

The TAGteam DNA motif controls the timing of *Drosophila* pre-blastoderm transcription

John R. ten Bosch, Joseph A. Benavides and Thomas W. Cline*

The *Drosophila* sex-determination switch gene *Sex-lethal* (*Sxl*) and the X-chromosome signal element genes (XSEs) that induce the female-specific expression of *Sxl* are transcribed extremely early in development when most of the genome of this organism is still silent. The DNA sequence CAGGTAG had been implicated in this pre-cellular blastoderm activation of sex-determination genes. A genome-wide computational search, reported here, suggested that CAGGTAG is not specific to early sex-determination genes, since it is over-represented upstream of most genes that are transcribed pre-cellular blastoderm, not just those involved in sex determination. The same search identified similarly over-represented, one-base-pair degenerate sequences as possible functional synonyms of CAGGTAG. We call these heptamers collectively, the TAGteam. Relevance of the TAGteam sequences to pre-cellular blastoderm transcription was established through analysis of TAGteam changes in *Sxl*, *scute* (an XSE), and the 'ventral repression element' of the pattern-formation gene *zerknüllt*. Decreasing the number of TAGteam sites retarded the onset of pre-blastoderm transcription, whereas increasing their number correlated with an advanced onset. Titration of repressors was thought to be the rate-limiting step determining the onset of such early transcription, but this TAGteam dose effect shows that activators must also play an important role in the timing of pre-blastoderm gene expression.

KEY WORDS: *Drosophila melanogaster*, Transcription, DNA motif, Midblastula transition, Enhancer, Temporal regulation

INTRODUCTION

Most animal embryos begin zygotic transcription only after a considerable delay following fertilization (Andeol, 1994; Davidson, 1986). For example, widespread activation of zygotic transcription in *Drosophila melanogaster* occurs only midway through nuclear cycle 14, some 2.5 hours after fertilization, during the cellular blastoderm stage, when division slows and the egg cell membrane invaginates to engulf peripheral somatic nuclei and form a cellularized embryo (Anderson and Lengyel, 1979; Foe and Alberts, 1983; Lamb and Laird, 1976; McKnight and Miller, Jr, 1976). The fact that the nucleo-cytoplasmic ratio influences the time of onset of general zygotic transcription in diverse species led to the idea that this delay in the onset of zygotic gene expression is a consequence of the action of maternally loaded repressor proteins that are titrated by zygotic DNA (Newport and Kirschner, 1982a; Newport and Kirschner, 1982b; Pritchard and Schubiger, 1996).

Even for the frog, however, where the concept of a 'midblastula transition' marking the onset of widespread zygotic transcription has perhaps been most thoroughly developed, a few genes are known to be transcribed earlier (Kimelman et al., 1987; Nakakura et al., 1987). The same is true for *Drosophila*, although even the earliest fly genes are not transcribed until 1 hour after fertilization (see Fig. 1). Remarkably little is known about the features of these unusually early genes that allow them to be expressed at a time when most of the genome is silent. Previous studies suggested that the time of onset of transcription for even these exceptional genes is determined by the disappearance of repressors that are titrated by DNA, ones presumably different from those proposed to suppress the majority of zygotic genes (Brown et al., 1991; Pritchard and Schubiger, 1996). Work we report here adds a new perspective on this process

by allowing us to infer that the binding of positively acting factors to specific sites near the promoters of such exceptionally early genes must also be an important factor in determining precisely when these genes become active.

Our attention was drawn to pre-cellular blastoderm (pre-CB) transcription by the fact that *Drosophila* sex-determination genes are among the handful of genes expressed so early. *Drosophila* sex and X-chromosome dosage compensation are determined by the number of X chromosomes, which are counted by the feminizing switch gene *Sex-lethal* (*Sxl*) prior to the cellular blastoderm stage (reviewed by Cline and Meyer, 1996). *Sxl* counts by measuring the level of zygotic gene products generated from at least four X-chromosome signal elements (XSEs): *sisterlessA* (*sisA*), *scute* (*sc*, a.k.a. *sisB*), *outstretched* (*os*, a.k.a. *upd* or *sisC*) (Sefton et al., 2000) and *run* (*run*). Diplo-X (female) embryos, but not haplo-X (male) embryos, generate a level of XSE products high enough to activate transcription at the *Sxl* 'establishment' promoter, *Pe*, in somatic cells, thereby producing a pulse of SXL protein. This female-specific protein pulse subsequently engages a positive feedback loop on *Sxl* pre-mRNA splicing that locks *Sxl* into its feminizing expression mode throughout the rest of development.

By counting X-chromosomes in nuclear cycle 12, embryos engage their dosage compensation machinery in time to avoid an imbalance in X-linked gene products that would otherwise arise between the sexes during cycle 14 when genome-wide transcription begins (Gergen, 1987; Tracey et al., 2000). But such early chromosome counting demands that expression of the genes that communicate X-chromosome dose to *Sxl* must begin even before cycle 12. As Fig. 1 illustrates, the two strongest XSEs, *sisA* and *sc*, are among the earliest expressed genes (Erickson and Cline, 1993).

The question of whether these sex-determination genes might have something in common that allows them to be expressed so early led us to the heptanucleotide motif, CAGGTAG. Three of the four XSEs in *D. melanogaster* (all but *run*) possess multiple copies of this sequence or its reverse complement within 500 bp of their transcription start sites (Erickson and Cline, 1998; Sefton et al.,

Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720-3204, USA.

*Author for correspondence (e-mail: sxcline@berkeley.edu)

2000). The cluster of three CAGGTAG sites upstream of *sc* was shown to be functionally significant by the demonstration that their elimination reduced *sc* XSE activity, and abolished it when combined with a deletion of a downstream regulatory element (Wrischnik et al., 2003).

We have used genome-wide computational analysis and species comparisons to show that CAGGTAG clustering is not unique to sex-determination signal genes, but is also found upstream of other genes whose transcription begins prior to cycle 14. 1-bp degenerate heptamers are identified that cluster with, and can probably substitute for CAGGTAG. We call these sequences the TAGteam. Functional analysis of TAGteam clusters in transgenic reporter constructs for *sc*, *Sxl_{pe}* and the patterning gene *zerknüllt* (*zen*) revealed that TAGteam site number influences the time of onset of pre-CB transcription.

MATERIALS AND METHODS

Fly stocks and culture

Drosophila were cultured in uncrowded conditions on a standard cornmeal, yeast, sucrose and molasses medium. Transgenes were tested for *sisB*⁺ activity only after five generations of backcrossing to a standard *sc*^{M6} stock to generate homogeneous genetic backgrounds. Wild-type *Sxl_{pe}* transgene lines were provided by J. Erickson, wild-type VRE600 transgene lines by M. Levine, and *hermaphrodite* alleles by B. Baker. Markers and balancers are described at <http://flybase.bio.indiana.edu>.

Bioinformatics

We created a simple algorithm to search for CAGGTAG clusters in the *D. melanogaster* genome (Adams et al., 2000). When the program FLY ENHANCER (Markstein et al., 2002) became available, we determined that it generated the same results. Pre-CB genes were chosen as genes transcribed prior to nuclear cycle 14, according to previously published reports in the literature (Fig. 1), and post-CB genes (Table S1 in supplementary material) were randomly chosen among those known not to be transcribed prior to gastrulation according to the BDGP expression database (Tomancak et al., 2002). For pre- and post-CB genes, the total number of occurrences of CAGGTAG and related heptamer sequences per interval was subjected to a χ^2 test incorporating Yate's correction. The same analysis was applied to the orthologous genes in *D. pseudoobscura* (Richards et al., 2005).

Identification of orthologous sequences and alignments

Sequences in *D. yakuba*, *D. ananassae*, *D. pseudoobscura*, *D. mojavensis* and *D. virilis* were downloaded from the Lawrence-Berkeley National Lab Vista Browser (Frazer et al., 2004). Promoter sequences for *bottleneck* (*bnk*) from *D. willistoni*, *D. hydei* and *Zaprionus tuberculatus* were obtained by PCR amplification with conserved *D. pseudoobscura* primers: (1) GGTGCGCGGAAAACACGTAATAATTCGCGTG; (2) GTGTTGGTC-AGCTTGTGAAGAAGTTGATTTTGTG. Sequences were aligned using ClustalW (Thompson et al., 1994), then manually adjusted to align motifs identified with MEME (Bailey and Gribskov, 1998).

Site-directed mutagenesis and germline transformation

Standard techniques were used for germline transformation (Spradling, 1986). All mutagenesis reactions were carried out with the Quick-Change site-directed mutagenesis kit (Stratagene). PCR reactions were performed under standard conditions and all products were sequence verified prior to subcloning into *P*-element vectors. Mutagenic primers used are listed below, with their altered TAGteam sites underlined and the base pairs in bold indicating changes from the wild-type site (shown in parentheses).

sc: GATTCTCTCGCCCGTTTCTCGGGAACAGTGTGTGCAGAGAGTTG (CTGCCTG);

Sxl_{pe}: (1) GATCAATTTCGGGATCTTCGTTCTCCACGCGACTGGCAC (TAGGTAG);

(2) CGAGAATGCGGAACATCTGAAGTACGGCGAAGATCCGCGCATCC (CTGCCTGCCTG), (3) GCCTCCTTCGATCTTAGAACGGG-CACCCAGCCACCGC (CTACCTG);

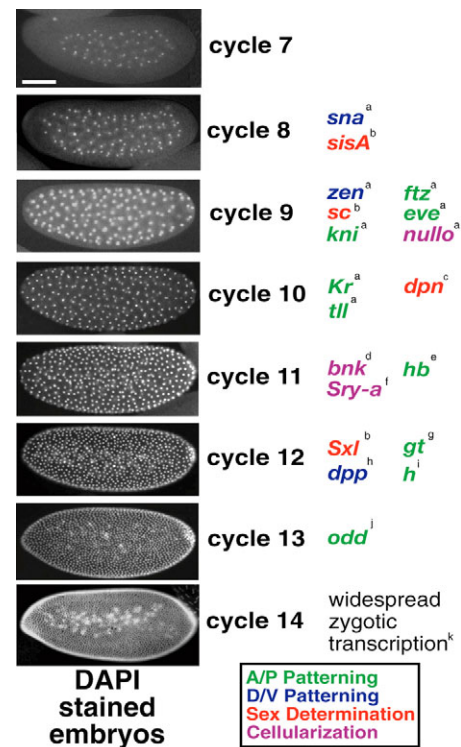


Fig. 1. Sex-determination genes are among the handful of *Drosophila* genes known to be transcribed prior to nuclear cycle 14, the cellular blastoderm stage. Pre-cycle 14 (pre-CB) genes whose precise time of onset of transcription is known to within one nuclear cycle are ordered with respect to that parameter. The biological function of their pre-CB expression is also indicated. Pre-CB genes for which less precise information on timing is available include *outstretched* (Harrison et al., 1998), *huckebein* (Bronner et al., 1994), *runt* (Klingler and Gergen, 1993), *twist* (Thisse et al., 1988), *brinker* (Jazwinska et al., 1999) and *short gastrulation* (Markstein et al., 2002). Scale bar: 100 μ m. ^a(Pritchard and Schubiger, 1996); ^b(Erickson and Cline, 1993); ^c(Barbash and Cline, 1995); ^d(Schejter and Wieschaus, 1993); ^e(Tautz et al., 1987); ^f(Ibnsouda et al., 1993); ^g(Eldon and Pirrotta, 1991; Kraut and Levine, 1991); ^h(St Johnston and Gelbart, 1987); ⁱ(Ingham et al., 1985); ^j(Coulter et al., 1990); ^k(Anderson and Lengyel, 1979; Lamb and Laird, 1976).

zen: (1) CATTTCGACCAGCGGACGGTGTGTTTATTCACCGAACGG-AAACCCATAC (CTGCCTG-CTACCTG);

(2) CGGTCGCACTATTCGTTTCGACACTGTACCCTCCGCACT-AGCGGG (TAGGTAG-CAGGCA).

For *scute*, a 1.65-kb *XhoI*-*Bam*HI genomic fragment in pBluescript (Stratagene) was used as a template for mutagenesis. Mutants were subcloned into the *scute* genomic DNA fragment described by Wrischnik et al. (Wrischnik et al., 2003). For *Sxl*, wild-type 1.4-kb *Sxl_{pe}*-*lacZ* vector (Yang et al., 2001) was digested with *Eco*RI-*Not*I and subcloned into pBluescript for use as a template to make mutants, which were subsequently cloned back into the original vector using the same digest. For *zen*, mutagenesis of VRE TAGteam sites was performed in a modified pGEM72f(-) vector, pGEM-BGS, in which the sequence from *Bsp*EI to *Kpn*I was replaced with *Bgl*III-600VRE-*Spe*I (a gift from M. Markstein). Following mutagenesis reactions, the mut600VRE pGEM-BGS was digested with *Bgl*III-*Spe*I and cloned into newE2G, a gypsy-insulated pCaSpeR vector containing an *eve* minimal promoter fused to a *lacZ* reporter (Markstein et al., 2004).

To generate the 250VRE transgene, the 600VRE in pGEM-BGS was digested with *Kas*I-*Hind*III, incubated with Klenow polymerase (New England Biolabs) to create blunt ends, and self-ligated. The following

primers were used to amplify the wild-type or mutant 111-bp fragment from the appropriate 600-bp VRE plasmid (lowercase sequences denote exogenous DNA added to introduce *Bam*HI or *Bgl*II sites):

- (1) atgcatcTTTGCACCAGCTGCCTGTGTTTA,
- (2) taagatctCCCGCTAGTGCTGCCTGTACAG,
- (3) atgcatcTTTGCACCAGCGGACGGTGTGTTA,

(4) taagatctCCCGCTAGTGCGGACGGTGTGTTA. The resulting products were digested with *Bam*HI-*Bgl*II and cloned into a modified pBluescript SK(+) that contains *Bam*HI/*Bgl*II sites at the original *Bam*HI/*Sma*I sites (a gift from M. Markstein). The 111VRE fragment was duplicated by digesting the 111VRE vector with *Bam*HI-*Sca*I or *Bgl*II-*Sca*I and ligating the VRE-containing vector fragments from the two different digests to each other. All VRE fragments were ligated into newE2G *Bgl*II-*Eco*RI sites following digestion with *Bam*HI-*Eco*RI except the 250VRE, which was cut with *Bgl*II-*Eco*RI prior to ligation.

In situ hybridization and nuclear dots

0- to 2-hour *D. melanogaster* embryos were collected at 25°C on standard molasses/agar plates smeared with live yeast paste. Embryo fixation and hybridization was performed with digoxigenin-labeled antisense *lacZ* RNA probes as described previously (Jiang et al., 1991; Tautz and Pfeifle, 1989), except for embryos used for nuclear dot analysis, which were stained with 0.5 µg/ml DAPI, washed with PBT, and mounted in 70% glycerol/1× PBS following alkaline phosphatase staining. Embryonic nuclear cycles were determined by the density of nuclei and their position relative to the periphery of the embryo.

We only counted dots in late interphase or prophase nuclei, since these stages appeared to have the greatest propensity for dot expression. For nuclear cycles 9-13, at least 80 nuclei per embryo were scored. For earlier cycles, all visible nuclei were scored. We defined the time of onset of transcription conservatively to be the first cycle with at least a 5-fold increase in the percentage of expressing nuclei over a previous cycle with at least 1% expressing nuclei.

Immunocytochemistry

Embryos grown at 25°C were stained with β-galactosidase antibody (5 Prime-3 Prime) and processed as described by Kuo et al. (Kuo et al., 1996), except that alkaline phosphatase-conjugated goat anti-rabbit secondary antibody was used and color was developed with 5-bromo-4-chloro,3-indolyl phosphate (BCIP)/nitroblue tetrazolium (NBT).

RESULTS

TAGteam sequences are over-represented upstream of genes transcribed pre-cellular blastoderm (pre-CB)

To determine whether CAGGTAG clusters are unique to genes that regulate *Sxl* in the early embryo, we wrote a computer algorithm to search the *D. melanogaster* genome for XSE-like CAGGTAG clusters containing three or more matches to CAGGTAG, or its reverse complement, within 500 bp upstream of a transcription start site. Only five genes had such clusters. Whereas three of them were XSEs, the other two were *bnk* and *Neurotactin* (*Nrt*), both autosomal genes with functions unrelated to sex determination, but that appear to share the feature of pre-CB transcription with XSEs (Lecuit and Wieschaus, 2000; Schejter and Wieschaus, 1993). Pre-CB transcription of *bnk* has been shown directly, but that for *Nrt* was inferred from immunostaining of its protein products in early cycle 14.

If CAGGTAG functions more generally to regulate pre-CB gene expression, perhaps in combination with other sequences, one might expect it to be over-represented in the promoters of genes expressed prior to the cellular blastoderm stage, even if it did not always appear in clusters. Moreover, CAGGTAG-related sequences might also be over-represented. To explore these possibilities, we compared the occurrence of CAGGTAG and all its 1-bp degenerate sequences in the region flanking the transcription start sites of genes expressed pre-CB to that for genes expressed post-CB (see Materials and

methods). CAGGTAG is indeed over-represented in the 500-bp region upstream of pre-CB genes relative to post-CB genes ($P=6.3 \times 10^{-7}$, Fig. 2A), even if *Sxl*_{pe} and the genes from the genomic cluster search are removed from the pre-CB set to avoid a possible ascertainment bias ($P=2.1 \times 10^{-4}$, data not shown). The degenerate sequence tAGGTAG, that had been suggested as a possible alternative to CAGGTAG in an earlier comparison of *Sxl* promoters (Erickson and Cline, 1998), is nearly as over-represented as CAGGTAG ($P=1.6 \times 10^{-6}$). The sequence CAGGcAG ($P=9.2 \times 10^{-4}$) also emerged in this analysis as the only other 1-bp degenerate with a *P*-value less than 0.01. Hereafter, we refer collectively to these over-represented heptamers as TAGteam sequences.

To determine whether the over-representation of TAGteam heptamers upstream of pre-CB genes is unique to *D. melanogaster*, we performed a similar analysis on *D. pseudoobscura* orthologs. Again, CAGGTAG ($P=1.5 \times 10^{-6}$) and tAGGTAG ($P=2.0 \times 10^{-5}$) were over-represented in the interval immediately upstream of pre-CB genes relative to post-CB genes, but CAGGcAG was not (Fig. 2B). CAGGTAG and/or tAGGTAG were also over-represented in regions beyond 500 bp in the two species (Fig. 2A,B), though less so than in regions closer to the promoter, suggesting that TAGteam members may be able to exert influence over distances greater than 500 bp.

This relative over-representation of TAGteam sequences in the interval directly upstream of the transcription start site was a consequence of pre-CB genes containing more sites than expected by chance, although the values for CAGGTAG in *pseudoobscura* and for tAGGTAG in both species were also increased by a depression in the number of sites upstream of post-CB genes (Fig. 2C,D). In contrast, although CAGGcAG in *D. pseudoobscura* was six times more prevalent than expected upstream of pre-CB genes, a strong positive bias among post-CB genes as well dropped this heptamer below the over-representation cutoff. An extension of the analysis of pre-CB genes shown in Fig. 2C to four more *Drosophila* species, two of which are more distant than *pseudoobscura*, showed that CAGGTAG was reliably the most over-represented of the three TAGteam members, and that the other two were consistently among the top three runners up among all 1-bp degenerates (Table S2 in supplementary material).

While this manuscript was under review, a genome-wide microarray-based developmental profile of relative poly(A)⁺ mRNA levels in *D. melanogaster* prior to gastrulation was published (Pilot et al., 2006) which gave us an opportunity to expand the data set for pre-CB genes in the analysis of TAGteam sequence occurrence. Our analysis of this much larger and differently biased new data set (Fig. S1 in supplementary material) showed remarkable agreement with our conclusions from Fig. 2A. Again the three TAGteam heptamers described above stood out, with CAGGTAG leading the pack, but they were now joined by two additional 1-bp degenerate sequences, CAGGTAA and CAGGTAT that rose far above the $P=0.01$ significance threshold. CAGGTAA had been just below that threshold in our original analysis, and was consistently among the top five most over-represented 1-bp degenerates in the six-species comparison of pre-CB genes (Table S2 in supplementary material).

Conservation of individual TAGteam sites is variable among early gene promoters

To explore the pattern of conservation of TAGteam sites in a pre-CB gene unrelated to sex determination, we performed an alignment of *bnk* sequence from nine species within the family *Drosophilidae* for the region in *melanogaster* that contains the triple CAGGTAG

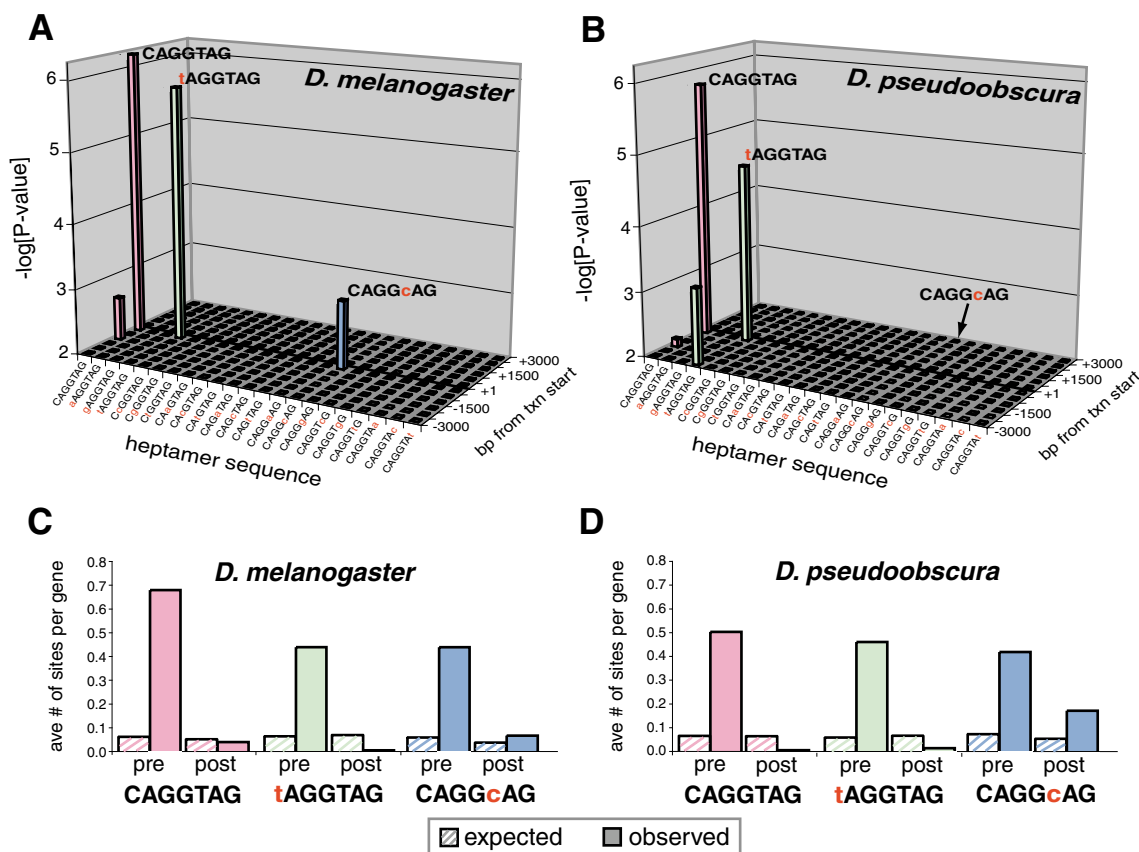


Fig. 2. Identification of the TAGteam group of heptanucleotide sequences based on their relative over-representation upstream of pre-cellular-blastoderm genes. (A,B) Probability that the observed difference between pre-CB and post-CB genes with regard to the number of heptamer sites found within a 500-bp window at the specified distance from the transcription (txn) start site (+1) is the result of chance. The baseline cutoff is $P=0.01$ ($-\log[0.01]=2$). (C,D) Comparison of the same pre-CB and post-CB genes with respect to the observed number of TAGteam sequences within 500 bp upstream of their transcription start sites versus the number expected based on %GC for the set. Gene sets are defined in Materials and methods (*melanogaster*: $n=25$ pre-CB versus 78 post-CB; *pseudoobscura* orthologs: $n=24$ pre-CB versus 70 post-CB).

cluster and, just upstream of that, a tAGGTAG sequence (Fig. 3A,B). Support for the functionality of the latter sequence was strengthened by the observation that each of the *melanogaster* CAGGTAG sites was replaced by tAGGTAG in at least one species. Moreover, the upstream tAGGTAG sequence was conserved among the three closest relatives of *melanogaster*.

Although the overall number of TAGteam sites in pre-CB genes was relatively conserved (Fig. 2C,D; Table S2 in supplementary material), the conservation of specific sites and overall number within *bnk* was somewhat variable (Fig. 3A,B). The lack of conservation of the distal *melanogaster* TAGteam sites in species beyond *pseudoobscura* appears to be compensated for at least in part by the appearance of TAGteam sites elsewhere: *D. willistoni* and *Z. tuberculatus* evolved a new CAGGTAG site between the two most highly conserved CAGGTAG sites, whereas *D. mojavensis* acquired a CAGGTAG sequence upstream of all others.

To determine if the findings for *bnk* were representative, we used the genome sequence available for six *Drosophila* species to align the 500-bp regions upstream of the promoters of five additional pre-CB genes (Fig. 3C). Three of these genes are related to sex determination (*sisA*, *sc*, *Sxl_{PE}*), but the other two [*snail* (*sna*) and *zen*] are not. For *sna* and *zen*, known enhancer elements further upstream of these genes were also examined, since they contained TAGteam clusters in *melanogaster*. As with *bnk*, these regions exhibited a variable degree of conservation. There was no obvious difference in conservation

between genes involved in sex determination and those that are not. The fact that TAGteam sites in the upstream enhancer elements of *zen* and *sna* seemed as conserved as those closer to the promoter argues for the functional significance of TAGteam site over-representation (Fig. 2A,B) regardless of its distance from the promoter.

There were several instances in the five-gene/six-species comparison in which de novo TAGteam sites appeared to replace lost sites, suggesting that stabilizing selection may influence TAGteam evolution (see Ludwig et al., 2000; Ludwig et al., 1998). In view of our recent analysis of the Pilot et al. (Pilot et al., 2006) data pointing to CAGGTAA as a member of the TAGteam, it seems significant that in the species comparisons shown in Fig. 3, this sequence replaced one of the conserved original three TAGteam heptamers even more frequently than any one of them replaced the two others (data not shown).

The least over-represented TAGteam member does contribute to pre-CB gene function

The functional significance of CAGGTAG clusters had previously been established using an assay for the XSE function of *scute* (*sc*) (Wrischnik et al., 2003). Since *sc* has a CAGGcAG site 53 bp upstream of its transcription start site (Fig. 4A), we could utilize the same assay to determine whether this least-over-represented member of the original TAGteam also contributes to pre-CB gene functioning.

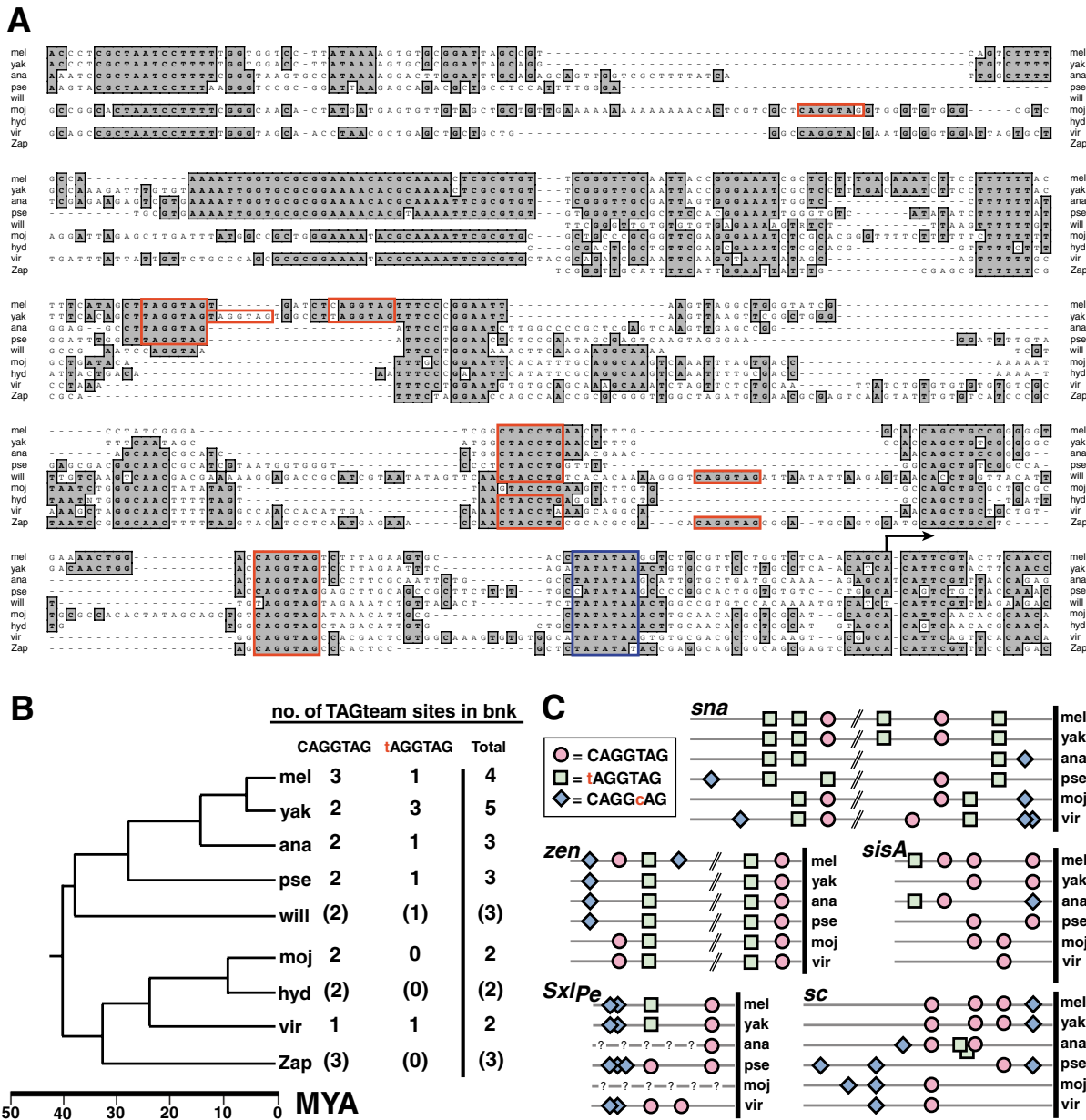


Fig. 3. Conservation of TAGteam sites in pre-cellular blastoderm genes is variable. (A) Alignment of *bnk* promoter sequences for *Drosophila melanogaster* (mel), *yakuba* (yak), *ananassae* (ana), *pseudoobscura* (pse), *willistoni* (will), *mojavensis* (moj), *hydei* (hyd), *virilis* (vir) and *Zaprionus tuberculatus* (Zap). Grey shading indicates at least 50% identity among nucleotides at sites present in at least two species. The TATA site is boxed in blue, while TAGteam sites are boxed in red. An arrow denotes the transcription start site. (B) Phylogenetic relationships, estimated divergence times (Kwiatowski et al., 1994; Powell and DeSalle, 1995; Russo et al., 1995), and number of TAGteam sites in *bnk* for the nine species. Parentheses indicate values from species for which less DNA sequence was available. (C) Schematic of a five-species alignment for the promoter-proximal upstream ~400 bp of five pre-CB genes, and for TAGteam-rich enhancer elements further upstream in two cases (distances not to scale). Circles represent CAGGTAG sites, squares tAGGTAG and diamonds CAGGcAG. Overlapping symbols indicate overlapping motifs. In the rare event that a TAGteam site appeared to be replaced by its reverse complement, that complement was diagrammed as a separate site.

This assay for *sc* XSE activity is based on the fact that females homozygous for the null allele *sc*^{M6} fail to activate *Sxl*_{Pe} and therefore die from dosage compensation upsets. A genomic transgene containing the *scute* protein coding region and only a few kb of DNA on either side (Fig. 4A) fully complements *sc*^{M6} with respect to female-specific lethality (Fig. 4B). Consequently, the extent to which mutations interfere with the pre-CB functioning of this transgene is reflected in the extent to which they reduce the ability of the transgene to rescue *sc*^{M6} mutant females.

Mutating the single CAGGcAG site by itself had no effect on the ability of this transgene to rescue mutant females (101% viability, relative to that of the heterozygotes); however, rescue was reduced to only 3% in combination with the triple CAGGTAG knockout, a value far below the 35% average rescue for the triple knockout alone (Fig. 4B). We could be confident of the biological significance of this synergism, even in the face of considerable variation in rescue among insertion lines, since values for individual transgene lines were reproducible, and

since the range of rescue values for the various quadruple knockout lines (1-5%) did not overlap that for the triple knockouts (7-76%).

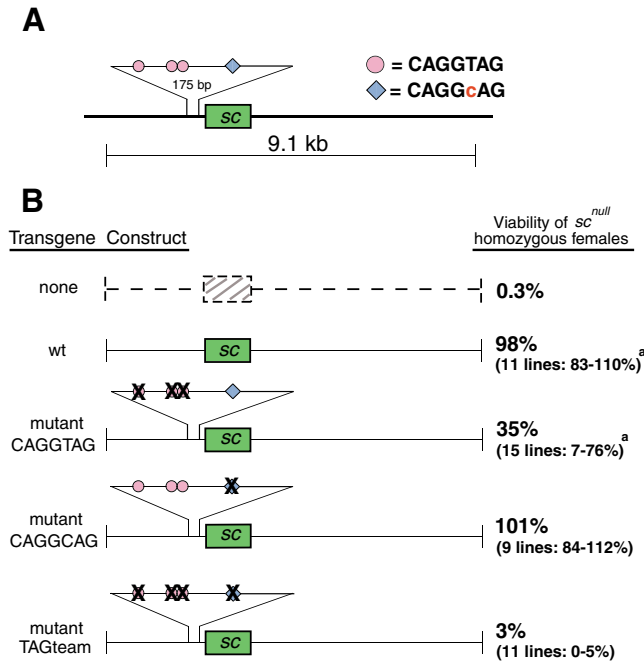


Fig. 4. CAGGcAG contributes to pre-cellular blastoderm *scute* function. (A) Schematic of the *sc* X-chromosome signal element (XSE) transgene. TAGteam sequences in the promoter region are indicated by colored symbols. (B) XSE activity of the indicated transgene based on its ability to rescue *sc*^{null} mutant females. Viability of *sc*^{M6}/*sc*^{M6} females was measured relative to that of their *sc*^{M6}/+ sisters (*n*>140 per line) from the cross *sc*^{M6}/*w*/*Binsinscy* X *sc*^{M6}/*w*/*Y*; *P[*sc*^{M6}]/+* at 25°C. ^a Values from Wrischnik et al. (Wrischnik et al., 2003). Extant representative lines from the previous study were tested in parallel with the new transgenes to confirm that the comparison was valid.

TAGteam sites affect the early expression of *Sxl*

The target of XSE action, the promoter *Sxl*_{Pe}, becomes active in female embryos during cycle 12, two cycles before cellularization of the blastoderm. Consistent with such pre-CB expression, *Sxl*_{Pe} has a TAGteam cluster within 250 bp of its transcription start site. The *melanogaster* cluster includes one CAGGTAG and a CAGGcAG doublet (Fig. 5A). From *melanogaster* to *virilis*, these sites are either identical or they are replaced by the reverse complement or another TAGteam member (see Fig. 3C). A 1.4-kb *Sxl*_{Pe}-*lacZ* reporter transgene had been reported to faithfully mimic endogenous *Sxl* sex-specific expression (Estes et al., 1995; Yang et al., 2001). To determine if *Sxl*_{Pe} expression depends on TAGteam sequences, we examined the effect of TAGteam mutations on the β-galactosidase levels generated by this reporter in stage 8-11 embryos.

As expected, the wild-type transgene gave a 1:1 ratio of stained to unstained embryos, indicating females and males, respectively, with nearly all females staining darkly (Fig. 5B,F). By contrast, almost no embryos carrying the triple TAGteam mutant showed dark or intermediate staining, and only an estimated 8% of females reached even the lightest staining category (Fig. 5C,G). Two additional transgenes established that the contribution of TAGteam sites to *Sxl*_{Pe} regulation is cumulative. Loss of just the two promoter-proximal sites, CAGGTAG and tAGGTAG (mutYAGGTAG), impaired expression (Fig. 5D,H), but not as severely as loss of all three. Again only a few embryos had dark or intermediate staining, but an estimated 50% of females were lightly stained. Even mutation of the CAGGcAG doublet by itself reduced staining (Fig. 5E,I), though not by as much as the loss of the two other TAGteam sites together. Inferences regarding the contribution of the CAGGcAG site are complicated by the fact that this double site partially overlaps an E-box sequence previously implicated in *Sxl*_{Pe} regulation (Yang et al., 2001). Although the E-box hexamer itself was purposely left intact in the CAGGcAG mutants, the possibility remains that transcription factor binding to this E-box site could be marginally affected by sequences outside the canonical hexamer motif.

We used in situ hybridization to RNA from the *Sxl*_{Pe}-*lacZ* transgene in 0- to 2-hour embryos both to establish that the effects of TAGteam mutations on β-galactosidase levels reflected effects on

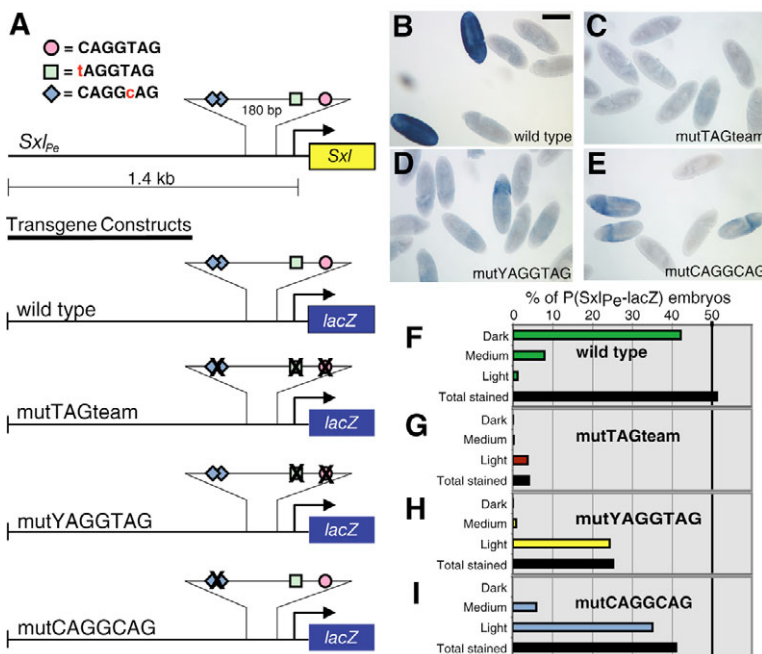


Fig. 5. TAGteam sequences affect *Sxl*_{Pe} responsiveness. (A) Schematic of wild-type and TAGteam-mutant versions of the 1.4-kb *Sxl*_{Pe}-*lacZ* reporter transgene. (B-E) Representative samples of β-galactosidase antibody-stained 3- to 5-hour embryos (25°C) carrying the indicated transgene. (F-I) Distribution of β-galactosidase antibody staining intensities among stage 8-11 embryos carrying the indicated transgene (*n*>2400 total embryos per construct, with four independent lines sampled for each). If all females and only females express *Sxl*_{Pe}, 50% of the embryos should stain. Scale bar: 250 μm.

transcript levels by the cellular blastoderm stage, and to determine whether TAGteam mutations delayed the onset of *Sxl* transcription. Fig. 6 shows that, as expected, the level of cytoplasmic mRNA from the TAGteam mutant *Sxl* reporter was far below that from the wild-type transgene at cycle 14 (compare Fig. 6A and C).

In pre-CB embryos, the time of onset of transcription for a given gene can be deduced through the analysis of ‘nuclear dots’, which reflect the hybridization of probe to nascent mRNA transcripts at the transcribed gene (Erickson and Cline, 1993; O’Farrell et al., 1989; Pritchard and Schubiger, 1996; Shermoen and O’Farrell, 1991). We defined the onset of transcription conservatively as the first nuclear cycle in which the average percentage of dot-containing nuclei exceeded 5% (see Materials and methods).

Dots from endogenous *Sxl* transcripts were reported to first appear at cycle 12, with the proportion of dot-containing nuclei quickly reaching 100% by the end of that cycle (Barbash and Cline, 1995; Erickson and Cline, 1993). We were surprised to find that the unmutated 1.4-kb *Sxl_{pe}-lacZ* reporter transgene was not as faithful a mimic of the endogenous gene as we had expected; nevertheless, it did reveal an unambiguous effect of the triple TAGteam knockout on both the timing and extent of transcription driven by *Sxl_{pe}*. The unmutated reporter initiated transcription during nuclear cycle 9, rather than cycle 12, with the proportion of active nuclei only gradually increasing during subsequent cycles to a maximum of just over 80% (Fig. 6B,E). Pritchard and Schubiger (Pritchard and Schubiger, 1996) had observed a similar gradual increase in nuclear

dots for many pre-CB genes, a behavior that had set these genes apart from endogenous *Sxl*. Moreover, we observed a small but reproducible amount of transcription in the nuclei of males, as deduced from the fact that after cycle 9, all embryos had at least a few nuclei with dots (data not shown). Nuclear dots of hybridization from the endogenous *Sxl_{pe}* had been observed previously in pre-CB male embryos, albeit not in all males examined (D. A. Barbash, PhD Thesis, University of California, 1995). In our study, the average percent of dot-containing nuclei in males always remained low, consistent with the fact that expression of the reporter appeared female-specific when assayed by β -galactosidase levels (Fig. 5F) or even by cytoplasmic mRNA accumulation (data not shown). In order to determine the onset of transcription of the 1.4-kb *Sxl_{pe}-lacZ* reporter specifically in females, we included only the top 50% of expressing embryos per cycle in our analysis.

The mutant *Sxl_{pe}* reporter initiated transcription three cycles later than the non-mutant, and fewer than 15% of the nuclei appeared to be active at any given time, even by cycle 13 (Fig. 6D,E). Thus, TAGteam sites not only advance the time of onset of *Sxl* transcription, they also increase the proportion of nuclei that transcribe *Sxl* during a given cell cycle. By contrast, the TAGteam mutations had no effect on the time at which transcription of the *Sxl_{pe}* reporter shuts off, a key event in the process of sex determination. Nuclear dot frequency plummeted for both the wild-type and mutant reporters shortly after nuclei began elongating early in cycle 14 (data not shown).

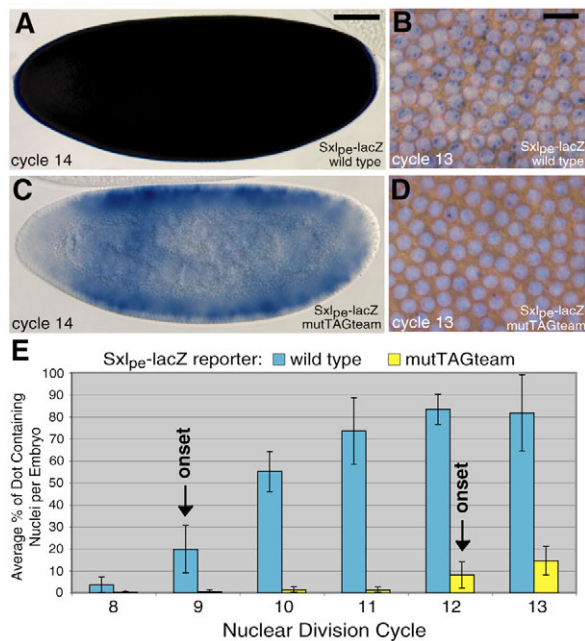


Fig. 6. TAGteam sequences affect the timing and penetrance of transcription from *Sxl_{pe}*. (A,C) In situ hybridization to *lacZ* RNA from a cycle 14 embryo carrying, respectively, the wild-type or the mutTAGteam 1.4-kb *Sxl_{pe}-lacZ* reporter transgene described in Fig. 5. (B,D) Nuclear dots of *lacZ* RNA hybridization in a cycle 13 embryo of the genotype as in A or C, respectively, co-stained for DNA (DAPI) to reveal nuclei. (E) Expression measured as the average number of dot-containing nuclei per presumed female embryo ($n \geq 30$ embryos per cycle per transgene). Arrows indicate onset of transcription (first cycle $\geq 5\%$ expressing nuclei). Dot analysis was performed on a single wild-type and TAGteam mutant transgene line, each determined to be representative based upon β -galactosidase staining. Scale bars: A,C, 50 μ m; B,D, 20 μ m.

TAGteam sites in the *zen* ventral repression element influence the time of onset of pre-CB transcription

The search for large TAGteam clusters led to *zen*, a gene expressed as early as the earliest XSEs (Pritchard and Schubiger, 1996), but one not involved in sex determination. The gene has six upstream TAGteam sites, four of which form an exceptionally compact cluster (91 bp) relatively far (1.4 kb) from the promoter (Fig. 7A) in a 600-bp region called the *zen* ventral repression element (VRE). The *zen* VRE had been shown to harbor pre-CB transcription activation sequences of unknown identity based on its ability to activate the *even-skipped* (*eve*) basal promoter in *lacZ* reporter constructs (Jiang et al., 1993; Kirov et al., 1993), a basal promoter that responds faithfully to pre-CB regulatory information (Markstein et al., 2002; Small et al., 1992). We assessed the potential role of the TAGteam in VRE functioning using the same reporter-gene strategy.

The intact 600-bp VRE begins to drive cytoplasmic mRNA accumulation in the soma prior to cellularization, with the level peaking early in cycle 14 (Fig. 7B). Accumulation is limited to the dorsal half of the embryo as a result of the action of Dorsal protein that represses *zen* in the ventral half of the embryo by binding to the VRE (reviewed by Stathopoulos and Levine, 2002). A 250-bp fragment of the VRE containing the four TAGteam sites had been shown to contain sequences necessary for transcriptional activation (Jiang et al., 1993). We found that this 250-bp fragment is not only necessary for activation, it is sufficient: it drove expression as effectively as the full-length 600-bp fragment (Fig. 7D). Expression was nearly uniform, since the 250-bp fragment lacked all but one Dorsal binding site.

We truncated the VRE even further to just a 111-bp fragment containing only the four TAGteam sites and 10 bp on each side. An alignment of this region from the species examined in Fig. 3C showed that the four TAGteam sites account for two-thirds of the nucleotides conserved in all six species. In contrast, the remaining 83 nucleotides of this fragment contained only eleven invariant base

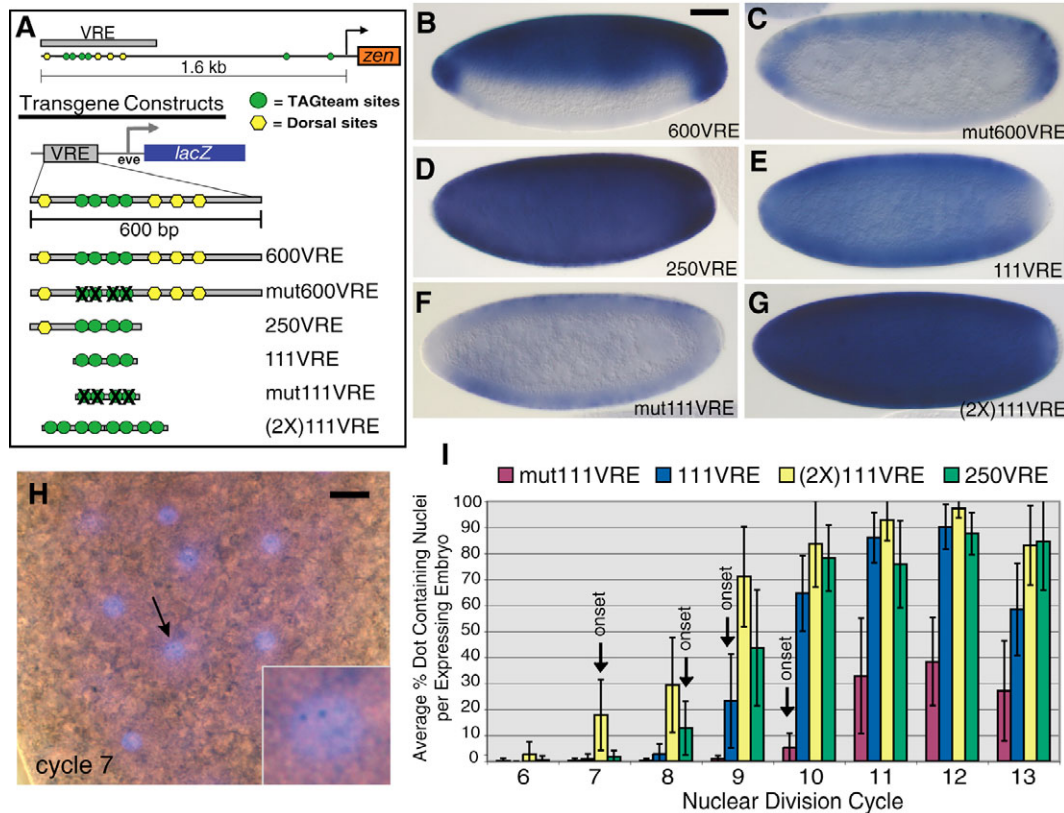


Fig. 7. TAGteam sites in the *zen* ventral repression element (VRE) help activate transcription. (A) Schematic of transgenes showing the segments of the VRE that were used to drive an *eve-lacZ* reporter. (B–G) *lacZ* RNA in situ hybridization to cycle 14 embryos carrying the 600VRE (B), mut600VRE (C), 250VRE (D), 111VRE (E), mut111VRE (F), or (2X)111VRE (G) transgene. (H) Nuclear RNA hybridization dots in a cycle 7 embryo carrying the (2X)111VRE transgene, co-stained with DAPI for DNA to reveal nuclei. The arrow indicates a dot-containing nucleus, which is magnified in the inset. (I) Expression measured as the average percentage of dot-containing nuclei per embryo. Transcription onset is defined as the first cycle $\geq 5\%$. Six lines were examined for the mut111VRE construct, three for all others. For each nuclear cycle, ≥ 18 embryos were assayed for the 250VRE construct, and ≥ 37 for all others. Scale bars: B–G, 50 μm ; H–J, 20 μm .

pairs, of which the longest contiguous run was only three base pairs that were adjacent to a TAGteam site. Even this minimal fragment drove ubiquitous expression in the early embryo, albeit at a reduced level, particularly at the posterior pole (Fig. 7E). Duplicating this minimal TAGteam cluster fragment increased transcript accumulation to levels comparable to those of the 250-bp VRE fragment (Fig. 7G). In contrast, mutating the four VRE TAGteam sites, whether in the original 600-bp fragment or in the minimal 111-bp fragment, reduced the accumulation of transcript at cycle 14 to low levels (Fig. 7C,F). The effects of the TAGteam mutations appear to be specific to pre-CB expression, since we saw no differences between the TAGteam mutant and wild-type 600-bp VRE fragment through stage (not cycle) 14 with respect to the highly patterned and dynamic post-CB embryonic expression pattern (data not shown).

Analysis of nuclear dots of *lacZ* RNA hybridization revealed that the minimal 111-bp VRE fragment begins to drive the *eve* basal promoter during nuclear cycle 9 (Fig. 7I), contemporaneous with the onset of endogenous *zen* transcription (Erickson and Cline, 1993; Pritchard and Schubiger, 1996). The minimal fragment was active in 23% of the nuclei during the onset cycle. That fraction increased gradually in subsequent cycles, peaking at 90% in nuclear cycle 12. In contrast, transcription driven by the TAGteam mutant 111-bp VRE transgene began only in cycle 10, one cycle later than for the non-mutant transgene, just barely reaching the 5% cutoff for onset

even in that cycle. The fraction of nuclei transcribing the mutant transgene peaked in cycle 12, just as it had for the nonmutant, but that peak was only 38%.

Whereas decreasing the number of TAGteam sites retarded initiation of pre-CB transcription, increasing their number appeared to advance it. The onset of transcription from the (2X)111VRE construct was in nuclear cycle 7 (Fig. 7H,I), two cycles earlier than that of the 111VRE construct and one cycle ahead of the earliest point that transcription of endogenous genes has been shown to begin (Erickson and Cline, 1993; Pritchard and Schubiger, 1996). This advance is unlikely to be an artifact of increased staining levels, since cytoplasmic mRNA staining for the 250-bp fragment was comparable to that for the (2X)111VRE construct in early cycle 14 (compare Fig. 7D with G), yet the 250-bp construct initiated expression one cycle later. The lack of concordance between onset of transcription and cycle 14 staining level was even greater between the (2X)111VRE construct and the wild-type 1.4-kb *Sxl_{P₆}* construct, which stained much more darkly by cycle 14, despite having begun transcription two cycles later (compare Fig. 6A with Fig. 7G).

***hermaphrodite* does not encode the TAGteam binding factor**

The *hermaphrodite* (*her*) gene seemed an attractive candidate as a source of the TAGteam binding factor, since its maternally encoded product is a zinc-finger transcription factor that had been reported to

positively regulate not only *Sxl_{pe}* but also non-sex-specific targets of unknown identity in the young embryo (Li and Baker, 1998; Pultz and Baker, 1995; Pultz et al., 1994). Consequently we determined whether the maternal effect of mutations in *her* would interfere with expression of the (2X)111VRE *zen* transgene. It did not, as the level of *lacZ* mRNA immunostaining from this transgene was the same for the progeny of *her^{(2)mat}* homozygous mutant mothers as for the progeny of their heterozygous sisters (data not shown).

Interpretation of this negative result is complicated by the fact that we did not observe the female-biased lethal maternal effect reported for *her^{(2)mat}* that had made the gene such an attractive candidate. For both *her^{(2)mat}* and the *her^{(2)mat}/her^l* heteroallelic combination, we did observe maternal-effect embryonic lethality, but both sexes of progeny were affected equally. Moreover, even when we tried to sensitize daughters to the maternal effect of *her^{(2)mat}/her^l* by reducing the dose of *Sxl⁺* to one copy, we saw no sex bias in lethality. Since we verified that the *her* alleles carried the appropriate molecular lesions, we can only surmise that the effect of *her* on *Sxl* is far weaker than originally indicated, and that some undefined and unsuspected aspect of genetic background or culture conditions sensitized *Sxl* to *her* in the previous studies. Hence the identity of the protein that recognizes TAGteam sites to drive pre-CB gene expression remains to be determined.

DISCUSSION

The combined bioinformatic and functional analysis reported here establishes that the TAGteam of three heptanucleotide sequences (CAGGTAG, tAGGTAG and CAGGcAG) plays an important, positive role in determining the time of onset of transcription for *Drosophila* genes whose expression begins prior to the point in the cellular blastoderm stage when widespread transcriptional activation of the genome of the embryo occurs. Previous studies had emphasized the role of maternally loaded repressor proteins in determining when zygotic transcription could begin (Newport and Kirschner, 1982a; Newport and Kirschner, 1982b; Pritchard and Schubiger, 1996). It was proposed that repression would ultimately be relieved by titration of the fixed amount of repressor protein present at fertilization by the increasing amount of zygotic DNA generated as development proceeds. Tramtrack is an example of one such repressor that was shown to control the timing of pre-cellular blastoderm (pre-CB) transcription for the *Drosophila* gene *fushi tarazu* (*ftz*) (Brown et al., 1991; Pritchard and Schubiger, 1996). The onset of transcription driven by a *ftz* regulatory DNA could be advanced either by eliminating Tramtrack binding sites or by reducing the amount of maternally encoded Tramtrack protein present in the young embryo. Increasing Tramtrack protein had the opposite effect.

In contrast, we show here that eliminating TAGteam sites has the opposite effect on the onset of pre-CB gene expression, retarding transcription driven either by a 111-bp minimal regulatory fragment of the pattern-formation gene *zerknüllt* (*zen*) or by the promoter of *Sex-lethal* (*Sxl*), the gene that counts fly X chromosomes to determine sex. Hence TAGteam sites must recruit activators, rather than repressors. Our observation that duplication of the minimal TAGteam VRE fragment advances the onset of transcription suggests that precisely when pre-CB genes are first expressed is determined by a balance between specific activator and repressor proteins. Reciprocal changes in the number of binding sites for an activator protein involved in *C. elegans* pharynx development had been shown to have reciprocal effects on the onset of expression of the mutated target gene (Gaudet and Mango, 2002); however, the gene studied in that case was expressed only after widespread zygotic transcription had already begun.

Analysis of the mechanism by which the TAGteam sites act to influence pre-CB transcription may provide unique insights into transcriptional regulation, since the nuclear environment during this early period appears to differ significantly from that later when general transcription begins. For example, during this early period in *Drosophila*, the high mobility group protein D (HMG-D) stands in for histone H1 (Ner et al., 2001; Ner and Travers, 1994). Moreover, it has been shown that *Xenopus* embryos have a uniquely large excess of core histones prior to the midblastula transition (MBT) that may play an important part in pre-MBT transcriptional quiescence (Prioleau et al., 1994). It is not known whether *Drosophila* embryos have a comparable shift in the level of core histones relative to DNA that correlates with the onset of widespread transcription, but measurements of histone mRNA levels are consistent with the possibility of a large shift (Anderson and Lengyel, 1980). The availability of TATA-binding factor also seems to be a factor limiting transcription in *Xenopus* only prior to the MBT (Veenstra et al., 1999), but it is not known whether a similar change occurs in *Drosophila*.

Although the TAGteam is highly over-represented upstream of the genes that serve as the sex-determination signal in *D. melanogaster*, our discovery that TAGteam sites can also be unusually abundant upstream of early expressed genes like *zen* and *bnk* that have nothing to do with sex determination shows that their function is not restricted to sex-determination genes. On the other hand, the TAGteam cannot be the only sequences that mark genes for pre-CB expression, since the mutant 111-bp *zen* VRE transgene with no TAGteam sequences was still able to initiate pre-CB transcription from the *even-skipped* promoter, albeit with a delay relative to the nonmutant transgene, and with a reduction in the fraction of nuclei that ultimately express. Moreover, pre-CB genes exist (e.g. *nullo*) that have no TAGteam heptamer within 2 kb of their transcription start site. As more information becomes available on various *Drosophila* species with respect not only to DNA sequence, but also to start times of transcription for specific genes, a clearer picture should emerge on the interchangeability of motifs that direct pre-CB gene expression. Already, our analysis of very new data that increased the list of recognized *melanogaster* pre-CB genes (Pilot et al., 2006) led us to two 1-bp degenerates of CAGGTAG that may earn the certified TAGteam label once their functional contribution to pre-CB gene expression is tested by mutation.

Thus the TAGteam seems to provide an opportunity to analyze how enhancer sequences change over evolutionary time under circumstances where the organism may have a variety of alternatives available to achieve the same functional goal. But if those alternatives are truly equivalent, it is puzzling why the X-chromosome signal elements (XSEs) in *D. melanogaster* seem to rely so heavily on the TAGteam. Even *run*, the only exception to the rule that XSEs in *D. melanogaster* have three CAGGTAG sites within 500 bp of their transcription start sites, has six TAGteam sites within 3.1 kb upstream. Previously published analysis of the regions we now know contain TAGteam sites showed that they direct the very early, broad expression of *run* that affects *Sxl_{pe}* (Klingler et al., 1996).

Because the duplicated 111-bp TAGteam fragment from *zen* drove expression earlier than any endogenous gene is known to be transcribed, it seems likely that any TAGteam binding protein that activates pre-CB transcription will be derived from maternal rather than zygotic gene expression. With our elimination of the genetically characterized *hermaphrodite* (*her*) gene as a potential source of this activity (and indeed perhaps even as a regulator of *Sxl_{pe}*), a straightforward biochemical approach may be the most practical

route for identifying the relevant protein. Information on the identity of this protein should further understanding of the molecular mechanisms that govern pre-CB transcription and reveal whether those mechanisms differ qualitatively from those governing the subsequent widespread activation of zygotic gene expression.

We thank M. Markstein, J. Erickson and B. Baker for providing stocks and reagents, and thank B. Meyer, M. Stern, and members of the Cline lab for helpful discussions and comments on the manuscript. This work was supported by US National Institutes of Health grant GM-23468 to T.W.C.

Supplementary material

Supplementary material for this article is available at <http://dev.biologists.org/cgi/content/full/133/10/1967/DC1>

References

- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195.
- Andeol, Y. (1994). Early transcription in different animal species: implication for transition from maternal to zygotic control in development. *Roux's Arch. Dev. Biol.* **204**, 3-10.
- Anderson, K. V. and Lengyel, J. A. (1979). Rates of synthesis of major classes of RNA in *Drosophila* embryos. *Dev. Biol.* **70**, 217-231.
- Anderson, K. V. and Lengyel, J. A. (1980). Changing rates of histone mRNA synthesis and turnover in *Drosophila* embryos. *Cell* **21**, 717-727.
- Bailey, T. L. and Gribskov, M. (1998). Methods and statistics for combining motif match scores. *J. Comput. Biol.* **5**, 211-221.
- Barbash, D. A. and Cline, T. W. (1995). Genetic and molecular analysis of the autosomal component of the primary sex determination signal of *Drosophila melanogaster*. *Genetics* **141**, 1451-1471.
- Bronner, G., Chu-LaGriff, Q., Doe, C. Q., Cohen, B., Weigel, D., Taubert, H. and Jackle, H. (1994). Sp1/egr-like zinc-finger protein required for endoderm specification and germ-layer formation in *Drosophila*. *Nature* **369**, 664-668.
- Brown, J. L., Sonoda, S., Ueda, H., Scott, M. P. and Wu, C. (1991). Repression of the *Drosophila* fushi tarazu (ftz) segmentation gene. *EMBO J.* **10**, 665-674.
- Cline, T. W. and Meyer, B. J. (1996). Vive la difference: males vs females in flies vs worms. *Annu. Rev. Genet.* **30**, 637-702.
- Coulter, D. E., Swaykus, E. A., Beran-Koehn, M. A., Goldberg, D., Wieschaus, E. and Schedl, P. (1990). Molecular analysis of odd-skipped, a zinc finger encoding segmentation gene with a novel pair-rule expression pattern. *EMBO J.* **9**, 3795-3804.
- Davidson, E. H. (1986). *Gene Activity in Early Development*. Orlando: Academic Press.
- Eldon, E. D. and Pirrotta, V. (1991). Interactions of the *Drosophila* gap gene *giant* with maternal and zygotic pattern-forming genes. *Development* **111**, 367-378.
- Erickson, J. W. and Cline, T. W. (1993). A bZIP protein, sisterless-a, collaborates with bHLH transcription factors early in *Drosophila* development to determine sex. *Genes Dev.* **7**, 1688-1702.
- Erickson, J. W. and Cline, T. W. (1998). Key aspects of the primary sex determination mechanism are conserved across the genus *Drosophila*. *Development* **125**, 3259-3268.
- Estes, P. A., Keyes, L. N. and Schedl, P. (1995). Multiple response elements in the *Sex-lethal* early promoter ensure its female-specific expression pattern. *Mol. Cell. Biol.* **15**, 904-917.
- Foe, V. E. and Alberts, B. M. (1983). Studies of nuclear and cytoplasmic behaviour during the five mitotic cycles that precede gastrulation in *Drosophila* embryogenesis. *J. Cell Sci.* **61**, 31-70.
- Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273-W279.
- Gaudet, J. and Mango, S. E. (2002). Regulation of organogenesis by the *Caenorhabditis elegans* FoxA protein PHA-4. *Science* **295**, 821-825.
- Gergen, J. P. (1987). Dosage compensation in *Drosophila*: evidence that daughterless and *Sex-lethal* control X-chromosome activity at the blastoderm stage of embryogenesis. *Genetics* **117**, 477-485.
- Harrison, D. A., McCoon, P. E., Binari, R., Gilman, M. and Perrimon, N. (1998). *Drosophila* unpaired encodes a secreted protein that activates the JAK signaling pathway. *Genes Dev.* **12**, 3252-3263.
- Ibnsouda, S., Schweisguth, F., de Billy, G. and Vincent, A. (1993). Relationship between expression of serendipity alpha and cellularisation of the *Drosophila* embryo as revealed by interspecific transformation. *Development* **119**, 471-483.
- Ingham, P. W., Howar, K. R. and Ish-Horowitz, D. (1985). Transcription pattern of the *Drosophila* segmentation gene hairy. *Nature* **318**, 439-445.
- Jazwinska, A., Rushlow, C. and Roth, S. (1999). The role of *brinker* in mediating the graded response to Dpp in early *Drosophila* embryos. *Development* **126**, 3323-3334.
- Jiang, J., Kosman, D., Ip, Y. T. and Levine, M. (1991). The dorsal morphogen gradient regulates the mesoderm determinant twist in early *Drosophila* embryos. *Genes Dev.* **5**, 1881-1891.
- Jiang, J., Cai, H., Zhou, Q. and Levine, M. (1993). Conversion of a dorsal-dependent silencer into an enhancer: evidence for dorsal corepressors. *EMBO J.* **12**, 3201-3209.
- Kimelman, D., Kirschner, M. and Scherson, T. (1987). The events of the midblastula transition in *Xenopus* are regulated by changes in the cell cycle. *Cell* **48**, 399-407.
- Kirov, N., Zhelnin, L., Shah, J. and Rushlow, C. (1993). Conversion of a silencer into an enhancer: evidence for a co-repressor in dorsal-mediated repression in *Drosophila*. *EMBO J.* **12**, 3193-3199.
- Klingler, M. and Gergen, J. P. (1993). Regulation of runt transcription by *Drosophila* segmentation genes. *Mech. Dev.* **43**, 3-19.
- Klingler, M., Soong, J., Butler, B. and Gergen, J. P. (1996). Disperse versus compact elements for the regulation of runt stripes in *Drosophila*. *Dev. Biol.* **177**, 73-84.
- Kraut, R. and Levine, M. (1991). Spatial regulation of the gap gene giant during *Drosophila* development. *Development* **111**, 601-609.
- Kuo, Y. M., Jones, N., Zhou, B., Panzer, S., Larson, V. and Beckendorf, S. K. (1996). Salivary duct determination in *Drosophila*: roles of the EGF receptor signalling pathway and the transcription factors fork head and trachealless. *Development* **122**, 1909-1917.
- Kwiatowski, J., Skarecky, D., Bailey, K. and Ayala, F. J. (1994). Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu,Zn Sod gene. *J. Mol. Evol.* **38**, 443-454.
- Lamb, M. M. and Laird, C. D. (1976). Increase in nuclear poly(A)-containing RNA at syncytial blastoderm in *Drosophila melanogaster* embryos. *Dev. Biol.* **52**, 31-42.
- Lecuit, T. and Wieschaus, E. (2000). Polarized insertion of new membrane from a cytoplasmic reservoir during cleavage of the *Drosophila* embryo. *J. Cell Biol.* **150**, 849-860.
- Li, H. and Baker, B. S. (1998). Her, a gene required for sexual differentiation in *Drosophila*, encodes a zinc finger protein with characteristics of ZFY-like proteins and is expressed independently of the sex determination hierarchy. *Development* **125**, 225-235.
- Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* **125**, 949-958.
- Ludwig, M. Z., Bergman, C., Patel, N. H. and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* **403**, 564-567.
- Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA* **99**, 763-768.
- Markstein, M., Zinzen, R., Markstein, P., Yee, K. P., Erives, A., Stathopoulos, A. and Levine, M. (2004). A regulatory code for neurogenic gene expression in the *Drosophila* embryo. *Development* **131**, 2387-2394.
- McKnight, S. L. and Miller, O. L., Jr (1976). Ultrastructural patterns of RNA synthesis during early embryogenesis of *Drosophila melanogaster*. *Cell* **8**, 305-319.
- Nakakura, N., Miura, T., Yamana, K., Ito, A. and Shiokawa, K. (1987). Synthesis of heterogeneous mRNA-like RNA and low-molecular-weight RNA before the midblastula transition in embryos of *Xenopus laevis*. *Dev. Biol.* **123**, 421-429.
- Ner, S. S. and Travers, A. A. (1994). HMG-D, the *Drosophila melanogaster* homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1. *EMBO J.* **13**, 1817-1822.
- Ner, S. S., Blank, T., Perez-Paralle, M. L., Grigliatti, T. A., Becker, P. B. and Travers, A. A. (2001). HMG-D and histone H1 interplay during chromatin assembly and early embryogenesis. *J. Biol. Chem.* **276**, 37569-37576.
- Newport, J. and Kirschner, M. (1982a). A major developmental transition in early *Xenopus* embryos: I. characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675-686.
- Newport, J. and Kirschner, M. (1982b). A major developmental transition in early *Xenopus* embryos: II. Control of the onset of transcription. *Cell* **30**, 687-696.
- O'Farrell, P. H., Edgar, B. A., Lakich, D. and Lehner, C. F. (1989). Directing cell division during development. *Science* **246**, 635-640.
- Pilot, F., Philippe, J. M., Lemmers, C., Chauvin, J. P. and Lecuit, T. (2006). Developmental control of nuclear morphogenesis and anchoring by charleston, identified in a functional genomic screen of *Drosophila* cellularisation. *Development* **133**, 711-723.
- Powell, J. R. and DeSalle, R. (1995). *Drosophila* molecular phylogenies and their uses. In *Evolutionary Biology*. Vol. 28 (ed. M. K. Hecht, R. J. MacIntyre and M. T. Clegg), pp. 87-138. New York: Plenum Press.
- Prioleau, M. N., Huet, J., Sentenac, A. and Mechali, M. (1994). Competition between chromatin and transcription complex assembly regulates gene expression during early development. *Cell* **77**, 439-449.

- Pritchard, D. K. and Schubiger, G. (1996). Activation of transcription in *Drosophila* embryos is a gradual process mediated by the nucleocytoplasmic ratio. *Genes Dev.* **10**, 1131-1142.
- Pultz, M. A. and Baker, B. S. (1995). The dual role of hermaphrodite in the *Drosophila* sex determination regulatory hierarchy. *Development* **121**, 99-111.
- Pultz, M. A., Carson, G. S. and Baker, B. S. (1994). A genetic analysis of hermaphrodite, a pleiotropic sex determination gene in *Drosophila melanogaster*. *Genetics* **136**, 195-207.
- Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P. et al. (2005). Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1-18.
- Russo, C. A., Takezaki, N. and Nei, M. (1995). Molecular phylogeny and divergence times of drosophilid species. *Mol. Biol. Evol.* **12**, 391-404.
- Schejter, E. D. and Wieschaus, E. (1993). bottleneck acts as a regulator of the microfilament network governing cellularization of the *Drosophila* embryo. *Cell* **75**, 373-385.
- Sefton, L., Timmer, J. R., Zhang, Y., Beranger, F. and Cline, T. W. (2000). An extracellular activator of the *Drosophila* JAK/STAT pathway is a sex-determination signal element. *Nature* **405**, 970-973.
- Shermoen, A. W. and O'Farrell, P. H. (1991). Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. *Cell* **67**, 303-310.
- Small, S., Blair, A. and Levine, M. (1992). Regulation of even-skipped stripe 2 in the *Drosophila* embryo. *EMBO J.* **11**, 4047-4057.
- Spradling, A. C. (1986). P-element-mediated transformation. In *Drosophila: A Practical Approach* (ed. D. B. Roberts), pp. 175-197. Oxford: IRL Press.
- St Johnston, R. D. and Gelbart, W. M. (1987). Decapentaplegic transcripts are localized along the dorsal-ventral axis of the *Drosophila* embryo. *EMBO J.* **6**, 2785-2791.
- Stathopoulos, A. and Levine, M. (2002). Dorsal gradient networks in the *Drosophila* embryo. *Dev. Biol.* **246**, 57-67.
- Tautz, D. and Pfeifle, C. (1989). A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* **98**, 81-85.
- Tautz, D., Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jackle, H. (1987). Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation genes. *Nature* **327**, 383-389.
- Thisse, B., Stoetzel, C., Gorostiza-Thisse, C. and Perrin-Schmitt, F. (1988). Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. *EMBO J.* **7**, 2175-2183.
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
- Tomancak, P., Beaton, A., Weiszmann, R., Kwan, E., Shu, S., Lewis, S. E., Richards, S., Ashburner, M., Hartenstein, V., Celniker, S. E. et al. (2002). Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biol.* **3**, RESEARCH0088.1-0088.14.
- Tracey, W. D., Jr, Ning, X., Klingler, M., Kramer, S. G. and Gergen, J. P. (2000). Quantitative analysis of gene function in the *Drosophila* embryo. *Genetics* **154**, 273-284.
- Veenstra, G. J., Destree, O. H. and Wolffe, A. P. (1999). Translation of maternal TATA-binding protein mRNA potentiates basal but not activated transcription in *Xenopus* embryos at the midblastula transition. *Mol. Cell. Biol.* **19**, 7972-7982.
- Wrischnik, L. A., Timmer, J. R., Megna, L. A. and Cline, T. W. (2003). Recruitment of the proneural gene scute to the *Drosophila* sex-determination pathway. *Genetics* **165**, 2007-2027.
- Yang, D., Lu, H., Hong, Y., Jinks, T. M., Estes, P. A. and Erickson, J. W. (2001). Interpretation of X chromosome dose at Sex-lethal requires non-E-box sites for the basic helix-loop-helix proteins SISB and daughterless. *Mol. Cell. Biol.* **21**, 1581-1592.

Table S1. Pre- and post-cellular blastoderm genes used in Fig. 2

Gene name	CG number	FB ID number
Pre-CB genes		
bnk	CG1480	FBgn0004389
brk	CG9653	FBgn0024250
dpn	CG8704	FBgn0010109
dpp	CG9885	FBgn0000490
eve	CG2328	FBgn0000606
ftz	CG2047	FBgn0001077
gt*	CG7952	FBgn0001150
h	CG6494	FBgn0001168
hb	CG9786	FBgn0001180
hkb	CG9768	FBgn0001204
kni	CG4717	FBgn0001320
Kr	CG3340	FBgn0001325
nullo	CG14426	FBgn0004143
odd	CG3851	FBgn0002985
os	CG5993	FBgn0004956
run	CG1849	FBgn0003300
sc	CG3827	FBgn0004170
sisA	CG1641	FBgn0003411
sna	CG3956	FBgn0003448
sog	CG9224	FBgn0003463
sry-alpha	CG17957	FBgn0003510
Sxl	CG33070	FBgn0029933
tll	CG1378	FBgn0003720
twi	CG2956	FBgn0003900
zen	CG1046	FBgn0004053
Post-CB genes		
Ast	CG13633	FBgn0015591
br	CG11514	FBgn0000210
caup	CG10605	FBgn0015919
ct	CG11387	FBgn0004198
Cyp304a1	CG7241	FBgn0038095
Cyp6a18*	CG13977	FBgn0039519
dei	CG5441	FBgn0008649
dif	CG6794	FBgn0011274
e	CG3331	FBgn0000527
ft	CG3352	FBgn0001075
gl	CG7672	FBgn0004618
Gld	CG1152	FBgn0001112
gol*	CG2679	FBgn0004919
kay	CG15507	FBgn0001297
ome	CG9370	FBgn0002997
ppk	CG3478	FBgn0020258
Pxn	CG12002	FBgn0011828
qtc	CG14039	FBgn0028572
CAP	CG18409	FBgn0033504
toe	CG10704	FBgn0036285
yellow-b*	CG17914	FBgn0032601
zfh2	CG1449	FBgn0004607
	CG10026	FBgn0032785
	CG10086	FBgn0037517
	CG10175	FBgn0039084
	CG10300	FBgn0039107
	CG10561	FBgn0002036
	CG1092	FBgn0037228
	CG11175	FBgn0034530
	CG1124	FBgn0037290
	CG11426	FBgn0037166
	CG12116	FBgn0030041
	CG12565	FBgn0031180
	CG12817	FBgn0037798
	CG1283	FBgn0051536
	CG13159	FBgn0033721
	CG13499	FBgn0034680
	CG13506	FBgn0034723
	CG13636	FBgn0039232

CG13699	FBgn0036781
CG14253	FBgn0039467
CG14879	FBgn0038419
CG14889	FBgn0038443
CG14948	FBgn0030723
CG1499	FBgn0039852
CG15105	FBgn0034412
CG15407	FBgn0053281
CG15512	FBgn0051038
CG15592	FBgn0037416
CG15598	FBgn0037427
CG15786	FBgn0029749
CG17104	FBgn0040496
CG2191	FBgn0039873
CG3074	FBgn0034709
CG31757	FBgn0051757
CG32048	FBgn0052048
CG3544	FBgn0031279
CG4576	FBgn0038366
CG4623	FBgn0035587
CG4678	FBgn0030778
CG4914	FBgn0036436
CG4984	FBgn0034267
CG5171	FBgn0031907
CG5476	FBgn0039441
CG5973	FBgn0031914
CG6113	FBgn0032264
CG6305	FBgn0033869
CG6337	FBgn0033873
CG6776	FBgn0035904
CG7227	FBgn0031970
CG8193	FBgn0033367
CG8713	FBgn0033257
CG8850	FBgn0033708
CG9192	FBgn0035193
CG9444	FBgn0037730
CG9460	FBgn0033115
CG9990	FBgn0039594
CG9993	FBgn0034553

*Genes not included in analysis for *D. pseudoobscura* because of a lack of promoter sequence.

Table S2. Conservation of TAGteam over-representation in pre-CB genes of the genus *Drosophila*

TAGteam sequence	Observed/expected frequency of TAGteam sites for each given species* (rank [†])					
	<i>mel</i>	<i>yak</i>	<i>ana</i>	<i>pse</i>	<i>moj</i>	<i>vir</i>
CAGGTAG	12.4 (1)	12.6 (1)	9.8 (1)	10.7 (1)	12.2 (1)	9.6 (1)
tAGGTAG	7.2 (2)	8.2 (2)	7.7 (2)	8.5 (2)	4.4 (4)	4.7 (3)
CAGGcAG	7.2 (3)	5.3 (3)	4.9 (4)	7.3 (3)	5.6 (2)	8.5 (2)
CAGGTAA	2.6 (5)	3.7 (4)	6.3 (3)	2.1 (5)	5.1 (3)	3.4 (5)
CAGGTAt	0.7 (9)	1.5 (6)	0.7 (10)	2.1 (5)	1.5 (7)	1.3 (7)
Total	30.1	31.3	29.5	30.7	28.8	27.5

**mel*, *D. melanogaster*; *yak*, *D. yakuba*; *ana*, *D. ananassae*; *pse*, *D. pseudoobscura*; *moj*, *D. mojavensis*; *vir*, *D. virilis* (in order of relation to *D. melanogaster*). The following genes had incomplete sequence: *dpn* (*vir* -142 to -1); *ftz* (*ana* -30 to -1, *moj* -73 to -54); *h* (*ana* -431 to -332); *hkb* (*moj* -500 to -1); *sog* (*yak* -430 to -1, *ana* -497 to -398, *pse* -119 to -68, *vir* -500 to -202); *Sxl_{pe}* (*ana* -187 to -88). Expected frequency was based on the GC% for all the sequence analyzed within that species.

[†]Ranking with respect to degree of overrepresentation in the given species (1=most) among the set of sequences that includes CAGGTAG and the 20 1-bp degenerates.