HOM/Hox genes of *Artemia*: implications for the origin of insect and crustacean body plans

Michalis Averof and Michael Akam

Wellcome/CRC Institute and Department of Genetics, Tennis Court Road, Cambridge CB2 1QR, UK.

Background: Insects and crustaceans are generally assumed to derive from a segmented common ancestor that had a distinct head but uniform, undifferentiated trunk segments. The subdivision of the body into functionally distinct regions (e.g. thorax and abdomen) is thought to have evolved independently in these two lineages. In insects, the differences between segments in the trunk are controlled by the *Antennapedia*-like genes of the homeotic gene clusters. Study of these genes in crustaceans should provide a basis for comparing body plans and assessing their evolutionary origin.

Results: Using a polymerase chain reaction (PCR) / inverse PCR strategy, we have isolated six genes of the *HOM/Hox* family from the crustacean *Artemia franciscana*. Five of these are clearly identifiable as specific homologues of the insect homeotic genes *Dfd*, *Scr*, *Antp*, *Ubx* and *abdA*. The sixth appears to have no close counterpart in insects.

Conclusion: All the homeotic genes that specify middle body regions in insects originated before the divergence of the insect and crustacean lineages, probably not later than the Cambrian (about 500 million years ago). A commonly derived groundplan may underlie segment diversity in these two groups.

Current Biology 1993, **3**:73–78

Background

Genes of the *HOM/Hox* family have been identified in a wide variety of animals. In insects and vertebrates these homeotic genes play an important role in specifying the identity of structures along the antero–posterior body axis [1]. This function may be common to most major metazoan lineages [2–4].

The *HOM/Hox* genes share a conserved sequence motif, the homeobox, and a complex organization in gene clusters, features that are presumed to reflect their common evolutionary ancestry from a single gene. The number of clusters and the precise array of genes within each cluster differ among various animal groups [5], suggesting that this gene family has evolved significantly during the diversification of the metazoa.

In insects, *HOM/Hox* genes operate within the framework of body segmentation, where they serve to specify segmental identity (Fig. 1a) [6–9]. Mutations in these genes transform structures of one segment to the corresponding structures of another (for example, the transformation of halteres to a second pair of wings by *Ubx* mutations in *Drosophila* [6]). This central role in specifying segment type implies that these homeotic genes provide the underlying information for tagmosis, the regional diversification and specialization of groups of segments to form integrated functional units (e.g. mouthparts, thorax and abdomen) [10].

The subdivision of the body into distinct units, or tagmata, is a characteristic feature of arthropods, and is one of the principal characters used to define the major taxonomic subdivisions within the group [11]. The evolution of these body plans must reflect genetic changes in the developmental mechanisms that control segment specification. Because the link between

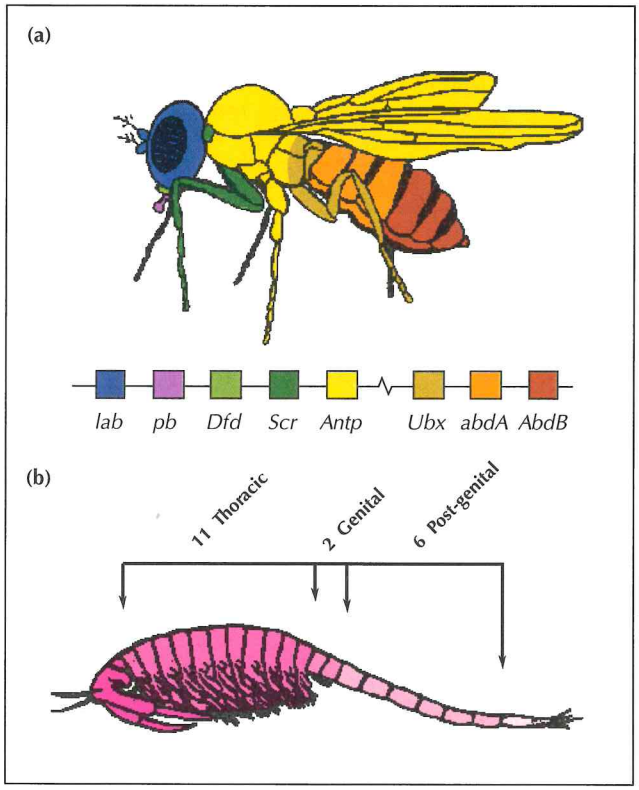

Fig. 1. The segmental pattern in *Drosophila* and *Artemia*. **(a)** *Drosophila melanogaster*. The cluster organization of insect homeotic genes and the segmental structures that they specify are indicated by colours. The genes are: *labial* (*lab*), *proboscipedia* (*pb*), *Deformed* (*Dfd*), *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*), *abdominal-A* (*abdA*) and *Abdominal-B* (*AbdB*). **(b)** *Artemia franciscana*. The segments that constitute the three major subdivisions of the trunk are indicated.

Correspondence to: Michalis Averof.

homeotic genes and segment identity is so explicit, it is attractive to investigate how changes in the structure or regulation of the *HOM/Hox* genes may be related to the observed morphological diversity of the arthropods [6,10]. To address this question we focus on Crustacea, a group of arthropods that display a huge diversity of body plans. Although the exact evolutionary relationships between insects and crustacea are still a matter of controversy [12,13], it is generally accepted that they have been derived from a segmented common ancestor. This view is supported by comparative morphology and, more recently, by descriptive and molecular studies of development [14–16].

In most crustaceans, as in insects, the trunk region of the body is divided into distinct regions (Fig. 1b), these being defined most obviously by the modification or suppression of limb development [17]. The distinctions between these body regions are traditionally thought to have evolved independently in insects and crustaceans, since presumed ancestral forms (resembling myriapods and remipedes respectively) are thought to have had a distinct head but a serially undifferentiated, homonomous trunk (Fig. 2, reviewed in [11]). It might be expected, therefore, that the 'homeotic' genes that determine trunk specializations would have arisen independently in these two lineages, after the divergence of the myriapod-like and remipede-like organisms from their last common ancestor. To test this hypothesis we examine representatives of these 'trunk' genes in a crustacean, the brine shrimp *Artemia franciscana* (Anostraca).

Results

We have developed a generally applicable method to isolate *HOM/Hox* genes, based on the PCR technique. Degenerate general homeobox primers (similar to those described in [18]) are used for PCR amplification on total genomic DNA. The amplified fragments are cloned and sequenced to provide information on a short stretch of homeobox sequence, which is then used to design divergent pairs of primers for inverse PCR [19]. This second amplification, carried out on restriction cut and circularized genomic DNA, produces clones that provide additional sequence information both upsteam and downstream of the previously identified region. This strategy can be used to isolate extensive homeobox sequences from a variety of sources, and may be particularly useful in cases where the amount and quality of the available material are limiting.

Using this method, we have isolated six homeobox-containing genes from *Artemia franciscana* (*Af*). Figure 3 shows the sequences of the protein-coding regions present on these clones. Comparisons at the amino acid level indicate that five of these genes are closely related to specific members of the insect *HOM* gene family (Fig. 4). Sequence similarity extends beyond the homeobox, into upstream and downstream flanking regions, where the observed divergence is comparable to that seen between homologous genes from different insect groups. We therefore suggest that these five genes represent specific homologues of the insect homeotic genes *Deformed* (*Dfd*), *Sex combs*

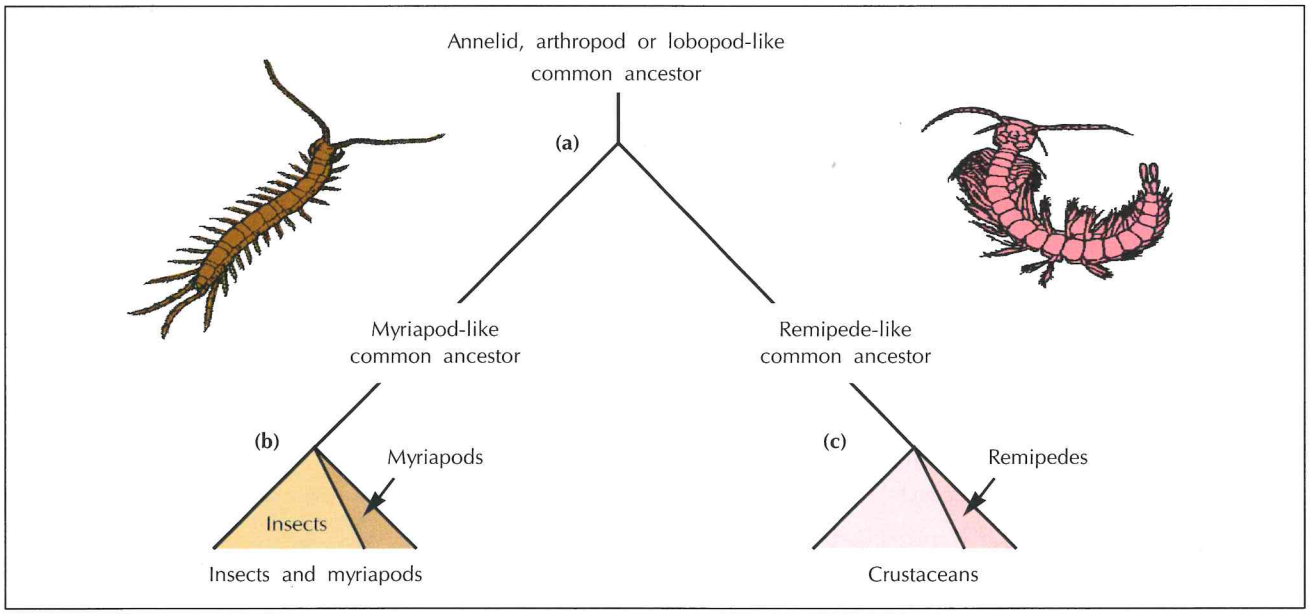

Fig. 2. Classical view of the evolution of insect and crustacean body plans. The uniform, serially undifferentiated trunk of remipedes and myriapods is thought to represent the 'primitive' condition, found in nodes **(a)**, **(b)** and **(c)**. Specialization of segments of the head is thought to have evolved earlier than the trunk, in some cases before node **(a)**. The nature of the common ancestor at node **(a)** and the positions of other arthropod groups within the context of this tree are a matter of controversy.

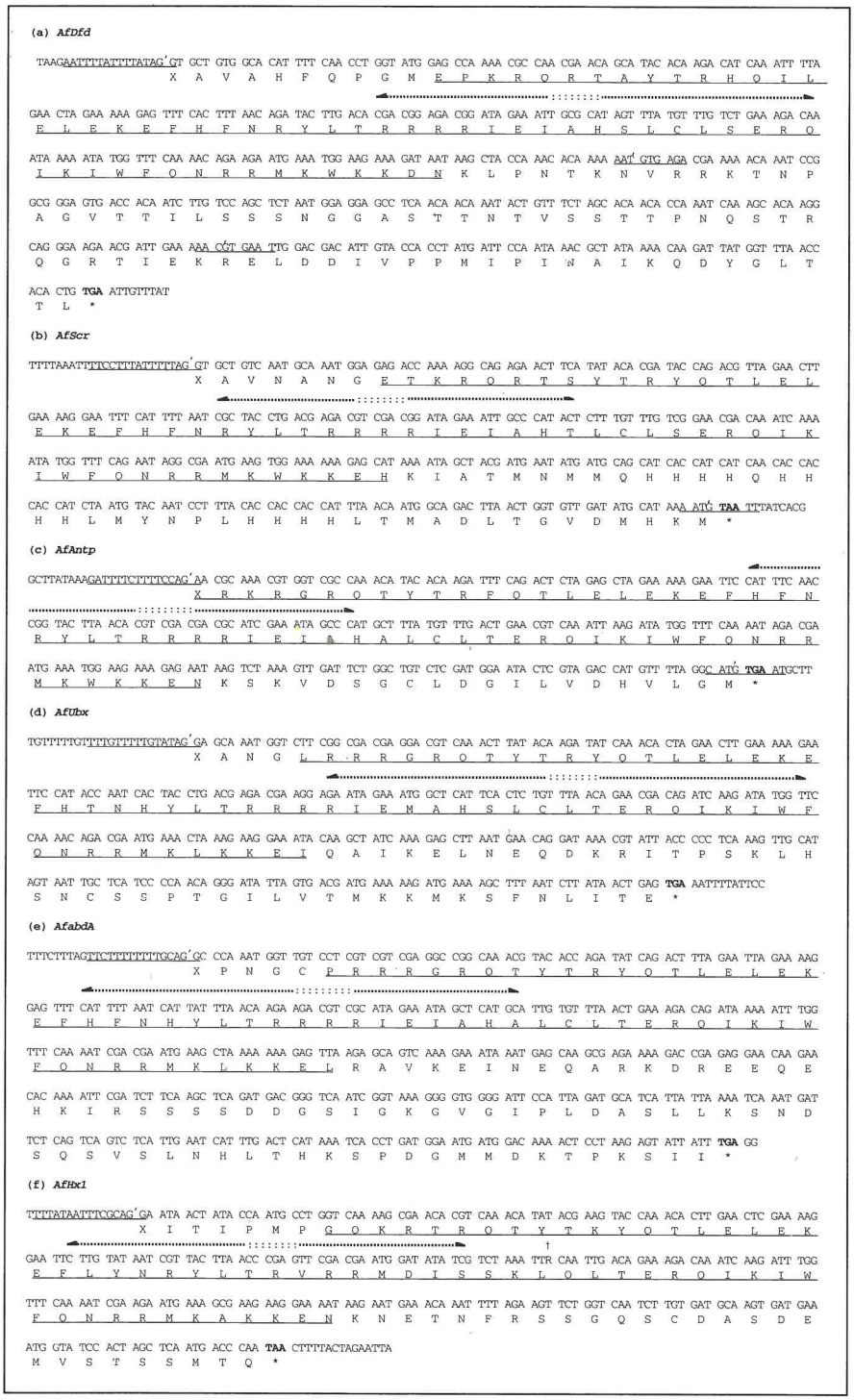

Fig. 3. DNA sequences from six *Artemia franciscana* genes. Translated amino acid sequences are shown in single letter code. Putative splice sites are indicated by underlining and an acute accent, the homeobox by underlining, and the position of the inverse PCR primers by overlining arrows. Sequences represent a consensus from three or more PCR or inverse PCR clones (except in the regions of the inverse PCR primers for *AfUbx*, which are derived from a single PCR clone). Termination codons are shown in bold print. The ambiguity at *AfHx1* position 153 (†; where R = A or G) reflects differences observed in a number of independent clones and may represent a silent polymorphism. Accession numbers for *AfabdA*, *AfAntp*, *AfDfd*, *AfHx1*, *AfScr* and *AfUbx* are X70076–81, respectively.

reduced (*Scr*), *Antennapedia* (*Antp*), *Ultrabithorax* (*Ubx*) and *abdominal-A* (*abdA*) and we therefore refer to these five genes as *AfDfd*, *AfScr*, *AfAntp*, *AfUbx* and *AfabdA*. The sixth sequence (provisionally named *AfHx1*) is more divergent. It appears to be related to the *Antp*-like *HOM/Hox* family, but it is not clearly identifiable as the specific homologue of any known gene. Preliminary observations suggest that this, and similarly divergent *Antp*-like homeoboxes isolated from insects, may comprise a family of rapidly diverging homeobox genes that includes the *Drosophila* gene *fushi tarazu* (R Dawes and I Dawson in preparation; S Brown and R Denell, personal communication.).

Putative splice acceptor sites [20] lie upstream of the homeobox in each of these genes (Fig. 3). In *AfScr*, *AfAntp*, *AfUbx*, and *AfabdA* these coincide precisely with intron positions conserved in their presumed insect counterparts (Fig. 4). The precise conservation of these intron–exon boundaries reinforces our view

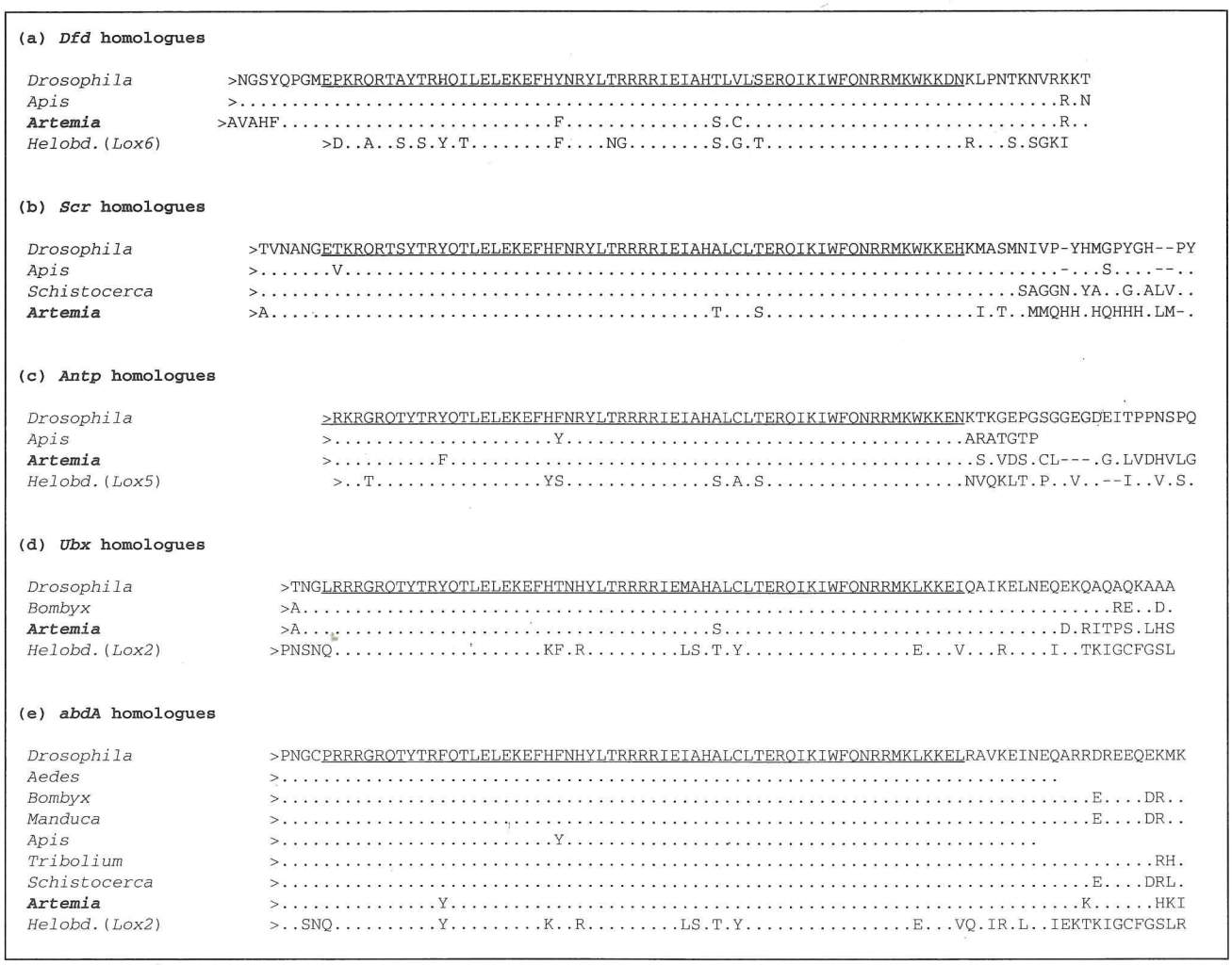

Fig. 4. Aminoacid alignments between *Artemia franciscana* sequences and their most closely related counterparts from annelids and insects. Dots indicate aminoacid identity to the *Drosophila* sequence, dashes indicate gaps introduced to optimize alignments. The homeodomain is underlined, putative intron positions upstream of the homeobox are indicated by a horizontal arrow. Organisms represented include insects *Drosophila melanogaster* and *Aedes aegypti* (Diptera), *Bombyx mori* and *Manduca sexta* (Lepidoptera), *Apis mellifera* (Hymenoptera), *Tribolium castaneum* (Coleoptera), *Schistecerca gregaria* (Orthoptera), the crustacean *Artemia franciscana* (Anostraca), and the annelid *Helobdella robusta* (Hirudinida) [4,9,10,24–35] (I Dawson, personal communication) (V Walldorf, personal communication). The *Lox2* gene of *Helobdella* is very similar to that previously isolated from another leech, *Hirudo medicinalis* [36].

that the proposed homologies are real. In *AfDfd* the intron has shifted by a single triplet and lies in the same reading frame.

Discussion

HOM/Hox genes in insects and vertebrates fall into a number of well defined classes [21]. The anteriorly expressed vertebrate genes of class 1, 2 and 4 clearly resemble the insect 'head' genes *lab*, *pb* and *Dfd*, respectively, suggesting that distinct homologues for each of these genes existed in some common ancestor. Similarly, posterior vertebrate genes of classes 9–13 are all most similar to the insect gene *Abd-B*, suggesting that a 'tail' gene was present in that common ancestor. Lying between these head and tail genes in the clusters, both vertebrates and insects have a set of *Antp* class genes, which specify the diversity of middle body regions. Sequence alone suggests no specific one-to-one correspondence between these 'trunk' genes in insects and vertebrates. Rather, it suggests that they may have arisen independently in the two lineages, by duplication of an ancestral *Antp*-like gene. This model is consistent with the observation that the *Hox* cluster in a nematode contains four genes, of which only one (*mab-5*) corresponds to the *Antp* class [22]. It therefore seems likely that specification of the antero–posterior body pattern in major metazoan lineages, and possibly all bilaterally symmetrical triploblastic animals, is commonly derived from a distant ancestor with distinct *HOM/Hox* genes for the head, trunk and tail (Fig. 5) [5,23].

It is still not clear when the duplications leading to specific trunk genes took place. At least one probably occurred before the annelid–arthropod split, since two distinct *Antp* class genes have been reported in annelids, one (*Lox 5*) more closely resembling *Antp* and the other (*Lox 2*) more similar to *Ubx* and *abdA*

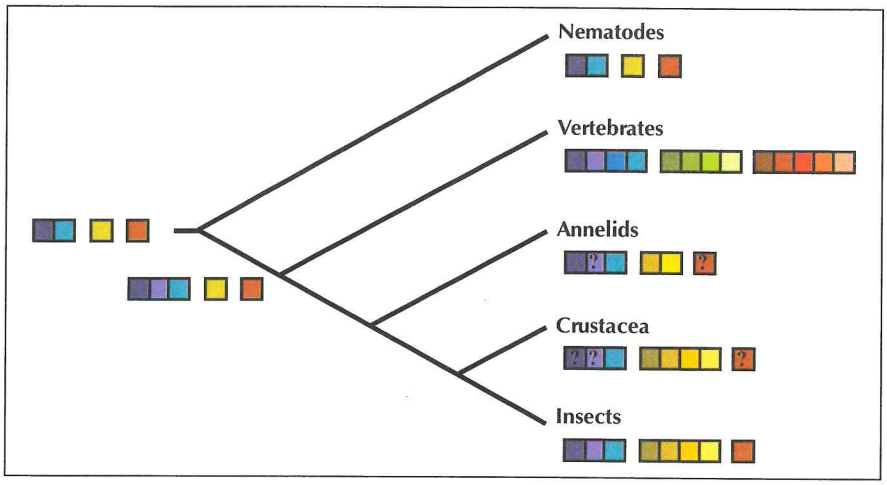

Fig. 5. The phylogenetic distribution of *HOM/Hox* genes among major metazoan lineages. A number of 'head' genes (indicated in blue) appear to have arisen early in the evolution of metazoans. In contrast, 'trunk' genes (indicated in shades of yellow) appear to have diversified independently in the lineages leading to vertebrates (yellow-green) and arthropods (yellow-brown). 'Tail' genes (indicated in red) appear to have diversified specifically in the vertebrate lineage. The four vertebrate clusters are represented here by a single set comprising all identified paralogous groups. Question marks indicate genes thought to be present but not yet isolated. The grouping of genes in this diagram does not reflect the physical structure of gene clusters.

[24]. Our results show that all specific gene duplications leading to the array of trunk genes in modern insects must have taken place before the divergence of insect and crustacean lineages. Both lineages share genes uniquely identifiable as *Scr*, *Antp*, *Ubx* and *abd-A*, as well as the more ancient head gene *Dfd*. Moreover, because each class of these trunk genes has acquired unique, conserved sequence differences, we infer that they must also have acquired distinct functions in the common ancestor.

This result is surprising, if we accept the traditional view that the specialization of trunk segments occurred independently in insect and crustacean lineages. It implies that a serially homonomous ancestor had already evolved an array of trunk homeotic genes. If so, we can only assume that these genes must have been used for some more limited function—for example, for patterning specific structures within the central nervous system or establishing differences between the segmentally repeated primordia for other internal organs. Only later would they have been co-opted for more overt segmental diversification. Alternatively, the conventional ideas about the morphological transitions of the insect and crustacean trunk may be incorrect. A specialized and subdivided trunk may be the primitive condition, and the homonomous trunk of myriapods and remipedes a derived and secondarily simplified state. In this respect, the study of *HOM/Hox* genes in myriapods or remipedes could be particularly informative.

In either case, it is implicit that the evolution of a diversified trunk has been based on a molecular groundplan that is common to both insects and crustaceans. Comparison of the expression of these genes during development should provide a crucial test for distinguishing the alternative evolutionary hypotheses.

Materials and Methods

Isolation of Genomic DNA

Dried diapause cysts of *Artemia franciscana* from the north arm of Great Salt Lake, Utah, USA were obtained from NT Laboratories (Kent, UK). Cysts were hatched in 3% artificial seawater. Hatched nauplii were homogenized in cold buffer (50 mM Tris-Cl pH 8.0, 15 mM EDTA, 5 mM EGTA, 10 mM NaCl, 0.4 M sucrose) and treated with 50 µg/ml proteinase K in 0.25% SDS at 55°C overnight. The homogenate was then extracted several times with phenol/chloroform, once with chloroform and then precipitated using 1/10th volume 3M NaAc and 2 volumes ethanol. Precipitated DNA was lifted from the solution using a sealed glass pipette tip, washed in 70% ethanol and resuspended in TE. Samples used for inverse PCR were further treated with 50 µg/ml RNAase, and then with proteinase K in SDS, phenol/chloroform, and precipitated as described earlier. The haploid genome size of *Artemia franciscana* was taken to be approximately 2 x 10^9 bp.

Degenerate PCR

Degenerate inosine-containing primers with 5′ non-specific tags were designed to amplify genes of the *lab*, *Dfd*, *Antp* and *AbdB* classes: forward primer (homeobox residues 16–21, ELEKEF): 5′ GGAATTC $GA^A/_G$ CTI $GA^A/_G$ $AA^A/_G$ $GA^A/_G$ TT-3′; reverse primer (homeobox residues 49–54, WFQNRR): 5′GCTCTAGA CG ICG $^A/_G$TT TTG $^A/_G$AA CCA-3′. These were used in 50 µl PCR amplifications containing 1x PCR buffer (10 mM Tris-Cl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1 mg/ml gelatin, ≤1% Triton), 200 µM each dNTP, 0.5 µM reverse primer, 1.25 µM forward primer, 50–100 ng *Artemia* genomic DNA, and 2 U Taq DNA polymerase, overlayed with paraffin oil. The temperature profile consisted of (a) initial denaturation of 5 minutes at 95°C, (b) 5 early cycles of 30 seconds at 94°C, 30 seconds at 33–50°C, 2–3 minutes ramp time, and 30 seconds at 72°C, (c) 30 late cycles of 30 seconds at 94°C, 30 seconds at 50–55 °C, and 30 seconds at 72°C, and (d) final extension of 5 minutes at 72°C. Aliquots of the reaction were tested by agarose gel electrophoresis for the presence of the expected 131 bp product.

Inverse PCR

Artemia genomic DNA was digested with restriction enzymes which are predicted to produce, on average, fragments 0.3–3.0 kb in length, and which do not cut within the target region in each gene. The DNA fragments were circularised using T4 DNA ligase (standard buffer without PEG) in dilute solution (1–3 ng/µl), and then concentrated by ethanol precipitation. Unique divergent primers, based on the sequenced region of each gene (see Fig. 3), were used in 50 µl PCR amplifications containing 1x PCR buffer (as before), 200 µM each dNTP, 0.5 µM each primer, 150 ng circularised genomic DNA, 5 µg/ml *E. coli* SSB protein, and 2 U Taq DNA polymerase, overlayed with paraffin oil. The temperature profile consisted of (a) initial denaturation of 5 minutes at 95°C, (b) 35 cycles of 2 minutes at

94°C, 1 minute at 55°C, 5 minutes at 72°C, and (c) final extension of 10 minutes at 72°C. Recovered fragments ranged from 0.5 to 1.6 kb in size.

Cloning and Sequencing

PCR-derived DNA fragments were treated with Klenow DNA polymerase, resolved by electrophoresis on low-melt agarose gels, and the required bands were excised. The DNA was recovered using glass beads (MerMaid or Geneclean). Fragments were then treated with polynucleotide kinase, ligated to a *Sma*I-cut and phosphatased plasmid vector (pBS) and electroporated into *E. coli* DH5α host cells. Recombinant plasmid clones were screened by restriction digests and sequenced by the dideoxy chain termination method (Sequenase).

Acknowledgements: We would like to thank Hans Georg Frohnhofer for stimulating discussions, Miguel Manzanares for practical advice, and the Artemia Reference Centre in Ghent for information on strains. Our work has been supported by the Wellcome Trust.

References

1. McGinnis W, Krumlauf R: **Homeobox genes and axial patterning**. *Cell* 1992, **68**:283–302.
2. Chisholm A: **Control of cell fate in the tail region of *C. elegans* by the gene *egl-5***. *Development* 1991, **111**:921–932.
3. Costa M, Weir M, Coulson A, Sulston J, Kenyon C: **Posterior pattern formation in *C. elegans* involves position-specific expression of a gene containing a homeobox**. *Cell* 1988, **55**:747–756.
4. Nardelli-Haefliger D, Shankland M: **Lox2, a putative leech segment identity gene, is expressed in the same segmental domain in different stem cell lineages**. *Development* 1992, **116**:697–710.
5. Schubert FR, Nieselt-Struwe K, Gruss P: **The Antennapedia-type homeobox genes have evolved from three precursors separated early in metazoan evolution**. *Proc Natl Acad Sci USA* 1993, in press.
6. Lewis EB: **A gene complex controlling segmentation in *Drosophila***. *Nature* 1978, **276**:565–570.
7. Akam M: **The molecular basis for the matameric pattern in the *Drosophila* embryo**. *Development* 1987, **101**:1–22.
8. Beeman R, Stuart JJ, Haas MS, Denell RE: **Genetic analysis of the homeotic gene complex (HOM-C) in the beetle *Tribolium castaneum***. *Dev Biol* 1989, **133**:196–209.
9. Ueno K, Hui C-C, Fukuta M, Suzuki Y: **Molecular analysis of the deletion mutants in the E homeotic complex of the silkworm *Bombyx mori***. *Development* 1992, **114**:555–563.
10. Akam M, Dawson I, Tear G: **Homeotic genes and the control of segment diversity**. *Development* 1988, **104 (suppl)**:123–133.
11. Brusca RC, Brusca GJ: *Invertebrates*. Massachusetts:Sinauer; 1990.
12. Willmer P: *Invertebrate Relationships; patterns in animal evolution*. Cambridge:Cambridge University Press; 1990.
13. Shear WA: **End of the 'Uniramia' taxon**. *Nature* 1992, **359**:477–478.
14. Dohle W, Scholtz G: **Clonal analysis of the crustacean segment: discordance between genealogical and segmental borders**. *Development* 1988, **104 (suppl)**:147–160.
15. Patel NH, Kornberg TB, Goodman CS: **Expression of *engrailed* during segmentation in grasshopper and crayfish**. *Development* 1989, **107**:201–212.
16. Scholtz G: **Cell lineage studies in the crayfish *Cherax destructor* (Crustacea, Decapoda): germ band formation, segmentation, and early neurogenesis**. *Rouxs Arch Dev Biol* 1992, **202**:36–48.
17. Schram FR: *Crustacea*. Oxford:Oxford University Press; 1986.
18. Frohman MA, Boyle M, Martin GR: **Isolation of the mouse *Hox-2.9* gene; analysis of embryonic expression suggests that positional information along the anterior-posterior axis is specified by mesoderm**. *Development* 1990, **110**:589–607.
19. Silver J: **Inverse polymerase chain reaction**. In *PCR; a practical approach*. Edited By McPherson MJ, Quirke P, Taylor GR. Oxford:IRL Press; 1991.
20. Lenstra JA, Van Vliet AV, Arnberg AC, Van Hemert FJ, Moller W: **Genes coding for the elongation factor EF-1α in *Artemia***. *Eur J Biochem* 1986, **155**:475–483.
21. Scott MP: **Vertebrate homeobox gene nomenclature**. *Cell* 1992, **71**:551–553.
22. Burglin TR, Ruvkun G, Coulson A, Hawkins NC, McGhee JD, Schaller D, Wittmann C, Muller F, Waterston RH: **Nematode homeobox cluster**. *Nature* 1991, **351**:703.
23. Akam M: ***Hox* and HOM X: Homologous gene clusters in insects and vertebrates**. *Cell* 1989, **57**:347–349.
24. Shankland M, Martindale MQ, Nardelli-Haefliger D, Baxter E, Price DJ: **Origin of segmental identity in the development of the leech nervous system**. *Development* 1991, **suppl 2**:29–38.
25. Eggleston P: **Identification of the *abdominal-A* homologue from *Aedes aegypti* and structural comparisons among related genes**. *Nucleic Acids Research* 1992, **20**:4095.
26. Fleig R, Walldorf U, Gehring WJ, Sander K: **In situ localization of the transcripts of a homeobox gene in the honeybee *Apis mellifera* L. (Hymenoptera)**. *Rouxs Arch Dev Biol* 1988, **197**:269–274.
27. Karch F, Bender W, Weiffenbach B: ***abdA* expression in *Drosophila* embryos**. *Genes Dev* 1990, **4**:1573–1587.
28. Kornfeld K, Saint RB, Beachy PA, Harte PJ, Peattie DA, Hogness DS: **Structure and expression of a family of *Ultrabithorax* mRNAs generated by alternative splicing and polyadenylation in *Drosophila***. *Genes Dev* 1989, **3**:243–258.
29. LeMotte PK, Kuroiwa A, Fessler LI, Gehring WJ: **The homeotic gene *Sex Combs Reduced* of *Drosophila*: gene structure and embryonic expression**. *EMBO J* 1989, **8**:219–227.
30. Nagy LM, Booker R, Riddiford LM: **Isolation and embryonic expression of an *abdominal-A-like* gene from the lepidopteran, *Manduca sexta***. *Development* 1991, **112**:119–129.
31. Regulski M, McGinnis N, Chadwick R, McGinnis W: **Developmental and molecular analysis of *Deformed*; a homeotic gene controlling *Drosophila* head development**. *EMBO J* 1987, **6**:767–777.
32. Schneuwly S, Kuroiwa A, Baumgartner P, Gehring WJ: **Structural organisation and sequence of the homeotic gene *Antennapedia* of *Drosophila melanogaster***. *EMBO J* 1986, **5**:733–739.
33. Stuart JJ, Brown SJ, Beeman RW, Denell RE: **The *Tribolium* homeotic gene *Abdominal* is homologous to *abdominal-A* of the *Drosophila* bithorax complex**. *Development* 1993, in press.
34. Tear G, Akam M, Martinez-Arias A: **Isolation of an *abdominal-A* gene from the locust *Schistocerca gregaria* and its expression during early embryogenesis**. *Development* 1990, **110**:915–925.
35. Walldorf U, Fleig R, Gehring WJ: **Comparison of homeobox-containing genes of the honeybee and *Drosophila***. *Proc Natl Acad Sci USA* 1989, **86**:9971–9975.
36. Wysocka-Diller JW, Aisemberg GO, Baumgarten M, Levine M, Macagno ER: **Characterization of a homologue of the bithorax-complex genes in the leech *Hirudo medicinalis***. *Nature* 1989, **341**:760–763.

Received: **23 December, 1992.**
Accepted: **5 January, 1993.**