Sequence of the *Tribolium castaneum* Homeotic Complex: The Region Corresponding to the *Drosophila melanogaster* Antennapedia Complex

Susan J. Brown,*,[1] John P. Fellers,† Teresa D. Shippy,* Elizabeth A. Richardson,* Mark Maxwell,‡ Jeffery J. Stuart‡ and Robin E. Denell*

**Division of Biology, Kansas State University, Manhattan, Kansas 66506, †United States Department of Agriculture-Agricultural Research Service, Kansas State University, Manhattan, Kansas 66506 and ‡Department of Entomology, Purdue University, West Lafayette, Indiana 47907*

Manuscript received July 30, 2001
Accepted for publication December 13, 2001

ABSTRACT

The homeotic selector genes of the red flour beetle, *Tribolium castaneum*, are located in a single cluster. We have sequenced the region containing the homeotic selector genes required for proper development of the head and anterior thorax, which is the counterpart of the ANTC in Drosophila. This 280-kb interval contains eight homeodomain-encoding genes, including single orthologs of the Drosophila genes *labial*, *proboscipedia*, *Deformed*, *Sex combs reduced*, *fushi tarazu*, and *Antennapedia*, as well as two orthologs of *zerknüllt*. These genes are all oriented in the same direction, as are the Hox genes of amphioxus, mice, and humans. Although each transcription unit is similar to its Drosophila counterpart in size, the Tribolium genes contain fewer introns (with the exception of the two *zerknüllt* genes), produce shorter mRNAs, and encode smaller proteins. Unlike the ANTC, this region of the Tribolium HOMC contains no additional genes.

HOX genes encode transcription factors that regulate developmental fate along the anterior-posterior (A-P) body axis of virtually all bilaterians (Finnerty and Martindale 1998). Hox genes, located in clusters, are spatially colinear in that their chromosomal organization reflects their functional domains along the A-P axis. Multiple Hox clusters have been described for several vertebrates including mice (Graham *et al.* 1989), humans (Acampora *et al.* 1989), and fish (Kim *et al.* 2000). In contrast, single Hox clusters have been identified in a number of invertebrates including amphioxus (Garcia-Fernàndez and Holland 1994), sea urchins (Popodi *et al.* 1996), and several insects such as mosquitoes (Devenport *et al.* 2000; Powers *et al.* 2000), beetles (Beeman 1987), and locusts (Ferrier and Akam 1996). In Drosophila, the single complement of Hox genes is divided into two clusters, the Antennapedia complex (ANTC) and the bithorax complex (BXC), separated by 7.5 Mb on the right arm of LG 3 (Adams *et al.* 2000). The ANTC contains genes required for proper development of the gnathal and thoracic segments [*labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), and *Antennapedia* (*Antp*); Kaufman *et al.* (1990)], while the BXC genes [*Ultrabithorax* (*Ubx*), *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*)] direct developmental fate in the posterior thorax and abdomen (Bender *et al.* 1983). The ANTC contains additional genes: *zerknüllt* (*zen*), a Hox gene that lost its function along the A-P axis and gained a function in the extraembryonic region of the insect egg (Falciani *et al.* 1996); *zerknüllt-related* (*zen2*), a recent duplication of *zen*; *bicoid* (*bcd*), another recent derivative of *zen* (Stauber *et al.* 2000) that encodes the anterior morphogen of the Drosophila embryo (Struhl *et al.* 1989); and *fushi tarazu* (*ftz*), a pair-rule gene related to the Antp-class homeobox genes (Scott and Weiner 1984). Eight cuticle genes, five lysine tRNA genes, and *amalgam* (*ama*), which encodes a member of the immunoglobulin superfamily, are also located in the Drosophila ANTC (Adams *et al.* 2000).

Sequence analysis of the relatively compact Hox clusters (80–120 kb) of several deuterostomes has revealed that the order and orientation of Hox genes is well conserved. However, among protostomes, only the considerably larger ANTC and BXC (~400 and ~350 kb, respectively) of *Drosophila melanogaster* (Martin *et al.* 1995; Adams *et al.* 2000) and the highly degenerate Hox cluster of *Caenorhabditis elegans* (*C. Elegans* Sequencing Consortium 1998) have been sequenced. Analysis of homeobox sequences from several arthropod classes (for a recent synthesis see Cook *et al.* 2001) and an onychophoran (Grenier *et al.* 1997) indicates that precursors of eight Hox genes, as well as *zen* and *ftz*, existed in the protoarthropod. Although these gene fragments are often aligned with their Drosophila counterparts for comparison, the number, order, and orientation of these genes, or whether they are actually clustered, is not fully understood for any other arthropod.

In the red flour beetle, *Tribolium castaneum*, we have previously identified and characterized the cDNA sequences, expression patterns, and, for some genes, the

[1] *Corresponding author:* Division of Biology, Ackert Hall, Kansas State University, Manhattan, KS 66506. E-mail: sjbrown@ksu.edu

mutant phenotypes of the orthologs of *pb* (Shippy *et al.* 2000a,b), *zen* (Falciani *et al.* 1996), *Dfd* (Brown *et al.* 1999, 2000), *ftz* (Brown *et al.* 1994), *Scr* (Curtis *et al.* 2001), *Ubx* (Bennett *et al.* 1999), and *abd-A* (Stuart *et al.* 1993; Shippy *et al.* 1998). We recently described a bacterial artificial chromosome (BAC) clone containing the *Tclab* transcription unit and the 3′ end of the Tribolium homeotic complex (HOMC; Nie *et al.* 2001). We found no genes in the interval between the 5′ end of *Tclabial* (*Tclab*) and the 3′ end of *maxillopedia* (*mxp*, the ortholog of *pb*), indicating that the cuticle genes found in the corresponding location in the Drosophila ANTC are likely to be a recent addition. We also noted that different genes are located downstream of *lab* in Tribolium and Drosophila, indicating a lack of synteny outside the HOMC between these two insects. Further, although *even-skipped* (*eve*) homologs are linked to homeotic gene clusters in mammals (D'Esposito *et al.* 1991; Faiella *et al.* 1991), we previously determined that, similar to *eve* in Drosophila, *Tceve* is not linked to the HOMC (Brown *et al.* 1997). We have sequenced the rest of the BAC clone containing *Tclab*, as well as two additional BAC clones containing the interval delimited by the 5′ half of *mxp* and *prothoraxless* (*ptl*, the ortholog of *Antp*). In a brief description of the second BAC clone (Brown *et al.* 2001) we noted that it contains *mxp*, *TcDeformed* (*TcDfd*), *Cephalothorax* (*Cx*, the ortholog of *Scr*), and *Tcfushi tarazu* (*Tcftz*), as well as two *Tczerknüllt* (*Tczen*) genes (more closely related to each other than to the Drosophila *zen* genes), but no other genes. The third BAC clone contains *ptl* and overlaps the second BAC clone in the region containing *Tcftz*. Here we describe the organization of the region of the Tribolium HOMC corresponding to the Drosophila ANTC, including gene order and orientation, as well as intergenic distances.

MATERIALS AND METHODS

BAC library screens: One BAC library was constructed in our laboratory by partially digesting GA-1 (wild-type) genomic DNA with *Hin*dIII and ligating it into pBeloBac11 (Zhu *et al.* 1997). This library contains inserts averaging 40–60 kb in length. The identification of Bmxp1 from this library was previously described (Shippy *et al.* 2000a). A second BAC library (kind gift of Exelixis Pharmaceutical, San Francisco), constructed from GA-2 genomic DNA that had been partially digested with *Eco*RI and ligated into pBACe3.6, contains inserts averaging 125 kb. GA-2 (a kind gift of Scott Thomson, University of Wisconsin, Parkside), was derived from GA-1 by 20 generations of single brother/sister matings and is also being used in the construction of a high-density molecular map of the Tribolium genome. Twenty-two BAC clones were identified by screening the second BAC library with radiolabeled *Tcftz* cDNA and a fragment from the 3′ end of an *Ultrathorax* (*Utx*) cDNA (*Utx*, the ortholog of *Ubx*). *Eco*RI fingerprint blots of each clone were hybridized under stringent conditions with radiolabeled fragments from *Tclab*, *mxp*, *TcDfd*, *Cx*, *Tcftz*, *ptl*, and *Utx*. A preliminary contig was assembled based on the hybridization patterns. Finally, the extent of overlap between individual clones was determined by radiolabeling selected BAC clone inserts and hybridizing them to the fingerprint blots. Clones with minimal overlap were chosen for sequencing.

Shotgun subcloning and sequencing: BAC clone DNA was isolated using the Qiagen large construct kit (QIAGEN, Valencia, CA), sheared with a nebulizer and subcloned into pCR4Blunt-TOPO using the TOPO Shotgun subcloning kit (Invitrogen, Carlsbad, CA). White colonies were restreaked to verify color selection. Colonies containing inserts were picked and transferred to 96-well microtiter plates. Plasmid DNA was purified and sequenced by automated protocols at the Sequencing and Genotyping Facility in the Department of Plant Pathology at Kansas State University. Files containing *Escherichia coli* DNA sequence were removed from the data set and BAC contigs were constructed using PHRED, PHRAP, and CONSED (Ewing *et al.* 1998; Gordon *et al.* 1998). The finished sequence averages greater than sixfold coverage. Gaps in the BAC insert sequences were filled by sequencing available subclones on both strands with primers designed from the ends of contigs.

Sequence analysis: The previously identified cDNAs were mapped on the BAC sequences by comparing cDNA and genomic sequence. The following GenBank accession numbers correspond to the cDNA sequences used: *Tclab*, AF231104 and AF230312; *mxp*, AF187068 and AF187069; *Tczen1*, X97819; *Tczen2*, AF452568; *TcDfd*, U81038; *Cx*, AF227628; *ftz*, U14734; and *ptl*, AF228509 and AF452569. The BAC contig sequences were also analyzed using BLASTX [National Center for Biotechnology Information (NCBI)], FGENESH (BCM Gene Finder), GENIE (Reese *et al.* 1997), and GENESCAN (Burge and Karlin 1997) to identify putative genes. MatInspector (Quandt *et al.* 1995) was used to search Tribolium sequences for conserved regulatory sites. Annotated sequences are available in GenBank under the following accession numbers: Bmxp1, AY043293; 35E10, AF321227; and 35F4, AY043292. These entries provide corrected and updated sequences for previous entries of *Tczen1*, *Tcftz*, and *ptl* with corresponding changes in the deduced amino acid sequences. The Tribolium sequence was analyzed for repeats and regions of simple sequence using RepeatMasker (University of Washington Genome Center).

RT-PCR: We used total RNA from 0- to 16-hr eggs as template for reverse transcription using the Thermoscript RT-PCR system (GIBCO BRL, Gaithersburg, MD). Primers designed to amplify the second Tribolium *zen* gene were located 56 bp downstream of the first methionine codon in the newly identified open reading frame (ORF) and 86 bp upstream of the translation stop codon in frame with the homeobox. The single RT-PCR product was cloned into the pGEM-T Easy Vector (Promega, Madison, WI) and sequenced at the Sequencing and Genotyping Facility in the Department of Plant Pathology at Kansas State University (GenBank accession no. AF452568).

Protein alignment: Protein sequences were aligned using Clustal W with the Blosum62mt score matrix (Vector NTI, Informax, North Bethesda, MD). The dendogram shown in Figure 2 was produced using the neighbor-joining algorithm (Saitou and Nei 1987). The GenBank accession numbers of the Zen protein sequences used in this analysis are the following: *Megaselia abdita*, CAB40893; *Schistocerca gregaria*, CAB61208; *D. melanogaster*, AAF54087 and AAF54088; and *T. castaneum*, AAK16425 and AAK16424.

RESULTS

Two Tribolium BAC libraries were screened for clones containing homeotic genes. We sequenced three clones

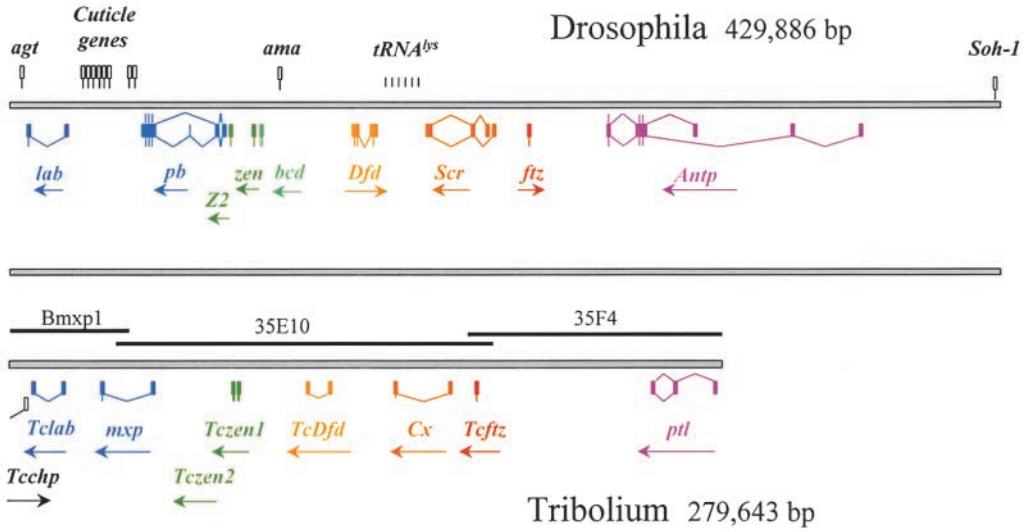

FIGURE 1.—Comparison of Drosophila ANTC and the corresponding region of the Tribolium HOMC. The Tribolium BAC clones are shown above the Tribolium map. The transcription units are shown with boxes denoting exons and vertical or diagonal lines denoting introns. Differential splicing is indicated by introns above and below the exons. Homeobox genes are shown below the chromosome and nonhomeobox genes above, with the exception of *Tcchp*. Arrows indicate the direction of gene transcription. The Tribolium and Drosophila orthologs are color coded. The Drosophila map was constructed using the annotations from GenBank entries AE001572, AE001573, AE001574.

with minimal overlap (Bmxp1, 35E10, and 35F4), which span a region corresponding to the Drosophila ANTC. The resulting 279,643-bp Tribolium genomic contig is delimited at one end by the 3′ end of *chaoptic* (*chp*), the gene flanking *Tclab*. The other end of the contig extends 25 kb upstream of *ptl* (Figure 1). This portion of the Tribolium HOMC is two-thirds the length of the Drosophila ANTC (as defined by the distance between the *agt* and *Soh-1* genes, which flank *lab* and *Antp*, respectively). The G/C content of the Tribolium contig is 33.5%. Fifteen simple sequence repeats comprise 0.3% of the total sequence, only two of which contain any G/C. A total of 203 regions of low complexity, which are dispersed randomly throughout the contig, are predominantly A, T, or A/T rich and comprise 3.25% of the total sequence.

Polymorphisms in Tribolium DNA: The sequences of Bmxp1 and 35E10 match perfectly in the 4.2 kb by which they overlap. In contrast, we found several single nucleotide polymorphisms (SNPs) within the 8-kb overlap of BAC clones 35E10 and 35F4, indicating that the inbred strain used as the source of DNA for this BAC library is not completely isogenic and still contains heterogeneous chromosomal regions (if not entire chromosomes). SNPs were detected at a rate of 1/85 bp, which is similar to the rate of SNP detection in Drosophila (MORIYAMA and POWELL 1996).

Genes within the Tribolium HOMC: The transcription units of the Tribolium homeotic genes *Tclab, mxp, TcDfd, Cx*, and *ptl*, as well as those of *Tczen* and *Tcftz*, were located within the contig by comparing genomic and previously identified cDNA sequences (Figure 1). In the following descriptions of Tribolium homeobox genes, the total length of the transcription units and the length of the large intron upstream of the homeobox (when present) were determined from the sequence described here (Table 1). Exon sizes reported here confirm previous analyses and are included for completeness. *Tclab* is located at the 3′ end of the HOMC. The 13-kb transcription unit is composed of two exons (658 and 816 bp) separated by an 11.6-kb intron located 125 bp upstream of the homeobox. The *mxp* gene, located 13.8 kb upstream of *Tclab*, contains three exons (561, 142, and 1430 bp). The first two are separated by a large intron (18.5 kb) located 11 bp upstream of the homeobox. The second intron (only 51 bp) is located within the homeobox. *TcDfd* is located 57.7 kb upstream of *mxp*. This 9.78-kb transcription unit contains a single intron (7.7 kb) located 29 bp upstream of the homeobox. The Tribolium *Cx* gene is located 24.6 kb farther upstream. It contains three exons (753, 328, and 633 bp). The first two are separated by a 20.1-kb intron located 20 bp upstream of the homeobox. The second exon (160 bp) is located in the 3′ untranslated region (UTR) 24 bp downstream of the translation stop codon. The 922-bp *Tcftz* transcription unit is located 9.7 kb upstream of *Cx*. It contains a single intron (49 bp) located 17 bp upstream of the homeobox.

The Tribolium *zen* genes: The previously described *Tczen* gene is located 31.5 kb upstream of *mxp* (Figure 1). Two short introns (53 and 57 bp) located upstream of the homeobox separate three exons (122, 96, and 613 bp). A homeobox encoding a *zen*-related homeodomain, identified by BLASTX (NCBI), was found be-

TABLE 1

Comparison of homeotic orthologs

Gene	Transcription unit length (kb)[a]	Exons	Encoded protein (aa)
labial	17.0	3	629
Tclabial	13.0	2	354
proboscipedia	33.8	8	798
maxillopedia	20.7	3	654
zen-related	1.0	2	292
zerknüllt	1.35	2	353
Tczen2[b]	0.977	3	246
Tczen1	0.938	3	292
Deformed	10.6	5	586
TcDeformed	9.785	1	412
Sex combs reduced	23.0	3	417
Cephalothorax	22.0	2	313
fushi tarazu	1.9	2	398
Tcfushi tarazu	1.4	2	358
Antennapedia p1	103.0	7	378
p2	36.0	6	378
prothoraxless p1	24.581	2	325
p2	10.342	1	325

aa, amino acids.

[a] Lengths are based on available cDNAs and GenBank annotations.

[b] The length of the 3′ UTR is based on the location of the polyadenylation signal.

tween *Tczen* and *mxp* (Brown *et al.* 2001). The ORF including this homeobox does not contain an initiation methionine. Therefore, we identified an additional upstream ORF, including a candidate initiation methionine, and designed primers to amplify a cDNA derived from this putative gene. The sequence of the resulting 730-bp fragment (GenBank accession no. AF452568) was compared with genomic DNA to identify introns. This newly identified *zen* gene contains three exons (180, 84, and 610 bp) separated by two small introns (55 and 47 bp) that are located upstream of the homeobox. We have changed the name of the previously identified *zen* gene to *Tczen1* and named the newly identified gene *Tczen2*. There are only 216 bp between the 3′ end of *Tczen1* [as determined by the sequence of a cDNA clone (Falciani *et al.* 1996)] and the first candidate initiation methionine within the *Tczen2* ORF. The proteins encoded by the Tribolium *zen* genes share regions of similar amino acids, one of which extends downstream of the homeodomain for several residues. The two Tribolium Zen proteins are more similar to each other than they are to either of the Drosophila Zen proteins, which do not contain sequence similarity to one another outside the homeodomain. We compared the Zen proteins from Tribolium with those from the Orthopteran *S. americana* and two Dipterans, *D. melanogaster* and *M. abdita*. In the dendogram shown in Figure 2, the Drosophila Zen proteins are most closely related

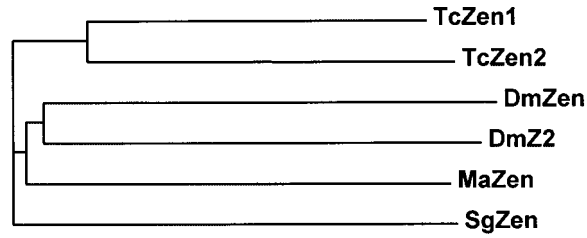

Figure 2.—The phylogenetic relationship of insect Zen proteins. We compared the Zen proteins from Tribolium (TcZen1 and TcZen2) with those from *S. americana* (SgZen), *D. melanogaster* (DmZen and DmZ2), and *M. abdita* (MaZen). The Tribolium and Drosophila Zen proteins form separate groups, suggesting that the *zen* gene pairs in Tribolium and Drosophila arose via independent duplication events in each lineage.

to one another within the group of dipteran Zen proteins, while the two Tribolium Zen proteins form a separate group. These results support the hypothesis that the *zen* gene pairs in Tribolium and Drosophila arose via independent duplication events in each lineage.

The Tribolium ortholog of *Antp*: Comparison of cDNA and genomic sequences provides evidence that at least four different messages are transcribed from the Tribolium *ptl* locus encoding the ortholog of *Antp*. These messages differ at the 3′ end, indicating the use of three different polyadenylation sites. The *ptl* locus contains three exons (512, 862, and 1613 bp) separated by two introns (14.4 and 7.2 kb). In addition to the single cDNA containing all three exons, we sequenced several different cDNAs that lack the first exon and instead contain sequence from the 3′ end of the first intron contiguous with the second exon, indicating the presence of a second promoter downstream of the first (Figure 1). Messages initiated from this promoter contain two exons (1539 and 1613 bp) and the 7.2-kb intron immediately upstream of the homeobox. All identified mRNAs from the Tribolium *ptl* gene encode the same protein, which contains the conserved motifs previously described for Antp orthologs in other hexapods. A detailed analysis of *ptl* will be required to assess differential expression from the two promoters and the use of the multiple polyadenylation sites during development.

DISCUSSION

Comparison of Tribolium Zen proteins: The amino and carboxy terminal sequences, as well as the homeodomain sequences of the eight Tribolium Hox proteins, are aligned in Figure 3. The putative Zen proteins show similarity to the central class proteins (group C in Figure 3) within the N-terminal octapeptide (Zhao *et al.* 1993). Zen1 is particularly interesting in that it contains the same inserted Q residue as the *Scr/Cx* orthologs. Another shared region includes the carboxy terminus. Three of the last four residues (LTxL) are shared not only with each other but also with TcDfd (Figure 3)

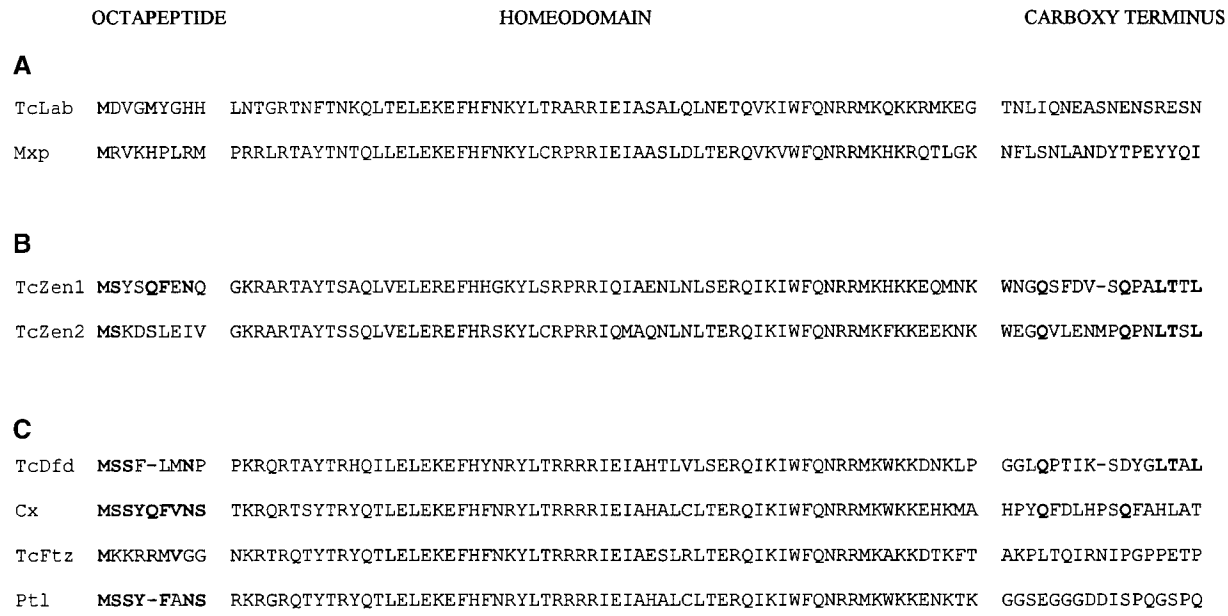

FIGURE 3.—Comparison of Tribolium Hox protein sequences. Sequences from the amino terminus, the homedomain, and the carboxy terminus of each protein are shown. (A) Proteins encoded by anterior-class genes. (C) Proteins transcribed by central-class genes. Residues outside the homeodomain that are shared by the Zen proteins (B) and proteins in C are in boldface type.

and several of its orthologs (data not shown). The Zen/Hox 3 proteins of grasshopper, *S. gregaria* (GenBank entry CAB61208), and a polychaete worm, *Chaetopterus variopedatus* (GenBank entry AAD55934), also end in LTxL. Short motifs, such as the WRPY motif located at the carboxy terminus of Hairy and Runt proteins (WAINWRIGHT and ISH-HOROWICZ 1992), can be functionally important. Although the function of the motifs in the Tribolium proteins is not known, such sequence conservation outside the homeodomain supports the hypothesis that the Hox 3 gene family arose via duplication of the central class precursor (ZHANG and NEI 1996) rather than the anterior class precursor (SCHUBERT *et al.* 1993).

Comparison of Tribolium and Drosophila genes: In Table 1 the transcription unit length, number of exons per gene, and size of the proteins encoded by the Tribolium genes are compared with their Drosophila orthologs. The length and gene content of the intergenic regions are compared in Table 2. *Tclab* is ~4 kb shorter than *lab*. The location of the first intron, between the YKWM motif-encoding region and the homeobox, is conserved. However, the *lab* homeobox contains an intron not found in *Tclab*. Sequence information from additional species will be required to determine whether *lab* gained this intron or *Tclab* lost it. In the 29.6-kb region between *lab* and *pb*, there are eight cuticle genes (GenBank accession no. AE003674). However, no genes are in the corresponding interval in Tribolium, which is only 13.8 kb (Table 2 and NIE *et al.* 2001). *pb* is ~13 kb larger than *mxp* and contains five more exons. The location of the two introns in *mxp*, one in the homeobox and one 11 bp upstream of the homeobox, are conserved in *pb*. Again, in the absence of additional data it is impossible to determine whether the Drosophila *pb* gene gained introns or the Tribolium *mxp* gene lost them.

Both insects contain two *zen* genes; however, these genes appear to have arisen via independent duplica-

TABLE 2

Comparison of intergenic distances

Interval	Organism	Distance (kb)[b]	Genes in interval
lab-pb[a]	Dm	29,623	8 cuticle
	Tc	13,853	—
pb-zen2	Dm	1,129	—
	Tc	31,362	—
zen2-zen1	Dm	8,711	—
	Tc	217	—
zen1-Dfd	Dm	37,637	CG1162,
	Tc	26,100	*bcd, ama*
Dfd-Scr	Dm	20,752	6 $tRNA^{lys}$
	Tc	24,601	—
Scr-ftz	Dm	14,291	—
	Tc	9,747	—
ftz-Antp	Dm	30,038	—
	Tc	66,838	—

[a] Drosophila gene names are used.
[b] Distances do not include transcription units.

tion events in each lineage and show species-specific features. For example, the two Drosophila *zen* genes contain two exons while the two Tribolium *zen* genes contain three. In Tribolium there is very little distance between the *zen* genes (217 bp), and they are located 31 kb upstream of *mxp* (Table 2). In contrast, *zen2* is within 1.2 kb of *pb* in Drosophila, and >8 kb separates *zen1* and *zen2*. Three genes (*bcd*, *ama*, and CG1162) are located in the 37-kb region between *zen1* and *Dfd* in Drosophila. The corresponding region is somewhat smaller in Tribolium (26 kb) and does not appear to contain any genes. *TcDfd* contains fewer exons than *Dfd* (two *vs.* five, Table 1; Brown *et al.* 1999). However, the transcription units are similar in size. This is due in large part to the presence of a long intron (>9 kb in each insect) immediately upstream of the homeobox. *TcDfd* is transcribed in the same orientation as the other genes. In contrast, *Dfd* is transcribed in the opposite orientation. Six lysine tRNA genes are in the 20.7-kb region between *Dfd* and *Scr*. Although slightly larger (24.6 kb), the corresponding region in Tribolium does not contain any genes. *Scr* and *Cx* contain four and two exons, respectively. *Scr* transcripts initiate at two promoters (Andrew 1995; Figure 1). However, we have found evidence only for a single promoter in *Cx* (Curtis *et al.* 2001; Figure 1). *Tcftz* and *ftz* are similar in that they are relatively short transcription units (1.4 and 1.9 kb) located upstream of the *Cx/Scr* orthologs (9.7 and 14.3 kb, respectively) and contain a single small intron upstream of the homeobox. However, *Tcftz* is transcribed in the same direction as the other genes, whereas *ftz* is transcribed in the opposite direction. The distance between *Tcftz* and *ptl* is more than twice that between *ftz* and *Antp* (66 *vs.* 30 kb, Table 2). *Antp* contains two promoters and eight exons that are differentially spliced to produce different mRNAs. *ptl* is similar to *Antp* in that it contains two promoters and produces multiple mRNAs. However, it contains only three exons.

Although the Tribolium genes encode smaller proteins (Table 1) and, with the exception of the *zen* genes, contain fewer introns, they are only slightly shorter than their Drosophila counterparts (Table 1). The *Antp/ptl* orthologs are an exception to this observation in that the longest *Antp* transcription unit is >100 kb in length, while the longest *ptl* transcription unit is only 24 kb. Most of this difference is due to the distance between promoters. Several of the Drosophila *Antp* gain-of-function (GOF) mutations result from translocation breakpoints located between the first and second promoter (Schneuwly *et al.* 1986). Assuming that corresponding breaks would produce similar mutant phenotypes in Tribolium, it is tempting to speculate that the lack of GOF mutations mapping to the *ptl* locus reflects the considerably smaller target region between the two promoters.

It has been suggested that the large transcription units in Drosophila prevent them from being expressed prematurely during rapid, long-germ development (counted in hours; Shermoen and O'Farrell 1991). Similar correlation in Tribolium seems unlikely, since the embryos require 3–4 days to develop at 30° and the early divisions alone are counted in hours (Handel *et al.* 2000 and our unpublished observations).

Search for a *bicoid* ortholog: Comparison of Bicoid and Zen homeodomain sequences from Megaselia, a basal cyclorrhaphan fly, and Drosophila suggests that *bicoid* may have resulted from a recent duplication of the *zen* gene (Stauber *et al.* 2000). Further, *bicoid* is located upstream of *zen* in both Megaselia and Drosophila (Brown *et al.* 2001), suggesting this might be its ancestral location. However, we find no evidence for any additional genes, including *bicoid*, in the Tribolium contig described here. Thus, if Tribolium does contain an ortholog of *bicoid* it must be located outside this portion of the HOMC.

Large intergenic distances: The distance between homeobox genes in Tribolium and Drosophila is compared in Table 2. Additional genes found in the Drosophila ANTC are noted in the appropriate intervals. As noted above, the Tribolium *zen* genes are quite distant from neighboring genes. In addition, the distance between *ftz* and *Antp/ptl* orthologs is twice as far in Tribolium (60 kb) as it is in Drosophila (30 kb). Such differences in intergenic spacing may contribute to the evolution of regulatory differences between the Drosophila and Tribolium orthologs.

The intergenic distances in Drosophila and Tribolium are considerably larger than those in any deuterostome Hox cluster. It has been suggested that the differences between Drosophila and vertebrate complexes (larger intergenic distances, longer transcription units, additional nonhomeotic genes, inversions, and breaks in the Drosophila complexes) reflect a loosening of structural and regulatory constraints in Drosophila that, in other organisms, might be required to maintain temporal, if not spatial, colinearity (Ferrier and Akam 1996). These large intervals are known to contain widely spaced regulatory elements in Drosophila (Gindhart *et al.* 1995) and may be important to regulation in Tribolium as well. In the mosquito *Anopheles gambiae*, the interval containing the orthologs of *lab* and *Antp* is >350 kb. Long distance Southern mapping indicates similarly distant spacing of Hox genes in the grasshopper, *S. gregaria* (Ferrier and Akam 1996). Thus, it is likely that large size and dispersed regulatory regions are common to hexapod Hox clusters.

Putative regulatory regions upstream of identified genes were searched for known binding sites. Several consensus homeodomain binding sites were identified by MatInspector in the 5-kb regions immediately upstream of *mxp*, *TcDfd*, and *Cx*. To understand the significance, if any, of these sites, it will be necessary to analyze these regions for their ability to direct gene-specific expression patterns of reporter genes. Such studies are

now feasible in Tribolium using a recently described transformation system (Berghammer *et al.* 1999)

The lack of inversions, breaks, and nonhomeobox genes in the region of the Tribolium HOMC corresponding to the Drosophila ANTC suggests that the structure of the Tribolium HOMC is more constrained than that of the Drosophila clusters. To determine whether such structural constraints have regulatory consequences, it will be necessary to determine whether Tribolium Hox genes can function outside the chromosomal context of the complex. This can be accomplished by examining homeotic mutants that might carry translocations or other chromosomal rearrangements or by attempting to rescue the effects of a homeotic mutation by reintroducing a copy of the wild-type gene into the Tribolium genome.

We thank Barb Van Slyke for expert technical assistance. We thank Scott Thomson for the inbred Tribolium strain and Exelixis for the kind gift of the Tribolium BAC library. This work was supported by grants from the National Institutes of Health and the National Science Foundation to R. Denell and S. Brown. E. Richardson was supported by a Howard Hughes Medical Institute undergraduate research fellowship. Contribution no. 02-7-J is from the Kansas Agricultural Experiment Station.

LITERATURE CITED

Acampora, D., M. D'Esposito, A. Faiella, M. Pannese, E. Migliaccio *et al.*, 1989 The human HOX gene family. Nucleic Acids Res. **17:** 10385–10402.

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

Andrew, D. J., 1995 The *Sex combs reduced* gene of *Drosophila melanogaster* has multiple transcripts. Gene **152:** 149–155.

Beeman, R. W., 1987 A homoeotic gene cluster in the red flour beetle. Nature **327:** 247–249.

Bender, W., M. E. Akam, F. Karch, P. A. Beachy, M. Peifer *et al.*, 1983 Molecular genetics of the bithorax complex in *Drosophila melanogaster*. Science **221:** 23–29.

Bennett, R. L., S. J. Brown and R. E. Denell, 1999 Molecular and genetic analysis of the *Tribolium Ultrabithorax* ortholog, *Ultrathorax*. Dev. Genes Evol. **209:** 608–619.

Berghammer, A. J., M. Klingler and E. A. Wimmer, 1999 A universal marker for transgenic insects. Nature **402:** 370–371.

Brown, S. J., R. B. Hilgenfeld and R. E. Denell, 1994 The beetle *Tribolium castaneum* has a *fushi tarazu* homolog expressed in stripes during segmentation. Proc. Natl. Acad. Sci. USA **91:** 12922–12926.

Brown, S. J., J. K. Parrish, R. W. Beeman and R. E. Denell, 1997 Molecular characterization and embryonic expression of the *even-skipped* ortholog of *Tribolium castaneum*. Mech. Dev. **61:** 165–173.

Brown, S. J., S. Holtzman, T. Kaufman and R. Denell, 1999 Characterization of the Tribolium *Deformed* ortholog and its ability to directly regulate Deformed target genes in the rescue of a Drosophila *Deformed* null mutant. Dev. Genes Evol. **209:** 389–398.

Brown, S. J., M. Decamillis, K. Gonzalez-Charneco, M. Denell, R. Beeman *et al.*, 2000 Implications of the *Tribolium Deformed* mutant phenotype for the evolution of Hox gene function. Proc. Natl. Acad. Sci. USA **97:** 4510–4514.

Brown, S., J. Fellers, T. Shippy, R. Denell, M. Stauber *et al.*, 2001 A strategy for mapping *bicoid* on the phylogenetic tree. Curr. Biol. **11:** R43–R44.

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

C. Elegans Sequencing Consortium, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012–2018.

Cook, C. E., M. L. Smith, M. J. Telford, A. Bastianello and M. Akam, 2001 Hox genes and the phylogeny of the arthropods. Curr. Biol. **11:** 759–763.

Curtis, C. D., J. A. Brisson, M. A. Decamillis, T. D. Shippy, S. J. Brown *et al.*, 2001 Molecular characterization of *Cephalothorax*, the Tribolium ortholog of *Sex combs reduced*. Genesis **30:** 12–20.

D'Esposito, M., F. Morelli, D. Acampora, E. Migliaccio, A. Simeone *et al.*, 1991 EVX2, a human homeobox gene homologous to the *even-skipped* segmentation gene, is localized at the 5′ end of HOX4 locus on chromosome 2. Genomics **10:** 43–50.

Devenport, M. P., C. Blass and P. Eggleston, 2000 Characterization of the Hox gene cluster in the malaria vector mosquito, Anopheles gambiae. Evol. Dev. **2:** 326–339.

Ewing, B., L. Hillier, M. C. Wendl and P. Green, 1998 Basecalling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:** 175–185.

Faiella, A., M. D'Esposito, M. Rambaldi, D. Acampora, S. Balsofiore *et al.*, 1991 Isolation and mapping of EVX1, a human homeobox gene homologous to *even-skipped*, localized at the 5′ end of HOX1 locus on chromosome 7. Nucleic Acids Res. **19:** 6541–6545.

Falciani, F., B. Hausdorf, R. Schröder, M. Akam, D. Tautz *et al.*, 1996 Class 3 Hox genes in insects and the origin of *zen*. Proc. Natl. Acad. Sci. USA **93:** 8479–8484.

Ferrier, D. E. K., and M. Akam, 1996 Organization of the Hox gene cluster in the grasshopper, *Schistocerca gregaria*. Proc. Natl. Acad. Sci. USA **93:** 13024–13029.

Finnerty, J. R., and M. Q. Martindale, 1998 The evolution of the Hox cluster: insights from outgroups. Curr. Opin. Genet. Dev. **8:** 681–687.

Garcia-Fernàndez, J., and P. W. H. Holland, 1994 Archetypal organization of the amphioxus *Hox* gene cluster. Nature **370:** 563–566.

Gindhart, Jr., J. G., A. N. King and T. C. Kaufman, 1995 Characterization of the *cis*-regulatory region of the Drosophila homeotic gene *Sex combs reduced*. Genetics **139:** 781–795.

Gordon, D., C. Abajian and P. Green, 1998 Consed: a graphical tool for sequence finishing. Genome Res. **8:** 195–202.

Graham, A., N. Papalopulu and R. Krumlauf, 1989 The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell **57:** 367–378.

Grenier, J. K., T. L. Garber, R. Warren, P. M. Whitington and S. Carroll, 1997 Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. Curr. Biol. **7:** 547–553.

Handel, K., C. G. Grünfelder, S. Roth and K. Sander, 2000 *Tribolium* embryogenesis: a SEM study of cell shapes and movements from blastoderm to serosal closure. Dev. Genes Evol. **210:** 167–179.

Kaufman, T. C., M. A. Seeger and G. Olson, 1990 Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. Adv. Genet. **27:** 309–362.

Kim, C.-B., C. Amemiya, A. Bailey, K. Kawasaki, J. Mezey *et al.*, 2000 Hox cluster genomics in the horn shark, *Heterodontus francisci*. Proc. Natl. Acad. Sci. USA **97:** 1655–1660.

Martin, C. H., C. A. Mayeda, C. A. Davis, C. L. Ericsson, J. D. Knafels *et al.*, 1995 Complete sequence of the bithorax complex of Drosophila. Proc. Natl. Acad. Sci. USA **92:** 8398–8402.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

Nie, W., B. Stronach, G. Panganiban, T. Shippy, S. Brown *et al.*, 2001 Molecular characterization of *Tclabial* and the 3′ end of the Tribolium homeotic complex. Dev. Genes Evol. **211:** 244–251.

Popodi, E., J. C. Kissinger, M. E. Andrews and R. A. Raff, 1996 Sea urchin Hox genes: insights into the ancestral Hox cluster. Mol. Biol. Evol. **13:** 1078–1086.

Powers, T. P., J. Hogan, Z. Ke, K. Dymbrowski, X. Wang *et al.*, 2000 Characterization of the Hox cluster from the mosquito *Anopheles gambiae* (Diptera; culicidae). Evol. Dev. **2:** 311–325.

Quandt, K., K. Frech, H. Karas, E. Wingender and T. Werner, 1995 MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res. **23:** 4878–4884.

Reese, M. G., F. H. Eeckman, D. Kulp and D. Haussler, 1997 Improved splice site detection in Genie. J. Comput. Biol. **4:** 311–323.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new

method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Schneuwly, S., A. Kuroiwa, P. Baumgartner and W. J. Gehring, 1986 Structural organization and sequence of the homeotic gene *Antennapedia* of *Drosophila melanogaster*. EMBO J. **5:** 733–739.

Schubert, F. R., K. Nieselt-Struwe and P. Gruss, 1993 The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution. Proc. Natl. Acad. Sci. USA **90:** 143–147.

Scott, M. P., and A. J. Weiner, 1984 Structural relationships among genes that control development: sequence homology between the *Antennapedia*, *Ultrabithorax*, and *fushi tarazu* loci of Drosophila. Proc. Natl. Acad. Sci. USA **81:** 4115–4119.

Shermoen, A. W., and P. H. O'Farrell, 1991 Progression of the cell cycle through mitosis leads to abortion of nascent transcripts. Cell **67:** 303–310.

Shippy, T. D., S. J. Brown and R. E. Denell, 1998 Molecular characterization of the Tribolium *abdominal-A* ortholog and implications for the products of the Drosophila gene. Dev. Genes Evol. **207:** 446–452.

Shippy, T. D., S. J. Brown and R. E. Denell, 2000a *maxillopedia* is the Tribolium ortholog of *proboscipedia*. Evol. Dev. **2:** 145–151.

Shippy, T. D., J. Guo, S. J. Brown, R. W. Beeman, M. S. Haas *et al.*, 2000b Analysis of *maxillopedia* expression pattern and larval cuticular phenotypes of wild type and mutant Tribolium. Genetics **155:** 721–731.

Stauber, B., H. Jackle and U. Schmidt-Ott, 2000 The anterior determinant bicoid of Drosophila is a derived Hox class 3 gene. Proc. Natl. Acad. Sci. USA **96:** 3786–3789.

Struhl, G., K. Struhl and P. M. Macdonald, 1989 The gradient morphogen *bicoid* is a concentration-dependent transcriptional activator. Cell **57:** 1259–1273.

Stuart, J. J., S. J. Brown, R. W. Beeman and R. E. Denell, 1993 The *Tribolium* homeotic gene *Abdominal* is homologous to *abdominal-A* of the *Drosophila* Bithorax complex. Development **117:** 233–243.

Wainwright, S. M., and D. Ish-Horowicz, 1992 Point mutations in the *Drosophila* hairy gene demonstrate in vivo requirements for basic, helix-loop-helix, and WRPW domains. Mol. Cell. Biol. **12:** 2475–2483.

Zhang, J. Z., and M. Nei, 1996 Evolution of Antennapedia-class homeobox genes. Genetics **142:** 295–303.

Zhao, J. J., R. A. Lazzarini and L. Pick, 1993 The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila Sex combs reduced* gene. Genes Dev. **7:** 343–354.

Zhu, H., S. Choi, A. K. Johnston, R. A. Wing and R. A. Dean, 1997 A large-insert (130 kbp) bacterial artificial chromosome library of the rice blast fungus Magnaporthe grisea: genome analysis, contig assembly, and gene cloning. Fungal. Genet. Biol. **21:** 337–347.

Communicating editor: T. C. Kaufman