Eukaryotic operon-like transcription of functionally related genes in Drosophila

Yehuda Ben-Shahar, Kishore Nannapaneni, Thomas L. Casavant, Todd E. Scheetz, and Michael J. Welsh

This information is current as of December 2006.

Supplementary Material	Supplementary material can be found at: www.pnas.org/cgi/content/full/0609683104/DC1
	This article has been cited by other articles: www.pnas.org#otherarticles
E-mail Alerts	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.
Rights & Permissions	To reproduce this article in part (figures, tables) or in entirety, see: www.pnas.org/misc/rightperm.shtml
Reprints	To order reprints, see: www.pnas.org/misc/reprints.shtml

Notes:

Eukaryotic operon-like transcription of functionally related genes in *Drosophila*

Yehuda Ben-Shahar*†, Kishore Nannapaneni‡§, Thomas L. Casavant‡§¶‖, Todd E. Scheetz‡§¶, and Michael J. Welsh*†**††

*Howard Hughes Medical Institute, and Departments of †Internal Medicine, **Physiology and Biophysics, and ¶Ophthalmology and Visual Sciences, Roy J. and Lucille A. Carver College of Medicine Center for Bioinformatics and Computational Biology, Departments of ‖Electrical and Computer Engineering and §Biomedical Engineering, University of Iowa, Iowa City, IA 52242

Complex biological processes require coordinated function of many genes. One evolutionary solution to the problem of coordinately expressing functionally related genes in bacteria and nematodes is organization of genes in operons. Surprisingly, eukaryotic operons are considered rare outside the nematode lineage. In *Drosophila melanogaster*, we found *lounge lizard* (*llz*), which encodes a degenerin/ENaC cation channel, cotranscribed with *CheB42a*, a nonhomologous gene of unknown function residing <100 bp upstream. These two genes were transcribed from a single promoter as one primary transcript and were processed posttranscriptionally to generate individual mRNAs. The mechanism did not involve alternative splicing, and it differed from the trans splicing used in nematode operons. Both genes were expressed in the same tissues, and previous work suggested that both may be involved in courtship behavior. A bioinformatic approach identified numerous additional loci as potential *Drosophila* operons. These data reveal eukaryotic operon-like transcription of functionally related genes in *Drosophila*. The results also suggest that operon-based transcription may be more common in eukaryotes than previously appreciated.

degenerin/epithelial Na⁺ channel | fruit fly | ion channels | *lounge lizard* | *CheB42a*

Fig. 1. The *CheB42a/llz* operon-like locus is conserved across *Drosophila* species. VISTA plots of the *CheB42a/llz* locus in various *Drosophila* species with *D. melanogaster* as the reference genome. Order of species represents phylogenetic distance from *D. melanogaster* with closest species (*D. simulans*) at the top. Curves represent % conservation in a continuous 100-bp interval.

Degenerin/epithelial Na⁺ channel (DEG/ENaC) proteins form non-voltage-gated cation channels. Their functions vary widely and include mechanosensation, nociception (sensing acidic stimuli), salt sensing, and detecting the peptidergic neurotransmitter FMRFamide (1–4). During our investigation of *Drosophila melanogaster* DEG/ENaC channels, we discovered a genomic locus containing a DEG/ENaC gene, which we named *lounge lizard* (*llz*, CG33349), and a nonhomologous gene of unknown function (*CheB42a*) (5) that appeared to be transcribed in the same direction with a predicted intergenic distance of <100 bp. The close proximity of these two genes and the lack of obvious promoter sequences in front of the downstream gene suggested that they might be cotranscribed from a single upstream promoter. This organization would resemble that of the nematode operon, an elegant evolutionary solution for coexpressing genes that are part of the same biochemical pathway or physiological process (6, 7).

Although abundant in nematodes, operons are thought to be rare in other eukaryotes. This is despite eukaryote's complicated cellular physiology and the intricacy of their genomes, which might predict that operon-like transcription would solve some of the problem of coordinately regulating genes (8). With this background, we hypothesized that the *CheB42a/llz* locus might be transcribed as a eukaryotic operon, and, hence, the two genes might be functionally related.

Results

Organization of the *CheB42a/llz* Locus is Conserved Across *Drosophila* Species. In the *D. melanogaster* genome, *llz* is located immediately downstream of a previously cloned gene, *CheB42a* (5). Transcription of the two genes appeared to be in the same direction,
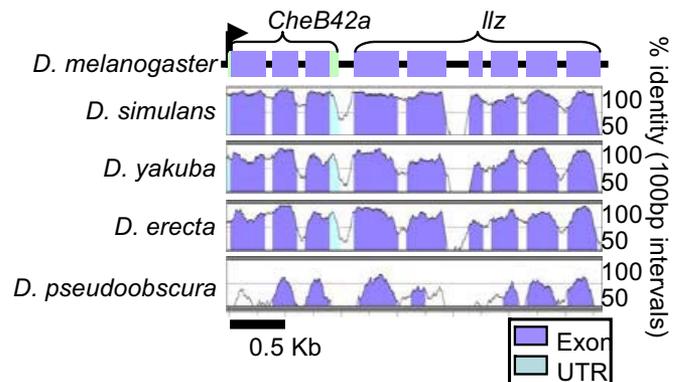
and both have multiple exons (Fig. 1). The two gene products belong to different protein families. Genomes of other *Drosophila* species contain the same two genes in a similar organization (Fig. 1), raising the possibility that the arrangement imparts some function. Examination of the genomic sequence revealed that the 3′ end of *CheB42a* contains a poly(A) addition sequence, predicting that *CheB42a* should generate an independent mRNA. However, we detected no obvious promoter elements in the 97-bp intergenic sequence between *CheB42a* and *llz*, suggesting that if it is expressed, then the *llz* gene may be controlled by upstream sequences.

These genomic features suggested that the *CheB42a/llz* locus might be transcribed as an operon. In nematode operons, one promoter/enhancer generates a primary transcript (or pre-mRNA) for two or more genes (6, 9). *Caenorhabditis elegans* process this primary transcript into individual mRNAs for each gene and attach a specific spliced leader sequence (SL2) to the 5′ ends of mRNAs from the downstream genes. Nematode operons often encode genes contributing to a common biological process, just as bacterial operons usually code for proteins in a shared pathway (10).

CheB42a and llz Are Cotranscribed from a Single Promoter. To test the hypothesis that the *CheB42a/llz* locus undergoes operon-like
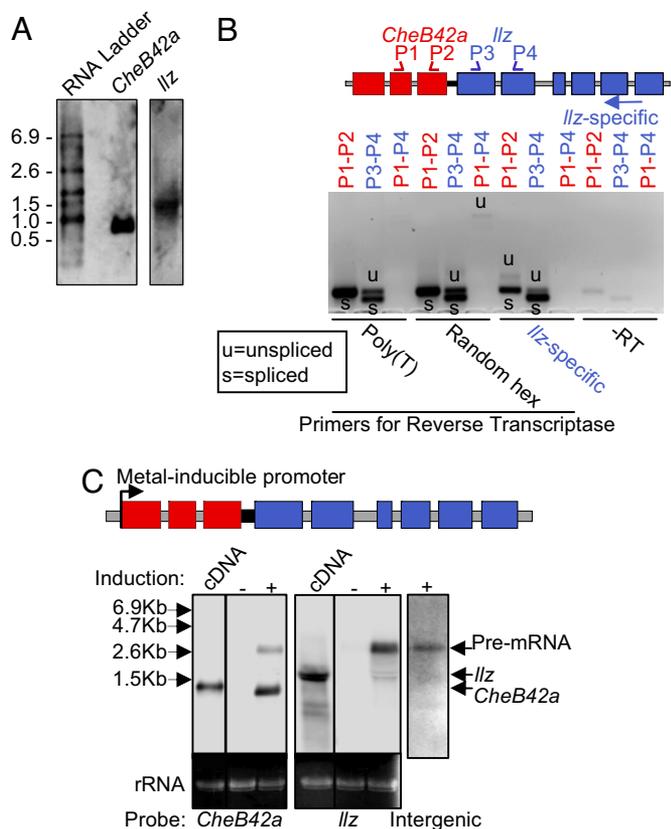
Fig. 2. The *CheB42a/llz* locus is transcribed as a single pre-mRNA and is processed posttranscriptionally. (*A*) Northern blot analysis of *CheB42a* and *llz* expressed in flies. Same membrane was probed for both transcripts. Left lane shows an RNA-size marker. (*B*) RT-PCR of total RNA extracted from male appendages. RT step was performed with poly(T), random hexamers, or *llz*-specific primers. "−RT" indicates reaction without reverse transcriptase. P1–P4 represents primers used in the PCR step. Primers P1 and P2 are *CheB42a*-specific and primers P3 and P4 are *llz*-specific. Multiple clones from all PCR products were sequenced to verify their identity. Some PCR products from P1–P4 primers had some of the introns variably spliced, further indicating that they were not due to genomic contamination. (*C*) Northern blot of *CheB42a* and *llz* transcribed from an exogenous inducible upstream promoter. A genomic fragment containing all exons and introns of *CheB42a* and *llz* was subcloned into a vector containing a metal-inducible promoter and transfected into *Drosophila* S2 cells. cDNAs of both genes were transfected as controls. RNA produced from the *CheB42a* cDNA is slightly larger than the spliced product from the genomic construct due to the presence of V5 and 6× His tags sequences in the *CheB42a* cDNA. The *llz*-specific signal showed two bands when expressed from both the cDNA and the genomic construct, which may indicate alternative lengths of poly(A) tails. Pre-mRNA also was identified with a probe to the intergenic region (right lane). Failure to detect the pre-mRNA on Northern blot analysis (*A*) is likely because of its low abundance, as has been observed in *C. elegans* where operon processing is very efficient (6).

transcription, we first asked whether both genes are transcribed. Northern blots of whole flies revealed independent transcripts for both *CheB42a* and *llz* (Fig. 2*a*).

We then used RT-PCR to determine whether the locus also produced a primary transcript containing both genes. Primers for the reverse transcription (RT) reaction included as follows: (*i*) poly(T), which allows reverse transcription of polyadenylated mRNA; (*ii*) random hexamers, which reverse transcribes all RNAs; and (*iii*) an *llz*-specific primer (Fig. 2*B*). PCR primers spanned the last intron of *CheB42a*, the first intron of *llz*, or both plus the intergenic region. With the poly(T) and random hexamer primers, PCR generated the predicted spliced products from *CheB42a* and *llz* (Fig. 2*B*), and products from the poly(T) reaction suggested polyadenylation of both *CheB42a* and *llz*

mRNAs. We also found PCR products extending from the *CheB42a* to the *llz* locus; cloning and sequencing of those products showed that they retained the intergenic region, but some had introns variably spliced out. These data suggested that a single pre-mRNA contained both genes. Importantly, when we primed the RT reaction with an *llz*-specific primer, PCR generated spliced *CheB42a* products and unspliced *CheB42a* and *CheB42a-llz* products (Fig. 2*B*). Finding a primary transcript spanning both genes suggested operon-like transcription.

If the *CheB42a/llz* locus functions as an operon, we reasoned that an exogenous promoter inserted upstream of the locus also should generate a primary transcript containing sequences from CheB42a and *llz* plus individual transcripts for each gene. To test this prediction, we cloned a complete genomic fragment of the *CheB42a/llz* locus without its endogenous promoter (i.e., upstream of *CheB42a*) into a metal-inducible expression vector and transfected *Drosophila* S2 cells (Fig. 2*C*). Without induction, we detected neither transcript. This result suggests that the intergenic region and *CheB42a* introns do not contain promoter activity that drives *llz* expression independent of *CheB42a*. After inducing transcription with copper sulfate, Northern blots probed for *CheB42a* and *llz* revealed an RNA species encoding both genes plus mature transcripts for each individual gene. The identity of the primary transcript also was verified by an intergenic probe (Fig. 2*C*). The ratio of mature mRNA to primary transcript was higher for *CheB42a* than for *llz*; this result suggests that processing of the downstream gene was less efficient than the upstream gene, a finding consistent with tissue expression levels (Fig. 2*A*) and operon processing in *C. elegans* (11).

As an additional test for an unrecognized promoter that might drive *llz* transcription, we cloned a genomic fragment extending from the start of the first *CheB42a* exon through the intergenic sequence and used it to drive a GFP reporter gene, using the UAS-GAL4 system. In flies transgenic for this construct, we failed to detect GFP expression (data not shown), consistent with the lack of a promoter downstream of the *CheB42a* transcriptional start site.

Together, these results indicate that an upstream promoter generated a single pre-mRNA that contained sequences for both genes. The pre-mRNA then was processed to two mature transcripts, one for *CheB42a* and one for *llz*. Thus, this locus exhibits processing that shares some features with *C. elegans* operons (6). In contrast, the *CheB42a/llz* locus behaves differently from the *Drosophila stoned* locus, in which a single mature (spliced) transcript encodes two different proteins (i.e., a dicistronic mRNA) (12). We excluded this alternative because Northern blots from fly tissues revealed two independent transcripts (Fig. 2*A*). In addition, the RT-PCR results (Fig. 2*B*) and the *CheB42a* 3′ RACE and *llz* 5′ RACE results (see below) failed to identify fully spliced transcripts containing both *CheB42a* and *llz* sequences.

Analysis of *CheB42a* and *llz* Transcripts. Transcripts encoding two unrelated proteins also could arise from one primary transcript by alternative splicing; an example is the *Drosophila* cholinergic locus (13). This mechanism predicts that the two mature mRNAs will share some 5′ UTR sequences. To test this possibility, we used 5′ RACE to define the transcriptional start sites of *CheB42a* and *llz*. We found that mature (all introns spliced out) *llz* transcripts shared no sequence with *CheB42a* transcripts (Fig. 3*A*). Lack of a shared 5′ UTR excludes alternative splicing as the mechanism for individual transcript processing in at this locus.

In *C. elegans* operons, primary transcripts are processed to individual mRNAs with trans splicing of a specific spliced-leader sequence (SL2) to the 5′ end of mRNA from the downstream gene (6). Our 5′ RACE analysis revealed no splice-leader sequence attached to the *llz* transcript, suggesting that a different mechanism was responsible for 5′ end formation of the *llz* gene.

Fig. 3. The downstream *llz* transcript does not contain traditional 5′ cap *llz* 5′ RACE analysis. (*A*) Schematic of genomic (*Upper*) and 5′ RACE (*Lower*) structure. (*A Lower*) Genomic sequence and sequence of 5′ end of typical llz products with the extra nontemplate G highlighted in yellow (product 1). Product 2 indicates sequence in same region of a partially spliced product that extended through *CheB42a*. (*B*) Immuoprecipitation of total RNA with an anti-cap antibody. Total RNA from male appendages was precipitated with an anti-cap or a control antibody. Relative amounts of mRNAs were detected with real-time quantitative RT-PCR with a gene-specific probe and primers. The ribosomal gene *rp49* served as a positive control for a capped mRNA, and the 28s ribosomal RNA served as an uncapped negative control. (*B Upper*) Relative amount of precipitated RNA compared with the amount of RNA remaining in the supernatant; low levels of precipitated *llz* were not simply the result of low levels of *llz* mRNA. (*C*) Analysis of polysomal RNA. S2 cells were transfected with the same genomic construct as in Fig. 2*C*. Cytosolic suspension was loaded on a 15–45% sucrose gradient, and total RNA was isolated from each fraction. (*C Top*) Absorbance at 254 nm for each fraction, which was used as a guide to indicate subpolysomal and polysomal fractions. (*Middle Top*, *Middle Bottom*, and *Bottom*) Real-time quantitative RT-PCR (shown as percentage of total mRNA) results for *CheB42a*, *llz*, and *rp49* (as a capped control), respectively.

In attempting 5′ RACE, on multiple occasions, we first selected capped mRNAs, but although we readily detected *CheB42a* products, we did not generate *llz*-specific products. However, when we omitted selection for capped mRNA, we recovered products from both genes. These results suggested that *llz* transcripts might not bear a traditional 5′ cap. To further test this hypothesis, we used an anti-cap antibody to immunoprecipitate capped RNAs from fly tissues. The anti-cap antibody precipitated *CheB42a*, but not *llz* transcripts, which remained in the supernatant (Fig. 3*B*). These results suggest that *llz* transcripts do not bear a conventional cap.

Interestingly, at the 5′ end of *llz* 5′ RACE products, we found a guanosine that was not encoded by genomic DNA (Fig. 3*A*). We did not find a guanosine in partial cDNAs, and genomic DNA sequencing of the *llz* 5′ UTR region showed this did not arise from a DNA polymorphism. These data indicate that the additional guanosine was added posttranscriptionally during 5′ end formation of *llz* and suggest it may be an important element in processing the *llz* transcript. We considered that the guanosine might have been added in a self-splicing reaction analogous to that which occurs in group I introns, although in that case, the guanosine is added to intron- rather than exon-encoded sequence (14). Nevertheless, we tested for group I intron-like activity but found no processing of the locus under *in vitro* conditions used to study group I introns (data not shown). Although we do not know the mechanism involved, these results suggest that the guanosine might be added posttranscriptionally, possibly as part of the 5′ end formation of *llz*.

We asked whether *llz* produces functional mRNA by testing for spliced *llz* transcripts in polyribosomes, the sites of protein synthesis (15). We expressed the *CheB42a/llz* genomic locus in *Drosophila* S2 cells (as described above for Fig. 3*C*) and collected sucrose gradient fractions to separate subpopulations of ribosome-bound cellular mRNAs. We found *llz* transcripts in the same fractions as *CheB42a* and *rp49* (a ribosomal protein used as a positive control). These data indicate that although *llz* may not contain a traditional cap, it attracted the translational machinery and, hence, is very likely a functional mRNA. Of note, cap-independent translation has been reported for eukaryotic mRNAs (16).

Other *D. melanogaster* Genes also May Be Transcribed as Operons. Operons are common in *C. elegans* (7). This suggested that the *CheB42a/llz* locus might not be a unique case in *Drosophila*, but rather represent a general pattern of chromosomal organization. To evaluate this possibility, we examined the *D. melanogaster* genome. Using an arbitrary and conservative maximum length of 700 bp for the intergenic region, we identified 2,468 pairs of genes transcribed in the same direction. Of these, 1,378 gene pairs had no predicted promoter sequences in the intergenic region [supporting information (SI) Table 1]. Requiring a stop codon in the upstream gene and polyadenylation signal sequences in the 3′ UTR of both genes further reduced the number to 409 gene pairs. Of these, 397 were pairs of nonhomologous genes. These features are hallmarks of operons in the *C. elegans* genome (7) and suggest that the *Drosophila* genome contains numerous loci that might be transcribed in a similar fashion.

To test whether any of these gene pairs are possible operons, we chose two candidate loci containing nonhomologous pairs of genes: *Tektin-C/CG10542* and *Cdk5/flotillin*. Both loci were conserved across *Drosophila* species (SI Fig. 6 *A* and *B*). We then used an RT-PCR analysis similar to what we reported in Fig. 2*B* for the *CheB42a/llz* locus. PCR primers crossed introns in each gene (SI Fig. 6 *A* and *B*). When we used a *CG10542* (the downstream gene) primer for the RT reaction, we found *Tektin-C* (the upstream gene) PCR products that had an intron spliced out (SI Fig. 6*C* and confirmed by DNA sequencing). This result indicates that the locus generated a single pre-mRNA that contained both genes. The functions of *Tektin-C* and *CG10542* encoded proteins are unknown, but both may be involved with cytoskeleton organization (www.flybase.org). We obtained similar RT-PCR results from the *Cdk5/flotillin* locus; RT-PCR products representing spliced transcripts were derived from both genes (SI Fig. 6*D*). *flotillin* is an integral membrane protein and *Cdk5* (cyclin-dependent kinase 5) is a soluble protein (17, 18). Although they are not known to participate in a common biochemical pathway, they both contribute to neuronal signal transduction, and both have been implicated in the pathogenesis of Alzheimer's disease in humans (19, 20). These data suggest

that these two additional loci also may show an operon-like organization.

Because *llz* transcripts begin with a guanosine not encoded by genomic DNA, we searched for this feature in ESTs from the downstream genes present in our candidate list. We found an EST for 392 of the 409 downstream genes. In 26 cases (6.6%), we discovered a leading guanosine not encoded by the genomic template. Other noncoded nucleotides were rare (1 T, 3 A, and 2 C) in these 392 ESTs. Moreover, two random EST sets had a nontemplate G in 4 of 392 (1.0%) and 6 of 392 (1.5%) ESTs ($P < 0.0005$ for both by Fisher's exact test). These data are consistent with the hypothesis that the *Drosophila* genome contains additional operon-like loci and that a specialized, although as-yet-unknown, mechanism may be involved in their processing.

Production of Separate *CheB42a* and *llz* Transcripts Requires Specific Sequences. Processing of *C. elegans* operons requires specific sequence elements in the intergenic region and the 3′ UTR of the upstream gene (11, 21, 22). Therefore, we mutated these regions to learn whether they also influence production of individual *CheB42a* and *llz* transcripts; we used the construct shown in Fig. 2*C*. The 3′ end of the *CheB42a-llz* intergenic region (Fig. 4*A*) is conserved across *Drosophila* species (data not shown). Although it is not conserved in the 409 gene pairs, it is similar to a critical 3′ end sequence of some *C. elegans* operons (22). Mutating this sequence (mutation 1; Fig. 4*B*) increased the amount of pre-mRNA and mature *llz* transcripts relative to the wild-type construct. These data suggest that intergenic sequences contribute to processing of the locus.

Previous work in *C. elegans* suggested that an interplay between the poly(A) signal of the upstream gene and downstream intergenic sequences is important for operon processing (22). We found that disrupting the *CheB42a* poly(A) signal eliminated mature *CheB42a* transcripts while increasing pre-mRNA and *llz* transcripts (mutant 2; Fig. 4*B*). Formation of *llz* from this mutant suggests that *llz* and *CheB42a* may mature independently of each other. The loss of mature *CheB42a* transcripts may have resulted from their reduced stability or failure to properly separate from the primary transcript. In contrast, when we converted the noncanonical, and much less common poly(A) signal (AT-TAAA) to the canonical AATAAA (mutant 3; Fig. 4*B*), we eliminated both the primary and the mature *llz* transcripts. This result is consistent with derivation of mature *llz* transcripts from the primary transcript rather than from a cryptic promoter or some form of transcriptional reinitiation. In addition, these data suggest that the poly(A) signal of the upstream gene is important for preventing premature transcriptional termination at the locus. Why eukaryotes have evolved the less common, noncanonical ATTAAA poly(A) signal is not known (23). Our data suggest one potential function for this polyA signal.

CheB42a and llz Show Similar Expression Patterns. Operon-like transcription from a single promoter predicted the two genes would show similar expression patterns. A previous study reported that *CheB42a* was expressed in the front legs of males (5). In agreement with that work, real-time quantitative RT-PCR and Northern blot analyses revealed *CheB42a* enriched in male appendages (Fig. 5 *A* and *B*); expression in females was less abundant. Importantly, *llz* showed the same pattern with enrichment in male appendages. Although the ratio of transcripts in appendages to those in bodies differed somewhat between *CheB42a* and *llz*, similar variations are observed with nematode operons (9). Such variations have been attributed to posttranscriptional processes that vary transcript levels for coordinately transcribed genes. We also found that ablating cells by expressing the apoptotic gene *rpr* (24) under control of the *CheB42a/llz* promoter reduced expression of both genes in male appendages
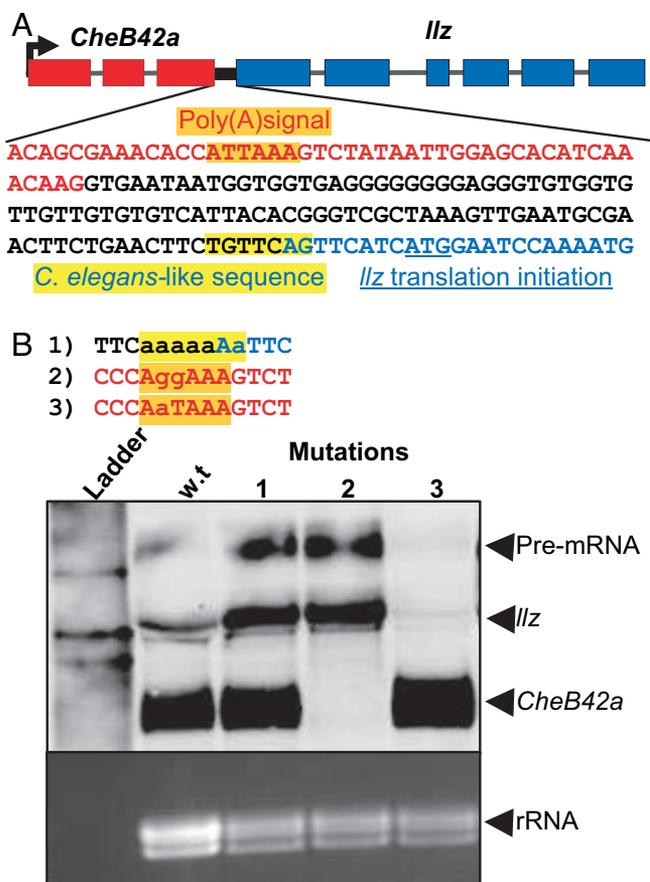

Fig. 4. Maturation of *CheB42a* and *llz* mRNA requires specific sequences. (*A*) Intergenic region of the *CheB42a/llz* locus. Red letters represent 3′ end of the last exon in *CheB42a*. Orange box indicates predicted polyadenylation signal for *CheB42a*. Blue letters indicate start of the first exon of *llz*. Blue underline indicates translation start site for *llz*. Yellow box represents conserved sequence found at 3′ end of intercistronic region in some *C. elegans* operons. Boundaries of the *CheB42a* and *llz* transcripts and the intergenic region were determined by 3′ and 5′ RACE. (*B*) Effect of mutagenesis on processing of the *CheB42a/llz* locus. Sequences represent the mutations studied. Colored boxes correspond to colored boxes in *A*. Mutations are in lowercase letters. Northern blot was incubated with probes for *CheB42a* and *llz* simultaneously.

(Fig. 5*C*). Together, these data suggest that both genes are expressed in male appendages.

In situ hybridization of late-stage embryos localized signal in two cells in the anterior portion of embryos (Fig. 5 *D–G*); based on their position and rarity, these cells are likely to be neurons. Importantly, we saw exactly the same pattern with both *CheB42a* and *llz* probes, suggesting that both genes are expressed in the same cells.

To identify cells expressing the locus, we drove a GFP reporter with a *CheB42a/llz* promoter by using the UAS/GAL4 binary system (25). In the heads of third-instar larvae, we detected a class II multidendritic (mechanosensitive) neuron (26) and two external sensory neurons that project to the chemosensitive terminal organ (Fig. 5*H*). These data are consistent with data from the embryos and suggest that the locus might play a sensory role.

Discussion

Prokaryotes and nematodes use operon-based transcription to coordinately regulate genes participating in common biological pathways. Our data indicate that this elegant process also occurs in *Drosophila*. Although we have investigated only one gene pair in detail and obtained less data for two other gene pairs, our
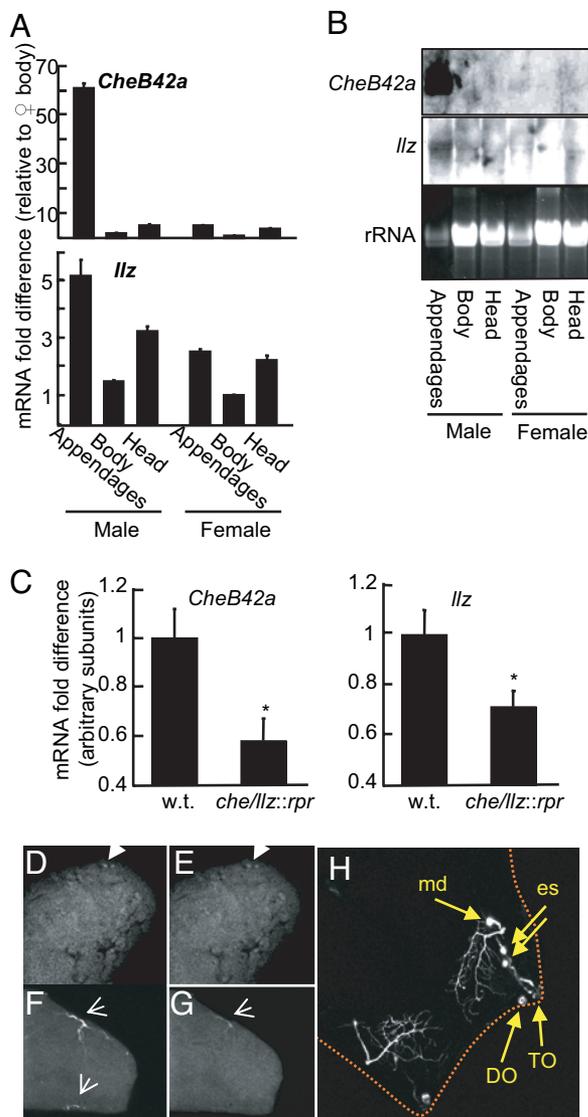
GENETICS

Fig. 5. Both *CheB42a* and *llz* transcripts are expressed preferentially in male-specific sensory structures. (*A*) Real-time quantitative RT-PCR. Transcripts of both *CheB42a* and *llz* were enriched in male appendages relative to other body parts or females by ANOVA ($P < 0.001$; $n = 6$ per group). Data are mean ± SEM. (*B*) Northern blot analysis of total RNA from adult tissues. The same membrane was probed for both genes with stripping between hybridizations. The blot of *llz* required much longer exposure of the membrane, suggesting lower-level expression than *CheB42a*. Data for male appendages are the same as those shown in Fig. 2*A*. (*C*) Genetic elimination of cells by expressing the cell death-inducer *rpr* under the *CheB42a/llz* promoter reduced abundance of both *CheB42a* and *llz* transcripts in male appendages (ANOVA $P < 0.02$; $n = 4$ per group). (*D–G*) *In situ* hybridization for *CheB42a* (*D*) and *llz* (*E*) in embryos. Individual embryos were hybridized simultaneously with *CheB42a* and *llz* antisense probes labeled with two different fluorophores or with the corresponding sense probes. The front of embryo is in upper right. Arrowheads indicate location of signal, likely in neurons. (*F* and *G*) *CheB42a* and *llz* sense controls. Arrows in *F* and *G* indicate nonspecific fluorescence in trachea. (*H*) The *CheB42a-llz* locus is expressed in third-instar larvae in a single lateral multidendritic neuron (md) plus two external sensory neurons (es) that project to the chemosensitive terminal organ (TO). Dorsal organ (DO), which shows autofluorescence, also is indicated. Larva is outlined by dotted line.

preliminary analysis suggests that the fly genome contains numerous gene pairs that may share a chromosomal arrangement like that of *CheB42a/llz*. It is likely that some of these also represent operon-like loci, an occurrence that would be consis-

tent with the abundance of operons in the *C. elegans* genome. That genome has >15% of its coding genes transcribed as operons, many with already known functional relationships (7).

Bacterial and archaeal operons consist of a single RNA species from which multiple proteins are translated. Nematode operons also transcribe a pre-mRNA from a single promoter, but the primary transcript then is processed to individual mature transcripts for each gene (27). Processing involves trans splicing a specific short RNA sequence to each 5′ UTR (28). In fact, it was recognition of the specialized spliced leader (SL2) RNA attached to downstream transcripts that led to the discovery that *C. elegans* genomes contained operons (28). As in nematode operons, in the *Drosophila CheB42a/llz* locus, a single promoter produces a pre-mRNA that is converted posttranscriptionally to individual mRNAs. But *Drosophila* does not seem to exhibit SL RNA trans splicing, and, thus, the mechanism is distinct. Therefore, we speculate that processing of operons may have evolved separately in nematodes and *Drosophila*.

Our data provide some insight into how mature *CheB42a* and *llz* transcripts may be generated. Splitting of the pre-mRNA occurs after transcription. As in *C. elegans* operons, the poly(A) sequence and sequences in the intergenic region are important for production of both mature upstream and downstream transcripts. Especially interesting is our finding that the 5′ end of the downstream transcript appears to be formed without a conventional 5′ cap. Instead, we were surprised to find that *llz* transcripts contain a 5′ guanosine not encoded by DNA. We also found this feature on some transcripts from other potential operons; it may provide an important clue to discover how the 5′ end is formed. Thus, we speculate that the *CheB42a/llz* locus may use a nontraditional capping mechanism for the downstream gene. Interestingly, novel capping-like mechanisms recently have been described for other eukaryotic genes (29), and cap-independent translation has been reported in eukaryotes (16).

An operon-like organization for the *CheB42a/llz* locus suggested a functional relationship between the gene products. Consistent with this suggestion, *CheB42a* and *llz* showed a similar pattern of expression. Moreover, they both seem to be involved in male courtship behavior (5, 30, 31), and the two proteins interact (Y.B.-S. and M.J.W., unpublished data).

Recognition of operon-based transcription in *Drosophila* may allow a better understanding of its chromosomal organization and transcriptome. We also speculate that genomes of other eukaryotic species may contain similar genome-level architecture to coregulate gene expression. Finally, identification of operons in *Drosophila* and other higher eukaryotes might provide a new way to identify genes contributing to the same biochemical process or physiological pathway. As an example, the protein products of *CheB42a* and *llz* would never have been predicted to interact without the recognition of the operon-like genomic organization.

Materials and Methods

***Drosophila* Stocks and Cultures.** See *SI Text* for more information.

RNA Analysis. Flies were separated by sex under CO_2 and kept at −80°C until processing. To separate body parts, microcentrifuge tubes with flies were dipped in liquid nitrogen and then separated by repeated vortexing. *Drosophila* S2 cells were transfected with the Effectene reagent (Qiagen, Valencia, CA) according to manufacturer instructions. Total RNA from tissues or cells was extracted with the RNeasy mini kit (Qiagen) or TRizol reagent (Invitrogen, Carlsbad, CA) according to manufacturer instructions. RT-PCR analysis was performed by using the SuperScript II reverse transcriptase (Invitrogen) with 1 μg of total RNA in a 20-μl reaction according to manufacturer instructions. PCRs were performed with the ACCUprime *pfx* supermix (Invitrogen) in 25-μl reactions and were subsequently separated on a 1.2%

agarose gel. Northern blots were performed with the Northern Max kit (Ambion, Austin, TX) and probed with DIG-labeled riboprobes (Roche, Indianapolis, IN). The same membranes were used for both probes and were stripped between hybridizations. Real-time quantitative RT-PCR assays were performed on an "ABI7000" or an "ABI7500 fast" with SYBR Green or PowerSybr as a reporter according to manufacturer instructions. For analysis, expression levels of the housekeeping gene *rp49* were used as an RNA-loading control. Data were transformed according to the $\Delta \Delta$ Ct method and are represented as relative values (32). Fold difference was calculated relative to the lowest-expressing sample, which represented 1 unit of expression. Primer sequences for all genes analyzed are contained in SI Table 2. The 5′ RACE analysis for capped transcripts was performed with the RLM-RACE kit (Ambion). 5′ RACE of noncapped mRNA was performed by using the cRACE protocol (33). Polysomal RNA extraction was performed as described in ref. 15 with the following modifications: Cells were treated with cyclohexamide for 15 min before lysis, and fractionated RNA was isolated with the TRIzol reagent (Invitrogen). Precipitation of capped mRNAs was performed with H2O anti-cap monoclonal antibody (Synaptic Systems) according to manufacturer's protocol. As a control, we incubated an aliquot of the same RNA with an unrelated antibody of identical isotype (IgG1). Both IPs were performed with protein-G Sepharose (Pierce) according to manufacturer's instructions. *In situ* hybridization analysis was performed as described in ref. 34; fluorescent probes were produced with the FISH-Tag kit (Invitrogen). Individual embryos were hybridized with a mixture of AlexaFluor 488-labeled *CheB42a* antisense and AlexaFluor 555-labeled *llz* antisense probes. For controls, individual embryos were labeled with the corresponding sense probes.

Operon Processing in S2 Cells. The genomic region spanning all of the *CheB42a/llz* exons was PCR-amplified with ACCUprime *pfx* supermix (Invitrogen) and was subcloned into the metal-inducible pMT-V5-HIS vector (Invitrogen). The *CheB42a* transcriptional start site was determined according to the 5′ RACE data. Cells were transfected transiently with Effectene Transfection Reagent (Qiagen) according to manufacturer instructions. RNA was harvested from cells (RNeasy Mini Kit; Qiagen) after they were induced with 500 μM CuSO$_4$ for 24 h.

Bioinformatics. See *SI Text* for more information.

1. Bianchi L, Driscoll M (2002) *Neuron* 34:337–340.
2. Krishtal O (2003) *Trends Neurosci* 26:477–483.
3. Welsh MJ, Price MP, Xie J (2002) *J Biol Chem* 277:2369–2372.
4. Liu L, Leonard AS, Motto DG, Feller MA, Price MP, Johnson WA, Welsh MJ (2003) *Neuron* 39:133–146.
5. Xu A, Park SK, D'Mello S, Kim E, Wang Q, Pikielny CW (2002) *Cell Tissue Res* 307:381–392.
6. Blumenthal T, Gleason KS (2003) *Nat Rev Genet* 4:112–120.
7. Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK (2002) *Nature* 417:851–854.
8. Lawrence JG (2002) *Cell* 110:407–413.
9. Lercher MJ, Blumenthal T, Hurst LD (2003) *Genome Res* 13:238–243.
10. Lawrence JG (2002) *Cell* 110:407–413.
11. Kuersten S, Lea K, MacMorris M, Spieth J, Blumenthal T (1997) *RNA* 3:269–278.
12. Andrews J, Smith M, Merakovsky J, Coulson M, Hannan F, Kelly LE (1996) *Genetics* 143:1699–1711.
13. Kitamoto T, Wang W, Salvaterra PM (1998) *J Biol Chem* 273:2706–2713.
14. Woodson SA (2005) *Curr Opin Struct Biol* 15:324–330.
15. Meyuhas O, Bibrerman Y, Pierandrie-Amaldi P, Amaldi F (1996) in *A Laboratory Guide to RNA*, ed Kreig PA (Wiley–Liss, New York), pp 65–81.
16. Merrick WC (2004) *Gene* 332:1–11.
17. Hoehne M, de Couet HG, Stuermer CA, Fischbach KF (2005) *Mol Cell Neurosci* 30:326–338.
18. Morabito MA, Sheng M, Tsai LH (2004) *J Neurosci* 24:865–876.
19. Han P, Dou F, Li F, Zhang X, Zhang YW, Zheng H, Lipton SA, Xu H, Liao FF (2005) *J Neurosci* 25:11542–11552.
20. Girardot N, Allinquant B, Langui D, Laquerriere A, Dubois B, Hauw JJ, Duyckaerts C (2003) *Neuropathol Appl Neurobiol* 29:451–461.
21. Huang T, Kuersten S, Deshpande AM, Spieth J, MacMorris M, Blumenthal T (2001) *Mol Cell Biol* 21:1111–1120.
22. Liu Y, Huang T, MacMorris M, Blumenthal T (2001) *RNA* 7:176–181.
23. MacDonald CC, Redondo, J-L (2002) *Mol Cell Endocrinol* 190:1–8.
24. White K, Tahaoglu E, Steller H (1996) *Science* 271:805–807.
25. Phelps CB, Brand AH (1998) *Methods* 14:367–379.
26. Grueber WB, Jan LY, Jan YN (2003) *Cell* 112:805–818.
27. Hastings KEM (2005) *Trends Genet* 21:240–247.
28. Spieth J, Brooke G, Kuersten S, Lea K, Blumenthal T (1993) *Cell* 73:521–532.
29. Nielsen H, Westhof E, Johansen S (2005) *Science* 309:1584–1587.
30. Lin H, Mann KJ, Starostina E, Kinser RD, Pikielny CW (2005) *Proc Natl Acad Sci USA* 102:12831–12836.
31. Park SK, Mann KJ, Lin H, Starostina E, Kolski-Andreaco A, Pikielny CW (2006) *Curr Biol* 16:1154–1159.
32. Ben-Shahar Y, Robichon A, Sokolowski MB, Robinson GE (2002) *Science* 296:741–744.
33. Maruyama IN, Rakow TL, Maruyama HI (1995) *Nucleic Acids Res* 23:3796–3797.
34. Kosman D, Mizutani CM, Lemons D, Cox WG, McGinnis W, Bier E (2004) *Science* 305:846.

GENETICS