCAAGCAATTCGAGCGCAAAGTGAACATTGTGATCTCTAGT
GTGGACATAACGTCTGACGTGAAGTTGCTGTACAATGTGATGAGCCTGGGTGGCTGCCTGGAACATCAGCACAAGAGCTTTCGCAG
TTTTGAGCGCAAATTTCATCTGCCCGCTGTTAAGGAGGTGCTGAGCAAACGTGTCGATCTTGAGGAGTACTACAAACAGCGGGGTG
TTCTTGGCGAATACATCAAAGACTTTCGCCTGCGACGTTATCGCCATCAGTTTGATTCATATCTACCACTTGTCACACCCGAACAA
TACAAGAGGAACTTAACGCTTTGGCTGGGACAAAACAATAGCAATAGTATATTAAGTGGTTCAACTATAGATTCGACTTCCACAGG
AGGCACCGATGAAATTCTCGAGTGTCTGATGAGCACGAAACGGGATAGACAATTGGCACCGAATCAGCAAATGGATCCAGATAAAA
TATCGCTGTCTGAGCACAGTGATGAAGTAATAGCAATGGAGCCTTTGGTTTTCGAAATGTCCGAATCGGAGCCCGAGGTGGAAGTC
ACACACGTGACAGATGCACAAGCCAAGGAGTCAAACACTAATGTGGTTATGCTCTCCAGCGATGATTGTGAGATAGTCACGCCGAC
AATCACTCCGACACATTCGGCTCTAAGCGCAACAAATTCAAAAACGTCCAAAACTAAGCGAAAATACACAAAACGTGCACAGACTG
CAAAGTCAGCCGAGCTTACTGATTCAGAGGCAGAGGAGCAGCGCCCAACTCCGAAAATAAAGGGGAAAAAACTGAATGTTAGATTA
AGTCGTGTTTGCTTACCCCAAAAGGATCTAAACATAATAGCAGTAACTAACTCGCACAAACCAAAAGAACAGGCAGCAACTAACTC
GCACAAACCAAAAGAACAGGCGTCAAAACCAGGGCAAGAAGTATCTAAAAAGCCGCCCGAAGAGATAACACCATTAGAAATAAAGT
TGTTGCGTGCTAAACCTAATAACAAGGTAAAACTTGACCCACCTGTAAAACCAACGCCTGAGCCGCAGACGCCTAAAGATGATAAA
CCTAAACCCTTGCCAATCACGCCCAAGACCGAGCCCAGGTCTAAACGCACTGACGCAAATATTTCAACTGGCGTCCAACTTAGAGA
GACGCGTGGAATGCAAAGAATGACGCGTTCCGCGGATGCACGGCCGTACAGCAAATACATGGGTCTCCGGAGGACATTGGACAACG
ATAAGAAAGTCACTCCCAAACGAAAACGCAACTCAGATACAAATCAAACGCCCACAACTAGCAAGCTCAAAAAAGTGGAAGTAAAA
CAAGAGGTCACAGTCACGGCTGAAAAGTCAAAGACAACACCAACTAAAGGAAACCCGAAAACTAAAAAGATAGGGCGAGCGCCGAA
GAGAAAATTGTCTAAAGAACCGAAATCTTTACCTAGCAATAAAAGAAGGGGAAAGATCAGTGAAGTAACCGTGAAGTCTGCGCCAG
CTGCCACGCCAGAGCTACTAAGCTCCGGTTCCCTGCAGTCCGATAATTCGTATTTGCAATGCGCACAAAAGCTGCCAGAGAACCTC
GCAGAACTCGGATCCTTGGAGCTGCCTCAGTTTCGTGTACCGCACGTGCCGCCGGCAACTCCACTCATGTTGCCCATCCCTTTGAA
TTTGTTTAGCGACTCGGAAGTGGCTATCGCACCGAGTCCCCAACAAAAGTCAGATGTAGTGGTCATCAACAGTTCGCATGACGAAA
GCTCCCAGCCATCTGCTAAACAATCACAAAGCCGACGCACTCGAGCATTAAAAAGAAAGCGTCCAGCTCCTTCCAAGGAGCAACAG
TCGCAGCCCACCACCTCCAGTTTTGGCATGTTGCTAGCACAGCAACGTGGATCGGTCAACAAATCTCCTGACATATTTAGCAACTG
CTCCGATTTATCGCAGCTCACGCTGGCACAGCCTGTGCCTTCAGTTTCACCGTTTGAGGGATTCAAGATCTTTGGCTCGGAAGTAA
AACAGCTTCAGCAGCAGCACGCCAAAACACAAGCTGGCATGCCGGTCAAAAAGAAGCGCGAACGCTCCTGCTTGGACATTCTGGAG
CAAATGTTTGAGCCGCAAAATACTAAGAATAGCGACCTAGGTGTTCAATTATTGCCAAATTTGCCCAGCCCCAAGCAACAGCTGCC
GCCCCGCCGCTTCACTTTGCTGGATGATGACATCTTTGAGATAACTAATAATGGCGAATTCGGCAGTCGGCTGCGTCTCGACTCCA
TGGGGAATGTGTCGCCCGTGCAGCAGCAACAGTCACAGCCCCAGCGCAACAAAATTACTAACTATTTGATTAGCTCCGGCGGCACT
CCAGAGGATGCTGGCCAGCGAACTCAAGCCAGTCAAGTGCTGCGCAAGTCACCCAAAAGCATCAAGTCCACACAGTCGACGAAACT
GACACGTTGGTTTGCCACTGCAGCCGCGGTGTCAGGGGAATCCCAAAGTGCGCCTAGTACGCCTGTAGTGCCAACCGACAAAATGG
CGCGCGCTCGCATGGCTCGCAGTGGTGGAGCTGGAAAACGCAAGCGACTGCAGCTCGATAAGTGA

***>D.grimshawi SuUR* ORF**

ATGTATCATTTTATCTCCGAGCGGACAGCGGAACTGCGCCTCTCCGAGAACGTGCTCATTCCCAGTCATGCGACGCAGTATTTGAA
GTGCTTTCAACTGGACGCTGTGCGTTTTCTCTACGAGCGGCTGTCCAAACAAGATTTTTGCATATTTAACGATGAGAGTGGCTTGG
GCAAAACTGCGACAATTGTAACACTCTTAAATGGTCTGGGTGCGTCGAAGAAAACGCTCATTGTGTTGCAGAACGATGATCAGTTG
CTGGCTGGCTGGCAGTTTCACCTGGGCATTCTATCCGATCTGCCCGTATGCATTCTAAAGGATGTCAATGACAGCACAGAGTCGGC
GCATAGCGTTTATCTGAGCAAATGGAGCGTGTTGCGCAGCATTGGGGATCTCAGCAAGCTGAAGTTTGACTATGTCATTGTTGATC
ATCGGGGCTACATGCTAAATAACAACTTCTGCACTTCCATGCTACTGCAGCAATACGAGCGCAAAGTTAACATTGTGATCTCTACT
GTGGACTTAACGTCCGATGTTAAGTTACTATATAATGTGCTACGCTTGGGTGGATGCCTGGAACATCAACACAAGAGCTTCCGTAT
ATTTAATCTGAAATTCAATCTGCCGGATGTTAAGGAGGTGCTTAACAAACGTGTTGATCTCGAGGATTACTACAAGCAGCGAGGTG
TTCTTGGCGAATACATCAAGGATTTCCGGCTGCGTCGATATCGCCACCAGTTCGAGTCATATTTACCTCTAGTCACGCCGGAACAA
TACAAGATCAATGTATCGCTATGGATGGGTGAAAATATAAGCAATTCTACCATAAGTGGCTCAAGTGTAGACCCACCATCGGAAGC
ACGCTCCACGGGCAGCACTGGTGAAATATTCGAGTGTCTGATGAGCATAAGACGAGAACGAGAATTGGGCCAACAACAGGATGCAG
AGAAAATATCGCTATCCGAGCACAGTGACGAAGTGGTAGCCATGGAGCCGTTGATTTTCGAAATGTCCGATTCAGAGGCGGAGGCA
GACGTAAACGTGCCGACGCAAGCCAAAGTGTCAACCACGGATGTGGTGGTGCTATCCAGCGATGATTGTGAAATAATTGCGACGCC
AAGCACTCCGCCACAGTCCTCACCAAAGGCAAGCACGTCACCAATAGCCAAAACCAAGCGAAAGTACACCAAACGTGTCAAGGCTG
CGAAGCCAACAGAGCTAACCGAATCGGAGAACGATGAACAGCCAAGCAGTCCCTCGAAGATATTCGCCAGAAAATTAAATGTTAAA
ATAAATCGTCTTCATTTACCCAAAATGCAAAAAAATGTACTAGCAGCAATTAGTTCGAAAATACCCACAAACCAGTTGACAAAACC
TGCGCCTGCGCCTGCGCCACGTGAGCAACCGACTCCATTGCCAGTTAAATCGCCTGTCAAAGCTAAGCCGATGCCAGAGCCTAACG
CACAAAATGTGAAACAGTTATCGCCAGAGCCACGAACGCCCAACCATCAGAAGACGTTGCCAATCACGCCAAAGACTGAGCCCAGA
TCAAAACGCGCACAGGAAAAAAGTGCTTCAGCTGTGTTGACGCCAAGAGTGACACGTGGAATGCAAAGAATGACGCGTTCAGCGGA
ATCGCGACACAACAGCAAATATATGACACCACATTTGGTATTAGGCTCAACGAAACGGAAACGCAACATAGACTCAAATCAAACGC
CCAAAACGAGCAAGCGGAGAAGGGAGCCAATAGTAATACAAGAGGAGGCAGCTCCGACAGCAAAGCTGGAGTCGACAGCAGAACCC
AAAATCAAAAAGCCAAGGGGACGGCCGCCAAAGAAAACAAAATTAAGCAACGGATCAACGTCTTTGGCCAGCAACGAAAGTAAGGC
AGTGGCACCCCGCATATTAAGAGGAAGACCGTCGAAGAAAACGATATCTAAGGATCCGAAGTCCTTACCAATTAATTCAAAGTCTC
CCATAATTCAGCCAGCTGCCACGCCAGAAATAATTAGTCCCGGCTCTTTCTTGTCGGACTATTCGTATATGCAATGCGCCCAGAAA
CTACCCGAGAATCTTGCAGAGCTTGGTTCAATGGAATTGCCAGAATTTCGTGTACCACTTGCGCCGCCGGCAACGCCACTAATGAC
GCCACTCACACTGAATTTCTTTAGTGACTCGGAGGTGGTCGTCGTGCCCAATGGCAATCAGCAGCAATCAGATGTTGTGGTGATCC
ACAGTTCGCTTGACGAGAGCTCGGATCAATCCGCTAAGCAGTCACAAAGTCGACGCACTAGAGCACTAAAGCGGAAACGTTCCACC
CCACTCAAAGAGCAGCCACAACCGAGCATCTCCAGTTTTGGGCAATTACTTGCACAGCAGCGCGCGATGGCCAACAAATCGCCCGA

CATATTCAGCAACTGCTCCGATTTGTCGCAGCTGACGCTGGCACAGCCAGTGCCCTTCGAAGGCTTCAAGATCTTTGGATCGGAAG
TAAAGCAACAGCAGCAACAACAGAAGCAGCAGCATTTGAAGGCACAACCAGGCACGATGGCTAAAAAGAAACGTGAACGCTCCTGC
CTGGACATCTTGGAGCAAATTTTTGAGACGAATCGGACTCAAGCCAACGAATCGGGAGTCCAATTGCTGCCAAATGTGTCCAGCCC
CAAACAGGTGTTAACGCAACGACGTTTCTCTTTGCTGGACGATGACATCTTTGAGATAACCAATAATGGTGAATTTGGCAGTCGAC
TGCGTCTCACCGCCACGGGGCATGTGTCGCCGGTGCTGCAGCCGCAGCGCAATAAAATCACAAACTATTTGATTGGCTCCAGCAGT
ACACAGGAAGATGCTGGCCAGCGCACTCAATGCAGTCAAGGGTTGCGAAAGTCACCCAAAAGCATCAAATCGACACAGTCAACGAA
GCTGACGCGATGGTTTGGTGCAGCAACATCATCAGCAGGAGGAGCCGGGGAATCCCAAAGTGTGCCCAACACACCTGTGGTGCCTG
GCAAGAAATCGGGCTGCGCTCGCATCGCTCGCAGCGGAGGAGCTGGAAAACGAAAGAGACTAGATCTCTACAAGTGA

FIGURE S1.—List of *SuUR* ORFs in different Drosophila species used.

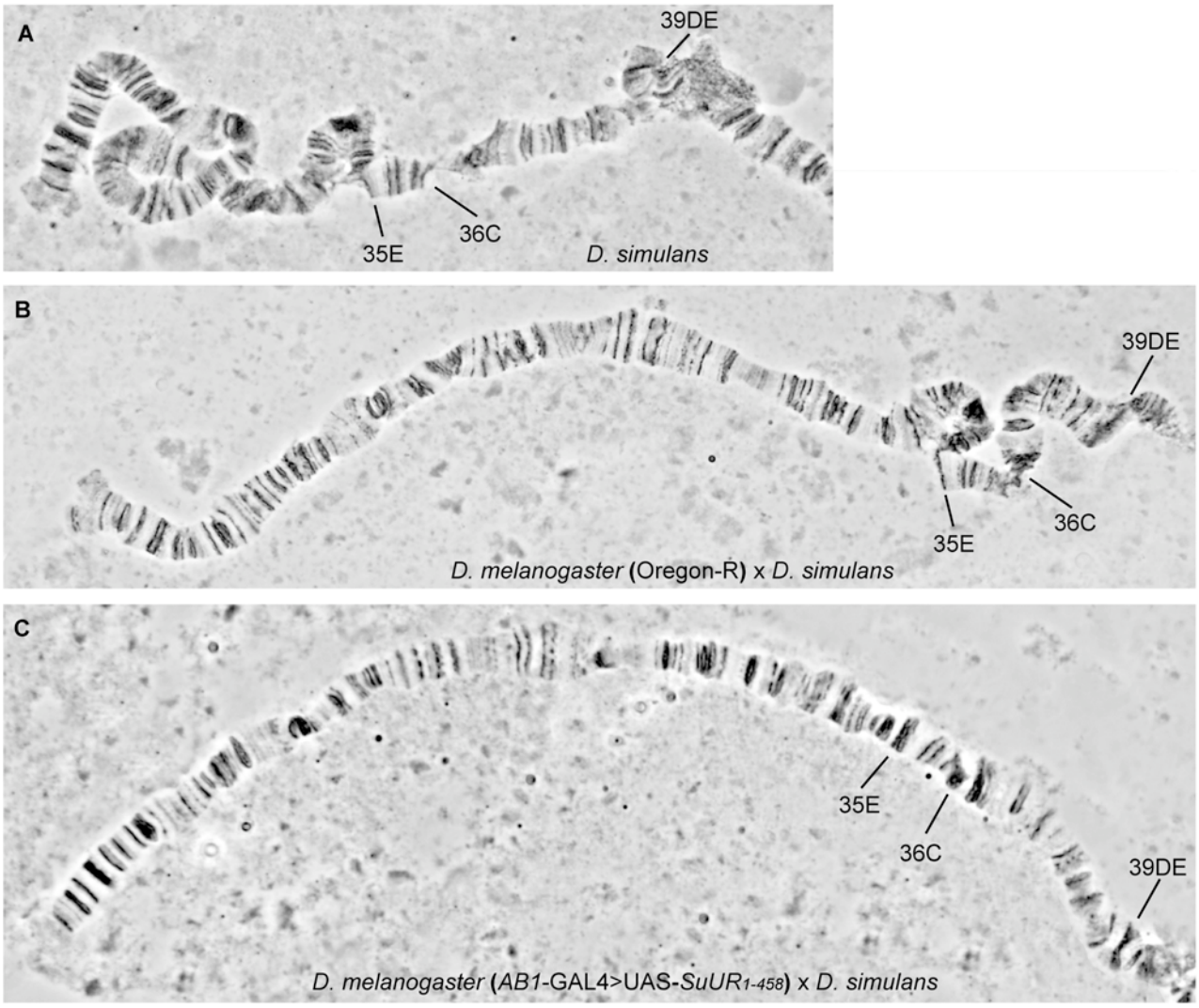
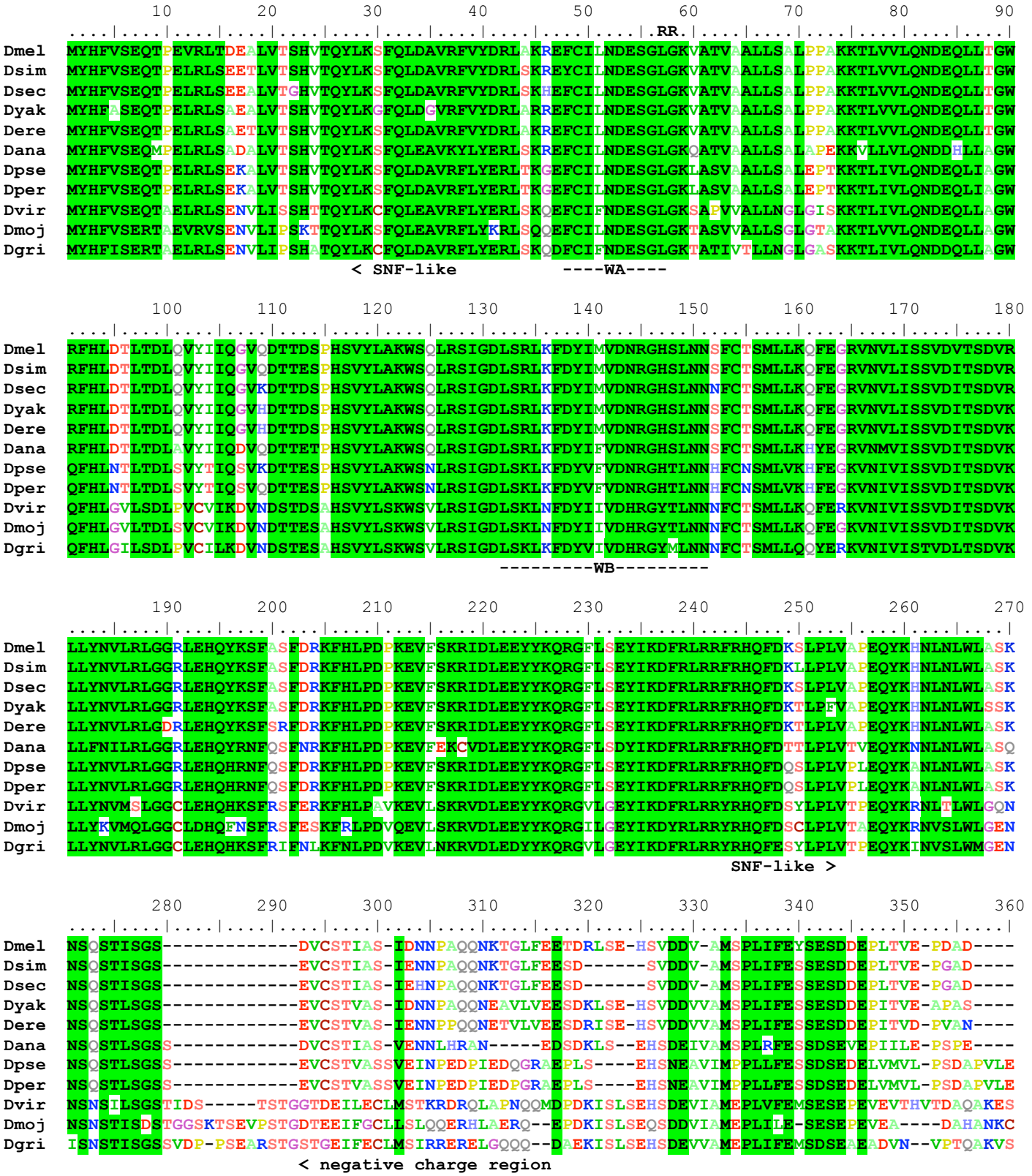
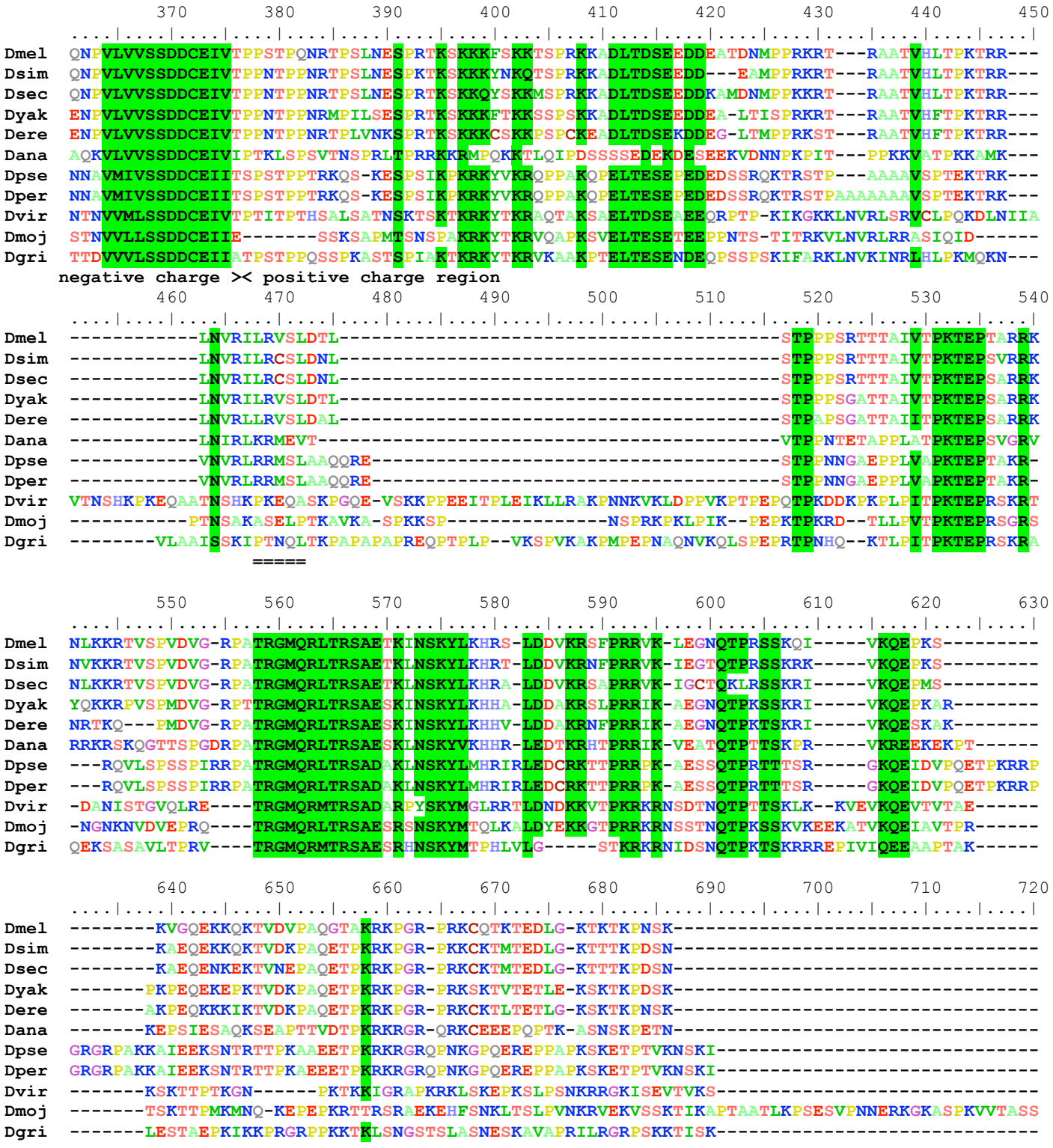
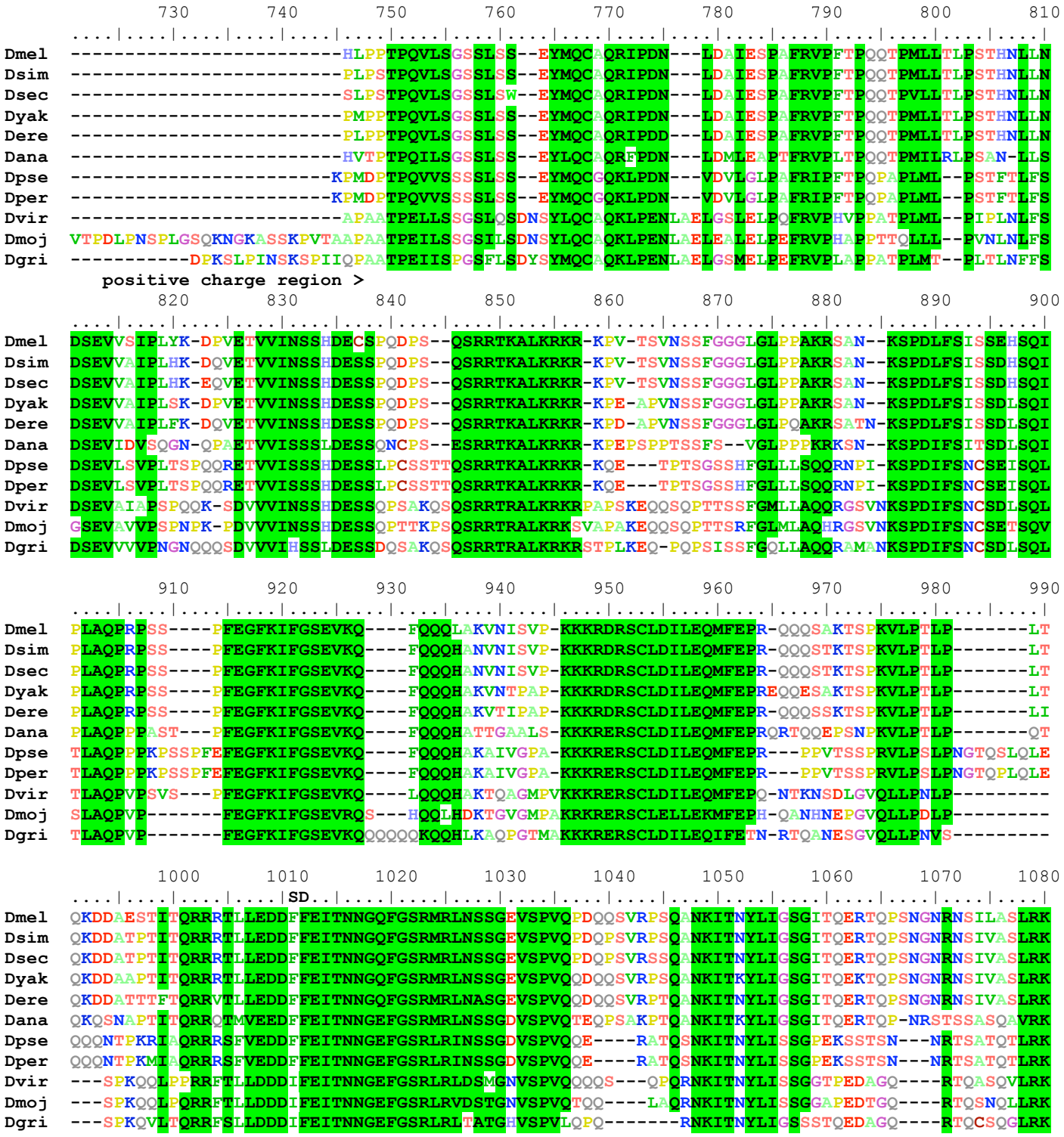


FIGURE S2.—Dominant-negative effect of ectopic $SUUR_{1\text{-}458}$ fragment on chromosome breaks in *D. melanogaster* x *D. simulans* hybrids. All panels show 2L chromosome, with typical chromosome breaks signed. (A) *D. simulans*. (B) *D. melanogaster* (Oregon-R) x *D. simulans* hybrids. (C) Breaks are absent in *D. melanogaster* x *D. simulans* hybrids with *AB1*-GAL4>UAS-$SuUR_{1\text{-}458}$ transgenes expressing dominant-negative fragment $SUUR_{1\text{-}458}$ under control of salivary gland specific driver.


10 20 30 40 50 60 70 80 90
RR.
Dmel MYHFVSEQTPEVRLTDEALVTSHVTQYLKSFQLDAVRFVYDRLAKREFCILNDESGLGKVATVAALLSALPPAKKTLVVLQNDEQLLTGW
Dsim MYHFVSEQTPELRLSEETLVTSHVTQYLKSFQLDAVRFVYDRLSKREYCILNDESGLGKVATVAALLSALPPAKKTLVVLQNDEQLLTGW
Dsec MYHFVSEQTPELRLSEEALVTGHVTQYLKSFQLDAVRFVYDRLSKHEFCILNDESGLGKVATVAALLSALPPAKKTLVVLQNDEQLLTGW
Dyak MYHFASEQTPELRLSAEALVTSHVTQYLKGFQLDGVRFVYDRLARREFCILNDESGLGKVATVAALLSALPPAKKTLVVLQNDEQLLTGW
Dere MYHFVSEQTPELRLSAETLVTSHVTQYLKSFQLDAVRFVYDRLAKREFCILNDESGLGKVATVAALLSALPPAKKTLVVLQNDEQLLTGW
Dana MYHFVSEQMPELRLSADALVTSHVTQYLKSFQLEAVKYLYERLSKREFCILNDESGLGKQATVAALLSALAPEKKVLLVLQNDDHLLAGW
Dpse MYHFVSEQTPELRLSEKALVTSHVTQYLKSFQLDAVRFLYERLTKGEFCILNDESGLGKLASVAALLSALEPTKKTLIVLQNDEQLIAGW
Dper MYHFVSEQTPELRLSEKALVTSHVTQYLKSFQLDAVRFLYERLTKGEFCILNDESGLGKLASVAALLSALEPTKKTLIVLQNDEQLIAGW
Dvir MYHFVSEQTAELRLSENVLISSHTTQYLKCFQLEAVRFLYERLSKQEFCIFNDESGLGKSAPVVALLNGLGISKKTLIVLQNDEQLLAGW
Dmoj MYHFVSERTAEVRVSENVLIPSKTTQYLKSFQLEAVRFLYKRLSQQEFCILNDESGLGKTASVVALLSGLGTAKKTLVVLQNDEQLLAGW
Dgri MYHFISERTAELRLSENVLIPSHATQYLKCFQLDAVRFLYERLSKQDFCIFNDESGLGKTATIVTLLNGLGASKKTLIVLQNDDQLLAGW
< SNF-like ----WA----
100 110 120 130 140 150 160 170 180
Dmel RFHLDTLTDLQVYIIQGVQDTTDSPHSVYLAKWSQLRSIGDLSRLKFDYIMVDNRGHSLNNSFCTSMLLKQFEGRVNVLISSVDVTSDVR
Dsim RFHLDTLTDLQVYIIQGVQDTTESPHSVYLAKWSQLRSIGDLSRLKFDYIMVDNRGHSLNNSFCTSMLLKQFEGRVNVLISSVDITSDVR
Dsec RFHLDTLTDLQVYIIQGVKDTTDSPHSVYLAKWSQLRSIGDLSRLKFDYIMVDNRGHSLNNNFCTSMLLKQFEGRVNVLISSVDITSDVR
Dyak RFHLDTLTDLQVYIIQGVHDTTDSPHSVYLAKWSQLRSIGDLSRLKFDYIMVDNRGHSLNNSFCTSMLLKQFEGRVNVLISSVDITSDVK
Dere RFHLDTLTDLQVYIIQGVHDTTDSPHSVYLAKWSQLRSIGDLSRLKFDYIMVDNRGHSLNNSFCTSMLLKQFEGRVNVLISSVDITSDVK
Dana RFHLDTLTDLAVYIIQDVQDTTETPHSVYLAKWSQLRSIGDLSRLKFDYIFVDNRGHSLNNSFCTSMLLKHYEGRVNMVISSVDITSDVK
Dpse QFHLNTLTDLSVYTIQSVKDTTESPHSVYLAKWSNLRSIGDLSKLKFDYVFVDNRGHTLNNHFCNSMLVKHFEGKVNIVISSVDITSDVK
Dper QFHLNTLTDLSVYTIQSVQDTTESPHSVYLAKWSNLRSIGDLSKLKFDYVFVDNRGHTLNNHFCNSMLVKHFEGKVNIVISSVDITSDVK
Dvir QFHLGVLSDLPVCVIKDVNDSTDSAHSVYLSKWSVLRSIGDLSKLNFDYIIVDHRGYTLNNNFCTSMLLKQFERKVNIVISSVDITSDVK
Dmoj QFHLGVLTDLSVCVIKDVNDTTESAHSVYLSKWSVLRSIGDLSKLNFDYIIVDHRGYTLNNNFCTSMLLKQFEGKVNIVISSVDITSDVK
Dgri QFHLGILSDLPVCILKDVNDSTESAHSVYLSKWSVLRSIGDLSKLKFDYVIVDHRGYMLNNNFCTSMLLQQYERKVNIVISTVDLTSDVK
---------WB---------
190 200 210 220 230 240 250 260 270
Dmel LLYNVLRLGGRLEHQYKSFASFDRKFHLPDPKEVFSKRIDLEEYYKQRGFLSEYIKDFRLRRFRHQFDKSLPLVAPEQYKHNLNLWLASK
Dsim LLYNVLRLGGRLEHQYKSFASFDRKFHLPDPKEVFSKRIDLEEYYKQRGFLSEYIKDFRLRRFRHQFDKLLPLVAPEQYKHNLNLWLASK
Dsec LLYNVLRLGGRLEHQYKSFASFDRKFHLPDPKEVFSKRIDLEEYYKQRGFLSEYIKDFRLRRFRHQFDKSLPLVAPEQYKHNLNLWLASK
Dyak LLYNVLRLGGRLEHQYKSFASFDRKFHLPDPKEVFSKRIDLEEYYKQRGFLSEYIKDFRLRRFRHQFDKTLPFVAPEQYKHNLNLWLSSK
Dere LLYNVLRLGDRLEHQYKSFSRFDRKFHLPDPKEVFSKRIDLEEYYKQRGFLSEYIKDFRLRRFRHQFDKTLPLVAPEQYKHNLNLWLASK
Dana LLFNILRLGGRLEHQYRNFQSFNRKFHLPDPKEVFEKCVDLEEYYKQRGFLSDYIKDFRLRRFRHQFDTTLPLVTVEQYKNNLNLWLASQ
Dpse LLYNVLRLGGRLEHQHRNFQSFDRKFHLPDPKEVFSKRIDLEEYYKQRGFLSEYIKDFRLRRFRHQFDQSLPLVPLEQYKANLNLWLASK
Dper LLYNVLRLGGRLEHQHRNFQSFDRKFHLPDPKEVFSKRIDLEEYYKQRGFLSEYIKDFRLRRFRHQFDQSLPLVPLEQYKANLNLWLASK
Dvir LLYNVMSLGGCLEHQHKSFRSFERKFHLPAVKEVLSKRVDLEEYYKQRGVLGEYIKDFRLRRYRHQFDSYLPLVTPEQYKRNLTLWLGQN
Dmoj LLYKVMQLGGCLDHQFNSFRSFESKFRLPDVQEVLSKRVDLEEYYKQRGILGEYIKDYRLRRYRHQFDSCLPLVTAEQYKRNVSLWLGEN
Dgri LLYNVLRLGGCLEHQHKSFRIFNLKFNLPDVKEVLNKRVDLEDYYKQRGVLGEYIKDFRLRRYRHQFESYLPLVTPEQYKINVSLWMGEN
SNF-like >
280 290 300 310 320 330 340 350 360
Dmel NSQSTISGS----------DVCSTIAS-IDNNPAQQNKTGLFEETDRLSE-HSVDDV-AMSPLIFEYSESDDEPLTVE-PDAD----
Dsim NSQSTISGS----------EVCSTIAS-IENNPAQQNKTGLFEESD------SVDDV-AMSPLIFESSESDDEPLTVE-PGAD----
Dsec NSQSTISGS----------EVCSTIAS-IEHNPAQQNKTGLFEESD------SVDDV-AMSPLIFESSESDDEPLTVE-PGAD----
Dyak NSQSTLSGS----------EVCSTVAS-IDNNPAQQNEAVLVEESDKLSE-HSVDDVVAMSPLIFESSESDDEPITVE-APAS----
Dere NSQSTLSGS----------EVCSTVAS-IENNPPQQNETVLVEESDRISE-HSVDDVVAMSPLIFESSESDDEPITVD-PVAN----
Dana NSQSTLSGSS---------DVCSTIAS-VENNLHRAN-----EDSDKLS--EHSDEIVAMSPLRFESSDSEVEPIILE-PSPE----
Dpse NSQSTLSGSS---------EVCSTVASSVEINPEDPIEDQGRAEPLS----EHSNEAVIMPLLFESSDSEDELVMVL-PSDAPVLE
Dper NSQSTLSGSS---------EVCSTVASSVEINPEDPIEDPGRAEPLS----EHSNEAVIMPLLFESSDSEDELVMVL-PSDAPVLE
Dvir NSNSILSGSTIDS-----TSTGGTDEILECLMSTKRDRQLAPNQQMDPDKISLSEHSDEVIAMEPLVFEMSESEPEVEVTHVTDAQAKES
Dmoj NSNSTISDSTGGSKTSEVPSTGDTEEIFGCLLSLQQERHLAERQ--EPDKISLSEQSDDVIAMEPLILE-SESEPEVEA----DAHANKC
Dgri ISNSTISGSSVDP-PSEARSTGSTGEIFECLMSIRRERELGQQQ--DAEKISLSEHSDEVVAMEPLIFEMSDSEAEADVN--VPTQAKVS
< negative charge region


370 380 390 400 410 420 430 440 450
Dmel QNPVLVVSSDDCEIVTPPSTPQNRTPSLNESPRTKSKKKFSKKTSPRKKADLTDSEEDDEATDNMPPRKRT---RAATVHLTPKTRR---
Dsim QNPVLVVSSDDCEIVTPPNTPPNRTPSLNESPKTKSKKKYNKQTSPRKKADLTDSEEDD---EAMPPRKRT---RAATVHLTPKTRR---
Dsec QNPVLVVSSDDCEIVTPPNTPPNRTPSLNESPRTKSKKQYSKKMSPRKKADLTDSEEDDKAMDNMPPKKRT---RAATVHLTPKTRR---
Dyak ENPVLVVSSDDCEIVTPPNTPPNRMPILSESPRTKSKKKFTKKSSPSKKADLTDSEEDDEA-LTISPRKRT---RAATVHFTPKTRR---
Dere ENPVLVVSSDDCEIVTPPNTPPNRTPLVNKSPRTKSKKKCSKKPSPCKEADLTDSEKDDEG-LTMPPRKST---RAATVHFTPKTRR---
Dana AQKVLVVSSDDCEIVIPTKLSPSVTNSPRLTPRRKKRMPQKKTLQIPDSSSSEDEKDESEEKVDNNPKPIT---PPKKVATPKKAMK---
Dpse NNAVMIVSSDDCEIITSPSTPPTRKQS-KESPSIKPKRKYVKRQPPAKQPELTESEPEDEDSSRQKTRSTP---AAAAVSPTEKTRK---
Dper NNAVMIVSSDDCEIITSPSTPPTRKQS-KESPSIKPKRKYVKRQPPAKQPELTESEPEDEDSSRQKTRSTPAAAAAAAVSPTEKTRK---
Dvir NTNVVMLSSDDCEIVTPTITPTHSALSATNSKTSKTKRKYTKRAQTAKSAELTDSEAEEQRPTP-KIKGKKLNVRLSRVCLPQKDLNIIA
Dmoj STNVVLLSSDDCEIIE-------SSKSAPMTSNSPAKRKYTKRVQAPKSVELTESETEEPPNTS-TITRKVLNVRLRRASIQID------
Dgri TTDVVVLSSDDCEIIATPSTPPQSSPKASTSPIAKTKRKYTKRVKAAKPTELTESENDEQPSSPSKIFARKLNVKINRLHLPKMQKN---
negative charge >< positive charge region
460 470 480 490 500 510 520 530 540
Dmel ------------LNVRILRVSLDTL---STPPPSRTTTAIVTPKTEPTARRK
Dsim ------------LNVRILRCSLDNL---STPPPSRTTTAIVTPKTEPSVRRK
Dsec ------------LNVRILRCSLDNL---STPPPSRTTTAIVTPKTEPSARRK
Dyak ------------LNVRILRVSLDTL---STPPPSGATTAIVTPKTEPSARRK
Dere ------------LNVRLLRVSLDAL---STPAPSGATTAIITPKTEPSARRK
Dana ------------LNIRLKRMEVT---VTPPNTETAPPLATPKTEPSVGRV
Dpse ------------VNVRLRRMSLAAQQRE---STPPNNGAEPPLVAPKTEPTAKR-
Dper ------------VNVRLRRMSLAAQQRE---STPPNNGAEPPLVAPKTEPTAKR-
Dvir VTNSHKPKEQAATNSHKPKEQASKPGQE-VSKKPPEEITPLEIKLLRAKPNNKVKLDPPVKPTPEPQTPKDDKPKPLPITPKTEPRSKRT
Dmoj -----------PTNSAKASELPTKAVKA-SPKKSP---------------NSPRKPKLPIK--PEPKTPKRD--TLLPVTPKTEPRSGRS
Dgri --------VLAAISSKIPTNQLTKPAPAPAPREQPTPLP--VKSPVKAKPMPEPNAQNVKQLSPEPRTPNHQ--KTLPITPKTEPRSKRA
=====
550 560 570 580 590 600 610 620 630
Dmel NLKKRTVSPVDVG-RPATRGMQRLTRSAETKINSKYLKHRS-LDDVKRSFPRRVK-LEGNQTPRSSKQI-----VKQEPKS---------
Dsim NVKKRTVSPVDVG-RPATRGMQRLTRSAETKLNSKYLKHRT-LDDVKRNFPRRVK-IEGTQTPRSSKRK-----VKQEPKS---------
Dsec NLKKRTVSPVDVG-RPATRGMQRLTRSAETKLNSKYLKHRA-LDDVKRSAPRRVK-IGCTQKLRSSKRI-----VKQEPMS---------
Dyak YQKKRPVSPMDVG-RPTTRGMQRLTRSAESKINSKYLKHHA-LDDAKRSLPRRIK-AEGNQTPKSSKRI-----VKQEPKAR--------
Dere NRTKQ---PMDVG-RPATRGMQRLTRSAESKINSKYLKHHV-LDDAKRNFPRRIK-AEGNQTPKTSKRI-----VKQESKAK--------
Dana RRKRSKQGTTSPGDRPATRGMQRLTRSAESKLNSKYVKHHR-LEDTKRHTPRRIK-VEATQTPTTSKPR-----VKREEKEKPT------
Dpse ---RQVLSPSSPIRRPATRGMQRLTRSADAKLNSKYLMHRIRLEDCRKTTPRRPK-AESSQTPRTTTSR-----GKQEIDVPQETPKRRP
Dper ---RQVLSPSSPIRRPATRGMQRLTRSADAKLNSKYLMHRIRLEDCRKTTPRRPK-AESSQTPRTTTSR-----GKQEIDVPQETPKRRP
Dvir -DANISTGVQLRE----TRGMQRMTRSADARPYSKYMGLRRTLDNDKKVTPKRKRNSDTNQTPTTSKLK--KVEVKQEVTVTAE------
Dmoj -NGNKNVDVEPRQ----TRGMQRLTRSAESRSNSKYMTQLKALDYEKKGTPRRKRNSSTNQTPKSSKVKEEKATVKQEIAVTPR------
Dgri QEKSASAVLTPRV----TRGMQRMTRSAESRHNSKYMTPHLVLG-----STKRKRNIDSNQTPKTSKRRREPIVIQEEAAPTAK------
640 650 660 670 680 690 700 710 720
Dmel --------KVGQEKKQKTVDVPAQGTAKRKPGR-PRKCQTKTEDLG-KTKTKPNSK-----------------------------------
Dsim --------KAEQEKKQKTVDKPAQETPKRKPGR-PKKCKTMTEDLG-KTTTKPDSN-----------------------------------
Dsec --------KAEQENKEKTVNEPAQETPKRKPGR-PRKCKTMTEDLG-KTTTKPDSN-----------------------------------
Dyak -------PKPEQEKEPKTVDKPAQETPKRKPGR-PRKSKTVTETLE-KSKTKPDSK-----------------------------------
Dere -------AKPEQKKKIKTVDKPAQETPKRKPGR-PRKCKTLTETLG-KSKTKPNSK-----------------------------------
Dana -------KEPSIESAQKSEAPTTVDTPKRKRGR-QRKCEEEPQPTK-ASNSKPETN-----------------------------------
Dpse GRGRPAKKAIEEKSNTRTTPKAAEETPKRKRGRQPNKGPQERE PPAPKSKETPTVKNSKI---------------------------
Dper GRGRPAKKAIEEKSNTRTTPKAEEETPKRKRGRQPNKGPQERE PPAPKSKETPTVKNSKI---------------------------
Dvir -------KSKTTPTKGN------PKTKKIGRAPKRKLSKEPKSLPSNKRRGKISEVTVKS-----------------------------
Dmoj -------TSKTTPMKMNQ-KEPEPKRTTRSRAEKEHFSNKLTSLPVNKRVEKVSSKTIKAPTAATLKPSESVPNNERKGKASPKVVTASS
Dgri -------LESTAEPKIKKPRGRPPKKTKLSNGSTSLASNESKAVAPRILRGRPSKKTISK-----------------------------


positive charge region >
SD..
Dmel
Dsim
Dsec
Dyak
Dere
Dana
Dpse
Dper
Dvir
Dmoj
Dgri

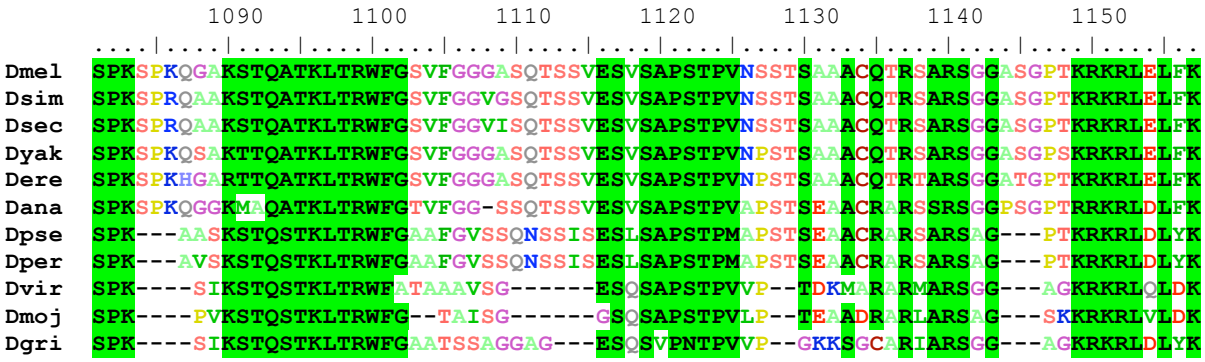
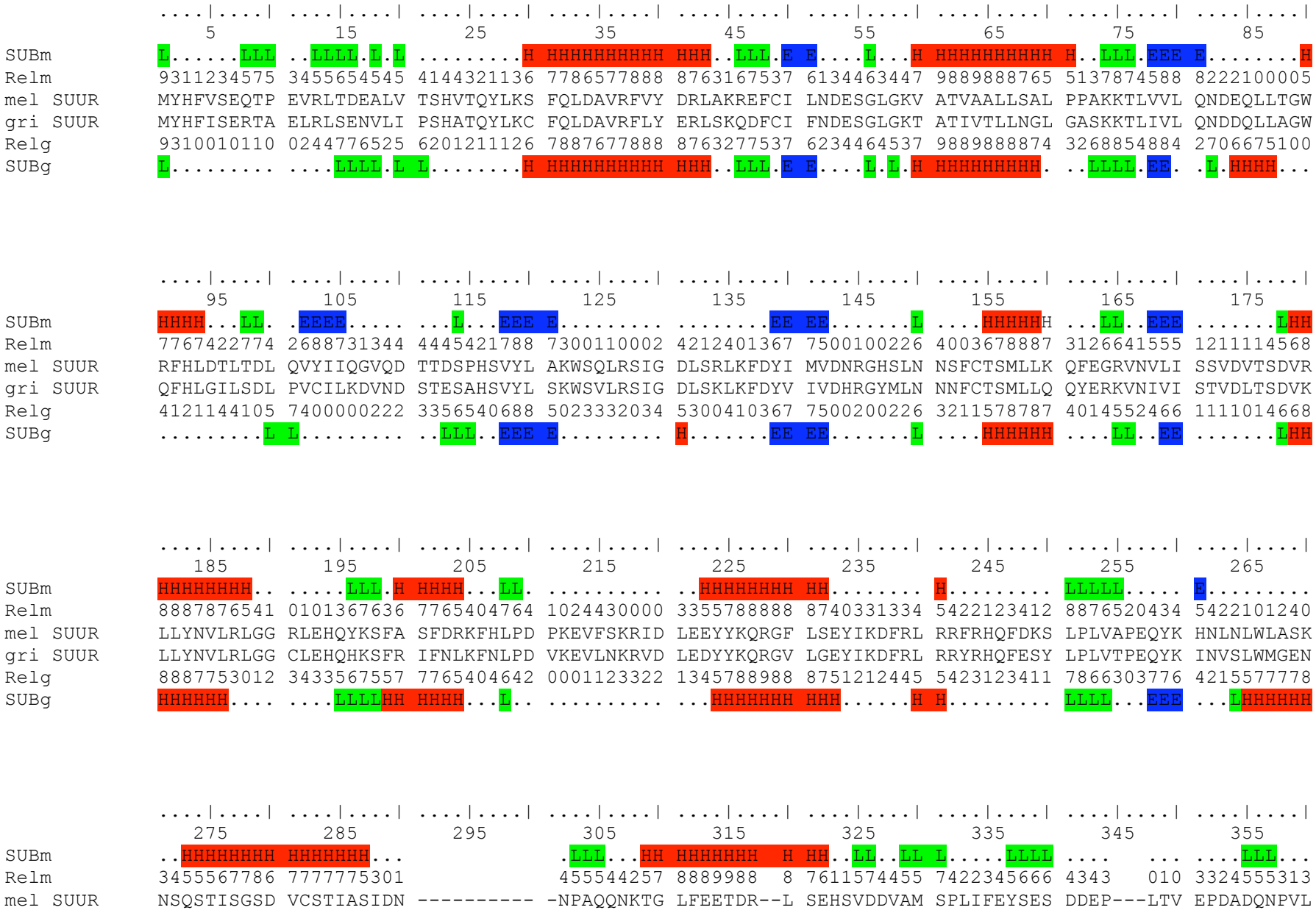
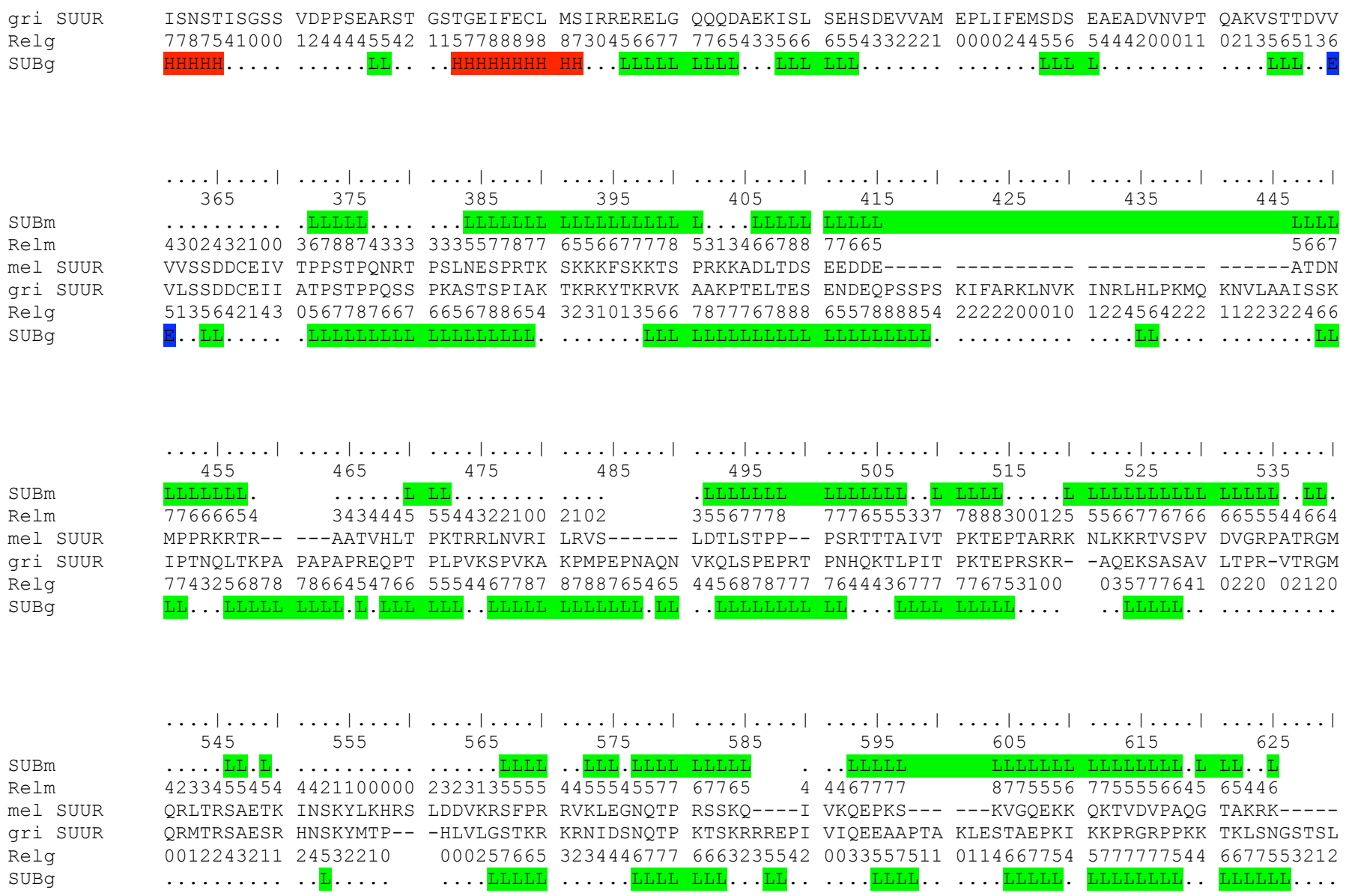
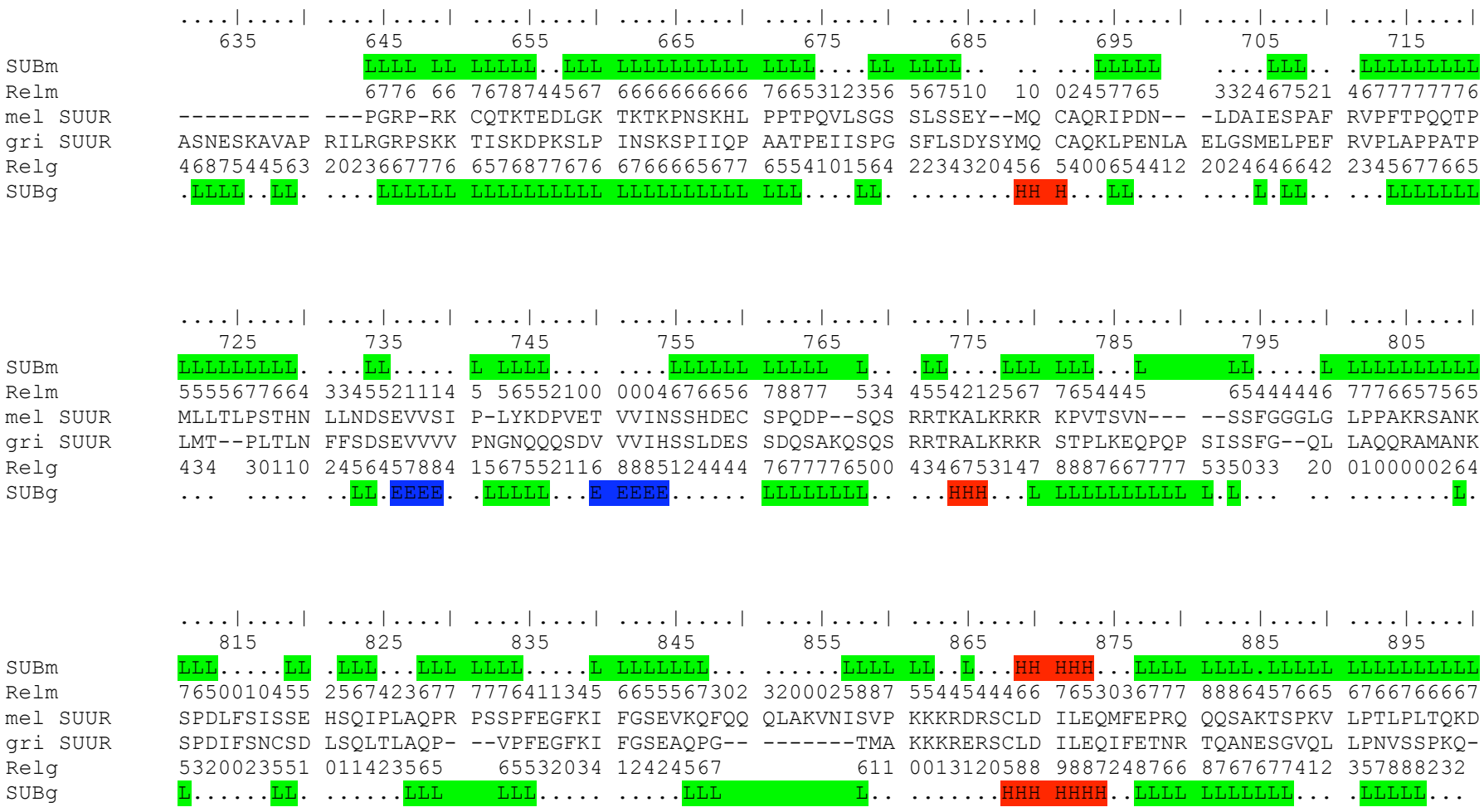


FIGURE S3.—Alignment of SUUR protein in different Drosophila species. Positions that are similar in at least 10 species out of 11 are shown in green. Regions with moderate similarity to the Walker A (WA) and Walker B (WB) motifs (MAKUNIN *et al.* 2002) are underlined, putative HP1-binding motif (PINDYURIN *et al.* 2008) and homologous sequences in other species are double-underlined. Boundaries of analyzed domains are shown below the alignment as "<" and ">". Introduced substitutions in *D. melanogaster* SUUR protein are shown above the alignment.


....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
5 15 25 35 45 55 65 75 85
SUBm L.....LLL ..LLLL.L.L ....H HHHHHHHHHH HHH..LLL.E E....L...H HHHHHHHHHH H..LLL.EEE E.....H
Relm 9311234575 3455654545 4144321136 7786577888 8763167537 6134463447 9889888765 5137874588 8222100005
mel SUUR MYHFVSEQTP EVRLTDEALV TSHVTQYLKS FQLDAVRFVY DRLAKREFCI LNDESGLGKV ATVAALLSAL PPAKKTLVVL QNDEQLLTGW
gri SUUR MYHFISERTA ELRLSENVLI PSHATQYLKC FQLDAVRFLY ERLSKQDFCI FNDESGLGKT ATIVTLLNGL GASKKTLIVL QNDDQLLAGW
Relg 9310010110 0244776525 6201211126 7887677888 8763277537 6234464537 9889888874 3268854884 2706675100
SUBg L......... ....LLLL.L L.....H HHHHHHHHHH HHH..LLL.E E....L.L.H HHHHHHHHH. ..LLLL.EE. .L.HHHH...
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
95 105 115 125 135 145 155 165 175
SUBm HHHH...LL. .EEEE..... ...L...EEE E..... .....EE EE.....L ....HHHHHH ...LL..EEE .....LHH
Relm 7767422774 2688731344 4445421788 7300110002 4212401367 7500100226 4003678887 3126641555 1211114568
mel SUUR RFHLDTLTDL QVYIIQGVQD TTDSPHSVYL AKWSQLRSIG DLSRLKFDYI MVDNRGHSLN NSFCTSMLLK QFEGRVNVLI SSVDVTSDVR
gri SUUR QFHLGILSDL PVCILKDVND STESAHSVYL SKWSVLRSIG DLSKLKFDYV IVDHRGYMLN NNFCTSMLLQ QYERKVNIVI STVDLTSDVK
Relg 4121144105 7400000222 3356540688 5023332034 5300410367 7500200226 3211578787 4014552466 1111014668
SUBg ....L L..... ..LLL..EEE E..... H.....EE EE.....L ....HHHHHH ....LL..EE .....LHH
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
185 195 205 215 225 235 245 255 265
SUBm HHHHHHHH.. .....LLL.H HHHH...LL. ....HHHHHHH HH..... H..... LLLLL..... E.....
Relm 8887876541 0101367636 7765404764 1024430000 3355788888 8740331334 5422123412 8876520434 5422101240
mel SUUR LLYNVLRLGG RLEHQYKSFA SFDRKFHLPD PKEVFSKRID LEEYYKQRGF LSEYIKDFRL RRFRHQFDKS LPLVAPEQYK HNLNLWLASK
gri SUUR LLYNVLRLGG CLEHQHKSFR IFNLKFNLPD VKEVLNKRVD LEDYYKQRGV LGEYIKDFRL RRYRHQFESY LPLVTPEQYK INVSLWMGEN
Relg 8887753012 3433567557 7765404642 0001123322 1345788988 8751212445 5423123411 7866303776 4215577778
SUBg HHHHHH.... ....LLLLHH HHHH...L.. ....HHHHHH HHH.....H H..... LLLL...EEE ...LHHHHHH
....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
275 285 295 305 315 325 335 345 355
SUBm ..HHHHHHH HHHHHHH... .LLL...HH HHHHHHH H HH..LL..LL L.....LLLL ....LLL...
Relm 3455567786 7777775301 455544257 8889988 8 7611574455 7422345666 4343 010 3324555313
mel SUUR NSQSTISGSD VCSTIASIDN ---------- -NPAQQNKTG LFEETDR--L SEHSVDDVAM SPLIFEYSES DDEP---LTV EPDADQNPVL


gri SUUR ISNSTISGSS VDPPSEARST GSTGEIFECL MSIRRERELG QQQDAEKISL SEHSDEVVAM EPLIFEMSDS EAEADVNVPT QAKVSTTDVV
Relg 7787541000 1244445542 1157788898 8730456677 7765433566 6554332221 0000244556 5444200011 0213565136
SUBg HHHHH..... ......LL.. ..HHHHHHHH HH...LLLLL LLLL...LLL LLL..... ....LLL L......... ....LLL..E
365 375 385 395 405 415 425 435 445
SUBm ..... .LLLLL.... ...LLLLLLL LLLLLLLLL L....LLLLL LLLLL LLLL
Relm 4302432100 3678874333 3335577877 6556677778 5313466788 77665 5667
mel SUUR VVSSDDCEIV TPPSTPQNRT PSLNESPRTK SKKKFSKKTS PRKKADLTDS EEDDE----- ---------- ---------- ------ATDN
gri SUUR VLSSDDCEII ATPSTPPQSS PKASTSPIAK TKRKYTKRVK AAKPTELTES ENDEQPSSPS KIFARKLNVK INRLHLPKMQ KNVLAAISSK
Relg 5135642143 0567787667 6656788654 3231013566 7877767888 6557888854 2222200010 1224564222 1122322466
SUBg E..LL..... .LLLLLLLLL LLLLLLLLL. ....LLL LLLLLLLLL LLLLLLLLL. ....LL.... LL
455 465 475 485 495 505 515 525 535
SUBm LLLLLLL. ......L LL..... .... .LLLLLLL LLLLLLL..L LLLL.....L LLLLLLLLLL LLLLL..LL.
Relm 77666654 3434445 5544322100 2102 35567778 7776555337 7888300125 5566776766 6655544664
mel SUUR MPPRKRTR-- ---AATVHLT PKTRRLNVRI LRVS------ LDTLSTPP-- PSRTTTAIVT PKTEPTARRK NLKKRTVSPV DVGRPATRGM
gri SUUR IPTNQLTKPA PAPAPREQPT PLPVKSPVKA KPMPEPNAQN VKQLSPEPRT PNHQKTLPIT PKTEPRSKR- -AQEKSASAV LTPR-VTRGM
Relg 7743256878 7866454766 5554467787 8788765465 4456878777 7644436777 776753100 035777641 0220 02120
SUBg LL...LLLLL LLLL.L.LLL LLL..LLLLL LLLLLLL.LL ..LLLLLLLL LL....LLLL LLLLL.... ..LLLLL.. .....
545 555 565 575 585 595 605 615 625
SUBm .....LL.L. ....LLLL ..LLL.LLLL LLLLL . ..LLLLL LLLLLLL LLLLLLLL.L LL..L
Relm 4233455454 4421100000 2323135555 4455545577 67765 4 4467777 8775556 7755556645 65446
mel SUUR QRLTRSAETK INSKYLKHRS LDDVKRSFPR RVKLEGNQTP RSSKQ----I VKQEPKS--- ---KVGQEKK QKTVDVPAQG TAKRK-----
gri SUUR QRMTRSAESR HNSKYMTP-- -HLVLGSTKR KRNIDSNQTP KTSKRRREPI VIQEEAAPTA KLESTAEPKI KKPRGRPPKK TKLSNGSTSL
Relg 0012243211 24532210 000257665 3234446777 6663235542 0033557511 0114667754 5777777544 6677553212
SUBg ....L..... ....LLLLL ......LLLL LLL...LL.. ....LLLL.. ....LLLLL. LLLLLLLL.. LLLLLL....


635 645 655 665 675 685 695 705 715
SUBm
Relm
mel SUUR
gri SUUR
Relg
SUBg
725 735 745 755 765 775 785 795 805
SUBm
Relm
mel SUUR
gri SUUR
Relg
SUBg
815 825 835 845 855 865 875 885 895
SUBm
Relm
mel SUUR
gri SUUR
Relg
SUBg
905 915 925 935 945 955 965 975 985
SUBm
Relm
mel SUUR

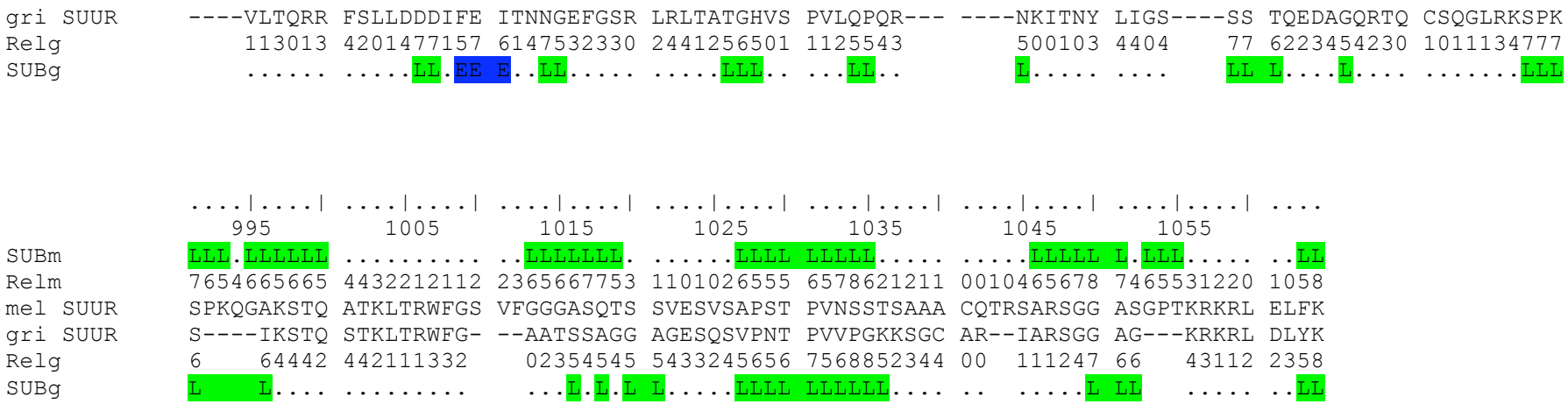


FIGURE S4.—PROF predicted secondary structure. SUBm and SUBg - Subset of the PROFsec prediction, for all residues in *D. melanogaster* and *D. grimshawi* SUUR respectively with an expected average accuracy greater than 82%. mel SUUR – *D. melanogaster* SUUR, gri SUUR – *D. grimshawi* SUUR. Relm and Relg - reliability index for PROFsec prediction (0=low to 9=high) for *D. melanogaster* and *D. grimshawi* respectively. H - helix, E - extended (sheet), L - loop, "." means that no prediction is made for this amino acid, as the is: Rel < 5 (http://www.predictprotein.org/).

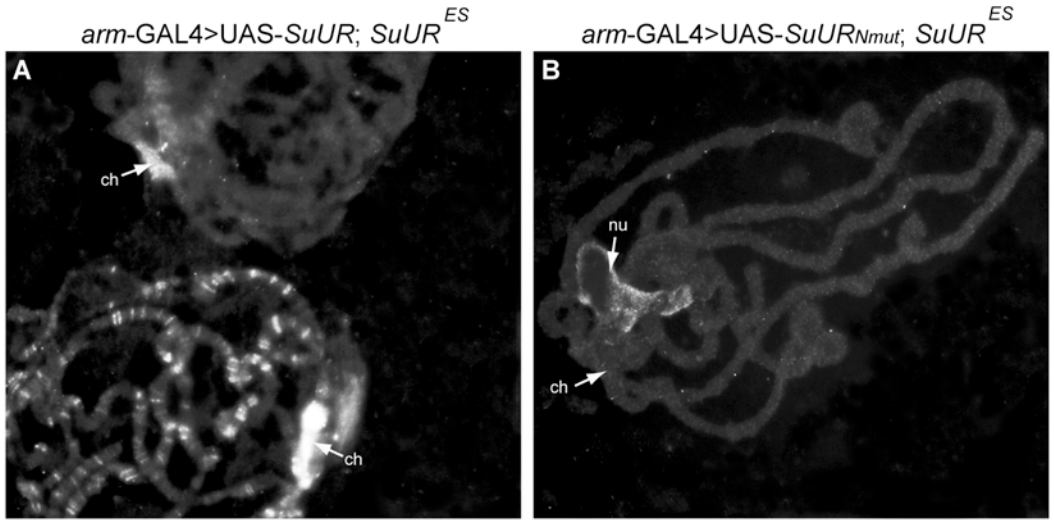


FIGURE S5.—Mutation in the N-terminus of SUUR abrogates the protein ability to bind the chromosomes. (A) Due to the variegated transgene expression in *arm*-GAL4>UAS-*SuUR; SuUR*^{ES} larvae, the immunostaining pattern varies from nucleus to nucleus on the same slide, ranging from signals confined only to the chromocenter (ch), to nuclei with signals in chromocenter and along the chromosome arms. (B) In polytene chromosomes from *arm*-GAL4>UAS-*SuUR*$_{Nmut}$; *SuUR*^{ES} larvae, the signal could only be detected in the nucleolus.