Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 9971–9975, December 1989
Evolution

Comparison of homeobox-containing genes of the honeybee and *Drosophila*

(gene isolation/homeobox sequence/evolutionary conservation)

UWE WALLDORF, RICHARD FLEIG*, AND WALTER J. GEHRING†

Department of Cell Biology, Biocenter, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Contributed by Walter J. Gehring, September 27, 1989

ABSTRACT **We report the isolation of seven homeobox-containing genes from the honeybee (*Apis mellifera*). Sequence analysis of all homeoboxes and some flanking sequences showed that six of seven genes are more than 90% identical to their corresponding *Drosophila* homologues within the homeobox and, with one exception, also in the flanking sequences. The homologues that were identified include three homeotic selector genes [Sex combs reduced (*Scr*), Antennapedia (*Antp*), and abdominal-A (*abd-A*); the two engrailed (*en*) genes; and the muscle segment homeobox (*msh*)]. Surprisingly, no homologue of the segmentation gene fushi tarazu was found in the honeybee. For the remaining bee gene, a *Drosophila* homologue is not known. This indicates that, with some exceptions, structurally homologous genes are involved in the control of bee and *Drosophila* development, although Hymenoptera differ significantly in their embryogenesis from Diptera and have evolved separately for some 250 million years.**

Bodies of insects consist of a series of segments, each of which gives rise to different and often highly specialized structures. In *Drosophila*, early embryonic events leading to this segmentation pattern have been found to require two classes of zygotic genes. Segmentation genes establish the basic metameric unit, and homeotic genes determine the identity and sequence of the individual segments (1, 2). Many of these genes include a short region of sequence similarity, the homeobox (3–5), which encodes a DNA-binding protein domain (6) and which has been found in a number of organisms (see refs. 7 and 8 for review). In the fruit fly, homeotic genes are organized in two major clusters, the Antennapedia complex (*ANT-C*) and the Bithorax complex (*BX*-C). *ANT-C* genes are involved in controlling the differentiation of more anterior segments (head and thorax) (9), whereas *BX*-C genes are required for more posterior ones (thorax and abdomen) (10). Since the Diptera are highly specialized insects, it cannot be assumed that the characteristics of these *Drosophila* genes are necessarily similar in other insects. To address this question, we isolated homeobox-containing genes from the honeybee (*Apis mellifera*), which belongs to the order of Hymenoptera. This organism has the same type of genomic organization as *Drosophila*—i.e., the long interspersed repeated DNA type (11)—but diverged about 250×10^6 years ago (12). It is also a "long-germ" insect in which the entire germ band becomes segmented more or less simultaneously but does not show germ band elongation and retraction as it does in *Drosophila*. Development of *Drosophila* and other Diptera is characterized by head involution and the reduction of the last two abdominal segments. In this respect the honeybee represents an insect with a less specialized developmental pattern than *Drosophila*.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Our aim is to compare in detail the homeobox genes of these two representatives of developmentally distinct insect orders. Earlier we identified a homologue of the *Drosophila* gene Deformed (*Dfd*) (13). Here we report the isolation of seven homeobox-containing genes of the honeybee. These genes belong to at least three different classes of homeoboxes; six of them can be identified as homologues of known *Drosophila* genes on the basis of their DNA sequence.

MATERIALS AND METHODS

Materials. Honeybees (*Apis mellifera*) were collected from bee colonies of the University of Freiburg, Germany. Radionucleotides were purchased from Amersham, and enzymes were from Biofinex, Boehringer Mannheim, and Amersham. DNA fragments used as probes for hybridization were derived from the following plasmids: *Antp* p903 (3); *Dfd* Hp0.25 (13); *ftz* pGEMF1 (14); *Scr* cY20 (15); *Ubx* p96 (3); *en* pS799-7 (16); *cad* cDNA 335 (17); *msh* pS135-4 (B. Jacq and W.J.G., unpublished data); *W13* Sa0.3 (U.W., unpublished data).

General Methods. Preparation of genomic *Drosophila* and honeybee DNA was as described (18). Restriction endonuclease digestions, gel electrophoresis of DNA fragments, screening of libraries, isolation of phage and plasmid DNA, and buffers were as described by Maniatis *et al.* (19).

Library Construction. Genomic *Apis mellifera* DNA isolated from adult worker bees was partially digested with *Sau*3A. After size selection on NaCl gradients, DNA of 15–20 kilobases (kb) was cloned into the *Bam*HI site of the EMBL-4 phage λ vector (20).

Screening and Hybridization Conditions. Usually, 1.5×10^5 phages (5 genome equivalents; see ref. 11) were screened under low-stringency conditions as described earlier (13).

DNA Sequencing. DNA sequencing procedures were carried out by using phage M13 cloning (21) and chain-termination sequencing (22). Both strands of the DNA were sequenced.

RESULTS AND DISCUSSION

Using Southern blot analysis, we tried to identify DNA fragments of honeybee DNA that hybridize with different *Drosophila* homeobox probes under low-stringency conditions. Nine different probes showed essentially five different patterns (Fig. 1). Homeobox probes of the Antennapedia class (*Dfd*, *ftz*, *Antp*, *Scr*, *Ubx*) always detected the same strong bands of 9.0, 5.5, 4.2, and 1.5 kb. Some other relatively strong bands were only revealed by the *Ubx* probe; additional bands seen with *ftz*, *Antp*, and *Scr* probes were much weaker. All the other probes used gave rise to unique patterns with only two bands (*en*), three bands (*W13*), or multiple bands with two being stronger than the rest (*msh*) or with only very weak ones (*cad*).

*Present address: Institut für Biologie I (Zoologie), Universität Freiburg, Albertstr. 21a, D-7800 Freiburg, F.R.G.
†To whom reprint requests should be addressed.
‡The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M29488–29494).

In an attempt to analyze some of these homeobox fragments of the honeybee, we constructed a genomic library of *Apis mellifera* and screened the library with *Antp*, *en*, *msh*, and *W13* probes. During this screen we focused only on very strong positive signals to start isolating clones with the highest degree of conservation. However, when we used the *en* probe, we also wanted to isolate the weakly hybridizing fragment to test whether the honeybee, like *Drosophila*, has two engrailed-like genes. Restriction analysis of all isolated clones showed that they belong to eight different types. These include the four bands seen with the Antennapedia-class probes, the two bands seen with *en*, the two strong bands seen with *msh*, and one band seen with *W13*. These are indicated in Fig. 1 and named according to their size. Fragments H17 and H10 are located in the same phage and belong to the same gene. Fragments of each type that cross-hybridize to homeobox probes were subcloned in plasmid vectors as *Eco*RI fragments except for the 9.0-kb fragment H90, of which a smaller *Bgl* II fragment was subcloned. Restriction maps were established for all fragments (Fig. 2), the regions of similarity were narrowed down by hybridization, and finally the relevant fragments were sequenced.

On the basis of its sequence, H55 is clearly homologous to *Scr* of *Drosophila*, since within a region of 86 amino acids of its derived protein sequence, only two amino acid differences are present—one within the homeobox and one within a region of 19 amino acids 3′ of the homeobox (Fig. 3). Also, the 7 amino acids 5′ of the homeobox up to a 3′ splice site of *Scr* are conserved. A *Scr* homologue was also found in the locust *Schistocerca gregaria* (23). This gene has an identical homeobox as that of *Scr* in *Drosophila*, but sequence conservation 3′ of the homeobox is less pronounced than in the bee H55 clone. Previously, we found an even better sequence conservation in clone H42, a homologue of *Dfd* (Deformed), in which 78 amino acids of the derived protein sequence including the homeobox are conserved as compared with *Dfd* in *Drosophila* (13). Furthermore, *in situ* hybridization experiments show that the bee gene H42 is expressed in the corresponding region of the embryo as *Dfd* in *Drosophila* (13).

Comparison of the homeobox sequence of H15 (Fig. 4) with that of *abd-A* (abdominal-A) (F. Karch, personal communication; ref. 23) revealed that it is the homologue of *abd-A*. Its product differs from the *Drosophila* gene product by one amino acid within the homeobox and two in a region of 22 amino acids 3′ of the homeobox. Beyond the 5′ end of the homeobox, six of seven amino acids are conserved.

The greatest similarity to *Antp* was found within the homeobox of H90 (Fig. 4). Only phenylalanine-22 in the *Antp* homeobox is changed to tyrosine. Outside the homeobox, the similarity is not very striking; here only four positions 3′ of

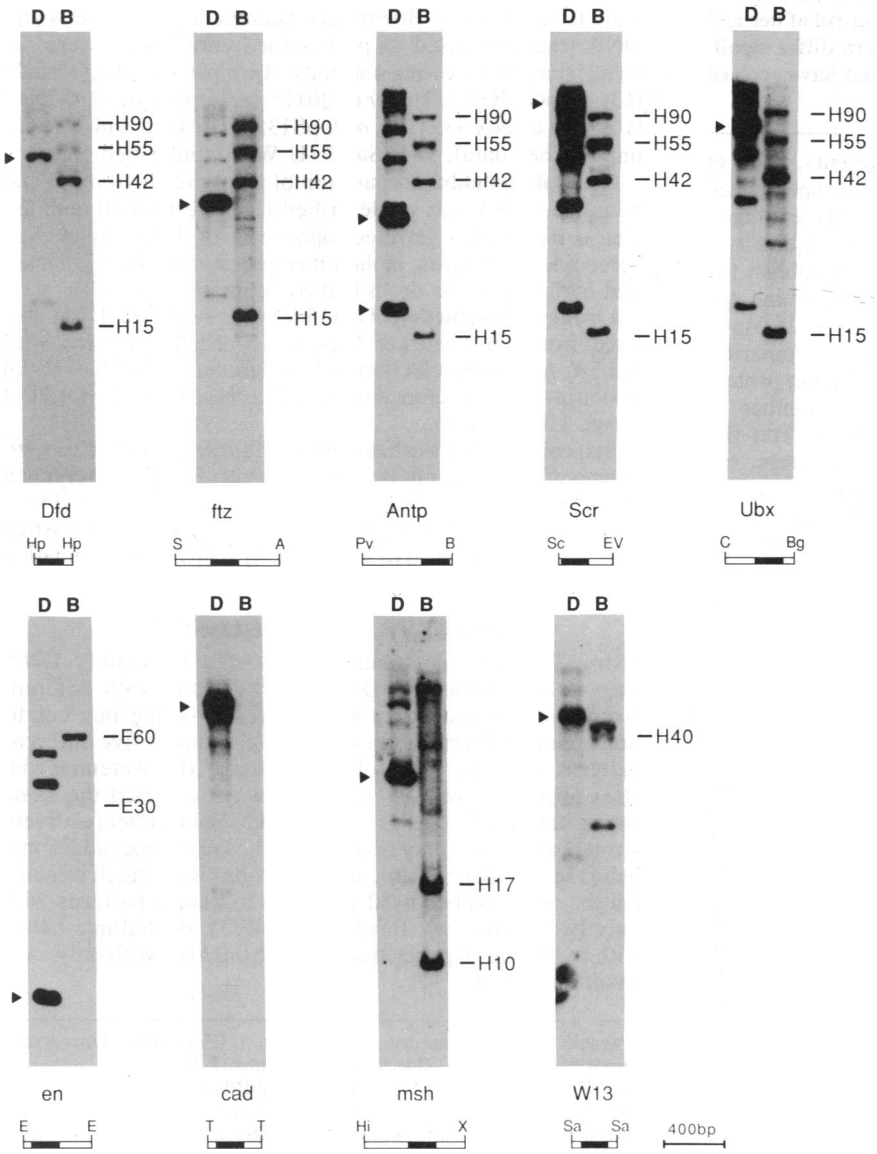

FIG. 1. Hybridization of *Drosophila* homeobox probes to genomic DNA from *Drosophila* and *Apis*. Genomic Southern blots with 2.5 μg of DNA digested with *Eco*RI from *Drosophila melanogaster* (lanes D) or 7.5 μg of such DNA from *A. mellifera* (lanes B) were hybridized under reduced-stringency conditions with different *Drosophila* gene probes. The probes are indicated at the bottom; the black bars show the localization and extent of the homeobox within each probe. A, *Ava* I; B, *Bam*HI; Bg, *Bgl* II; C, *Cla* I; E, *Eco*RI; EV, *Eco*RV; Hi, *Hin*cII; Hp, *Hpa* II; S, *Sal* I; Sa, *Sac* I; Sc, *Sca* I; T, *Taq* I; and X, *Xho* I.

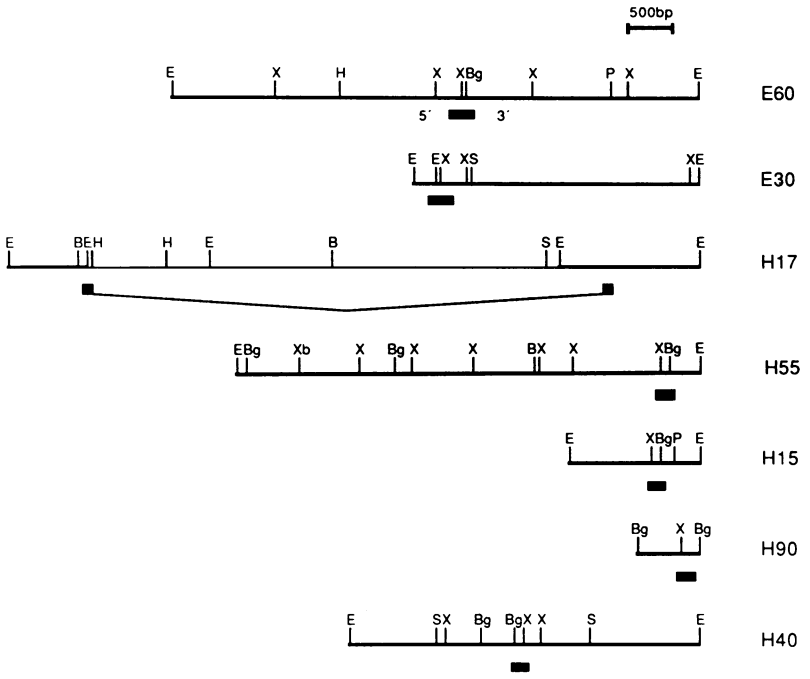

FIG. 2. Restriction maps of isolated homeobox fragments from *Apis mellifera*. DNA maps of seven *Apis* fragments that cross-hybridize with *Drosophila* homeobox probes were shown. The location of the homeobox within these fragments is shown by the black box. The 5'–3' orientation is always from left to right. The fragments are named according to their size, except for H90, where the map of a 0.7-kb *Bgl* II fragment derived from the original 9.0-kb *Eco*RI fragment is shown. The homeobox of clone H17 includes an intron, and each homeobox exon is located on a different *Eco*RI fragment. Both are separated by 5.4 kb of intron sequences indicated by the thin line. B, *Bam*HI; Bg, *Bgl* II; E, *Eco*RI; H, *Hin*dIII; P, *Pst* I; S, *Sal* I; X, *Xho* I; and Xb, *Xba* I.

the box are conserved, while in the 5' direction, no similarities were found. From the sequence comparison it is not clear whether H90 is the *Antp* homologue, but its hybridization pattern in tissue sections greatly resembles that of *Antp* (R.F., unpublished data), which suggests that it is indeed a homologue. Since we analyzed only the homeobox-containing genomic exon in our experiments, we did not identify possible sequence conservation in other exons.

Because our probes for *Dfd*, *ftz*, *Antp*, *Scr*, and *Ubx* all hybridized strongly with the same four fragments, and since all of these were identified, there appears to be no *ftz* homologue with the same level of sequence conservation in the honeybee genome. Another screening of the genomic library with a *ftz* probe only ended with the reisolation of clones already isolated with the *Antp* probe. Since it is possible that an *Eco*RI site within the homeobox of a putative *ftz* homologue might lead to a signal reduction from the resulting two *Eco*RI fragments when these were hybridized with the *ftz* probe, Southern blots of genomic DNA digested with different enzymes were performed. All four enzymes tested consistently yielded a pattern of four strong bands (data not shown). There was no sign of a split signal. If there is a *ftz* homologue, the homeobox-containing exon of this gene must have diverged considerably from the other genes of the Antennapedia class. This could mean that a segmentation gene like *ftz* has diverged much more during evolution than the homeotic genes analyzed or that the *ftz* function is performed by another gene. We do not know if the other segmentation genes behave similarly. Individual components of the developmental program in the two species might be different, which could prevent identification of such components by homology studies.

In contrast to the Antennapedia class genes, the two *Drosophila* genes of the engrailed class [engrailed (*en*) and invected (*inv*)] show a much stronger conservation during evolution. Homologues of these genes have been identified in the mouse (24) and in the sea urchin *Tripneustes gratilla* (25). By using the *Drosophila en* probe, clones E60 and E30 were isolated from the honeybee library. Comparison of the derived amino acid sequences of these two clones reveals strong conservation at the amino acid level in the homeobox (97%) and in a region of 22 amino acids 3' of the box (100%) (Fig. 5). In contrast to this, similarity 5' of the putative homeodomain within a region of 19 amino acids is less distinct but still significant (69%). The DNA

```
                                        Thr Val Asn Ala Asn Gly Glu
H55                                     ACG GTG AAC GCG AAC GGC GAG
Scr                                     ACT GTG AAT GCC AAT GGC GAG
                                        Thr Val Asn Ala Asn Gly Glu

    Val Lys Arg Gln Arg Thr Ser Tyr Thr Arg Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe
H55 GTG AAA CGG CAG AGG ACA TCC TAC ACC AGG TAC CAG ACG TTG GAG CTC GAG AAG GAG TTC
Scr ACG AAA CGA CAA CGG ACC TCA TAC ACC CGC TAC CAG ACG CTG GAG CTG GAG AAG GAG TTC
    Thr Lys Arg Gln Arg Thr Ser Tyr Thr Arg Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe

    His Phe Asn Arg Tyr Leu Thr Arg Arg Arg Arg Ile Glu Ile Ala His Ala Leu Cys Leu
H55 CAC TTC AAC CGA TAC TTG ACC AGG CGG CGT CGG ATC GAG ATC GCG CAC GCC CTC TGC CTC
Scr CAC TTC AAC CGC TAC CTG ACC CGC CGC CGC AGA ATC GAG ATC GCG CAT GCC CTG TGT CTC
    His Phe Asn Arg Tyr Leu Thr Arg Arg Arg Arg Ile Glu Ile Ala His Ala Leu Cys Leu

    Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys Glu His
H55 ACG GAG CGG CAG ATC AAG ATC TGG TTC CAG AAT CGG CGG ATG AAG TGG AAA AAG GAG CAC
Scr ACG GAG CGG CAG ATC AAG ATC TGG TTC CAT AAC CGG CGC ATG AAG TGG AAG AAG GAG CAC
    Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys Glu His

    Lys Met Ala Ser Met Asn Ile Val Pro Tyr His Met Ser Pro Tyr Gly His Pro Tyr
H55 AAG ATG GCG AGC ATG AAC ATT GTA CCG TAC CAC ATG TCG CCT TAC GGC CAC CCT TAC
Scr AAG ATG GCC TCG ATG AAC ATC GTA CCC TAC CAC ATG GGT CCA TAT GGC CAC CCG TAC
    Lys Met Ala Ser Met Asn Ile Val Pro Tyr His Met Gly Pro Tyr Gly His Pro Tyr
```

FIG. 3. Comparison of H55 and *Drosophila Scr* sequences. DNA and putative amino acid sequences of the homeobox regions in H55 are aligned with homologous regions from *Drosophila Scr*. The homeobox is boxed; amino acids of H55 that differ from those derived from *Scr* are shown by small boxes.

```
H15                                                     CCA GGG CCG AAC GGG TGT CCG
H40                                                     CGG CGG TGG GAC AGG CGG GAA
H90                                                     CTC TCG CAT TGT GTT CCA GAG

H15   AGG CGC AGG GGG CGG CAG ACG TAC ACG CGT TTC CAG ACC CTC GAG CTC GAG AAG GAG TTC
H40   GCG AGG AGG GCC AGG ACG GCG TTC ACT TAC GAG CAA CTG GTC GCC CTC GAG AAC AAG TTC
H90   AGG AAA CGA GGC CGG CAA ACG TAC ACC CGA TAC CAA ACC CTC GAG CTC GAG AAG GAG TTC

H15   CAC TAC AAC CAC TAC CTG ACG CGA CGG CGG CGA ATA GAA ATC GCG CAC GCG CTC TGC CTT
H40   AAG ACG ACG AGA TAC CTG TCT GTG TGC GAG CGG CTC AAC CTG GCC CTC TCG CTC TCG TTG
H90   CAC TAC AAC CGA TAC CTG ACC AGG CGG CGT CGC ATC GAG ATC GCG CAC GCC CTC TGC CTT

H15   ACC GAG CGG CAG ATC AAG ATC TGG TTC CAG AAT CGG CGG ATG AAA TTG AAG AAG GAG TTG
H40   ACC GAG ACC CAG GTG AAG ATC TGG TTC CAG AAC AGG CGC ACC AAG TGG AAG AAG CAG AAC
H90   ACC GAG CGG CAA ATC AAA ATC TGG TTT CAA AAC AGA CGG ATG AAA TGG AAG AAG GAG AAC

H15   AGG GCG GTG AAA GAG ATA AAC
H40   CCG GGC CTC GAC GTG ATC AGC
H90   GCT CGG GCG ACG GGG ACA CCG


H15                                                     Asn Gly Pro Asn Gly Cys Pro
H40                                                     Arg Arg Trp Asp Arg Arg Glu
H90                                                     Leu Ser His Cys Val Pro Glu

H15   Arg Arg Arg Glu Arg Gln Thr Tyr Thr Arg Phe Gln Thr Leu Glu Leu Glu Lys Glu Phe
H40   Ala Arg Arg Ala Arg Thr Ala Phe Thr Tyr Glu Gln Leu Val Ala Leu Glu Asn Lys Phe
H90   Arg Lys Arg Gly Arg Gln Thr Tyr Thr Arg Tyr Gln Thr Leu Glu Leu Glu Lys Glu Phe

H15   His Tyr Asn His Tyr Leu Thr Arg Arg Arg Arg Ile Glu Ile Ala His Ala Leu Cys Leu
H40   Lys Thr Thr Arg Tyr Leu Ser Val Cys Glu Arg Leu Asn Leu Ala Leu Ser Leu Ser Leu
H90   His Tyr Asn Arg Tyr Leu Thr Arg Arg Arg Arg Ile Glu Ile Ala His Ala Leu Cys Leu

H15   Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Leu Lys Lys Glu Leu
H40   Thr Glu Thr Gln Val Lys Ile Trp Phe Gln Asn Arg Arg Thr Lys Trp Lys Lys Gln Asn
H90   Thr Glu Arg Gln Ile Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Trp Lys Lys Glu Asn

H15   Arg Ala Val Lys Glu Ile Asn
H40   Pro Gly Leu Asp Val Ile Ser
H90   Lys Ser Lys Gly Ala Pro Ala
```

FIG. 4. H15, H40, and H90 homeobox sequences. (*Upper*) DNA sequences from the homeobox regions of H15, H40, and H90 are aligned. Sequences are shown from −21 to +201. Codons of the common open reading frame are aligned. (*Lower*) A conceptual translation of the regions in *Upper* is shown from amino acid residues −7 through +67.

sequences of E30 and E60 do not allow us to identify them as homologues of either *en* or *inv* since the sequences of these two pairs of genes are more similar within either species than they are between species (see Fig. 5). If one includes the genes *En1* and *En2* from mouse and an engrailed-like gene from a sea urchin in the comparison, the degree of conservation declines in the order *Drosophila*, honeybee, sea urchin, and mouse, reflecting an increasing divergence of these species from a common ancestor.

Since like *Drosophila* and mouse, the honeybee has two

```
      GGT CCG CGG ACG AGG CGG GTG AAG CGA TCG CAC AAC  -  GGG AAG AAT GGC TCG CCG GAG
E30       Pro Arg Thr Arg Arg Val Lys Arg Ser His Asn  -  Gly Lys Asn Gly Ser Pro Glu
en        Pro Arg Tyr Arg Arg Pro Lys Gln Pro Lys Asp Lys Thr Asn Asp  -   -   -   -
inv       Pro Arg Ala Arg Lys Pro Lys Lys Pro Ala Thr Ser  -   -   22AA   -  Pro Glu
E60       Pro Arg Thr Arg Arg Val Lys Arg Ser Asp Gly Arg Gly Asn Gly Gly Thr Pro Glu
      GGT CCG CGG ACG AGA AGG GTG AAG AGG TCG GAT GGC CGT GGC AAT GGC GGC ACC CCG GAG

      GAG AAG CGG CCA AGG ACC GCG TTC AGC GCG GAA CAA CTG GCA AGA TTG AAG AGG GAA TTC
E30   Glu Lys Arg Pro Arg Thr Ala Phe Ser Ala Glu Gln Leu Ala Arg Leu Lys Arg Glu Phe
en    Glu Lys Arg Pro Arg Thr Ala Phe Ser Ser Glu Gln Leu Ala Arg Leu Lys Arg Glu Phe
inv   Asp Lys Arg Pro Arg Thr Ala Phe Ser Gly Thr Gln Leu Ala Arg Leu Lys His Glu Phe
E60   Glu Lys Arg Pro Arg Thr Ala Phe Ser Gly Glu Gln Leu Ala Arg Leu Lys Arg Glu Phe
      GAG AAA CGT CCG AGG ACG GCA TTT AGC GGG GAG CAA CTG GCC AGG CTG AAG AGG GAG TTC

      GCG GAA AAT CGA TAT CTG ACT GAG AGA AGG AGG CAA CAA CTC TCG AGA GAT CTG GGA TTG
E30   Ala Glu Asn Arg Tyr Leu Thr Glu Arg Arg Arg Gln Gln Leu Ser Arg Asp Leu Gly Leu
en    Asn Glu Asn Arg Tyr Leu Thr Glu Arg Arg Arg Gln Gln Leu Ser Ser Glu Leu Gly Leu
inv   Asn Glu Asn Arg Tyr Leu Thr Glu Lys Arg Arg Gln Gln Leu Ser Gly Glu Leu Gly Leu
E60   Ala Glu Asn Arg Tyr Leu Thr Glu Arg Arg Arg Gln Gln Leu Ser Arg Asp Leu Gly Leu
      GCG GAG AAT CGA TAC CTG ACG GAG CGG AGG AGG CAG CAG CTC TCG AGG GAT CTG GGC CTG

      ACC GAG GCG CAG ATC AAG ATC TGG TTC CAA AAT AAA AGA GCG AAG ATC AAA AAG GCG AGC
E30   Thr Glu Ala Gln Ile Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Ile Lys Lys Ala Ser
en    Asn Glu Ala Gln Ile Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Ile Lys Lys Ser Thr
inv   Asn Glu Ala Gln Ile Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Leu Lys Lys Ser Ser
E60   Asn Glu Ala Gln Ile Lys Ile Trp Phe Gln Asn Lys Arg Ala Lys Ile Lys Lys Ala Ser
      AAC GAG GCG CAA ATC AAG ATC TGG TTT CAG AAC AAG AGG GCG AAG ATC AAG AAG GCG AGC

      GGG CAG AAA AAT CCG CTC GCT CTT CAA TTG ATG GCC CAA GGC CTG TAC AAT CAC TCG ACG
E30   Gly Gln Lys Asn Pro Leu Ala Leu Gln Leu Met Ala Gln Gly Leu Tyr Asn His Ser Thr
en    Gly Ser Lys Asn Pro Leu Ala Leu Gln Leu Met Ala Gln Gly Leu Tyr Asn His Thr Thr
inv   Gly Thr Lys Asn Pro Leu Ala Leu Gln Leu Met Ala Gln Gly Leu Tyr Asn His Ser Thr
E60   Gly Gln Lys Asn Pro Leu Ala Leu Gln Leu Met Ala Gln Gly Leu Tyr Asn His Ser Thr
      GGG CAG AAG AAT CCG CTC GCC CTT CAG CTG ATG GCC CAG GGT CTT TAC AAT CAT TCG ACG

      GTC CCC GTT GAC GAG GAC GGG GAG GAA ATC
E30   Val Pro Val Asp Glu Asp Gly Glu Glu Ile
en    Val Pro Leu Thr Lys Glu Glu Glu Glu Leu
inv   Ile Pro Leu Thr Arg Glu Glu Glu Glu Leu
E60   Val Pro Leu Thr Lys Glu Glu Glu Glu Gln
      GTT CCG TTG ACG AAG GAG GAG GAG GAA CAA
```

FIG. 5. Comparison of E30 and E60 sequences with *Drosophila en* and *inv* sequences. Putative amino acid sequences of the homeobox region in E30 and E60 are aligned with homologous regions derived from *en* and *inv* of *Drosophila*. The homeobox is boxed; amino acids not identical with the amino acids found in the corresponding region of E60 are boxed (*en* and *inv*) or shaded (E30).

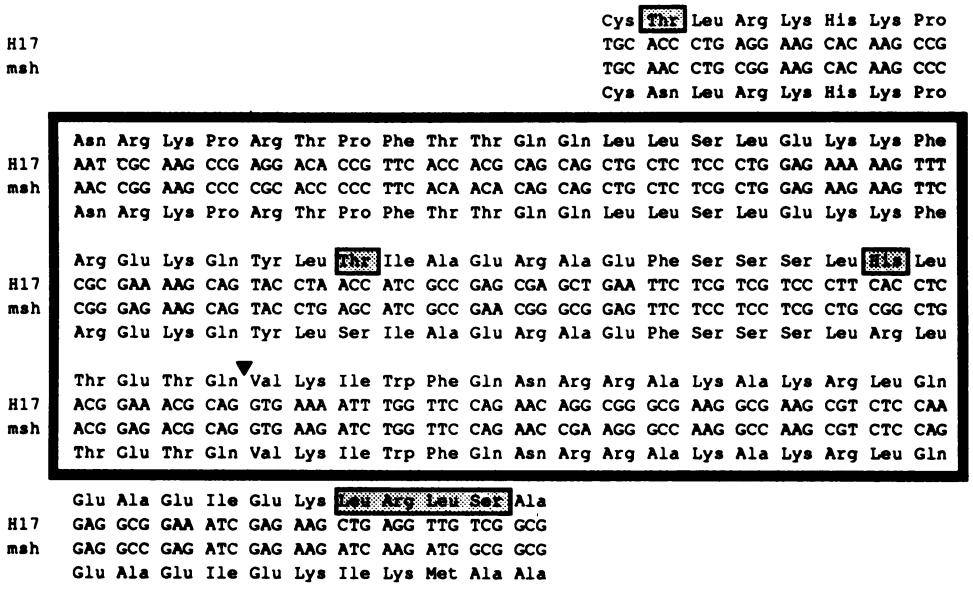

FIG. 6. Comparison of H17 and *Drosophila msh* sequences. DNA and putative amino acid sequences of the homeobox regions in H17 are aligned with homologous regions from *Drosophila msh*. The homeobox is boxed, nonidentical amino acids of H17 compared to *msh* are shown by small boxes, and the position of the intron in H17 is indicated by an arrowhead.

engrailed-like genes, the sea urchin is the only known organism with only one such gene. If this reflects the primary situation, this single gene would have to have undergone duplication on three separate occasions during evolution. The first must have taken place after the separation of echinoderms and vertebrates (25), the second and third after the separation of Hymenoptera and Diptera, since the honeybee genes have no introns, whereas the *Drosophila* genes have an intron in an identical position (26).

With a muscle segment homeobox (*msh*) probe belonging to yet a different class of homeobox sequences, it was possible to isolate a phage (H17) showing hybridization to two *Eco*RI fragments of 1.7 kb and 1.0 kb. Detailed analysis of H17 revealed the existence of an intron in the homeobox at position 44 (Fig. 6) as in the *Drosophila* genes labial (*lab*) (27) and Distal-less (*Dll*) (28). Due to high sequence conservation, both *Eco*RI fragments gave relatively strong signals in Southern blots (Fig. 1). Since within the putative homeodomain only one amino acid differs and in the 5′ and 3′ regions, the conservation is significant, H17 seems to be a homologue of the *Drosophila msh* gene. Although there is no mutant known for this gene and it therefore has not been classified, a homologue has been reported in the mouse (29, 30). Like the *Drosophila msh* gene and in contrast to H17, the mouse gene has no intron in the homeobox.

The sequence of the homeobox region of the recently isolated new homeobox gene *W13* of *Drosophila*, which corresponds to the empty spiracle (*ems*) gene (U.W., unpublished data), has strongly diverged from all other known *Drosophila* homeobox genes. By using a homeobox probe from *W13*, the honeybee clone H40 was isolated (Fig. 4). This clone has only 52% similarity with the *W13* probe. A *Drosophila* homologue of H40 has not yet been identified.

Our results show that the homeobox sequences of various genes of *Drosophila* show different degrees of conservation in the honeybee. Genes with Antennapedia-like homeoboxes have highly conserved homologues in the honeybee, while for the segmentation gene *ftz*, such a highly conserved homologue could not be identified. In general, we can now begin to make a case for a relatively high degree of conservation of genes involved in embryonic development of holometabolous insects. Whether some of the differences reflect specific alterations of parts of the developmental program and to what extent homologues identified by sequence perform the same function in the two species remain to be analyzed.

We thank Klaus Sander for his continued interest in this project and Paul Schroeder, Ken Cadigan, and Greg Gibson for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft (Fl 151/1-1 and 1-2; Wa 556/1-1 and 1-2) and the Swiss National Science Foundation.

1. Nüsslein-Volhard, C., Wieschaus, E. & Jürgens, G. (1982) *Verh. Dtsch. Zool. Ges.* **1982,** 91–104.
2. Akam, M. (1987) *Development* **101,** 1–22.
3. McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. (1984) *Nature (London)* **308,** 428–433.
4. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. (1984) *Cell* **37,** 403–408.
5. Scott, M. P. & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4115–4119.
6. Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. & Gehring, W. J. (1988) *EMBO J.* **7,** 4299–4304.
7. Gehring, W. J. & Hiromi, Y. (1986) *Annu. Rev. Genet.* **20,** 147–173.
8. Scott, M. P., Tamkun, J. W. & Hartzell, G. W., III (1989) *BBA* **989,** 25–48.
9. Kaufmann, T. C. (1983) in *Time, Space and Pattern in Embryonic Development*, eds. Raff, R. A. & Jeffreys, W. (Liss, New York), pp. 365–383.
10. Lewis, E. B. (1978) *Nature (London)* **276,** 565–570.
11. Crain, W. R., Davidson, E. H. & Britten, R. J. (1976) *Chromosoma* **59,** 1–12.
12. Riek, E. F. (1979) in *The Insects of Australia* (Melbourne University Press, Carlton, Victoria), Vol. I, pp. 168–186.
13. Fleig, R., Walldorf, U., Sander, K. & Gehring, W. J. (1988) *Roux's Arch. Dev. Biol.* **197,** 269–274.
14. Krause, H. M., Klemenz, R. & Gehring, W. J. (1988) *Genes & Dev.* **2,** 1021–1036.
15. Kuroiwa, A., Kloter, U., Baumgartner, P. & Gehring, W. J. (1985) *EMBO J.* **4,** 3757–3764.
16. Fjose, A., McGinnis, W. J. & Gehring, W. J. (1985) *Nature (London)* **313,** 284–289.
17. Mlodzik, M. & Gehring, W. J. (1987) *Cell* **48,** 465–478.
18. Walldorf, U., Richter, S., Ryseck, R.-P., Steller, H., Edström, J. E., Bautz, E. K. F. & Hovemann, B. (1984) *EMBO J.* **3,** 2499–2504.
19. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
20. Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842.
21. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
22. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
23. Akam, M., Dawson, I. & Tear, G. (1988) *Development Suppl.* **104,** 123–133.
24. Joyner, A. L. & Martin, G. R. (1987) *Genes & Dev.* **1,** 29–38.
25. Dolecki, G. J. & Humphreys, T. (1988) *Gene* **64,** 21–31.
26. Poole, S. J., Kauvar, L. M., Drees, B. & Kornberg, T. (1985) *Cell* **40,** 37–43.
27. Mlodzik, M., Fjose, A. & Gehring, W. J. (1988) *EMBO J.* **7,** 2569–2578.
28. Cohen, S. M., Brönner, G., Küttner, F., Jürgens, G. & Jäckle, H. (1989) *Nature (London)* **338,** 432–434.
29. Robert, B., Sassoon, D., Jacq, B., Gehring, W. J. & Buckingham, M. (1989) *EMBO J.* **8,** 91–100.
30. Hill, R. E., Jones, P. F., Rees, A. R., Sime, C. M., Justice, M. J., Copeland, N. G., Jenkins, N. A., Graham, E. & Davidson, D. R. (1989) *Genes & Dev.* **3,** 26–37.