# Intron Loss Is Associated with Gain of Function in the Evolution of the Gloverin Family of Antibacterial Genes in *Bombyx mori*[*][S]

**Nirotpal Mrinal**[1] **and Javaregowda Nagaraju**[2]

*From the Laboratory of Molecular Genetics, Centre for DNA Fingerprinting and Diagnostics, Hyderabad-500076, India*

**Gene duplication is a characteristic feature of eukaryotic genomes. Here we investigated the role of gene duplication in the evolution of the *gloverin* family of antibacterial genes (*Bmglv1*, *Bmglv2*, *Bmglv3*, and *Bmglv4*) in *Bombyx mori*. We observed the following two significant changes during the first duplication event: (i) loss of intronV, located in the 3′-untranslated region (UTR) of the ancestral gene *Bmglv1*, and (ii) 12-bp deletion in exon3. We show that loss of intronV during *Bmglv1* to *Bmglv2* duplication was associated with embryonic expression of *Bmglv2*. Gel mobility shift, chromatin immunoprecipitation, and immunodepletion assays identified chorion factor 2, a zinc finger protein, as the repressor molecule that bound to a 10-bp regulatory motif in intronV of *Bmglv1* and repressed its transcription. *gloverin* paralogs that lacked intronV were independent of chorion factor 2 regulation and expressed in embryo. These results suggest that change in cis-regulation because of intron loss resulted in embryonic expression of *Bmglv2-4*, a gain of function over *Bmglv1*. Studies on the significance of intron loss have focused on introns present within the coding sequences for their potential effect on the open reading frame, whereas introns present in the UTRs of the genes were not given due attention. This study emphasizes the regulatory function of the 3′-UTR intron. In addition, we also studied the genomic loss and show that "in-frame" deletion of 12 nucleotides led to loss of amino acids IHDF resulting in the generation of a prepro-processing site in BmGlv2. As a result, the N-terminal pro-part of BmGlv2, but not of BmGlv1, gets processed in an infection-dependent manner suggesting that prepro-processing is an evolved feature in Gloverins.**

Insects depend on humoral (antimicrobial peptide (AMP)[3] synthesis) and cellular responses to effectively kill the invading microbes (bacteria and fungi) as they lack adaptive immunity capable of producing antibodies (1). Typically AMPs have low molecular weight, are water-soluble, and possess broad spectrum antibacterial activity (2). Cecropins, attacins, drosocins, and duptericins have antibacterial activity against Gram-negative bacteria, whereas defensins and metchnikowin kill Gram-positive bacteria (2, 3). Specificity in the immune response against a particular class of microbes is because of specific interaction between pathogen-associated molecular patterns present on the microbes and pathogen recognition receptors of the insect (1, 4). Because of the polyvalent recognition of pathogen-associated molecular patterns, any microbial infection leads to production of the same battery of AMPs through two evolutionarily conserved pathways *toll* or *imd* (1, 4).

The Toll and IMD pathways mediate regulation of innate immune response in insects, in response to fungal and bacterial infections, respectively (1, 4). Dorsal was the first transcription factor to be identified as regulator of innate immunity in *Drosophila*. Ip (5) identified Dif (Dorsal related immunity factor), a second *Drosophila* Rel protein, in the larval fat body. *Drosophila* Rel proteins are found in the cytoplasm but are translocated to the nucleus upon activation of the immune pathway in a signal-dependent manner. An extracellular signal, encoded by the *spatzle* gene, binds to Toll, a membrane-bound receptor. Spatzle binding to Toll causes activation of a signal cascade involving Tube and Pelle, a serine/threonine kinase. This leads to phosphorylation and subsequent degradation of Cactus, an IkB homolog (1, 4). In the absence of Cactus, Dorsal and DIF are free to translocate to the nucleus, where they act as transcription factors and initiate expression of AMP genes. A third Rel factor, Relish, was isolated in a molecular screen for genes whose expression is altered after infection (6). Relish is a compound protein containing both Rel (activating) and IkB (inhibitory) domains (7). Upon infection, the Relish inhibitory domain is proteolytically cleaved by dredd, a *Drosophila* caspase, to release the active Rel protein (8). Genetic epistasis studies and molecular analysis of gene function show that *imd*, *Relish* components of a signaling pathway are distinct from the Toll pathway and are essential for combating Gram-negative bacterial infection (8–15).

Although the mechanism of AMP activation has remained conserved across species, the repertoire of AMPs present in

---

[3] The abbreviations used are: AMP, antimicrobial protein/peptide; CF2, chorion factor 2; ChIP, chromatin immunoprecipitation; RT, reverse transcription; AEL, after egg laying; ORF, open reading frame; RNAi, RNA interference; UTR, untranslated region; EMSA, electrophoretic mobility shift assay; DTT, dithiothreitol; GST, glutathione S-transferase; PBS, phosphate-buffered saline.

different organisms varies (1–4) *e.g. hemolin*, an AMP with an immunoglobulin fold (16), has been reported only from lepidopteran insects. This prompted us to look for Lepidoptera-specific AMPs in *Bombyx mori* with a broader aim to understand the evolution of immune system in insects. *B. mori* is the only lepidopteran insect for which whole genome sequence (17, 18) and EST data base (19) are available. An analysis of *B. mori* genome and EST sequence revealed the presence of AMPs like cecropins, moricins, attacins, lebocin, enbocin, hemolin, and gloverins. Our analysis, based on sequence information available currently, suggests that like hemolin gloverins are also restricted to lepidopteran insects.

Gloverins are glycine-rich (16–20%) antibacterial proteins and have been reported from lepidopteran insects *Hyalophora gloveri, Helicoverpa armigera*, and *Trichoplusia ni* (20–22). They are basic, heat-stable proteins with random coil structure in solution and take α-helical structure upon interaction with lipopolysaccharide (20). We found that silkworm has four *gloverin* genes (*Bmglv1, Bmglv2, Bmglv3,* and *Bmglv4*). The derived genes *Bmglv2–4* have evolved as a result of three gene duplication events. A significant difference was observed in the embryonic expression profile of these genes, whereas, derived genes *Bmglv2–4* express in all embryonic stages but not the ancestral gene *Bmglv1*. This suggested that embryonic expression of derived genes was gained during duplication of *Bmglv1*. Embryonic regulation of AMP genes is not well studied; hence we set out to study the genetic changes that led to evolution of embryonic expression in the Lepidoptera-specific AMP gene family, *gloverin*. Molecular analysis suggested that embryonic expression of *Bmglv1* was regulated by an intron present in the 3′ UTR of the gene. Further characterization of the regulatory role of intronV led to identification of CF2, a zinc finger transcription factor that regulates oogenesis as suppressor of *Bmglv1* in the embryo. We also tested the significance of embryonic expression of daughter gene *Bmglv2,* by RNAi which led to reduced hatching of embryos. This indicated that *Bmglv2* has a role in embryonic development. We show that this gain of function in embryonic development was linked to loss of intronV.

Introns, noncoding sequences interrupting protein-coding genes, are the hallmark of eukaryotic gene organization (23). However, the role of intron in AMP gene regulation was previously not known. This is the first study demonstrating the regulatory role of an intron in an AMP gene and association between intron loss and gain of function. This study also emphasizes evolution of functional divergence in AMP gene paralogs.

## EXPERIMENTAL PROCEDURES

*Animals*—*B. mori* strains Pure Mysore and Nistari were collected from the sericulture station at Hindupur, Andhra Pradesh, India. *Escherichia coli* (K12 strain), cultured in antibiotic free LB media, was used for infecting silkworm larvae.

*Gel Shift Assay (EMSA)*—Embryonic nuclear extracts were prepared by homogenizing embryos (40–72 h AEL) in extraction buffer (20 mM Hepes, pH 7.9, 5 mM MgCl$_2$, 0.1 mM EGTA, 12.5% sucrose, 25% glycerol, 0.5 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride and protease inhibitor mixture) using a Dounce homogenizer, followed by centrifugation at 3300 × g

for 20 min at 4 °C. The precipitated nuclei were suspended in 1 ml of the extraction buffer. Embryonic extracts were also prepared from $w^{1118}$ and Df(2L)$\gamma^{27}$ flies (these flies lack the *cf2* locus). For EMSA, 100 ng of double-stranded *cf2* oligonucleotide (AGTAA<u>ATATATATAT</u>TTAAA) was labeled with 3 μl of [γ-$^{32}$P]ATP (4 × $10^5$cpm) and 1 μl of polynucleotide kinase (10 units/μl) in 1 μl of PNK buffer (New England Biolabs) for 1 h at 37 °C. The labeled DNA was purified on a G-50 column. The binding reaction was performed for 30 min at room temperature by mixing 1 ng of purified $^{32}$P-labeled double strand synthetic oligonucleotide probe (4000 cpm/μl), 10 μg of nuclear extracts, 300 ng of poly(dI-dC), and 5 mM Zn$^{2+}$ in the presence of a protease inhibitor mixture (Sigma). Cold competition was performed by preincubating the extracts with a 40-fold excess of unlabeled oligonucleotide at room temperature for 15 min. Anti-CF2 monoclonal antibody was added to the binding reaction for 30 min to perform supershift experiments. The binding reaction was analyzed by electrophoresis on native 6% polyacrylamide gels.

*In Vitro Transcription and Translation*—*In vitro* transcription was done essentially as mentioned in Suzuki *et al.* (24, 25). For *in vitro* translation, different expression constructs were incubated with embryonic extracts prepared from *Drosophila* or silkworm embryos. Embryos were collected, dechorionated by bleaching, washed 3–5 times with 0.1% Triton X-100, and transferred to hypotonic buffer at 4 °C. Embryos were further washed in 3–5 volumes of cold hypotonic buffer (10 mM Hepes-KOH, pH 7.4, 15 mM KCl, 1.5 mM Mg(OAc)$_2$, 2 mM DTT) on ice. Next, embryos were Dounce-homogenized, and homogenate was centrifuged for 15 min at 15,000 rpm at 4 °C. The supernatant was transferred to a fresh microcentrifuge tube and centrifuged again under the same conditions to remove any residual debris. Extracts were centrifuged through Sephadex G-25 Superfine columns prepared in buffer A (30 mM Hepes-KOH, pH 7.4, 100 mM KOAc, 2 mM Mg(OAc)$_2$, 2 mM DTT, protease inhibitor mixture). The column was transferred to a fresh collection tube to which a volume of cold buffer A equal to the extract volume was added, and the column was centrifuged for 3 min at 200 × g at 4 °C. 100 μl of eluate was applied to a P6 (Bio-Rad) desalting column to remove salt and other contaminants, and the desalted eluate was used for *in vitro* translation as mentioned in Gebauer *et al.* (26). Entire protocol was first standardized with CantonS *Drosophila* embryonic extracts (data not shown), and then the standardized protocol was used for experiments with *B. mori* embryonic extracts. *B. mori* embryos were collected 40 and 56 h AEL and pooled.

*Immunodepletion of Embryonic CF2*—Immunodepletion was done largely following the protocol as described previously (27) with the following modifications. Immunodepletion of CF2 was performed in embryonic extracts in a final volume of 500 μl of buffer A. CF2-depleted supernatant was collected and used immediately for *in vitro* translation.

*Western Blotting*—*In vitro* translation reaction product was separated on a 10% minigel SDS-polyacrylamide gel (Bio-Rad). Protein samples separated by SDS-PAGE were electrophoretically transferred using a Trans-blot cell (Bio-Rad), at 200 mA overnight at 4 °C to Hybond-P polyvinylidene difluoride transfer membrane (Amersham Biosciences). The blots were stained

for total protein by Ponceau S (Sigma) and blocked in 10% nonfat dry milk in 0.5% Tween 20, 0.05% SDS in PBS (Blocking buffer). The blots were incubated for 6 h at room temperature in primary antibodies and then washed four times (10 min each) in Tween 20 + PBS followed by a 2-h incubation at room temperature with the secondary antibody (Sigma). The blots were then washed three times for 30 min in Tween 20 + PBS and rinsed once in PBS. The protein bands were detected using horseradish peroxidase-enhanced chemiluminescence (Amersham Biosciences). Anti-CF2 (mouse) and anti-GST (rabbit) antibodies were used in 1:10,000 and 1:1000 dilutions, respectively, for probing.

*RT-PCR, in Vitro Transcription, and RNAi*—Total RNA was isolated using TRIzol reagent (Invitrogen), dissolved in 50 $\mu$l of RNase-free water, and quantified in a spectrophotometer. 1 $\mu$g of RNA was used in 20 $\mu$l of RT reaction by using SupercriptII (Invitrogen) and oligo(dT) or gene-specific primers. For synthesizing dsRNA *Bmglv1* and *Bmglv2*, PCR products were cloned between BamHI and KpnI sites of pB-SK+ vector, and *in vitro* transcription was done with linearized template using T3 and T7 RNA polymerase, separately. Single strand RNA was purified after DNase treatment to remove template plasmid and quantified in a spectrophotometer. An equal amount of sense and antisense RNA was used for annealing in annealing buffer by heating at 95 °C for 5 min followed by slow cooling to room temperature. Respective dsRNA was injected on the 1st day of 5th instar larvae. Male and female moths eclosed from dsRNA injected larvae were allowed to mate (15 experiments each for *Bmglv1-dsRNA* and *Bmglv2-dsRNA*), and egg hatching was calculated. An equivalent volume of nontarget baculoviral *ie1-dsRNA* was injected in control larvae. Different concentrations of dsRNA were used for standardization of RNAi, and the data presented here are from larvae injected with 10 $\mu$g of dsRNA. A list of primers used for cloning of dsRNA target site has been provided in the supplemental material.

*Chromatin Immunoprecipitation*—The protocol followed for ChIP was essentially as mentioned on line with the following modifications. Instead of *Staphylococcus aureus* cells, protein A beads and silkworm embryo were used. Fluorescence real time PCR was done with double-stranded DNA dye SYBR green (PerkinElmer Life Sciences) on an ABI PRISM 7700 system (PerkinElmer Life Sciences) to quantify the enrichment of the *cf2*-binding element upon ChIP. PCR specificity was confirmed by the molecular size of the PCR product and $\Delta Ct$ analysis. Reactions were done in triplicate and compared with input DNA also (in triplicate) and nontemplate control (in duplicate).

*Plasmid Constructions*—Full-length *Bmglv1* cDNA was cloned between BamHI and KpnI sites in pFB-His-GST expression vector to generate expression plasmid pFB-His-GST-Glv1. This plasmid was later used for transforming DH10Bac bacterial cells to obtain recombinant pFB-His-GST-Glv1 bacmid that was subsequently transfected into Sf9 cells to get the recombinant virus expressing His-GST-Glv1 fusion protein. Maximum expression of the protein was observed between 70 and 96 h postinfection. His-GST-Gloverin1 fusion protein was purified using a glutathione column. To generate expression vectors with different promoters to be used in the *in vitro* translation reaction, *Autographa californica* nuclear polyhedrosis

virus (AcNPV) *polyhedrin* promoter of pFB-His-GST basic vector was replaced with *B. mori* cytoplasmic actin (*BmA3-Actin*) promoter (for control plasmid) or *Bmglv1* or *Bmglv2* promoters. For cloning the intronV of *Bmglv1*, genomic PCR with exon5- and exon6-specific primers (see supplemental material for primer sequences), which amplified exon5-intronV-exon6, was done, and subsequently this PCR product was cloned downstream to *gst* only or *gst* + *glv1* fusion ORF (*Bmglv1::GST-glv1-intronV* plasmid). Control plasmid lacking intronV was also generated by RT-PCR using same set of primers and cloned downstream to *gst* or other ORFs (*Bmglv1::GST-glv1-ΔintronV* plasmid). To study the role of CF2-binding motif of intronV, a deletion construct lacking *cf2* motif was generated by MseI digestion that deleted 53 nucleotides just before intronV-exon6 junction. MseI-digested product was blunt-ended, ligated, and then cloned downstream to *gst* to generate *Bmglv1::GST-glv1-IntronVΔcf2* plasmid. For immunodepletion, a plasmid expressing GST-BmGlv1 fusion protein (*Bmglv1::Gst-Bmglv1*(*ORF*)-*InVcf2*) was generated by cloning *Bmglv1* ORF between *gst* and the *Ex5-InV-Ex6* cassette in the plasmid *Bmglv1::gstEx5-InV-Ex6*.

*RNase Protection Assay*—Exon5-intronV-Exon6 and Exon5-intronVΔcf2-Exon6 fragments were PCR-amplified from the respective plasmid (described in preceding paragraph) constructs using 3′-UTR forward and reverse primers (see supplemental Methods) and cloned in pCR2.1 vector (Invitrogen). The radiolabeled antisense strand was synthesized using T7 RNA polymerase (New England Biolabs) in an *in vitro* transcription reaction. 5 $\mu$g of RNA sample was hybridized with *in vitro* transcribed antisense RNA probe ($3 \times 10^5$ cpm) at 45 °C overnight. RNA samples were dissolved in 25 $\mu$l of 75% formamide, 0.5 M NaCl, 10 mM Tris-HCl, pH 7.5. RNase A (100 $\mu$g/ml) was added to the reaction mix along with 200 $\mu$l of 200 mM NaCl, 5 mM EDTA and incubation was done for 1 h at 37 °C. After the RNase incubation, proteinase K (250 $\mu$g/ml) digestion was done followed by phenol/chloroform extraction and ethanol precipitation and finally subjected to denaturing (8 M urea) PAGE.

*Antibacterial and Prepro-processing Assay*—Antibacterial activity of BmGlv1 and BmGlv2 was determined by incubating $\sim 10^5$–$10^6$ cells/ml of *E. coli* with 5 mM of either BmGlv1 or BmGlv2 in 1× PBS, pH 7.1; and 1, 2, 4, and 6 h postincubation, the optical density of the respective cultures was taken, and a graph was plotted. Each experiment was repeated a minimum of three times. Antibacterial activity of BmGlv1 and BmGlv2 was also quantified by counting the number of surviving bacteria after 6 h of incubation in the presence of BmGlv1 or BmGlv2 (colony-forming units/ml) on antibiotic-free LB agar plates. In a control experiment, bacteria were incubated with an equivalent amount of PBS.

Bacterial challenge activates different proteases that in turn cleave the N-terminal prepro part and release mature AMPs. Hence, fat body extracts from *E. coli*-challenged 5th instar larvae were prepared 3, 6, 9, and 12 hours post infection and pooled. Later, purified GST-Glv1 and GST-Glv2 proteins were incubated with the pooled fat body extract for 2 h, and then the complete reaction mixture was separated on SDS-polyacrylamide gel followed by Western blotting with anti-GST antibody.
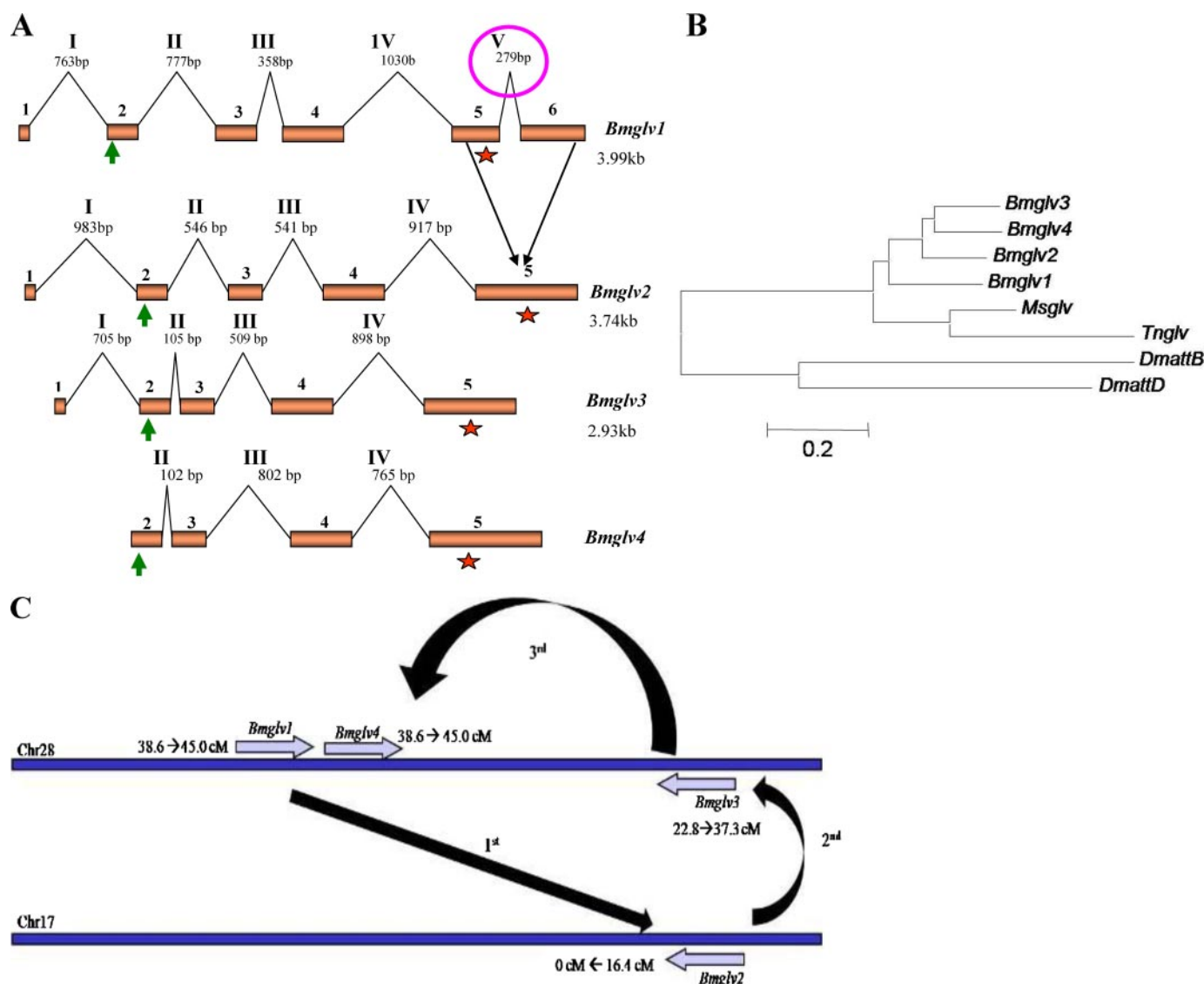
FIGURE 1. **Location, organization, and evolution of *Bombyx* gloverins.** *A, Bmglv1* has six exons whereas other paralogs have five. During the first duplication event loss of intronV led to fusion of exons 5 and 6 resulting in smaller *Bmglv2* gene. Intron loss does not change ORF. Start codon (*green arrow*) lies in the 2nd exon, whereas stop codon (*red star*) lies in the 5th exon in all paralogs. *B,* phylogenetic analysis suggests *Bmglv1* as the ancestral *gloverin* of *B. mori. AttacinB* and *AttacinD* of *Drosophila* were used as outgroup. *Bm, B. mori; Dm, D. melanogaster; Ms, Manduca sexta; Tn, Trichoplusia ni. C,* chromosomal location of four *Bombyx gloverins* as determined by physical mapping. *Arrows* (in *black*) indicate the sequence of duplication, and *1–3* 3 indicate first, second, and third duplication events, respectively.

Release of GST-specific band was an indication of processing of the AMP.

*Phylogenetic Analysis of Bombyx Gloverins*—Best fit model was tested using Model Test as implemented in HyPhy. GTR+G model was selected by both Hierarchical Model Testing and Akilike Information Criteria (AKI score = 13034.9) and $\alpha = 2.00475$. Using these parameters, a neighbor joining tree was constructed.

## RESULTS

*Identification of Gloverin Family of AMP Genes in B. mori*—We have identified four *gloverin* genes (named as *Bmglv1, Bmglv2, Bmglv3,* and *Bmglv4*) in the silkworm *B. mori.* The gene structures of silkworm *gloverins* revealed *Bmglv1* (3.99 kb) as the largest *gloverin* (Fig. 1*A*). Gene structure of *Bmglv4* could not be predicted reliably because of the presence of repetitive

elements in the scaffold corresponding to exon1. Exons 2–4, which constitute ORFs, are highly conserved in all four *gloverins*. Phylogenetic analysis based on ORFs of the four *gloverins* suggested *Bmglv1* as the ancestral gene (Fig. 1*B*, supplemental Fig. 1, and supplemental Table 1). Clearly, *Bmglv1* is orthologous to other *gloverins* as it shares a common ancestor with *Manduca sexta* and *Trichopusia ni gloverins* (Fig. 1*B*). The presence of intronV is the one major difference between *Bmglv1* and the other three silkworm *gloverins* (Fig. 1*B*). However, intronV of *Bmglv1* was lost during first duplication leading to fusion of exon5 and exon6 as a result of which the derived gene *Bmglv2* has only five exons. In subsequent duplication events, lengths of different exons largely remained conserved but not of introns (Fig. 1*A*). Although no clear pattern is seen in intron length dynamics, overall the gene size has become smaller with each round of duplication, precisely because of
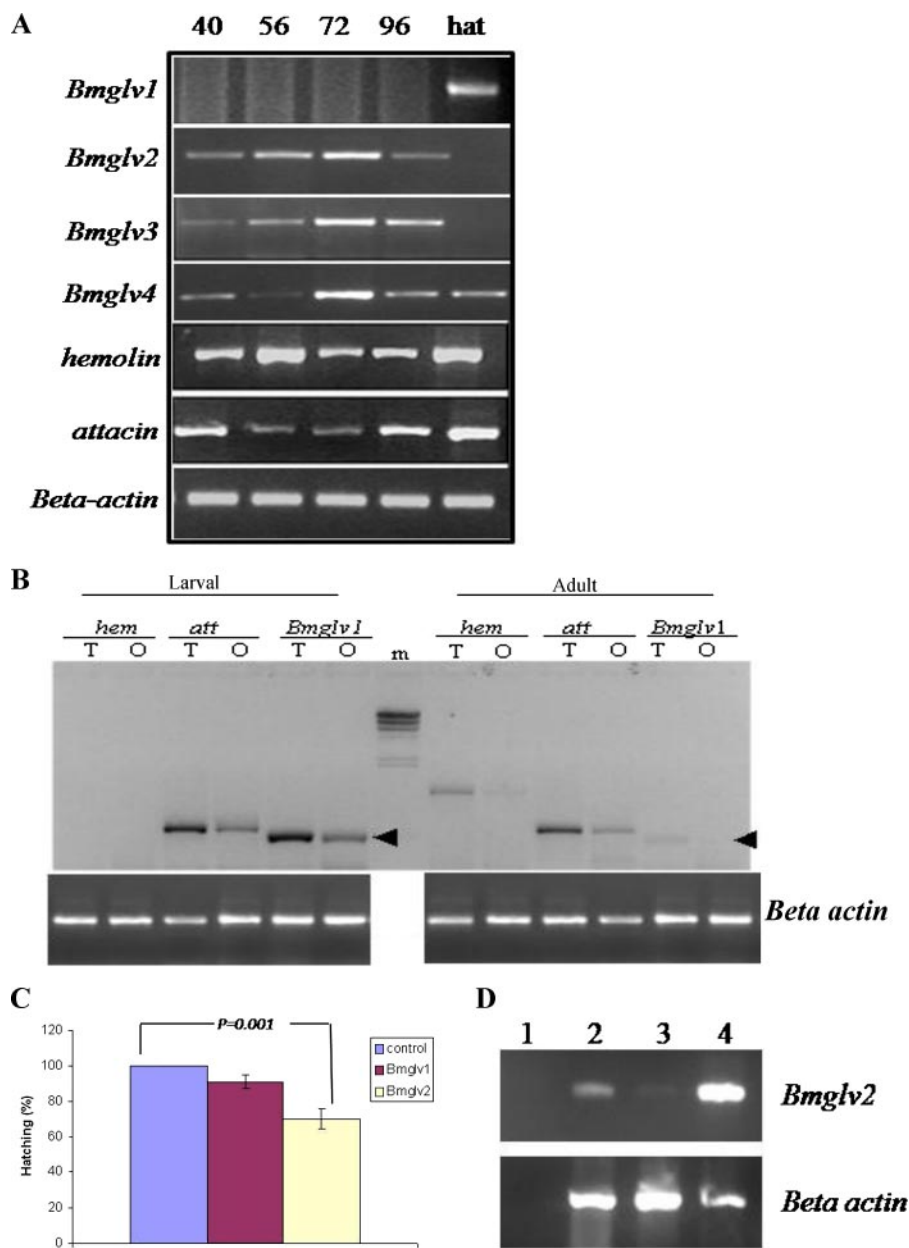
FIGURE 2. **Expression dynamics of *gloverin* and other AMP genes of *B. mori*.** *A,* all AMPs express in embryonic stages except for *Bmglv1*. (*Numbers* at the *top* indicate hours AEL; *hat* indicates hatched larvae.) *B, Bmglv1* (*arrowhead*), which expresses in larval gonads, is down-regulated in adult gonads. Contrary to *Bmglv1* down-regulation, *hemolin* is up-regulated in adult gonads, and its expression was not seen in larval gonads. *Attacin* expression did not change appreciably. (*m,* λHindIII marker; *T,* testes; *O,* ovary.) *C,* knockdown of *Bmglv2* by RNAi leads to ~30% reduction in embryo hatching ($p = 0.001$), whereas knockdown of *Bmglv1* has no effect on hatching. Data shown here are the average of five independent experiments. *D,* RT-PCR was done with embryonic RNA (40 h AEL) to show inhibition of *Bmglv2* upon RNAi (*lane 3*). *Lane 4* shows enhanced expression of *Bmglv2* (with respect to *lane 2*) where larvae were injected with *ie-1 dsRNA* followed by *E. coli* infection thus confirming that *ie-1 dsRNA* injection does not target *Bmglv2* transcript. (*Lane 1,* without RT control; *lane 2,* nontarget *ie-1* dsRNA control; *lane 3, Bmglv2*dsRNA; *lane 4, ie-1* dsRNA + Bacterial injection.). *Beta actin,* loading control. *Numbers above* the figure represent respective lanes.

exon as well as part of 3′-UTR. *Bmglv3* has the longest 3′-UTR (318 bp) and shows five unique insertions not shared by other *gloverins* (supplemental Fig. 1 and supplemental Table 1). Apart from this, 3′-UTRs of four *gloverins* also share low homology at the sequence level. Lack of conservation in A/T-rich 3′-UTRs of *gloverins* possibly indicates lack of conserved regulatory function.

We also confirmed the physical location of the four *gloverins*. Although *Bmglv1*, *Bmglv3,* and *Bmglv4* were physically mapped to chromosome 28, *Bmglv2* was mapped to chromosome 17 (Fig. 1*C*).[4] *Bmglv1* and *Bmglv4* are present as tandemly duplicated genes on positive strand at 38.6 → 45.0 centimorgans, and *Bmglv2* and *Bmglv3* are present on the complementary strand at 0 centimorgans ←16.4 centimorgans on chromosome 17 and at 22.8 → 37.3 centimorgans on chromosome 28, respectively (Fig. 1*C*).

*Bmglv1 Is Not Expressed in Embryo Unlike Other AMP Genes*—Induction of expression of AMP genes in larval and adult tissues, mainly fat body and mid-gut, upon immune challenge is an established fact (1). However, we observed basal expression of *Bmglv2–4* in embryonic stages (Fig. 2*A*). Other than *Bmglv2-4*, *attacin* and *hemolin* genes are also expressed in all embryonic stages. Significantly *Bmglv1* does not express in embryos (Fig. 2*A*). A more interesting developmental regulation of AMP expression was observed in gonads as we found that *Bmglv1* is expressed in larval but not in adult gonads, whereas *hemolin* and other *gloverins* expressed in adult but not in larval gonads (Fig. 2*B*). Clearly, down-regulation of *Bmglv1* starts in adult gonads, and complete repression is achieved in embryonic stages, and upon hatching its expression is restored. It also suggested that embryonic expression may be a general feature of AMPs. The study raised two questions as follows. (i) What is the significance of embryonic expression of AMPs in general? (ii) How is suppression of *Bmglv1* achieved in embryonic stages?

erosion in intronic regions of the genes (*Bmglv1*(3.9 kb) → *Bmglv3* (2.9 kb)). Among the conserved introns (introns I–IV), a significant deletion in intron length is seen for intronII, during *Bmglv2* to *Bmglv3* duplication; however, significance of this deletion remains to be elucidated (Fig. 1*B*).

Sizes of different *gloverins* may also differ because of size variations in their UTRs. Fifth exon contributed to 3′-UTR in all the *gloverins* except *Bmglv1*, which has an additional 6th

---
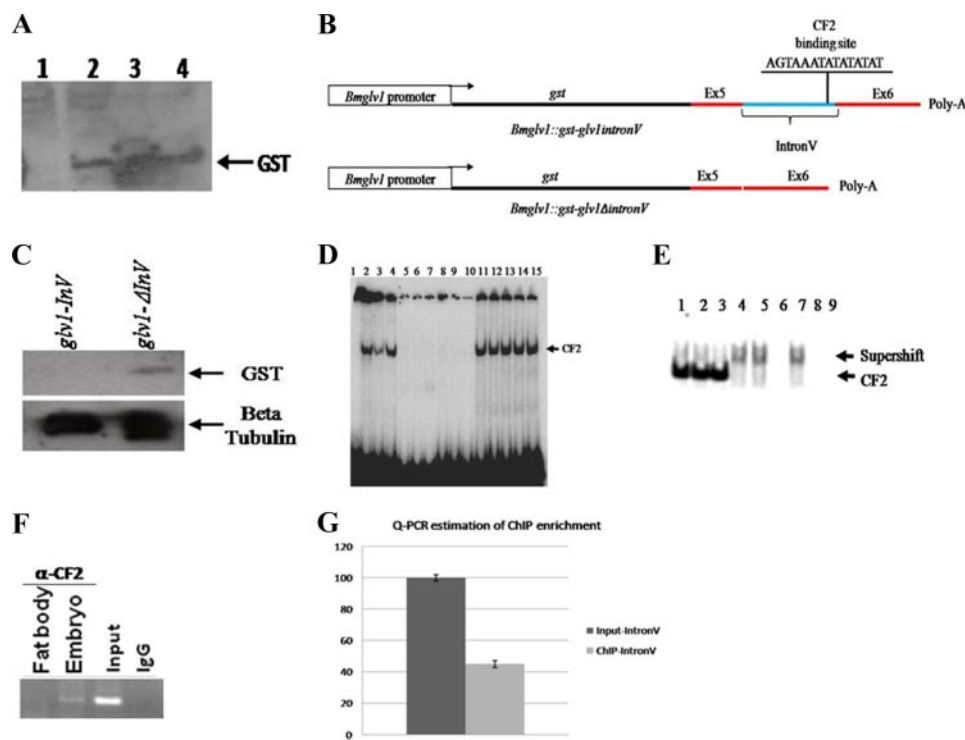
[4] K. Mita, personal communication.

FIGURE 3. **Characterization of repressor element in the 5th intron of *Bmglv1*.** *A, in vitro* synthesis of GST from plasmid driven by *Bmglv1* (*lane 2*) and *Bmglv2* promoters (*lane 3*) suggests that the *Bmglv1* promoter is functional in the embryo. Constitutively active A3-actin promoter (*lane 4*) was used as positive control, whereas AcNPV polyhedrin (*Polh*) promoter (*lane 1*), which requires viral factors for expression, was used as negative control. *B,* schematic diagram shows organization of expression plasmids used to study the interaction between promoter and intronV. *Upper panel* shows plasmid with intronV, and the *lower panel* shows plasmid without intronV. In both constructs Ex5-InV-Ex6 or Ex5-Ex6 cassette was cloned downstream to the *gst* ORF. *C, in vitro* translation product is synthesized with *Bmglv1::gst-Ex5-Ex6* template lacking intronV but not with *Bmglv1::gst-ex5-inV-ex6* template thus suggesting inhibitory role of intronV. β-Tubulin was used as control (*lower panel*). *D,* nuclear extracts from different developmental stages of silkworm were used for retardation with CF2-specific probe to check tissue-specific expression of CF2 protein. Expression profile of CF2 and that of *Bmglv1* are inversely correlated (compare with Fig. 2, *A* and *B*). *Lane 1,* cold competition; *lane 2,* nonhomologous mutant competition; *lane 3,* adult testes; *lane 4,* adult ovary; *lane 5,* homologous mutant CF2 probe; *lane 6,* mid gut; *lane 7,* fat body; *lane 8,* larval ovary; *lane 9,* larval testes; *lane 10,* free probe; *lane 11,* embryo 40-h AEL; *lane 12,* embryo 60-h AEL; *lane 13,* embryo 96-h AEL; *lane 14, drosophila* ovary; *lane 15, Drosophila* embryo. *E,* supershift with *Drosophila* CF2 antibody was done to check the specificity of the complex retarded with CF2 oligonucleotide. *Lane 1, Drosophila* embryo; *lane 2,* silkworm embryo 40-h AEL; *lane 3,* nonhomologous competition; *lane 4,* CF2-antibody + silkworm embryo 40-h AEL; *lane 5,* CF2-antibody + *Drosophila* embryo extract; *lane 6,* CF2-antibody + cold competition; *lane 7,* CF2-antibody + nonhomologous competition; *lane 8,* embryonic extract from *Df(2L)γ²⁷* stock that lacks CF2 locus; *lane 9,* free probe. *F,* ChIP with CF2 monoclonal antibody suggests *in vivo* interaction of CF2 with *Bmglv1-intronV*. Larval fat body where CF2 is not expressed was used as negative control. *G,* real time PCR was done to quantify the enrichment of *Bmglv1-intronV* in ChIP performed with embryonic extract and shows only 32% enrichment with respect to input. Weak enrichment of intronV upon ChIP with respect to control is probably because *Drosophila* anti-CF2 monoclonal antibody has been used to precipitate silkworm CF2 protein with which the antibody may not interact with the same efficiency. *Error bars* represent standard deviation of three independent experiments.

well. Knockdown effect was analyzed in the next generation embryo. Although RNAi of *Bmglv2* led to reduced hatching, knockdown of *Bmglv1,* which does not express in embryo, had no such effect (Fig. 2, *C* and *D*). These results suggest a role for *Bmglv2* in embryonic development that indicates a gain of function in *Bmglv2* with respect to *Bmglv1.* Thus our results point out that embryonic expression of AMPs (Fig. 2*A*) is an essential and developmentally regulated process.

*Bmglv1 Promoter Is Functional in Embryo*—Developmental regulation of AMP genes is not well studied. Because *Bmglv2 and Bmglv1* genes were quite distinct in their embryonic expression pattern, we examined the differences between the two genes to gain insight into their embryonic regulation (Fig. 2*A*). Promoters are the most important cis-elements that regulate spatio-temporal expression of downstream genes. One possibility was that during the duplication process, certain regulatory motifs in the promoter of *Bmglv2* might have been deleted/gained leading to its expression in embryo and was further investigated. Comparative analysis of *gloverin* promoters revealed binding sites of all essential transcription factors, which regulate AMP genes, *viz.* Rel and GATA (data not shown), and we did not find any striking difference between the two promoters. Hence, we tested whether the *Bmglv1* promoter was functional in embryonic stages or not. *Bmglv1::gst* plasmid,

*Significance of Embryonic Expression of Bmglv2*—Embryos are naturally protected against microbial infection because of the presence of the impregnable chorion layer that provides a strong physical barrier. Hence, embryos are not prone to infection nor has any natural infection been reported in insect embryos. If the threat of microbial infection is low/absent in embryos, then to what purpose does expression of AMPs in embryonic stages serve? We hypothesized that AMPs may have some role other than killing microbes. To test this hypothesis, we knocked down the embryonically expressing *gloverin* paralog *Bmglv2* and compared the effect of its knockdown with that of *Bmglv1* knockdown. *Bmglv1* and *Bmglv2* dsRNA were administered to 4th and 5th instar larvae and later injected into adult moths as

where *gst* is driven by the *Bmglv1* promoter, was incubated with embryonic extract for coupled *in vitro* transcription and translation experiments. Synthesis of GST from *Bmglv1::gst* plasmid was indicative of the fact that the *Bmglv1* promoter was capable of expressing in embryonic stages similar to that of *Bmglv2* (Fig. 3*A, lanes 2* and *3*). BmA3-Actin promoter (cytoplasmic actin promoter of *B. mori*) construct was used as reaction control (Fig. 3*A, lane 4*). Because both the promoters were functional during embryonic stages, we looked for other differences between the two genes that can account for different embryonic expression profile of the two genes. It is evident that the gene structure of *gloverin* paralogs has largely remained unchanged with the exception of intronV, which is present only in *Bmglv1* (Fig. 1*B*). As loss of intronV and gain of embryonic expression

took place during the 1st duplication, we investigated whether the two events were linked.

*IntronV of Bmglv1 Acts Like a Repressor*—Introns are non-coding parts of the gene, but they are essential segments of the genome as many of the introns are known to have regulatory functions such as enhancers and suppressors. To ascertain the embryo-specific regulatory role, if any, of intronV of *Bmglv1*, we constructed *gst* reporter plasmids under the control of the native *Bmglv1* promoter. IntronV is located downstream to the ORF in the 3′-UTR region of the *Bmglv1* gene; hence, to retain this natural organization, intronV along with flanking regions of exons 5 and 6 was cloned downstream to the reporter *gst* ORF (Fig. 3B). In the control plasmid intronV was not cloned. Both plasmids, driven by *Bmglv1* promoters, were incubated with *Bombyx* embryonic extracts for coupled *in vitro* transcription and translation of the reporter gene. GST synthesis took place in the reaction where control plasmid was used, and no GST was detected in the reaction performed with the plasmid that harbored intronV (Fig. 3C), thus suggesting that intronV had inhibitory action on GST synthesis. Next we set out to dissect the mechanism of embryonic suppression of *Bmglv1* gene by intronV.

*Identification of Repressor Element in IntronV of Bmglv1*—IntronV of *Bmglv1* is 279 bp long, and to characterize the motif regulating embryonic expression of upstream ORF, EMSA was done with different fragments of intronV generated by restriction digestion (data not shown). The fragment corresponding to the last 40 bp of the intron, which has a putative CF2 binding site, showed specific shift in EMSA. Although there are two additional CF2-binding motifs in the intron, only the one present near the intron5-exon6 boundary was found to be functional (supplemental Fig. 2). Later oligonucleotide (2× AGTA-AAATATATAT) corresponding to the "functional CF2-binding motif" was used as probe for gel shift. The CF2 complex could be retarded with nuclear extracts from adult ovary, testes, and embryonic extracts but not from tissues of larval origin (Fig. 3D). These results are consistent with the observation that *Bmglv1* expression is not seen in the tissues where CF2 is expressed (Fig. 2A and Fig. 3D). Supershift with *Drosophila* CF2 antibody confirmed that the retarded complex contained CF2 protein (Fig. 3E). No gel shift was seen when embryonic extract from $Df(2L)\gamma^{27}$ flies, which lack *cf2* locus, was used (Fig. 3E, lane 8). This further confirmed the specific interaction of CF2 to the intronic element. To test the interaction of CF2 to intronV under *in vivo* conditions, ChIP assay was performed (Fig. 3, F and G). The enrichment of intronV was seen with embryonic but not with fat-body extract (Fig. 3F). These results point out the presence of a CF2-mediated active regulatory element in the intronV of *Bmglv1* and also the apparent correlation between *Bmglv1* repression and expression of CF2.

*Binding of CF2 to IntronV of Bmglv1 Represses the Native Promoter*—The physical interaction between CF2 and intronV suggested that CF2 might be important for intron-mediated suppression of the *Bmglv1* gene. To establish that CF2 was required for intronV-mediated repression of *Bmglv1,* we performed *in vitro* translation of *Bmglv1::Gst-Bmglv1*(ORF)-InV and *Bmglv1::Gst-Bmglv1*(ORF)-ΔInV plasmids, with wholesome and CF2-depleted embryonic extracts (CF2 protein was
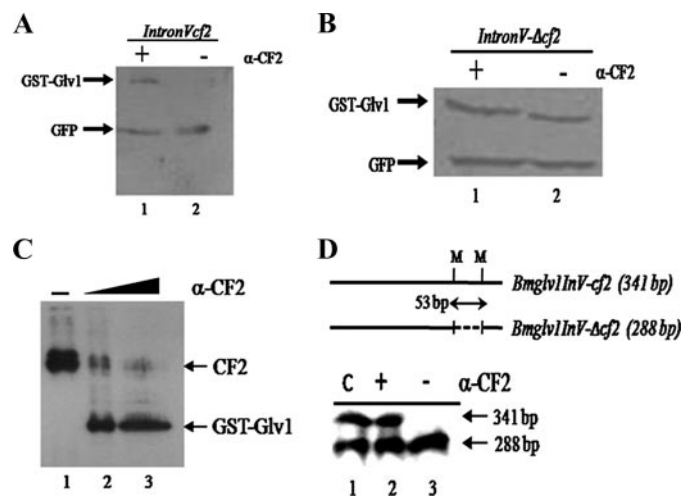


FIGURE 4. **CF2 represses *Bmglv1* expression.** *A,* immunodepletion of CF2 leads to synthesis of GST-Glv1 fusion protein in the presence of complete intronV (*lane 1*). In the mock-treated reaction, fusion protein is not synthesized when CF2-binding motif is intact (*lane 2*). *B,* deletion of the *cf2* binding motif makes *Bmglv1::gst-glv1InV-Δcf2* plasmid independent of CF2 regulation thus suggesting that repressor function of CF2 is mediated by its binding to the intronV. *BmA3-actin::GFP* control plasmid was also added in the same reaction as reaction control in all experiments (*A* and *B*). Clearly, green fluorescent protein synthesis is not affected by the presence or absence of CF2. *C,* immunodepletion of CF2 has inverse effect on expression of *Bmglv1* promoter and proves that CF2 inhibits *Bmglv1* promoter in a concentration-dependent manner. *Lane 1,* mock treatment where CF2 antibody was not added; *lane 2,* CF2 and PBS in 40:60 ratio; *lane 3,* CF2 and PBS in 60:40 ratio. *D,* RNase protection assay suggests that binding of CF2 to *Bmglv1-intronV-cf2* plasmid represses transcription in the presence of CF2 (*lane 3*) but not when CF2 is depleted out (*lane 2*) or in mock control (*C, lane 1*). *Bmglv1::gst-glv1InV-Δcf2* plasmid was added as reaction control, and its transcription is not affected by CF2 presence/absence (*lower band*). *Upper panel* shows the diagrammatic representation of the expected sizes of transcripts protected from RNase digestion (RNase protection). *M* represents the two MseI sites that are 53 bases apart and encompass the functional CF2-binding motif. Restriction digestion with MseI deletes these 53 bases and thus generating a 288-bp-long *glv1InV-Δcf2* transcript, whereas the *glv1InV-cf2* transcript is 343 bp long. *Numbers below* the figures indicate respective lanes.

immunodepleted using CF2 antibody). Western blot of *in vitro* translation product of *Bmglv1::Gst-Bmglv1*(ORF)-InV plasmid did not detect GST-BmGlv1 fusion protein in the reaction where complete embryonic extract was used (Fig. 4A, lane 2), but the fusion protein was detected in the reaction where CF2-depleted embryonic extract was used (Fig. 4A, lane 1). Constitutively expressing *piggyBac*-based *BmA3Actin::GFP* plasmid (28) was used as reaction control (Fig. 4A). Synthesis of GST-BmGlv1 fusion protein in the CF2-depleted extract suggested that native *Bmglv1* promoter was active in the embryo and also that the presence of CF2 led to *Bmglv1* suppression (Fig. 4A).

Furthermore, to prove that CF2-mediated repression of *Bmglv1* required binding of CF2 to the intronV, we repeated the immunodepletion experiment as mentioned in Fig. 4A, using *Bmglv1::Gst-Bmglv1*(ORF)-ΔInV plasmid that lacks the functional CF2-binding motif in the intronV (*intronV-Δcf2*). It is evident that *in vitro* translation was not affected by the presence or absence of CF2 in the embryonic extract if the CF2-binding motif was deleted (Fig. 4B, lanes 1 and 2). Taken together, data shown in Fig. 4, *A* and *B*, suggest that silencing of the *Bmglv1* promoter required physical interaction between CF2 and intronV of *Bmglv1*. This result was further confirmed when the CF2 protein was differentially depleted by using different

amounts of CF2 antibody. Synthesis of GST-BmGlv1 protein was found to be dependent on the extent of CF2 depletion in the embryonic extract (Fig. 4C). These results establish that CF2 recruitment to intronV is essential for *Bmglv1* suppression and also suggest that CF2- mediated intronic regulation is dominant over native *Bmglv1* promoter.

*Cf2 Blocks Transcription of Bmglv1*—We have shown that CF2 binding to intronV was required for suppression of GST-BmGlv1 fusion protein synthesis (Fig. 4, *A*–*C*). Lack of protein synthesis could be due to either suppression of transcription or of translation. If intronV acted like a *cis*-regulatory element, then transcription will be blocked (cis-regulation), and if the regulation was at the RNA level, then transcript will be formed but not the translation product. First, we tested whether CF2 binding to intronV suppressed transcription of *Bmglv1*. To elucidate the mechanism of suppression by CF2, if intronV acted like a cis-regulatory element, we performed *in vitro* transcription experiment with *Bmglv1::Gst-InVcf2* and *Bmglv1::Gst-InVΔcf2* plasmids. *In vitro* transcription was done with CF2-depleted and control extracts. RNA synthesis was checked by RNase protection assay (Fig. 4D). Because of deletion of the *cf2* motif, transcript formed from *Bmglv1::Gst-Bmglv1(ORF)-InVΔcf2* template is shorter by 53 nucleotides compared with transcript synthesized from *Bmglv1::Gst-Bmglv1(ORF)-InV* template (Fig. 4D, *upper panel*). When *Bmglv1::Gst-InVcf2* plasmid was used as template in the reaction where complete embryonic extract was used, no RNA could be protected indicating absence of transcript in this reaction (Fig. 4D, *lower panel, lane 3*). Protection of *gst-Bmglv1-inV*-specific transcript in *lane 2* indicates transcription of *Bmglv1::Gst-Bmglv1(ORF)-InV* template with CF2-depleted extract (Fig. 4D, *lower panel*). Synthesis of RNA was not affected by the presence or absence of CF2 protein when the *Bmglv1::Gst-inVΔCF2* template, which lacks CF2-binding motif, was used (Fig. 4D, *lower panel, lower band*). CF2 bound to the intronV represses transcription from the *Bmglv1* promoter, thus confirming that intronV acts like a cis-regulatory repressor element and not like a translational repressor.

This also explains that because of transcriptional repressor action of intronV, mediated by CF2, *Bmglv1* transcript is not expressed in tissues where CF2 is expressed. On the other hand *Bmglv2– 4* genes are independent of CF2 regulation as they lack intronV. As loss of intronV was responsible for paradigm shift in embryonic regulation of *gloverin* paralogs, we therefore believe that this intron loss was a critical event in the evolution of the *gloverin* family of genes in *B. mori*.

*The Genomic Deletion*—First duplication was also characterized by a genomic deletion in an exon of *Bmglv1*. Deletions in exons have a direct effect on the nature and function of the duplicated gene as most often it leads to a frameshift in ORF that results in change in protein sequence or truncation of the original protein.

Comparative analysis of silkworm *gloverins* revealed an in-frame deletion of 12 bp in the exon3 of *Bmglv1* (Fig. 5A). These nucleotides code for amino acids Ile, His, Asp, and Phe. ClustalW alignment of all known gloverins suggests that the presence of IHDF is a unique feature of BmGlv1 as this sequence motif is not present in other reported gloverin orthologs or paralogs

(Fig. 5B). RHPRDVTWD sequence motif, which has the signal-processing motif, is conserved in all gloverins except for BmGlv1, which has an insertion of amino acids IHDF between Asp and Val. This insertion might have potentially split/abrogated the processing site (Fig. 5B). Hence, we set out to study the functional consequences of the presence/absence of IHDF residues close to the prepro-processing site in BmGlv1.

*BmGlv1 Is Not Processed upon Immune Challenge*—AMPs in general are synthesized with an N-terminal prepro region, which keeps them in an inactive state (1–3). These prepro regions usually contain a signal sequence, which probably helps in their secretion. All the reported gloverins are known to have a precursor region that has a cleavage site between arginine and aspartate in the sequence RHPRDVTWD (Fig. 5B). However, the presence of amino acids IHDF between aspartate and valine of the cleavage recognition sequence has changed the sequence motif next to the cleavage site in BmGlv1. To elucidate the functional consequences of the presence of IHDF, recombinant AcNPV expressing *Bmglv1* and *Bmglv2* genes containing $His_6$-GST tag at the N terminus was expressed in Sf9 cells. Purified GST-BmGlv1 and GST-BmGlv2 proteins were incubated with fat body extracts prepared from bacteria challenged and unchallenged larvae.

The fat body extracts prepared from bacteria-challenged larvae are rich in proteases that were either absent or inactive in the extracts prepared from unchallenged larvae. These proteases process AMPs by cleaving the N-terminal prepro part. We designed an assay to test the processing of GST-BmGlv1 and GST-BmGlv2 proteins. GST-specific band, being upstream to AMP, will only be released if the prepro part of the fused AMP is processed. No GST-specific band was released from either of the proteins when incubated with fat body extracts prepared from unchallenged larvae (Fig. 5C, *lanes 1* and *4*). However, the GST band was released from BmGlv2 but not BmGlv1 upon incubation with fat body extracts prepared from challenged larvae. Thus, release of the GST band in *lane 3* indicates N-terminal processing of BmGlv2 and not of BmGlv1 (Fig. 5C). These results demonstrate that insertion of IHDF has abrogated the processing site in BmGlv1. However, lack of N-terminal processing had no effect on antibacterial activity of BmGlv1 as confirmed by zone inhibition and bacterial clearance assay. In fact, BmGlv1 has stronger antibacterial activity than BmGlv2 (0.05 $\mu$M) and cleared bacteria faster than BmGlv2 (0.1 $\mu$M) (Fig. 5, *D* and *E*) implying that lack of processing is not critical for antibacterial activity of BmGlv1.

## DISCUSSION

In the study reported here, we have explored the effect of genome dynamics on the evolution of the *gloverin* family of AMP genes. Our analysis suggests that *Bmglv1* is the ancestral *gloverin*, and other silkworm *gloverins* evolved in due course of time as a result of three gene duplication events. One notable feature of the first gene duplication was two gain of function phenotypes associated with two deletion events. The first was in-frame genomic deletion of 12 bp that led to gain of prepro cleavage site, and the second was loss of an intron that changed the expression pattern of the duplicated gene *Bmglv2.*
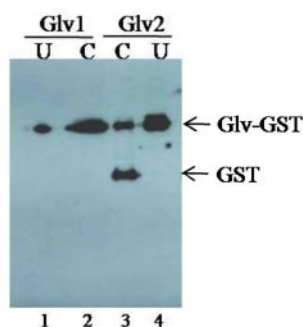
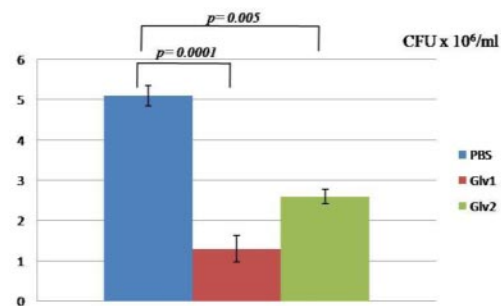# Intron Loss and Evolution of Embryonic Regulation in gloverin

FIGURE 5. **Prepro-processing of BmGlv1.** *A,* exon3 of ancestral *Bmglv1* has unique 12 nucleotides coding for amino acids IHDF (shown in *red*), but the same is deleted in exon3 of *Bmglv2*. *B,* multiple alignment of gloverins, as known now, reveals that IHDF motif, next to prepro cleavage site (*downward arrow*), is unique to BmGlv1. Otherwise BmGlv1 and BmGlv2 proteins are 92% similar. *C,* to check prepro-processing GST-BmGlv1 and GST-BmGlv2 both were incubated with fat body extract prepared from unchallenged (*U*) (*lanes 1* and *4*) and *E. coli* challenged (*C*) 5th instar larvae (*lanes 2* and *3*). Release of GST band in *lane 2* indicates processing of GST-BmGlv2 into GST and BmGlv2 upon immune challenge, whereas absence of GST band in *lane 2* indicates lack of processing of BmGlv1. No processing of either of the proteins is seen with extract prepared from unchallenged fat body. *D,* lack of prepro-processing does not affect antibacterial activity of BmGlv1 as seen in bacterial clearance assay. Shown here is the result of one representative experiment of the three such experiments done under identical conditions. *E, bar diagram* shows number of bacteria surviving (number of colony-forming units $\times$ $10^6$/ml) after 6 h of treatment with equivalent concentrations of BmGlv1, BmGlv2, or PBS (control). Post-treatment the bacterial culture was pelleted down, washed once with PBS, and then dissolved in 200 $\mu$l of sterile plain LB broth. 100 $\mu$l of the soup was plated on antibiotic-free LB-Agar plates and incubated overnight after which the number of colonies were counted. Five replicates of each experiment were done, and *p* value was calculated.

*Evolution of Prepro Domain in Gloverin Proteins*—It is known that precursor AMPs are produced with an N-terminal prepro part containing signal sequence, which is important for their activation (2, 3). Processing of the N-terminal prepro part was considered as a property inherent to all AMPs. Because the processing site sequence has remained conserved even in orthologous gloverins (Fig. 1*B*), we reason that deletion of four amino acids was not a random event but the result of an evolutionary pressure to acquire the processing ability. In view of these results it is tempting to propose that processing of AMPs

is an evolved character in gloverin*s of B. mori*. It will be interesting to investigate whether N-terminal processing in other AMPs has also evolved in a similar manner. We believe that it may be true for other AMPs as well and would most probably require study of humoral immunity in primitive insects. However, based on our data, we hypothesize that *Bmglv1* may be the relic of an ancestral and more primitive immune system.

*Regulation of Bmglv1 Gene by 3'-UTR Intron*—Another significant finding of this study is the functional characterization of a regulatory element in intron of an AMP gene. The 5'- and 3'-untranslated regions (UTRs) that bracket CDSs are fundamental structural and regulatory regions of eukaryotic genes (29–32). UTRs are known to contain large numbers of introns (33), yet there is a lack of hypotheses specifically addressing the evolution of introns within UTRs. A study by Pesole *et al.* (33) suggests that intron density is higher in 5'-UTRs than 3'-UTRs. The observation that fewer 3'-UTRs carry introns is surprising as 3'-UTRs are generally longer than 5'-UTRs and would thus be expected to have higher intron density. The reason for the lack of introns in 3'-UTRs could be intron loss, which is most often restricted to 3' introns (34, 35). The most common mechanism of intron loss is the gene conversion of original gene by reverse transcription of spliced RNA (36–41).

Introns have been shown to affect the expression of different genes at different levels like mRNA export, stability, and translation efficiency (42–44). But the role of introns in the regulation of AMP genes is not reported. Here we have reported a CF2-dependent intronic regulation of an AMP gene. In *Drosophila* the CF2 protein exists in two isoforms CF2I and CF2II. The 113-amino acid-long zinc finger domain of CF2 consists of five to seven contiguous zinc fingers of the C2H2 type (45–48). The zinc finger motif of CF2 resembles zinc finger domains of the developmentally regulated *Drosophila* transcription factors Kruppel and hunchback. CF2 is basically a transcriptional activator and possesses transcription activation domain consisting of 17 glutamines interspersed with 7 acidic residues (49). However, this study demonstrates that CF2 can act as a repressor as well. This also suggests that action of CF2 as activator or repressor is context-dependent. This is the first study where CF2 has been shown to bind to an intron element and repress the transcription of the native gene. CF2 bound to intronV may lead to looping back of the DNA, which in turn can silence the promoter (Fig. 6). Involvement of factors other than CF2 in intronV-mediated, promoter silencing cannot be ruled out.

*Intron Loss in 3'-UTR of the Gene Is Positively Selected*—Fifth intron of *Bmglv1* splits the 3'-UTR, implying that loss of intronV would not have affected the BmGlv1 protein; hence this intron loss could have been inconsequential from the evolution point of view. We show that loss of intronV did not alter the gene product, but it changed developmental regulation of *Bmglv2* resulting in acquisition of a new function in embryonic development. However, this gain of function, in the derived gene *Bmglv2*, in embryonic development was because of its ability to express in the embryonic stages. We have shown that the embryonic regulation of promoters of both the ancestral and the derived gene is the same; still the ancestral *Bmglv1* does not express in the embryo because of repression by intronV in embryonic stages (Fig. 3, *A* and *B*, and Fig. 4*C*). As intronV
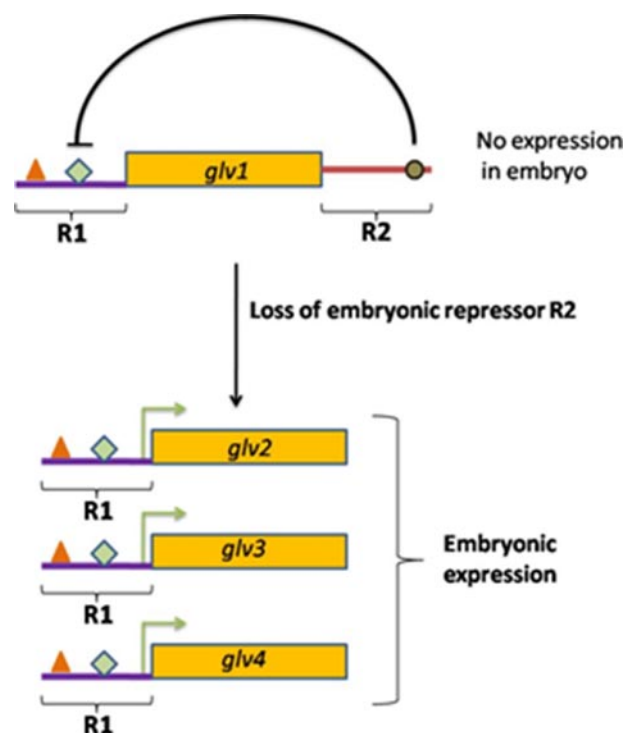


FIGURE 6. **Model to explain evolution of *gloverin* gene family by subneofunctionalization.** Our results suggest the presence of two regulatory elements as follows: (i) promoter (R1) and (ii) intronV (*R2*) in ancestral *gloverin* (*Bmglv1*). We have shown that promoter regulation is same for both the ancestral (*Bmglv1*) and the duplicated copy (*Bmglv2*); still expression of *Bmglv1* was not observed in embryonic stages because of inhibition of *Bmglv1* transcription in these tissues by CF2. However, this CF2-mediated repression was mediated by intronV, which is present only in *Bmglv1*, the ancestral copy. Thus we identify intronV as the second regulator (*R2*) and also show that R2 is dominant over R1 in embryonic stages. During the first duplication, intronV (*R2*), was lost resulting in embryonic expression of daughter gene *Bmglv2*. Interestingly, embryonically expressing paralog *Bmglv2* also controls embryonic development, a feature not observed in *Bmglv1*, suggesting gain of function for *Bmglv2* (neofunctionalization). Loss of R2 is an example of regulatory subfunctionalization which led to neofunctionalization of *Bmglv2*, so the changes during first duplication event can be summed up as subneofunctionalization.

regulation is dominant over the *Bmglv1* promoter, the only way to achieve embryonic expression was by losing intronV. Thus, *Bmglv1* would have experienced strong pressure to lose intronV, which was eventually lost during the first duplication process. In other words loss of intronV of *Bmglv1* was positively selected to achieve embryonic expression, and it was not lost randomly for being a 3'-UTR intron. Thus, our study suggests that loss of 3'-UTR introns may be associated with distinct phenotypes, and hence these introns could have experienced positive selection.

*AMPs Have a Role in Embryonic Development*—The significance of embryonic expression of AMPs is not clear. There are very few reports on the role of AMPs in embryonic development. Recently, *hemolin* expression has been shown to be important for embryonic diapause and development (50, 51). Embryonic diapause is a special physiological state that is developmentally controlled and is observed in many insects. A diapausing embryo, where most of the physiological processes are suppressed, is considered to be under stress, and it is suggested that expression of *hemolin* in such embryos could be part of a broad stress response (50, 51). We speculate that other AMPs,

like *gloverin* paralogs, which express in the embryonic stages, might be part of the same broad stress response. However, embryonic expression of AMPs has also been reported from *Drosophila,* which does not undergo diapause and hence diapause may not be the only function to be affected by AMPs (52). Embryonic expression of *cecropinA* in *Drosophila* embryo was detected in tissues like yolk and embryonic epidermis but not in embryonic fat body, mid gut, or hemocytes. This is in stark contrast to *CecropinA* expression in larval and adult stages where it is expressed in fat body and mid gut (1). Furthermore, GATA factor Serpent is needed for expression of AMPs in embryonic yolk but not in embryonic epidermis (52) suggesting that expression of AMPs in different tissues of the embryo is regulated by different transcription factors.

Here we have shown negative regulation of ancestral *gloverin* by embryonic protein CF2, which is expressed in yolk and controls oogenesis (53). This study adds one more dimension to the embryonic regulation of AMPs, and more precisely it reveals the evolution of embryonic regulation in an AMP gene family.

During the course of evolution, there have been episodes of extensive intron loss and gain because of selective forces that affect the rate of intron dynamics (54, 55). For evolutionarily conserved genes, intron insertion supposedly had adaptive effect like increasing stability of RNA, whereas intron loss was found to be deleterious (43, 44, 56). This apparent functional importance of introns could be due in part because of their effects on gene regulation. Our study demonstrates the role of intron as cis-regulators in gene evolution and thus adds another dimension to genome plasticity. The fact that intron loss achieved embryonic expression for *gloverin* paralogs, a property which appears to be common feature of all AMP genes except for *Bmglv1*, suggests that this intron loss might have experienced positive selection. In an earlier study positive selection of intron loss in *Drosophila* has been shown (57). To the best of our knowledge this is the first report where positive selection for intron loss in an AMP gene has been functionally validated. In summary, our study suggests that intron loss or gain may not be a passive/random feature of genome dynamics but a result of selection pressure.

### REFERENCES

1. Hoffmann, J. A. (2003) *Nature* **426,** 33–38
2. Boman, H. G., and Hultmark, D. (1987) *Annu. Rev. Microbiol.* **41,** 103–126
3. Boman, H. G. (1995) *Annu. Rev. Immunol.* **13,** 61–92
4. Lemaitre, B. (2004) *Nat. Rev. Immunol.* **4,** 521–527
5. Ip, Y. T. (1993) *Cell* **75,** 753–763
6. Dushay, M. S., Asling, B., and Hultmark, D. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 10343–10347
7. Hedengren, M., Asling, B., Dushay, M. S., Ando, I., Ekengren, S., Wihlborg, M., and Hultmark, D. (1999) *Mol. Cell* **4,** 827–837
8. Leulier, F., Rodriguez, A., Khush, R. S., Abrams, J. M., and Lemaitre, B. (2000) *EMBO Rep.* **1,** 353–358
9. Rutschmann, S., Jung, A. C., Zhou, R., Silverman, N., Hoffmann, J. A., and Ferrandon, D. (2000) *Nat. Immunol.* **1,** 342–347
10. Vidal, S., Khush, R. S., Leulier, F., Tzou, P., Nakamura, M., and Lemaitre, B. (2001) *Genes Dev.* **15,** 1900–1912
11. Silverman, N., Zhou, R., Stoven, S., Pandey, N., Hultmark, D., and Maniatis, T. (2000) *Genes Dev.* **14,** 2461–2471
12. Stöven, S., Ando, I., Kadalayil, L., Engström, Y., and Hultmark, D. (2000) *EMBO Rep.* **1,** 347–352
13. Georgel, P., Naitza, S., Kappler, C., Ferrandon, D., Zachary, D., Swimmer, C., Kopczynski, C., Duyk, G., Reichhart, J. M., and Hoffmann, J. A. (2001) *Dev. Cell* **1,** 503–514
14. Leulier, F., Vidal, S., Saigo, K., Ueda, R., and Lemaitre, B. (2002) *Curr. Biol.* **12,** 996–1000
15. Naitza, S., Rosse, C., Kappler, C., Georgel, P., Belvin, M., Gubb, D., Camonis, J., Hoffmann, J. A., and Reichhart, J. M. (2002) *Immunity* **17,** 575–581
16. Sun, S. C., Lindstorm, I., Boman, H. G., Faye, I., and Schmidt, O. (1990) *Science* **250,** 1729–1732
17. Mita, K., Kasahara, M., Sasaki, S., Nagayasu, Y., Yamada, T., Kanamori, H., Namiki, N., Kitagawa, M., Yamashita, H., Yasukochi, Y., Kadono-Okuda, K., Yamamoto, K., Ajimura, M., Ravikumar, G., Shimomura, M., Nagamura, Y., Shin-I, T., Abe, H., Shimada, T., Morishita, S., and Sasaki, T. (2004) *DNA Res.* **11,** 27–35
18. Xia, Q., Zhou, Z., Lu, C., Cheng, D., Dai, F., Li, B., Zhao, P., Zha, X., Cheng, T., Chai, C., Pan, G., Xu, J., Liu, C., Lin, Y., Qian, J., *et al.* (2004) *Science* **306,** 1937–1940
19. Mita, K., Morimyo, M., Okano, K., Koike, Y., Nohata, J., Kawasaki, H., Kadono-Okuda, K., Yamamoto, K., Suzuki, M. G., Shimada, T., Goldsmith, M. R., and Maeda, S. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 14121–14126
20. Axen, A., Carlsson, A., Engstrom, A., and Bennich, H. (1997) *Eur. J. Biochem.* **247,** 614–619
21. Mackintosh, J. A., Gooley, A. A., Karuso, P. H., Beattie, A. J., Jardine, D. R., and Veal, D. A. (1998) *Dev. Comp. Immunol.* **22,** 387–399
22. Lundstrom, A., Liu, G., Kang, D., Berzins, K., and Steiner, H. (2002) *Insect Biochem. Mol. Biol.* **32,** 795–801
23. Lynch, M., and Conery, J. S. (2003) *Science* **302,** 1401–1404
24. Suzuki, Y., Tsuda, M., Takiya, S., Hirose, S., Suzuki, M., Kameda, M., and Ninaki, O. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83,** 9522–9526
25. Sakurai, H., Susumu, I., and Tomino, S. (1995) in *In Vitro Transcription and Translation Protocols* (*Methods in Molecular Biology*) (Martin, J. T., ed) pp. 233–244, Humana Press Inc., Totowa, NJ
26. Gebauer, F., Corona, D. F. V., Preiss, T., Becker, P. B., and Hentze, M. W. (1999) *EMBO J.* **18,** 6146–6154
27. Lanoix, J., Ouwendijk, J., Lin, C. C., Stark, A., Love, H. D., Ostermann, J., and Nilsson, T. (1999) *EMBO J.* **18,** 4935–4948
28. Kanginakudru, S., Royer, C., Edupalli, S. V., Jalabert, A., Mauchamp, B., Chandrashekaraiah Prasad, S. V., Chavancy, G., Couble, P., and Nagaraju, J. (2007) *Insect Mol. Biol.* **16,** 635–644
29. Ptashne, M., and Gann, A. (2001) *Essays Biochem.* **37,** 1–15
30. Larizza, A., Makalowski, W., Pesole, G., and Saccone, C. (2002) *Comput. Chem.* **26,** 479–490
31. Mignone, F., Gissi, C., Liuni, S., and Pesole, G. (2002) *Genome Biol.* **3,** 3
32. Wilkie, G. S., Dickson, K. S., and Gray, N. K. (2003) *Trends Biochem. Sci.* **28,** 182–188
33. Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., and Liuni, S. (2001) *Gene* (*Amst.*) **276,** 73–81
34. Mourier, T., and Jeffares, D. C. (2003) *Science* **300,** 1393
35. Sakurai, A., Fujimori, S., Kochiwa, H., Kitamura-Abe, S., Washio, T., Saito, R., Carninci, P., Hayashizaki, Y., and Tomita, M. (2002) *Gene* (*Amst.*) **300,** 89–95
36. Lewin, R. (1983) *Science* **219,** 1052–1054
37. Bernstein, L. B., Mount, S. M., and Weiner, A. M. (1983) *Cell* **32,** 461–472
38. Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986) *Annu. Rev. Biochem.* **55,** 631–661
39. Fink, G. R. (1987) *Cell* **49,** 5–6
40. Long, M., and Langley, C. H. (1993) *Science* **260,** 91–95
41. Derr, L. K. (1998) *Genetics* **148,** 937–945
42. Bourdon, V., Harvey, A., and Lonsdale, D. M. (2001) *EMBO Rep.* **2,** 394–398

43. Le Hir, H., Nott, A., and Moore, M. J. (2003) *Trends Biochem. Sci.* **28,** 215–220

44. Nott, A., Meislin, S. H., and Moore, M. J. (2003) *RNA (N. Y.)* **9,** 607–617

45. Shea, M. J., King, D. L., Conboy, M. J., Mariani, B. D., and Kafatos, F. C. (1990) *Genes Dev.* **4,** 1128–1140

46. Hsu, T., Gogos, J. A., Kirsh, S. A., and Kafatos, F. C. (1992) *Science* **257,** 1946–1950

47. Gogos, J. A., Jin, J., Wan, H., Kokkinidis, M., and Kafatos, F. C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 2159–2164

48. Hsu, T., Bagni, C., Sutherland, J. D., and Kafatos, F. C. (1996) *Genes Dev.* **10,** 1411–1421

49. Mitchell, P. J., and Tjian, R. (1989) *Science* **245,** 371–378

50. Lee, K. Y., Horodyski, F. M., Valaitis, A. P., and Denlinger, D. L. (2002) *Insect Biochem. Mol. Biol.* **32,** 1457–1467

51. Bettencourt, R., Terenius, O., and Faye, I. (2002) *Insect Mol. Biol.* **11,** 267–271

52. Tingvall, T. O., Roos, E., and Engström, Y. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3884–3888

53. Mantrova, E. Y., and Hsu, T. (1998) *Genes Dev.* **12,** 1166–1175

54. Jeffares, D. C., Mourier, T., and Penny, D. (2006) *Trends Genet.* **22,** 16–22

55. Carmel, L., Wolf, Y. I., Rogozin, I. B., and Koonin, E. V. (2007) *Genome Res.* **17,** 1034–1044

56. Carmel, L., Rogozin, I. B., Wolf, Y. I., and Koonin, E. V. (2007) *Genome Res.* **17,** 1045–1050

57. Llopart, A., Comeron, J. M., Brunet, F. G., Lachaise, D., and Long, M. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8121–8126