Development 104 Supplement, 123–133 (1988)
Printed in Great Britain © The Company of Biologists Limited 1988

Homeotic genes and the control of segment diversity

MICHAEL AKAM, IAIN DAWSON and GUY TEAR

Department of Genetics, Downing Street, Cambridge CB2 3EH, UK

Summary

Homeotic genes control the diversity of segment development, but the domains of action of homeotic genes do not obviously correspond with the major morphological subdivisions of the insect body. We suggest that this lack of correspondence is misleading, because the spatial domains defined by genetics mask fundamental differences in the roles played by individual genes in different regions. In one or more parasegments, each homeotic gene is expressed 'metamerically'; that is, it is expressed from blastoderm stages onwards in all or virtually all cells of the parasegment primordium. Elsewhere, the same homeotic gene may be deployed adventitiously, only in subsets of cells and at later stages of development.

We argue that the early 'metameric' domains of gene expression do correlate with the major morphological subdivisions of the fly. This suggests a relatively direct relationship between the expression of particular homeotic genes and the establishment of the 'ground plan' that characterizes segments within each major tagma of the body. This relationship allows us to suggest a scenario for the evolution of homeotic genes in relation to the evolving morphological organization of the arthropod body plan in the insect–myriapod lineage.

Key words: homeotic genes, segmentation, evolution, tagmata, development.

Introduction

The bodies of insects, and of most other arthropods, are visibly subdivided into distinct regions, or tagmata (Snodgrass, 1935). In insects, the organization of segments within tagmata is very constant. There are three gnathal or mouthpart segments (often incorporated into the head), three thoracic segments and an abdomen with a maximum of 11 segments (Fig. 1).

The tagmata reflect functional subdivisions of the body, but in varying degree they also reflect domains of the body where each of several segments follows a similar developmental pathway. In *Drosophila*, for example, the thoracic segments will each establish imaginal disc primordia – groups of cells destined to proliferate during the larval period – whereas abdominal segments establish histoblasts that will be quiescent until puparium formation (Bodenstein, 1950).

In *Drosophila*, and presumably in all insects, the differences between segments are controlled by homeotic genes (Lewis, 1963, 1978; see Mahaffey & Kaufman, 1987). We might therefore expect that these major morphological divisions of the body would be reflected in the domains of action of homeotic genes. Little trace of this is evident, however, in the descriptions of the domains of homeotic gene activity defined by genetic analysis (Lewis, 1978; Sanchez-Herrero *et al.* 1985). In the bithorax complex, for example, the gene *Ultrabithorax* (*Ubx*) is assigned a role controlling both thorax and abdomen, whereas the very similar abdominal segments A4 and A5 are assigned to the control of different homeotic genes (Fig. 1).

These genetic subdivisions of the fly are clearly correct – the *Abdominal-B* gene (*Abd-B*) does play a visible role in the development of A5 but not A4. However, we believe that they are misleading in our attempts to understand the control of development. Analysis of the deployment of homeotic genes during development (for review, see Akam, 1987) suggests that a single gene may play qualitatively different roles in different segments. Typically, each homeotic gene is activated 'metamerically' in certain segments (or more precisely, parasegments (Martinez-Arias &

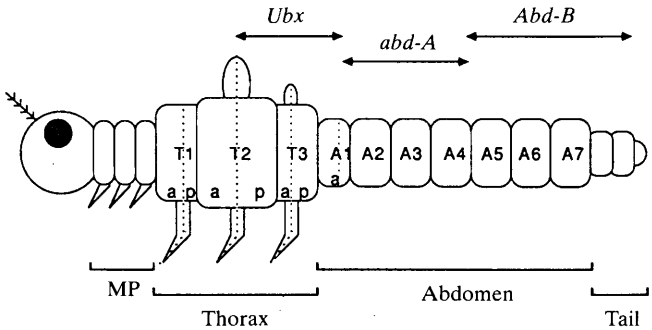

Fig. 1. Morphological and genetic domains in *Drosophila*. Domains of action of genes in the bithorax complex are shown above the schematic illustration of a fly (after Sanchez-Herrero *et al.* 1985). Dotted lines separate anterior (a) and posterior (p) compartments of thoracic (T1–3) and first abdominal (A1) segments. The genetically defined domain of *Ubx* activity extends from T2p to A1a; that of *abd-A* from A1p to A4; and of *Abd-B* from A5 to A9. The extent of each tagma within the segmented region of the body is shown below the fly. The three mouthpart segments (MP) are shown as separate from the head, although in the adults of Diptera and many other insects these segments are entirely incorporated into the head.

Lawrence, 1985)). By this, we mean that it is activated at the blastoderm stage, in all or virtually all cells of the parasegment primordium. In these parasegments the activity of the gene is probably required to establish the basic characteristics of the segment. In other segments, the same homeotic gene may be deployed adventitiously, only in subsets of cells and at later stages of development. In these segments, the gene must play a secondary role in defining developmental pathways.

We argue below that the early 'metameric' domains of gene expression do correlate with the major morphological subdivisions of the fly, suggesting a relatively direct relationship between the expression of particular homeotic genes and the decisions taken by individual cells in early development. This relationship allows us to suggest a scenario for the evolution of homeotic genes in relation to the evolving morphological organization of the arthropod body plan in the insect myriapod lineage. To approach this question, we first review how the molecular analysis of homeotic gene expression has altered our understanding of homeotic gene function.

Combinatorial and mosaic models of homeotic gene activity

Current models relating the expression of homeotic genes to the control of segment development are based on the pioneering work of Lewis (1963, 1978, 1981), and relate primarily to the role of genes in the bithorax complex in the control of development of the posterior thorax and abdomen.

Lewis suggested that the bithorax complex contained a series of genes, each of which would be turned on at a different position along the anteroposterior axis of the fly, in partially overlapping domains (Lewis, 1963, 1978). These partially overlapping domains of gene expression would give each segment a different combination of active homeotic genes, and this combination would define segment identity.

This model has been integrated with the idea of 'selector' genes, or genes serving to provide a genetic address for a polyclonal group of cells (Garcia-Bellido, 1975). In this form (Struhl, 1982; Lawrence & Morata, 1983) the model focuses on the assumption that within a domain – be it segment (Lewis, 1978), parasegment (Martinez-Arias & Lawrence, 1985) or compartment (Garcia-Bellido, 1975) – every cell takes the same decision about which homeotic gene(s) to turn on at an early stage of development, and remembers that decision through all subsequent cell divisions.

For example, all cells of T2 (parasegment 4) were assumed to express *Antennapedia (Antp)* alone, whereas all cells of T1 (parasegment 3) would express both *Antp* and *Sex combs reduced* (*Scr*), and all cells of T3 (parasegment 5) would express *Antp* and *Ubx* (Fig. 2). By extension, each segment of the embryo would express a unique combination of homeotic genes that constituted a specific codeword, defining segment identity.

Three aspects of the available molecular data suggest that this model needs to be revised. First, there are not enough genes, and probably not enough different gene products, to give a unique combination of products in every segment. All functions of the bithorax complex depend on the activity of just three homeotic genes – *Ubx, abdominal-A* (*abd-A*) and *Abd-B* (Sanchez-Herrero *et al.* 1985; for review see Duncan, 1987). Together these control the different identities of at least nine parasegments.

A more serious revision is required by the observation that, at the level of protein expression, the cells of a segment do not all express the same combination of homeotic genes. In the blastoderm and very early gastrula, parasegments do appear as domains of uniform homeotic gene expression (Akam & Martinez-Arias, 1985; Martinez-Arias, 1986), but once a pattern of different cell types has been generated within each segment, most segments become mosaics of cells expressing different homeotic proteins (White & Wilcox, 1985; Carroll *et al.* 1988) and, indeed, by late stages of embryogenesis, most individual cells express only a single homeotic protein at high level, as the expression of one protein

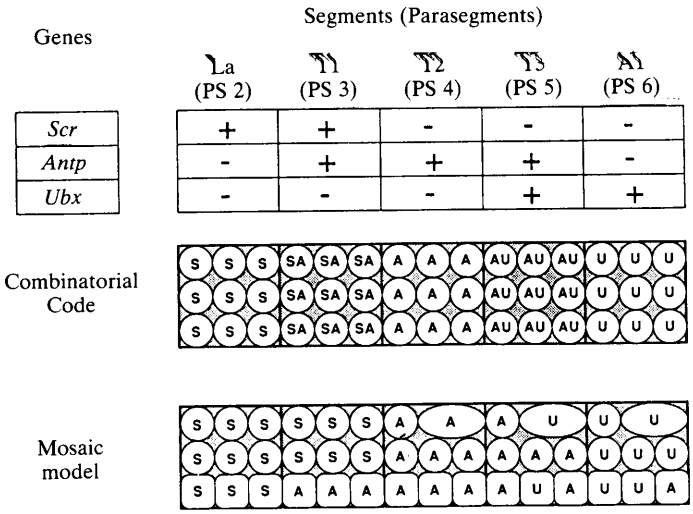

Fig. 2. Combinatorial and mosaic models of homeotic gene activity. The table at the top of the figure summarizes the requirement for each of the homeotic genes *Sex combs reduced* (*Scr*), *Antennapedia* (*Antp*) and *Ultrabithorax* (*Ubx*), based on genetic evidence (adapted from Struhl, 1982). According to the combinatorial model for homeotic gene function (middle panel), these requirements would reflect expression of an identical homeotic codeword in all cells of each segment. The observed patterns of protein expression suggests a different model (lower panel). By late stages of embryogenesis, segments that depend on a combination of two or more homeotic genes exhibit a mosaic of cells expressing different homeotic gene products. The differential expression of homeotic genes within a segment (or parasegment) reflects control of protein expression by cell type, as well as by segment identity.

represses the transcription of other homeotic genes in a fixed hierarchy (Hafen *et al.* 1984; Struhl & White, 1985).

Finally, certain cells switch from expressing one homeotic gene in very early development to expressing another, or none at all, in later stages of development, thus confounding the strict lineage model of homeotic gene expression. One clear example of this is seen in the ectoderm of parasegment 3, which expresses only *Antp* protein in the very early extended germ band, but then expresses *Scr* at later stages (Martinez-Arias *et al.* 1987; Carroll *et al.* 1988).

Many of these data can be reconciled with Lewis' model, if for the idea of 'turning on a gene' we substitute the phrase 'make accessible an enhancer' or region of DNA containing many independent enhancers. This is essentially the 'open-for-business' model developed by Welcome Bender and his colleagues (Peifer *et al.* 1988).

This model is based on the observation that, although there appear to be only three protein-coding 'homeotic genes' in the bithorax complex, each of these genes is controlled by multiple, independent regulatory regions (Bender *et al.* 1983; Karch *et al.* 1985). Thus the *Ubx* gene is controlled by the *bithorax* (*bx*) regulatory region in parasegment (PS) 5, and the *bithoraxoid* (*bxd*) regulatory region in PS6 (Fig. 3). Similarly the *abd-A* and *Abd-B* genes are controlled by a series of '*infra-abdominal*' (*iab*) regulatory regions – approximately one per segment (or parasegment; see Duncan, 1987).

The essence of the open-for-business model is that each cell decides, early in development, which regulatory elements or regulatory domains are 'open for business' according to position in the embryo. Each regulatory domain contains elements that interact independently with *trans*-acting factors to control the temporal and cell-specific activity of each homeotic gene. So for example, the '*bx*' regulators are 'open for business' in PS5, and are the only elements to control *Ubx* expression in this parasegment. The '*bxd*' regulators are 'open for business' in PS6 and more posteriorly, and act, perhaps with the PS5 regulators, to control *Ubx* expression in more posterior regions (Peifer *et al.* 1987).

While this model conforms to the formal structure of Lewis' model, its implications for the control of development are very different from those of the combinatorial selector gene model. For example, in PS6, *Ubx* protein is expressed in all or virtually all cells from shortly after the time of gastrulation (White & Wilcox, 1985). Thus, *Ubx* can play a role in early development throughout this parasegment. However, in PS5 the regulators of *Ubx* are 'off' at very early stages of development. *Ubx* expression first appears in a limited number of cells around the forming tracheal pits. Later, it appears in the nervous system, though at the extended germ band stage the precursors for the neural cells of PS5 show no trace of

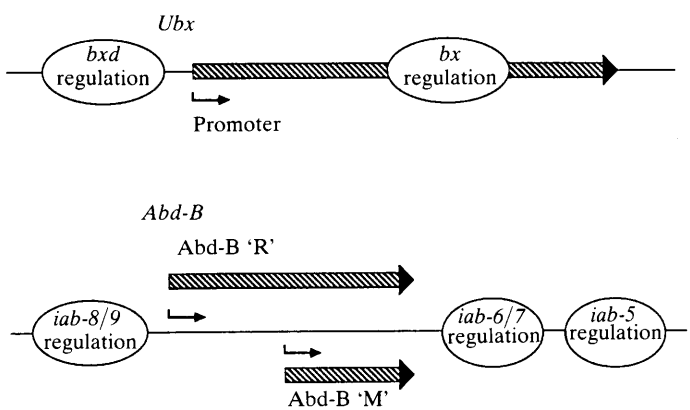

Fig. 3. Structural and regulatory elements of the *Ubx* and *Abd-B* genes. At *Ubx*, a single transcription unit is controlled by two regulatory regions, the *bithorax* (*bx*) region active in parasegment 5 and the *bithoraxoid* (*bxd*) region active in parasegment 6. At *Abd-B*, multiple regulatory elements control at least two distinct sets of transcripts utilizing the same homeobox.

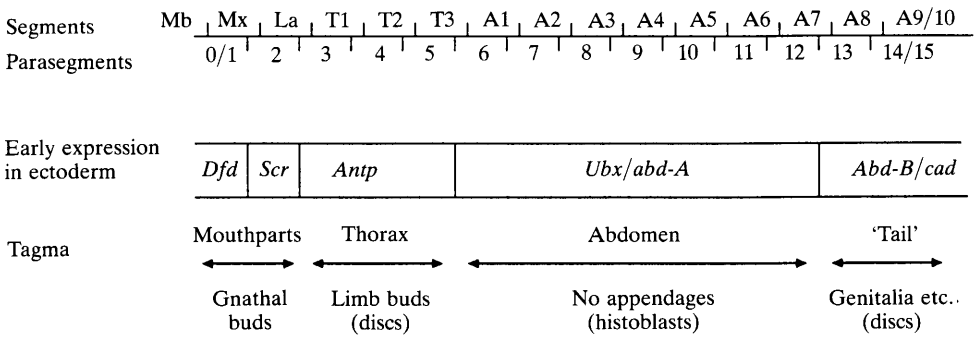

Fig. 4. Correspondence of the 'metameric' domains of homeotic gene expression with distinct morphological divisions of the body. Domains of expression are shown for each homeotic gene in the very early germ band, when boundaries of expression are parasegmental. For details see text.

Ubx transcription (Akam & Martinez-Arias, 1985; White & Wilcox, 1985).

The deployment of *Abd-B* illustrates the difference between these two models particularly clearly. *Abd-B* gives rise to at least two distinct sets of transcripts, probably from multiple promoters (Sanchez-Herrero & Crosby, 1988; DeLorenzi *et al.* 1988). One set are present only in the extreme posterior (PS14–15) under the control of the *iab-8,9* regulators, whereas the other transcripts are expressed more anteriorly, under the control of different regulators, some of which (*iab 5,6,7*) lie 3′ to the *Abd-B* gene (Fig. 3).

In early development, in the blastoderm and early extended germ band, *Abd-B* transcripts are expressed only in the extreme posterior of the presumptive abdomen, from parasegments 13–14. As development proceeds, detectable levels of transcripts appear in PS12, then 11, then 10 (Sanchez-Herrero & Crosby, 1988). Although the *iab 5,6,7* regulatory elements may perhaps be 'open for business' from blastoderm stages onwards in PS10 to 12, the *Abd-B* gene can play no role in controlling the development of these segments until a later stage of development. Thus the developmental role of *Abd-B* in PS10–12, like that of *Ubx* in PS5, must be more limited than the role of these same genes in their 'metameric' domains – PS6 for *Ubx*; PS13–14 for *Abd-B*.

The relationship between homeotic gene expression and segment development

If we consider, not the entire domain of action of each homeotic gene, but rather these metameric domains, then we see a closer relationship between homeotic gene expression and segment development. This is illustrated in Fig. 4, which summarizes homeotic gene expression in the ectoderm when the germ band first forms. To make this correlation clear we must look in more detail at the morphology of the Diptera.

The segmental organization of the Diptera

The tagmata of the insect body cannot simply be equated with groups of segments sharing common developmental origins. The gnathal segments, for example, probably did not evolve by the diversification of an archetypal gnathal segment. Comparative morphology suggests rather that they evolved successively, presumably by the independent evolution of mechanisms to modify the most anterior remaining trunk segments (Anderson, 1973).

There are better grounds for believing that the ancestors of modern insects possessed multiple, near identical, thoracic and abdominal segments, as such regions are still present in the trunk of myriapods, and in the abdomen of many insects. The diversity of segments within these tagmata may therefore be regarded as deriving secondarily through the diversification of generalized thoracic and abdominal progenitors.

Using the term tagma in this developmental and evolutionary sense, we would limit the abdomen to include only the progenital abdominal segments, A1–A7 (Snodgrass, 1935; Matsuda, 1975). Segments of the posterior abdomen (PS13–15, A8–A11) or tail (Jurgens, 1987) differ from the anterior abdominal segments, and from one another, in many respects, and give rise to the specialized reproductive and cercal appendages. Although functionally part of the abdomen, there are few grounds for regarding these tail segments as derived by modification of a segment resembling those in the preabdomen. Differentiated genital and terminal segments exist in the myriapods, so the equivalent structures of insects are more likely to derive from independent modification of an unspecialized trunk segment.

Within the abdomen, the first segment (A1) is generally unique, and clearly distinguishable even in early embryonic stages from A2–A7 (Anderson, 1972). These remaining segments of the preabdomen are strikingly similar during early embryogenesis. Their similarity is retained even by the adults of a

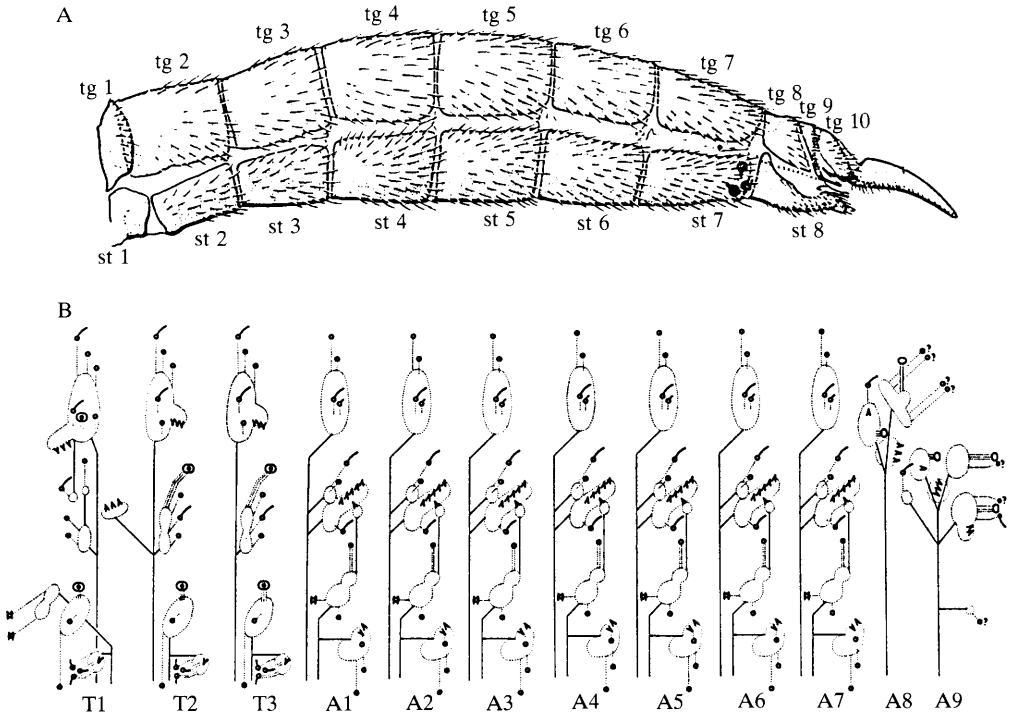

Fig. 5. Similarity of segments in the preabdomen of Diptera. (A) The abdomen of a female of the nematoceran *Trichocera columbiana*. In this and many other lower Diptera, the pregenital segments of the abdomen (A2–A7) remain very similar in external morphology. tg, tergite; st, sternite (from McAlpine *et al.* 1981). (B) Schematic diagram of the peripheral nervous system of the *Drosophila* embryo. An identical set of sensory cells are present on each of abdominal segments A1–A7. Peripheral innervation of the more posterior abdominal segments is very different (from Dambly-Chaudiere & Ghysen, 1986).

typical lower dipteran (e.g. *Trichocera*, Fig. 5), but in *Drosophila* and other higher Diptera, segments A5–A7 of the adult have become specialized (McAlpine *et al.* 1981). None the less, in the *Drosophila* embryo, segments A2–A7 remain remarkably similar. Even in such complex patterns as the somatic musculature (Hooper, 1986) and the peripheral nervous system (Ghysen *et al.* 1986; Dambly-Chaudiere & Ghysen, 1986), these segments are indistinguishable.

Metameric domains of gene expression

These morphological domains are clearly correlated with the metameric domains of homeotic gene expression established by early germ band stages. For example, *Ubx* and/or *abd-A* are expressed throughout PS6–12, which is the parasegmental equivalent of the preabdomen. The early domain of *Abd-B* expression is limited exclusively to the postabdomen (Fig. 4).

Ubx alone is expressed in PS6 (A1), but both *Ubx* and *abd-A* are expressed in the more posterior parts of the preabdomen. Where their expression overlaps, at least some of their functions are redundant. For example, the activity of either gene will result in normal development of the tracheal system (Lewis, 1981) and of certain characteristic structures of the nervous system (Ghysen & Lewis, 1986) in the abdominal segments. These two genes are the most similar of the *Drosophila* homeotic genes in the sequence of their homeoboxes (Fig. 6). Most remarkably, a mutation that fuses the amino terminus of the *abd-A* protein-coding region with the homeobox and carboxy terminus of *Ubx* allows expression of a functional product that rescues many aspects of both the *Ubx* and *abd-A* mutant phenotypes (Rowe & Akam, 1988). All of these observations suggest that both of these genes are able to lay down the ground plan for the preabdominal segments – albeit each is 'fine tuned' to make different structures.

The equivalent early role of *Abd-B* is probably to define the genital and postgenital regions of the abdomen, but development of this region depends also on the activity of another homeobox-containing gene, *caudal* (Macdonald & Struhl, 1986; Mlodzik & Gehring, 1987). The early expression domains of these two genes may overlap, though in later development they appear to be expressed in different structures (Sanchez-Herrero & Crosby, 1988), with *caudal* required for the development of the most posterior segmental structures (anal pads, A10 or A11).

In the more anterior regions, *Deformed* (*Dfd*) is

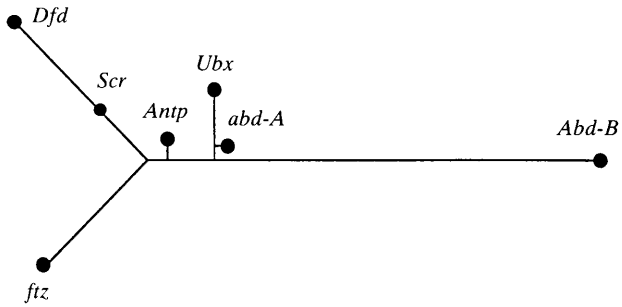

Fig. 6. Relatedness of homeobox sequences among *Antennapedia*-like genes. The diagram is an unrouted tree (Wagner network) based on amino acid differences between the different homeobox sequences. By this criterion, *Abd-B* is no more closely related to *Antp* than are several homeobox genes not generally included in the *Antennapedia*-like family (e.g. *labial, caudal*). All sequences are taken from published references cited in Akam (1987) except that for *abd-A*, which was provided by G. Tear (see Fig. 8).

expressed metamerically in and anterior to parasegment 1, while *Scr* is expressed initially throughout parasegment 2 (Martinez-Arias *et al.* 1987; Chadwick & McGinnis, 1987; Riley *et al.* 1987). These genes together specify the development of the gnathal segments. *Antp* is expressed metamerically in parasegment 4, and slightly later throughout the ectoderm of parasegments 3 and 5, thus extending throughout the thoracic parasegments (Martinez-Arias, 1986 and personal communication). However, the conceptual 'metameric' expression of *Antp* is fleeting because, in PS5, limited expression of *Ubx* appears shortly after germ band extension and, in PS3, the expression of *Antp* is 'overwritten' by *Scr* in the epidermis at about the same time (Carroll *et al.* 1988).

The correlation presented above ignores the inevitable overlaps that arise in aligning tagma, defined in segmental terms, with the early parasegmental domains of homeotic gene expression. Effectively, it assigns the identity of each parasegment to the tagma containing its A compartment. There is some justification for this. The primordia for the P compartments, defined by *engrailed* expression, include only about a quarter of the cells in each parasegment (DiNardo *et al.* 1985), and include few of the specific pattern elements on the trunk (though more on the appendages). The mesoderm of each segment derives entirely from the same parasegment as its A compartment (Lawrence, 1985). Moreover, in cases where a shift occurs from parasegmental to segmental expression of a homeotic gene (as for *Dfd* and *Scr* in the gnathal buds (Martinez-Arias *et al.* 1987; Jack *et al.* 1988)) it is the P compartment that switches to match expression of the anteriorly adjacent A compartment.

The relationship presented here suggests that the expression of homeotic genes in the very early germ band may directly control developmental decisions that establish what is best-called 'tagmatic identity' for each parasegment; not by a complicated decoding process, but as a relatively direct consequence of the homeotic gene first expressed. Once this decision has been taken, changing patterns of homeotic gene expression modify the development of segments within tagmata.

The evolution of segment diversity

We can relate these suggestions to the evolution of segment diversity by making two assumptions. First, we follow Lewis (1978) in assuming that the evolution of the homeotic gene family parallels, and in some degree directs, the evolution of segment diversity in the myriapod–insect lineage. Secondly, we suggest, in what is essentially a molecular version of Von Baer's rule (see Gould, 1977), that the patterns of gene deployment seen in early development – or more properly in the earliest phylotypic stage (Sander, 1983) – are likely to represent phylogenically ancient roles for the gene. Conversely, patterns of deployment that first appear in later development are more likely to represent recent uses of the gene.

The origins and relationships of homeotic genes

Although the homeobox was first characterized in homeotic and segmentation genes, it is becoming clear that proteins containing this structural motif are involved in a very wide range of developmental processes, in both segmented and nonsegmented organisms (e.g. Saint *et al.* 1988; Way & Chalfie, 1988).

On the basis of gene structure, this diverse family falls into a number of more or less well-defined subfamilies. One of these, the *Antennapedia*-like subfamily, includes most of the homeotic, segment-selector genes. *Antp, Ubx, abd-A* and *Scr* share particularly similar homeobox sequences (Fig. 6), and also share short conserved peptide motifs in other regions of the protein.

Genes sharing all of these structural features are also found in vertebrates (Krumlauf *et al.* 1987), indicating that this *Antp*-like subfamily arose before the divergence of the major animal phyla. At least within the homeobox, the sequence of the modern *Antp* gene appears to be very close to the ancestral sequence for this subfamily (Gehring, 1986).

The sequence of the *Dfd* homeobox is somewhat more diverged from the *Antp*-like concensus, although *Dfd* shares with the *Antp*-like genes at least one of the conserved peptides outside of the homeobox. However, homeotic genes that are clearly re-

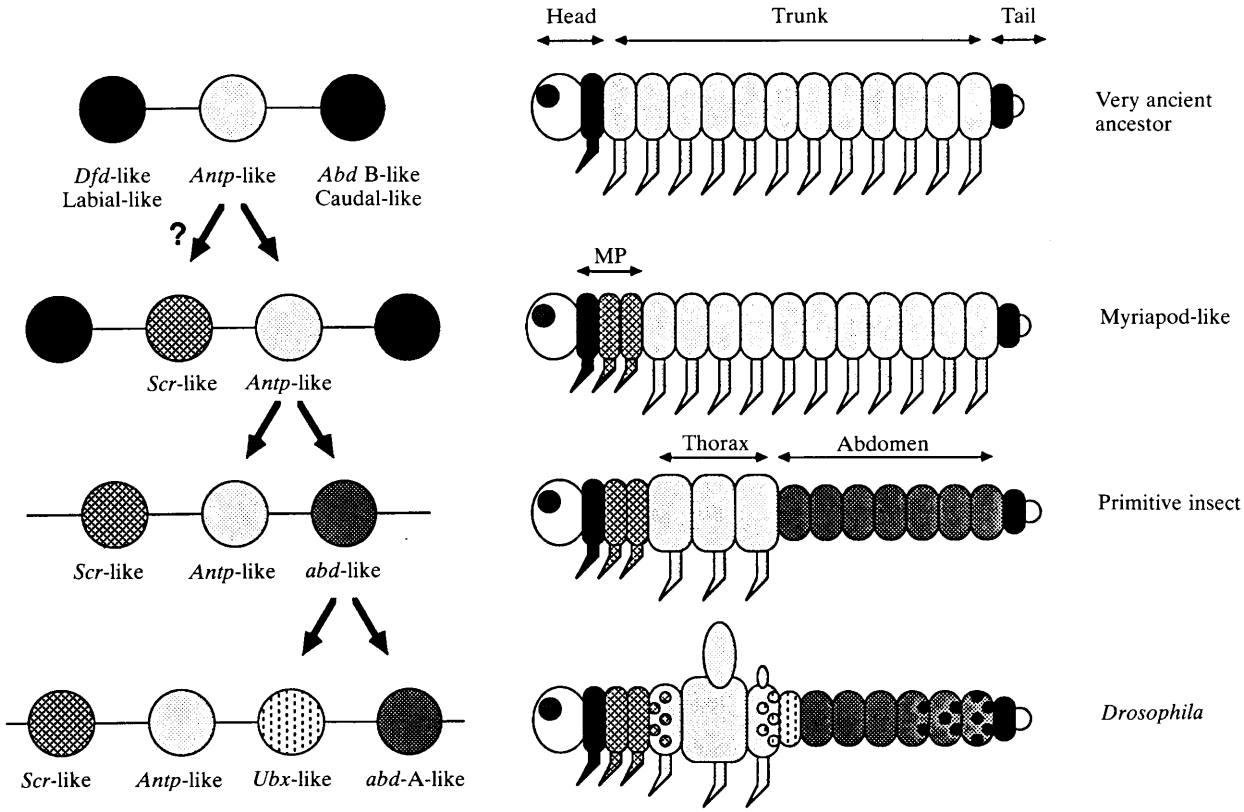

Fig. 7. A scheme for the origin of homeotic genes during evolution of the myriapod–insect lineage. The panel on the left shows the proposed set of *Antp*-like genes existing at each stage in the evolution of the myriapod–insect lineage. The black circles (shown only for stages 1 and 2) indicate the existence of other homeobox genes that probably arose prior to the isolation of this lineage. The arrows indicate potential gene duplication events. Corresponding shading on the diagrams at the right shows the proposed domains of expression for each of these genes, corresponding to the progressive specialization of trunk segments. Complex patterns of gene expression in the head and in the tail are depicted only by black boxes.

lated to *Dfd* rather than *Antp* are found in vertebrates, indicating that the distinction between *Dfd*-like and *Antp*-like genes predates the vertebrate–arthropod divergence (Regulski *et al.* 1987).

The same is probably true for the *Abd-B* gene of the bithorax complex, although no very close homologues of *Abd-B* have yet been identified in vertebrates. The sequence of the *Abd-B* gene places it outside the close *Antp* family; its homeobox sequence is no more closely related to *Antp* than are the *Drosophila* genes *caudal* and *labial*, and, of these, at least *labial* is represented in vertebrates by distinct and well-conserved homeobox homologue (Mlodzik *et al.* 1988).

In *Drosophila*, *Abd-B*, *caudal* and *labial* are all expressed in defined spatial domains either at the ends of, or beyond the limits of, the metameric region (Hoey *et al.* 1986; Akam, 1987). In vertebrates, the homeobox genes are also expressed in different regions along the A–P axis of the animal, suggesting that their role in regionalization may be a very ancient one (Holland & Hogan, 1988).

The role of homeotic genes in the evolution of the trunk segments

These data suggest that the earliest ancestor we can reasonably imagine for the myriapod–insect lineage already possesed distinct homeobox genes related to *labial*, *Deformed*, *Antp* and probably also *Abd-B* and/or *caudal* (Fig. 7; see also Martinez-Arias, 1987). We suggest that this animal utilized the *Antp* gene to control various aspects of the development of a region that has become the trunk in arthropods, and that this region of the body was already bounded by domains expressing homeobox genes related to *Deformed* and *caudal*. We would guess, however, that this common ancestor did not have specific homologues for the different genes of the *Antp*-like family that now specify the diversity of parasegments 2–12 in *Drosophila*.

In the line that led to the myriapods, a mechanism evolved that allowed the most anterior trunk segments to develop specialized mouthparts, quite different from the typical trunk segments. The deployment of homeotic genes in *Drosophila* suggests that

this process utilized the *Dfd* gene, and may have been associated with the origin of an *Scr*-like gene, but it is not possible from what we know at present to guess when, or from what ancestral gene(s), *Scr* arose. At some stage after this, distinct thorax/abdomen tagmosis evolved in the insect lineage. Our view of the role of homeotic genes leads us to suggest that a functional equivalent of *Ubx* or *abd-A* arose at this point.

With this array of homeotic genes, and with appropriate mechanisms to control segment number, our ancient ancestor could have been transformed into a primitive insect, with three similar thoracic segments and an array of similar preabdominal segments. To make a typical modern insect requires a sophisticated mechanism to make each segment different. Two rather early innovations may have been the use of *Scr* to make T1 different from the other thoracic segments, and also the origin of distinct *Ubx* and *abdominal-A* genes to make A1 different from the rest of the preabdomen. The evolution of the higher Diptera, leading to *Drosophila*, may have been achieved without any additional homeotic genes, but by the evolution of more complex regulatory mechanisms to modulate the expression of existing genes.

Suggested modes of evolutionary change

The scenario present above points to mechanisms whereby the current elaboration of segment specialization may have arisen. We are suggesting that one of the more rapid ways a homeotic gene may mediate evolutionary change is by acquiring new regulatory signals that work on existing products, independent of, and in addition to, existing regulation. We may be seeing the results of such evolutionary changes in the patterns of deployment of *Ubx* in PS5, and *Abd-B* in PS10–12. It is perhaps worth noting that in these two cases the *cis*-acting regulatory elements for the 'non-metameric' expression of these genes are clearly distinct from the DNA regions required for the metameric domain of expression, and in both cases lie distant from (and 3′ to) the promoter (Fig. 3).

Clearly, the evolution of novel regulatory mechanisms can proceed with less constraint if gene duplication occurs, allowing closely related proteins that have extensively overlapping functions to come under totally independent regulation. To some extent the *Ubx* and *abd-A* genes of *Drosophila* may be viewed as being still in this state. However, once genes have duplicated, functional divergence can create products with different regulatory effects. Such processes presumably occurred in the origin of the *abdominal*-like genes from an *Antp*-like ancestor, in association with the origin of thorax-abdomen tagmosis.

(5) An experimental approach

The scheme presented above makes predictions about which organisms should share genes with specific *Scr*-like or *Ubx*-like properties, and what features of the expression of these genes we should expect to see conserved, which different. For instance, it 'predicts' that the *bx* functions of *Ubx* and the *iab 5–7* functions of *Abd-B* may not be conserved among all insects, whereas the early embryonic PS6–12 domains of *Ubx*, and PS13–14 domains of *Abd-B* should be.

To test these predictions, we are cloning the homeotic genes of the *Antp*-like family from the locust *Schistocerca gregaria*, a representative of one of the older and embryologically more primitive groups of the Neoptera (I. Dawson, G. Tear, A. Martinez-Arias and M. Akam, in preparation).

The locust is a short-germ insect. That is to say, at the onset of gastrulation only the anterior part of the body plan is defined; posterior segments are generated by a growth process during embryogenesis. This difference raises questions concerning the establishment of segment pattern, intimately related to the problems of segment specification. These are discussed in the article by Tear, Bate and Martinez-Arias in this volume.

We have sequenced the homeobox regions of three homeotic genes from the locust. On the basis of sequence, two are clearly the locust homologues of specific *Drosophila* genes – *Scr* and *abd-A*. The homeoboxes are identical or almost identical (Fig. 8) and there is enough homology in the regions flanking the homeobox to identify the genes unambiguously, independent of the homeobox sequence itself.

We have only just started to analyse the pattern of expression of these genes. The most striking result is how similar are the patterns of expression between *Drosophila* and *Schistocerca*, at least in relatively late germ band stages. In both species, *Scr* is expressed in part of the mandibular segment, in the labium and in the anterior part of T1. The *abd-A* gene is expressed in the epidermis of the abdominal segments from posterior A1 backwards (I. Dawson, G. Tear, A. Martinez-Arias and M. Akam, in preparation).

The third homeobox is much less easy to identify. Its sequence places it firmly among the *Antp*-like family, but it differs in nine amino acids from *Antp* and more from all other known *Drosophila* homeoboxes. At present, we know only that this gene is expressed in the nervous system, without obvious segment specificity; in this respect it most resembles the neural expression of *fushi-tarazu*, the only segmentation gene that shares an *Antp*-like homeobox. If a perfect homologue of this third homeobox existed in *Drosophila*, the sequence would almost certainly

```
                    5'                          Homeobox                                    3'
Scr
Drosophila      TVNANGE  TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKEH  KMASMNIVP
                                                                                          ****
Locust          TVNANGE  TKRQRTSYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKEH  KMASMSAGG
                         *  * **                                                  *   *
Antp                     RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN

abd-A
Drosophila      FNGPNGCP RRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKEL  RAVKEINEE
                ***
Locust          ALQPNGCP RRRGRQTYTRFQTLELEKEFHFNHYLTRRRRIEIAHALCLTERQIKIWFQNRRMKLKKEL  RAVKEINEE
                          *       *          *                                  *   *
Antp                     RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
```

Fig. 8. Comparison of homeobox and flanking amino acid sequences from *Schistocerca gregaria* and *Drosophila*. The figure showing the amino acid sequence (in Dayhoff code) predicted from the nucleotide sequence of genomic DNA fragments spanning the homeobox of the *Drosophila Scr* and *abd-A* genes, and of two genes from the locust *Schistocerca gregaria*. The amino acid sequence of the *Drosophila Antp* homeobox is shown below each gene for comparison. Asterisks denote mismatches between the *Drosophila* and locust sequences. *Drosophila Scr* and *Antp* sequences are from Kuroiwa *et al.* (1985), as corrected by U. Walldorf and W. Gehring (personal communication). Sequences for the locust and for *Drosophila abd-A* are from I. Dawson, G. Tear, A. Martinez-Arias and M. Akam, in preparation.

have been detected by hybridization with *Antp* probes. It is therefore likely that this gene represents either a function that is no longer present in the *Drosophila* genome, or a homeobox subject to less stringent evolutionary constraint than the *abd-A* and *Scr* genes.

It remains to be seen whether such changes in gene structure, and changes in the pattern of deployment of genes during development, can be correlated with the evolution of new developmental strategies.

We acknowledge many discussions with Alphonso Martinez-Arias that have contributed to the development of the ideas presented above. We thank Michael Bate for teaching us the embryology of locusts, and Adrian Friday for calculating the network illustrated in Fig. 6. The evolutionary speculation presented above was prompted by an invitation to speak on the topic of 'Evolution and Development' at the Markey Symposium held in honour of E. B. Lewis at Caltech in April, 1988. Experimental work from our Laboratory is supported by the Medical Research Council.

References

Akam, M. (1987). The molecular basis for metameric pattern in the *Drosophila* embryo. *Development* **101**, 1–22.

Akam, M. & Martinez-Arias, A. (1985). The distribution of *Ultrabithorax* transcripts in *Drosophila* embryos. *EMBO J.* **4**, 1689–1700.

Anderson, D. T. (1972). The development of hemimetabolous insects. In *Developmental systems: Insects*. Volume 1 (ed. S. J. Counce & C. H. Waddington). pp. 96–165. London: Academic Press.

Anderson, D. T. (1973). *Embryology and Phylogeny in Annelids and Arthropods*. Oxford: Pergamon.

Bender, W., Akam, M. E., Karch, F., Beachy, P. A., Peifer, M. Spierer, P., Lewis, E. B. & Hogness, D. S. (1983). Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science* **221**, 23–29.

Bodenstein, D. (1950). The post-embryonic development of *Drosophila*. In *The Biology of Drosophila* (ed. M. Demerec), pp. 275–364. New York: Hafner.

Chadwick, R. & McGinnis, W. (1987). Temporal and spatial distribution of transcripts from the *Deformed* gene of *Drosophila*. *EMBO J.* **6**, 779–789.

Carroll, S. B., DiNardo, S., O'Farrell, P. H., White, R. A. H. & Scott, M. P. (1988). Temporal and spatial relationships between segmentation and homeotic gene expression in *Drosophila* embryos: Distribution of the *ftz*, *en*, *Scr*, *Antp* and *Ubx* proteins. *Genes and Development* **2**, 350–360.

Dambly-Chaudiere, C. & Ghysen, A. (1986). The sense organs in the *Drosophila* larva and their relation to the embryonic pattern of sensory neurons. *Wilhelm Roux Arch. devl Biol.* **195**, 222–228.

DeLorenzi, M., Ali, N., Saari, G., Henry, C., Wilcox, M. & Bienz, M. (1988). Evidence that the *Abd-B* R element function is conferred by a trans-regulatory homeoprotein. *EMBO J.* **7**, 3223–3231.

Duncan, I. (1987). The Bithorax complex. *A. Rev. Genet.* **21**, 285–319.

Garcia-Bellido, A. (1975). Genetic control of wing disc development in *Drosophila*. In 'Cell patterning', *CIBA Foundation Symposium* **29**, pp. 161–178. Amsterdam: Associated Scientific.

Gehring, W. J. (1986). The homeobox: Structural and evolutionary aspects. In *Molecular Approaches to Developmental Biology*, pp. 115–129. New York: Liss.

Ghysen, A., Dambly-Chaudiere, Ch., Aceves, E., Jan,

L. Y. & JAN, Y. N. (1986). Sensory neurons and peripheral pathways in *Drosophila* embryos. *Wilhelm Roux Arch. devl Biol.* **195**, 281–289.

GHYSEN, A. & LEWIS, E. B. (1986). The function of bithorax genes in the abdominal nervous system of *Drosophila. Wilhelm Roux Arch. devl Biol.* **195**, 203–209.

GOULD, S. J. (1977). *Ontogeny and Phylogeny.* Cambridge, Mass: Harvard.

HAFEN, E., LEVINE, M. & GEHRING, W. J. (1984). Regulation of *Antennapedia* transcript distribution by the bithorax complex of *Drosophila. Nature, Lond.* **307**, 287–289.

HOEY, T., DOYLE, H. J., HARDING, K., WEDEEN, C. & LEVINE, M. (1986). Homeo box gene expression in anterior and posterior regions of the *Drosophila* embryo. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4809–4813.

HOLLAND, P. W. H. & HOGAN, B. L. M. (1988). Expression of homeobox genes during mouse development: a review. *Genes and Development* **2**, 773–782.

HOOPER, J. (1986). Homeotic gene function in the muscles of *Drosophila* larvae. *EMBO J.* **5**, 2321–2329.

JACK, T., REGULSKI, M. & MCGINNIS, W. (1988). Pair rule segmentation genes regulate the expression of the homeotic selector gene *Deformed. Genes and Development* **2**, 635–651.

JURGENS, G. (1987). Segmental organization of the tail region in the embryo of *Drosophila melanogaster.* A blastoderm fate map of the cuticle structures of the larvae tail region. *Wilhelm Roux Arch. devl Biol.* **196**, 141–157.

KARCH, F., WEIFFENBACH, B., PEIFER, M., BENDER, W., DUNCAN, I., CELNIKER, S., CROSBY, M. & LEWIS, E. B. (1985). The abdominal region of the bithorax complex. *Cell* **43**, 81–96.

KRUMLAUF, R., HOLLAND, P. W. H., MCVEY, J. H. & HOGAN, B. L. M. (1987). Developmental and spatial patterns of expression of the mouse homeobox gene *Hox 2.1. Development* **99**, 603–618.

LAWRENCE, P. A. (1985). Notes on the genetics of pattern formation in the internal organs of *Drosophila. Trends in Neurosciences* **6**, 267–269.

LAWRENCE, P. A. & MORATA, G. (1983). The elements of the bithorax complex. *Cell* **35**, 595–601.

LEWIS, E. B. (1963). Genes and developmental pathways. *Am. Zool.* **3**, 33–56.

LEWIS, E. B. (1978). A gene complex controlling segmentation in *Drosophila. Nature, Lond.* **276**, 565–570.

LEWIS, E. B. (1981). Developmental genetics of the bithorax complex in *Drosophila.* In *Developmental Biology using Purified Genes. ICN-UCLA symposia* **23** (ed. D. D. Brown & C. F. Fox), pp. 189–208. New York: Academic Press.

MACDONALD, P. M. & STRUHL, G. (1986). A molecular gradient in early *Drosophila* embryos and its role in specifying the body pattern. *Nature, Lond.* **324**, 537–545.

MARTINEZ-ARIAS, A. (1986). The *Antennapedia* gene is required and expressed in parasegments 4 and 5 of the *Drosophila* embryo. *EMBO J.* **5**, 135–141.

MARTINEZ-ARIAS, A. (1987). On the developmental and evolutionary role of some genes from the *Antennapedia* complex. In *Molecular Approaches to Developmental Biology*, pp. 131–145. New York: Liss.

MARTINEZ-ARIAS, A., INGHAM, P. W., SCOTT, M. & AKAM, M. (1987). The spatial and temporal deployment of *Dfd* and *Scr* transcripts throughout development of *Drosophila. Development* **100**, 673–683.

MARTINEZ-ARIAS, A. & LAWRENCE, P. A. (1985). Parasegments and compartments in the *Drosophila* embryo. *Nature, Lond.* **313**, 639–642.

MAHAFFEY, J. W. & KAUFMAN, T. C. (1987). The homeotic genes of the Antennapedia complex and the bithorax complex of *Drosophila.* In *Developmental Genetics of Higher Organisms* (ed. G. M. Malacinski), pp. 329–359. New York: Macmillan.

MATSUDA, R. (1975). *Morphology and Evolution of the Insect Abdomen.* Oxford: Pergamon.

MCALPINE, J. F., PETERSON, B. V., SHEWELL, G. E., TESKEY, H. J., VOCKEROTH, J. R. & WOOD, D. M. (1981). Manual of nearctic Diptera. Volume 1. Research Branch Agriculture Canada. Monograph 27. Quebec.

MLODZIK, M., FJOSE, A. & GEHRING, W. J. (1988). Molecular structure and spatial expression of a homeobox gene from the *labial* region of the Antennapedia complex. *EMBO J.* **7**, 2569–2578.

MLODZIK, M. & GEHRING, W. J. (1987). Expression of the *caudal* gene in *Drosophila*: Formation of an RNA and protein gradient during early embryogenesis. *Cell* **48**, 465–478.

PEIFER, M., KARCH, F. & BENDER, W. (1987). The bithorax complex: Control of segment identity. *Genes and Development* **1**, 891–898.

REGULSKI, M., MCGINNIS, N., CHADWICK, R. & MCGINNIS, W. (1987). Developmental and molecular analysis of *Deformed*; a homeotic gene controlling *Drosophila* head development. *EMBO J.* **6**, 767–777.

RILEY, P. D., CARROLL, S. B. & SCOTT, M. P. (1987). The expression and regulation of *Sex combs reduced* protein in *Drosophila* embryos. *Genes and Development* **1**, 716–730.

ROWE, A. & AKAM, M. (1988). The structure and expression of a hybrid homeotic gene. *EMBO J.* **7**, 1107–1114.

SAINT, R. B., KALIANIS, B., LOCKETT, T. J. & ELIZAR, A. (1988). Pattern formation in the developing eye of *Drosophila melanogaster* is regulated by the homeobox gene *rough. Nature, Lond.* **334**, 151–154.

SANCHEZ-HERRERO, E. & CROSBY, M. A. (1988). The *Abdominal-B* gene of *Drosophila melanogaster*: Overlapping transcripts exhibit two different spatial distributions. *EMBO J.* **7**, 2163–2173.

SANCHEZ-HERRERO, E., VERNOS, I., MARCO, R. & MORATA, G. (1985). Genetic organization of *Drosophila* bithorax complex. *Nature, Lond.* **313**, 108–113.

SANDER, K. (1983). The evolution of patterning mechanisms: Gleanings from insect embryogenesis and

spermatogenesis. In *Development and Evolution* (ed. B. C. Goodman, N. Holder & C. C. Wylie), pp. 137–159. Cambridge.

SNODGRASS, R. E. (1935). *Principles of Insect Morphology*. London and New York: McGraw-Hill.

STRUHL, G. (1982). Genes controlling segmental specification in the *Drosophila* thorax. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7380–7384.

STRUHL, G. & WHITE, R. A. H. (1985). Regulation of the *Ultrabithorax* gene of *Drosophila* by other bithorax complex genes. *Cell* **43**, 507–519.

WAY, J. C. & CHALFIE, M. (1988). *Mec-3*, a homeobox containing gene that specifies differentiation of the touch receptor neurons in *Caennorhabditis elegans*. *Cell* **54**, 5–16.

WHITE, R. A. H. & WILCOX, M. (1985). The distribution of *Ultrabithorax* proteins in *Drosophila*. *EMBO J.* **4**, 2035–2043.