Clusters of Master Control Genes Regulate the Development of Higher Organisms

Edward B. Lewis, PhD

THIS article is an account of how basic research on the nature of the gene has led to the discovery of two clusters of master control genes that play a central role in programming the development of the fly. These clusters, known as the bithorax complex (BX-C) and Antennapedia complex (ANT-C), are believed to have evolved from an ancestral gene by a process of tandem duplication and divergence in function by mutation. So successful during evolution have cognates of these clusters been in programming development that they are now found in vertebrates, including humans, as well as in invertebrates.

The strategy of achieving complexity by duplicating and then diversifying is used at the developmental level, as well as at the genetic level. Thus, early in embryonic life, most higher organisms consist of little more than a tandem array of duplicated body segments. During the course of development, the segments diverge from one another and diversify to produce the variety of forms that these organisms achieve. It now seems likely that it is precisely the genes of the BX-C and ANT-C, or their cognates in other organisms, that initiate and control the diversification of the body segments.

This article is concerned primarily with the BX-C, how it came to be discovered, and how we think it functions during development.

From the Division of Biology, California Institute of Technology, Pasadena.

The 1991 Albert Lasker Medical Award was presented to Dr Lewis on September 27, 1991, in New York City.

Reprint requests to Division of Biology, California Institute of Technology, Pasadena, CA 91125 (Dr Lewis).

EARLY HISTORY

The view of genes as beads on a string, randomly arranged by function, was forever altered in 1925 by Sturtevant's discovery that the function of a gene is dependent on its position in the chromosome.[1] He came on this phenomenon of "position effect" in the course of analyzing one of the first cases of unstable genes in *Drosophila*, the famous *Bar*-eyed mutant, found by Tice in 1913. Sturtevant showed that the occurrence of two exceptional types of progeny from homozygous *Bar* flies—non-*Bar* (normal-eyed) and double-*Bar* (tiny-eyed) types—was associated with crossing-over of an unusual kind. He assumed that non-*Bar* types had lost the *Bar* gene and that double-*Bar* was a tandem duplication of that gene. (A detailed account of this early work and of relevant crossing-over mechanisms is given by Maeda and Smithies.[2])

In 1936, using the then recently discovered giant salivary gland chromosomes of the *Drosophila* larva, Bridges[3] and Muller et al[4] showed, independently, that the *Bar* mutation is itself a tandem duplication of seven bands, and double-*Bar* is a tandem triplication, rather than a duplication, for those bands. The origin of this triplication can be easily pictured if, for present purposes, we represent *Bar* as having the sequence ABCABC and if we introduce a crossing-over between two homologous ABC regions that are unequally paired. Thus,

$$\begin{array}{l} \text{ABCABC} \\ \qquad \times \longrightarrow \text{ABCABCABC} \\ \text{ABCABC} \end{array}$$

Non-*Bar* is, of course, readily accounted for as the reciprocal crossover product, ABC.

From homozygotes for the triplication a similar kind of crossing-over event is expected to generate a single copy of ABC and a fivefold tandem repetition of ABC. Just such a quintuplication was generated by Rapoport[5] who called it quadruple-*Bar*, since at the time he would not have known that *Bar* itself was a duplication. Later, Rapoport obtained a sevenfold to ninefold repetition from homozygous quadruple-*Bar* females, as described by Lindsley and Grell.[6] These early studies showed for the first time that, starting with a tandem duplication, multiple linear repetition can be generated by successive rounds of crossing-over.

In a second development, Bridges[7] called attention to several types of salivary gland chromosome banding patterns, which he interpreted as duplications that have become established in the species. Genetic evidence for such duplications was however lacking.

In 1940, Oliver[8] reported that he could recover apparent reverse mutations associated with crossing-over between two *lozenge* eye mutants of *Drosophila*. An alternative interpretation was that the reversions were not mutations but instead were wild-type crossovers between two separate but closely linked loci. Failure to find the reciprocal crossover having both mutants in the same chromosome left the validity of the latter interpretation in limbo.

Another pair of rough-eyed mutants, *Star* (*S*) and *asteroid* (*ast*), proved more tractable to analysis. (This mutant, originally called *Star-recessive*,[9] was found by E. Novitski, who sent it to me in 1937, suggesting that I test to see if it was an allele of *S*. We were under-

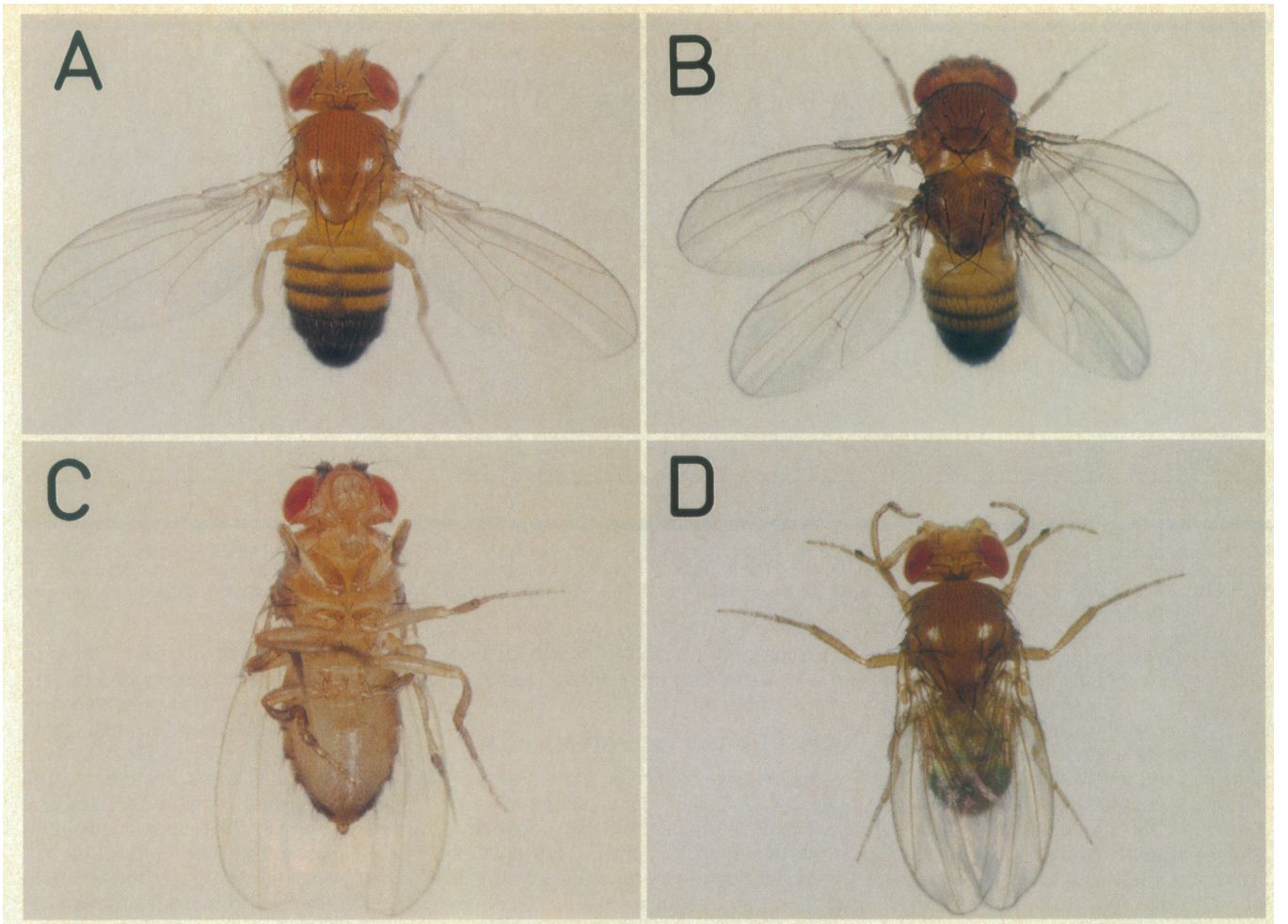

Fig 1.—Photographs of adult wild-type and homeotic mutant flies. A, Wild-type male (genotype: *heldout,* in order to produce spread wings). B, Dorsal view of a four-winged fly homozygous for the triple mutant combination *anterobithorax bithorax*^{3} *postbithorax*. The third thoracic segment is transformed into one resembling the second, resulting in two sets of nearly identical thoraces and wings. C, Ventral view of an eight-legged fly of the genotype *bithoraxoid / Df,* where *Df* is a deficiency for the *bithoraxoid* gene. D, Dorsal view of an *Antennapedia* male (genotype: *Antp*^{Yu} / *Pc*^{3} [where the *Pc* mutant enhances the *Antp* phenotype]).

graduates at the time, he at Purdue University and I at the University of Minnesota, where C. P. Oliver had kindly given me space in his laboratory to work on *Drosophila*. Novitski and I had worked together on *Drosophila* while in high school in Wilkes-Barre, Pa.) From the *trans* heterozygote, *S* + / + *ast*, not only wild-type crossovers but the reciprocal double mutant, *S ast*, could be derived as well.[10] Moreover, there is a striking position effect. Whereas *S* + / + *ast* animals are nearly eyeless and have wing defects, *S ast* / + + animals are wild-type except for a slightly roughened eye identical with that of *S* / + animals.[10,11] Thus, only when the two wild-type alleles (symbolized by + signs) are located together in the same chromosome or in *cis* do they function normally. The *cis* heterozygotes for *lozenge* mutants were eventually obtained by Green and Green[12]; whereas the *trans* heterozygotes have small eyes, the *cis* have normal eyes. A comparison of the *cis* and *trans* phenotypes has come to be known as the *cis-trans* test, and the position effect, if present, as the *cis-trans* effect (see Hayes[13] for a discussion of the history of these concepts).

The *S* and *ast* loci proved to be within a double-banded structure known as 21E1-2 near the tip of the second chromosome.[14] Not only had Bridges[7] interpreted such structures as tandem duplications that had become established in the species, he had actually cited 21E1-2 as the type example. Although this correlation of genetic and cytological observations seemed to lend support to the tandem gene duplication hypothesis, we still do not know the true significance of such doublet structures, and proof that they consist of separate, homologous bands has not been forthcoming.

THE BX-C

In 1946, a deliberate search was begun to find additional cases of the *S-ast* type to test the theory that a common mode of gene evolution would involve tandem gene duplication. The resultant genetic redundancy would allow one of the genes to evolve a new function while the other gene continued to carry out the old function.[11]

Of several cases that we explored, the most illuminating from a developmental standpoint was that of a series of mutants that transform the third thoracic segment (T3) of the fly toward the second (T2). The type mutant, *bithorax* (*bx*), was discovered by Bridges in 1915 (as reported by Lindsley and Grell[6]); it is a weak and variable mutant. Fortunately, a very strong and constant mutant, *bx*^{3}, was found later by Stern (also as reported by Lindsley and Grell[6]) and is the one that has been most useful in dissecting the *bx* function and contributing to the creation of the four-winged fly (Fig 1, B). The details of the early work on these mutants need not concern us herein except to note that three types of mutants, *bx*, *Ultrabithorax* (*Ubx*) (formerly *bx*D), and *bithoraxoid* (*bxd*) mapped to three closely linked loci on the genetic map. Cytologically, they mapped to two adjacent doublet struc-

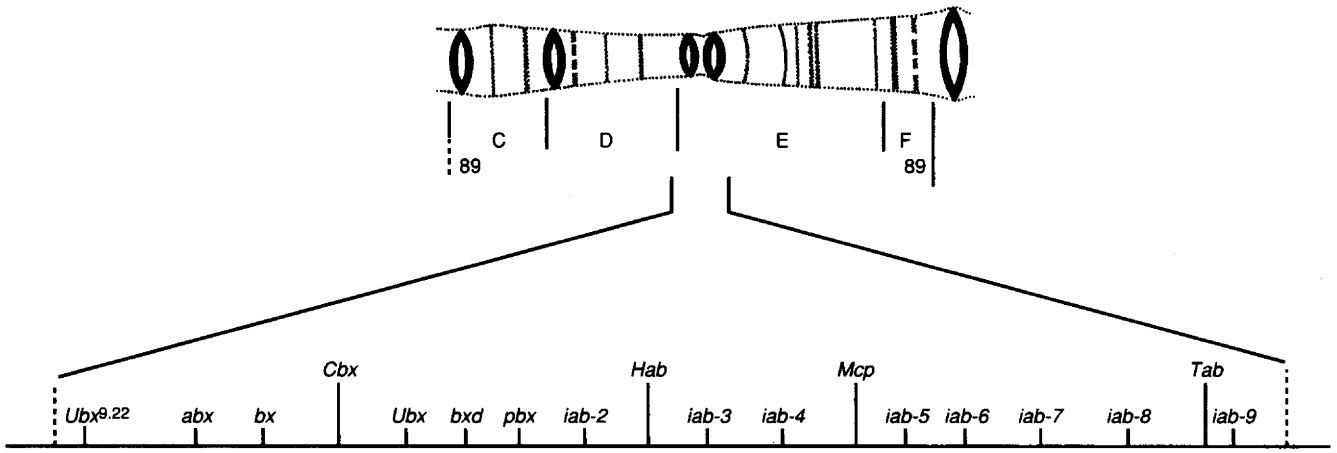

Fig 2.—A correlation of the cytological and genetic maps of the bithorax complex. The bithorax complex maps to the two doublets in 89E of the third chromosome of the salivary gland chromosomes. Dominant gain-of-function mutants are in the row above that of the loss-of-function mutants. *Ubx* indicates *Ultrabithorax*; *abx, anterobithorax*; *bx, bithorax*; *Cbx, Contrabithorax*; *bxd, bithoraxoid*; *pbx, postbithorax*; *iab, infra-abdominal*; *Hab, Hyperabdominal*; *Mcp, Miscadastral pigmentation*; and *Tab, Transabdominal.*

tures that often fuse into one structure, presumably as the result of pairing of partially homologous bands (Fig 2).[15] Again, Bridges' prediction that such structures represent tandem duplications seemed well supported.

From the start, we found dramatic *cis-trans* effects between almost every double mutant combination tested.[11] Furthermore, these effects were soon found to be conspicuously polarized. Thus, if we let *a* and *b* represent any two recessive mutants that are qualitatively different in phenotype, then in every case in which *a b* / + + is wild-type and *a* + / + *b* is mutant, the animal shows either the *a* phenotype or the *b* phenotype, but not both. For example, bx^3 + / + *pbx* (*postbithorax*) has a weak *pbx* phenotype (posterior T3 transformed toward posterior T2) but no trace of the *bx* phenotype (anterior T3 transformed toward anterior T2). Only much later were such polarized effects discovered in the operons of bacteria.[16]

ANALYSIS BY GENE DELETION

A new dimension to the analysis came with the derivation of x-ray–induced deletions for the BX-C, or for portions thereof. Thus, animals lacking all of the BX-C (homozygous for a deletion of the two 89E doublets) die as late embryos but not before they have developed the intricate cuticular and tracheal systems that enable the body segments to be distinguished from one another. Such animals show a striking transformation of the abdominal segments transformed into a row of thoraciclike segments.[15] Thus, complete loss of the BX-C results in a more extreme homeotic transformation than that produced by the spontaneous mutants, all of which turned out to represent only partial losses of gene function.

WHAT DO THE BX-C GENES DO?

In 1964, we reported that mutants of the BX-C are autonomously expressed in somatic mosaics.[17] We therefore knew that the developmental abnormalities were not due to substances, such as hormones, that diffuse readily between cells. Our mosaic studies further demonstrated that even late in development, a bx^3 mutant expressed the *bx* transformation even in a single cell of an otherwise wild-type fly.[18] Thus, the wild-type gene could be inferred to function throughout development, at least until the pupal stage.

It has always been somewhat treacherous to deduce the normal function of a gene from the effects on the organism of mutations in that gene. The problem is especially severe when the mutations result in morphological changes. To assess the normal or wild-type functions of the BX-C, we adopted a new procedure. We constructed a set of genotypes in which the BX-C was entirely lacking except for the presence of one dose of various wild-type regions of the complex. In this way, we could show that all of the regions studied are involved in initiating, or in some cases suppressing, the formation of specific organs and structures, a more profound and basic type of developmental function than had been inferred from considering the homeotic effects shown by the mutants.[15,18]

Analysis of the function of the wild-type BX-C genes during development revealed two basic properties of the system. First, the genes are arranged in the chromosome in the same order in which they are expressed along the anteroposterior axis of the organism (Fig 3).[15,19] This colinearity has now been found in all clusters of the BX-C type in which the order in the chromosome has been established, as will be discussed herein. Second, the morphological evidence showed that once expressed in a given segment, the gene tends to remain expressed in more posterior segments (Fig 4).[15] Thus, the more posterior a segment in the organism, the more advanced in development it is as far as genes of the BX-C are concerned, the ground state being that of a thoracic state.[15]

Further subdivision of the complex was achieved not only by means of overlapping deletions but by rearrangement of breakage points in the complex.[15,20] The latter have proved useful in identifying a set of *cis*-regulatory regions that direct the control of gene expression in each of the abdominal segments from A2 to A9, inclusive. A loss-of-function mutation in one of these regions, designated *infra-abdominal* (*iab*), causes the segment involved to transform to the next most anterior type of segment, ie, toward the thoracic or ground state.

Before recombinant DNA methodology was applied to the BX-C, we interpreted the wild-type action of the BX-C genes in terms of the operon model[16] for gene clusters in prokaryotes. Thus, many of the genetic properties of the BX-C, including polarized *cis-trans* effects and the presence of dominant constitutive mutants (Fig 2), appeared to be remarkably similar to the genetic properties of operons. This led us to

suggest that the BX-C might constitute an inducible system in which the function of its genes would be to "repress certain systems of cellular differentiation and thereby allow other systems to come into play."[17] However, the molecular basis of the BX-C, which will be discussed, proves to be entirely different from that of the operon. Our earlier concept of the function of the BX-C genes was clearly too narrow. Quite different and only partially understood mechanisms induce the expression of the BX-C genes, and we now suppose that these genes can activate, as well as repress, systems of cellular differentiation.

One of the most challenging problems remaining will be to identify the genes that the BX-C genes control. Such "downstream" or "target" genes are those that we assume will bring about the differentiation of tissues and organs of the third thoracic and abdominal regions of the organisms.

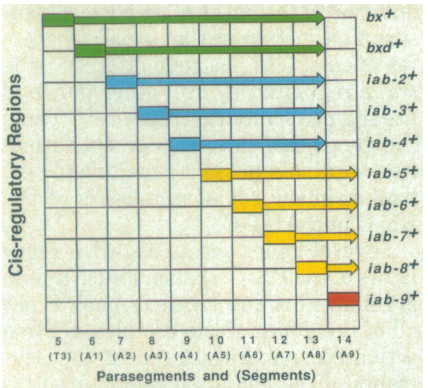

Fig 3.—Summary of *Drosophila* bithorax complex functions and their spatial expression. The parasegmental (PS) morphological function and protein expression for PS5 to PS14 (segments T3 to A9) are shown in parentheses and are inclusive. The extent of *Ultrabithorax* function and protein expression is shown in green; that of *abdominal-A* (*abd-A*) in blue, *Abd-BI* in yellow, and *Abd-BII* in orange; *bx* indicates *bithorax*; *bxd, bithoraxoid*; and *iab, infra-abdominal*.

THE MOLECULAR GENETICS OF THE BX-C

In 1978, Prof David Hogness and his collaborators at Stanford University began a molecular analysis of the BX-C. They used chromosome walking and jumping to clone the BX-C.[20] The current status of the molecular map of the complex is shown in the lower half of Fig 4 based on this early work and subsequent work.[21,22] Three domains of function have come to be recognized: *Ubx*, *abdominal-A* (*abd-A*), and *Abdominal-B* (*Abd-B*), named after the protein class coded by each domain.[18] The *Ubx* domain has four exons and a very large intron within which lie the *anterobithorax* (*abx*) and *bx* loci. As the result of alternative splicing, a family of proteins results.[23] For the sake of simplicity, only the largest transcription unit is shown in Fig 4.

The colinearity rule, first noticed for morphological effects, as already discussed, has been fully supported by studies of the expression of proteins coming from each of the three domains of the BX-C using immunohistochemical techniques (Fig 4). Thus, in wild-type embryos, the *Ubx* proteins are first detected in posterior T2 and anterior T3[24,25] (or in terms of the parasegmental [PS]) units[26] of the embryonic ectoderm, in PS5). The expression declines but remains in the abdominal segments.[25,27] Analysis of mutant embryos indicates that the bx and bxd regions direct *Ubx* protein expression proteins in PS5 and PS6, respectively.[26,27]

The level of *abd-A* protein in wild-type embryos remains elevated in these segments and those following up to and including PS13.[28,29] Double labeling with antibodies to *Ubx* and *abd-A* proteins has shown that *Ubx* proteins, although reduced in amounts, are on in different cell types than is the *abd-A* protein.[29] Analysis of mutants in the *iab-2*, *iab-3*,

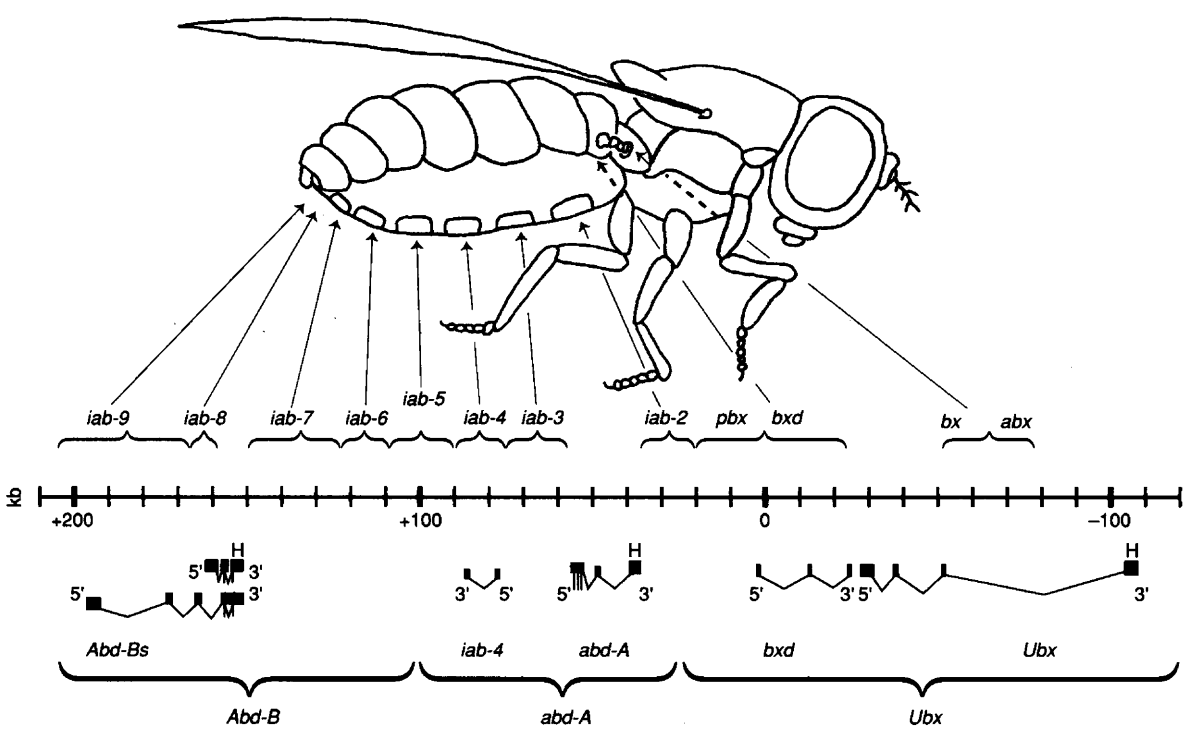

Fig 4.—Colinearity of order of gene expression in the adult fly and the order on the molecular map of the *cis*-regulatory regions. The bithorax complex encompasses 300 kb (from +200 kb to −100 kb, where the start of the walk was 0.0). Below the map, which is approximately to scale, are five characterized transcription units; three containing homeoboxes, *Abd-B, abd-A,* and *Ubx*; and two noncoding, *iab-4* and *bxd*; *iab* indicates *infra-abdominal*; *pbx, postbithorax*; *bxd, bithoraxoid*; *bx, bithorax*; *abx, anterobithorax*; *abd, abdominal*; and *Ubx, Ultrabithorax.*

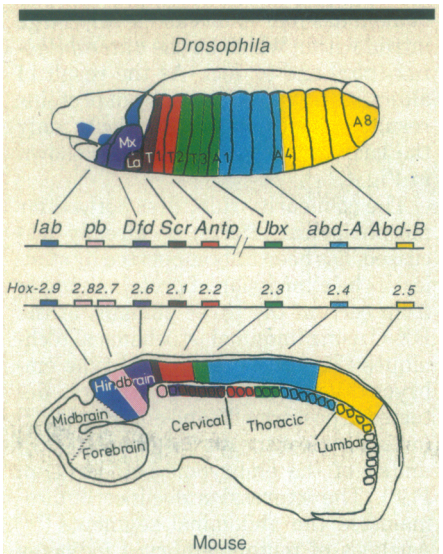

Fig 5.—Correlation of homeotic gene expression in the *Drosophila* embryo with that of the mouse embryo. The anterior ends of the embryos are to the left. The *Drosophila* homeotic genes form two clusters in the right arm of the third chromosome, the separation being between the *Antp* and *Ubx* genes. The homologous mouse genes are in a single cluster; among the four such clusters only that of *Hox2* in chromosome 11 is shown. The genes are color-coded; however, the homeobox sequences of the *Antp*, *Ubx*, and *abd-A* genes of *Drosophila* are so similar that homology with the mouse genes is based only on the expression patterns. In the *Drosophila* and mouse embryos, the anterior boundary of expression of each gene is shown by the corresponding color for each gene. For simplicity, the posterior boundary for a given color is not shown since it would usually continue beyond the next most posterior gene's anterior boundary. Only expression in the cuticle of the *Drosophila* embryo is shown, while in the mouse, expression is shown in the nervous system and in the prevertebrae. lab indicates labial; *pb, Proboscipedia*; *Dfd, Deformed*; *Scr, Sex-combs reduced*; *Ubx, Ultrabithorax*; and *abd, abdominal.*

and *iab-4* regions show that *abd-A* protein is expressed in a colinear fashion for PS7, PS8, and PS9.[29,30]

The *Abd-B* region produces a large and small protein.[30-32] Our analysis of the distribution of these proteins has been indirect since our monoclonal antibody is directed to an epitope in the 3′ exon common to both proteins of this region. To infer their distribution, therefore, we have used mutants that eliminate the messenger RNA for either the small or the large protein. The latter is first detected in the posterior fourth abdominal segment (PS10), while the small protein coming from the more distally located promoter is first detected in the posterior eighth abdominal segment (PS14). Using rearrangement break points disrupting the *iab-5* to *iab-9* regions, inclusively, we have shown that the pattern of expression of the large protein is colinear with the order of those regions.[33]

Thus, the expression of proteins from the three domains of the complex parallels the morphological expression. It confirms and establishes the rule that the more posterior the segment, the more functions controlled by the *cis*-regulatory regions become expressed up to and including PS13. Thereafter, there is loss of detectable expression of all but the small *Abd-B* protein.

The colinearity rule also applies to RNAs coming from the *iab* regions of the complex, as shown by Sanchez-Herrero and Akam.[34] They point out that these transcripts may be functionless, as discussed for the *bxd* case by Lipshitz et al,[35] or the *iab* transcripts "might function in *cis* by some unprecedented mechanism."[34]

THE ANT-C

It is likely that the BX-C was once part of a single cluster that included another larger complex, the ANT-C.[36] The latter lies in the same chromosome arm but at a considerable distance proximal to the BX-C. The type mutant for which this complex was named was first found in *Drosophila affinis* by Sturtevant.[37] The phenotype is that of an antenna transformed toward a leg. A mutant with a strong transformation of this type (Fig 1, D) was induced with x-rays in 1948 by Yu as described by Lindsley and Grell.[6] In the same year, Les Calves reported the occurrence of a neutron-induced mutant that he named *Aristapedia* and suggested that it was a dominant allele of a recessive mutant and causes the antenna to transform toward a tarsus, rather than a full leg. From his determination of its cytological location in 84A, the location of the ANT-C (and of *Aristapedia* in 89C[16]), it is now clear that it is in fact an allele of the *Antennapedia* locus.

Extensive genetic and cytological analyses of these and other related mutants by Kaufman and colleagues[38] established the existence of this second cluster of homeotic mutants and their colinearity (Fig 5). Extensive molecular analyses have been made, particularly of the *Antp* gene itself.[39,40]

REGULATION OF THE GENES OF THE BX-C IN *TRANS*

A class of genes that initiates the axes of the body and divides the body into segments has been identified by Nüsslein-Volhard and Wieschaus[41] based on massive screenings for mutants affecting early development of *Drosophila*.[42,43] Some of these genes appear to act as major *trans*-regulators of the BX-C and ANT-C. A possible relationship of one such gene, *hunchback* (*hb*), to the BX-C was first detected genetically when, *Regulator-of-pbx* (*Rg-pbx*) a dominant *trans*-regulator of the *postbithorax* function of the BX-C,[44] was found by Bender et al[45] to be a gain-of-function mutant of the *hb* gene. Garcia-Bellido and Capdevila[46] had earlier made extensive studies of the role of *Rg-pbx* in regulating the BX-C. Recently, Qian et al[47] have shown the *hb* protein binds to a specific motif in the intron of the *Ubx* gene and interacts to repress transcription of *Ubx* in regions anterior to the onset of its normal transcription in PS5 of that gene. The expression of *Ubx* has been inferred to be *trans*-regulated by still other genes of the segment-initiation class.[43,48-51] Thus, the genes that initiate segment formation seem to represent a major and perhaps primary set of regulators of the BX-C and ANT-C.

Several genes act as negative regulators of the BX-C; that is, if such a gene is removed or inactivated, many if not all of the genes of the complex become activated, resulting in all of the segments of the body being converted toward the eighth-abdominal segment.[16] The first of these genes was *Polycomb*, found by P. H. Lewis as reported by Lindsley and Grell.[6] Zink and Paro[52] have shown that the protein coded for by the normal *Polycomb* gene is specifically associated in the salivary gland chromosomes to the BX-C and ANT-C regions, as well as other sites. Zink et al[53] have shown in an elegant way that it binds to upstream (5′) sequences of the *Antp* gene. Paro and Hogness[54] find that the sequence of the *Polycomb* gene shows some homology, not as might have been expected to a DNA-binding class of proteins but to a nonhistone chromosomal protein. Hence, the remarkably precise binding of *Polycomb* protein to bands in the salivary gland chromosomes associated with the BX-C and ANT-C suggests that the protein is complexing with one or more other proteins that do bind to those bands. If so, the *Polycomb* class of genes[18] would be involved in maintaining a repressed state in regions of the embryo where it is not activated.[55,56] Ingham and Whittle[57] have shown that another gene, *trithorax*, probably allelic to *Regulator-of-bx* (*Rg-bx*), acts as a positive regulator. Ingham[58] suggests that it may be involved in maintaining the activated state of the BX-C genes.

THE HOMEOBOX

Support from molecular studies for the hypothesis that the BX-C and the ANT-C owe their origin to repeated tandem gene duplication finally came with the discovery of the homeobox by McGinnis et al[59] and independently by Scott and Weiner.[60] This region of 160 base pair codes for 60 amino acids that con-

stitute the DNA binding or "homeodomain" of each of the proteins coded by the BX-C and ANT-C genes. It is now thought likely that these proteins bind to downstream genes and thereby regulate them.

So similar are the homeobox sequences between adjacent genes that it is exceedingly improbable that they arose by convergent evolution. Only the homeodomain of the protein shows obvious conservation, possibly because most of the specificity required for DNA binding lies in the homeodomain. We would expect, however, that many motifs in enhancerlike elements of the *cis*-regulatory regions may be conserved. At present, there are insufficient data on DNA sequence in the *cis*-regulatory regions to test for conservation of such motifs.

HOMEOBOX GENE CLUSTERS IN OTHER ORGANISMS

Genetic and morphological evidence for cognates of the homeobox-containing genes of the ANT-C and BX-C have also been discovered in beetles, where they appear to lie within a single homeotic complex (HOM-C).[61,62] (In the silkworm *Bombyx*, a fascinating set of mutants have striking parallels to many of the mutants of the BX-C and ANT-C. They were designated as alleles of an *E* locus within which, however, crossing-over and *cis-trans* effects have been reported,[63-65] suggesting that they are in fact a gene complex; however, in the absence of a genetic or molecular map, the colinear rule obviously cannot be tested. One of the E mutants produces an eight-legged moth [a bronze sculpture of which is in an ancient temple in Japan]. The exact cognate of that mutant in *Drosophila* is *bxd* [Fig 1, C]. Another of the E mutants was reported to lack gonads and led us to examine its cognates in *Drosophila*, namely *iab-4* mutants. When homozygous, the latter mutants are viable and appear virtually wild-type, but internally they too lack gonads in both sexes.[15])

Astonishingly, mice and humans not only have cognates of the BX-C and ANT-C genes in a single HOM-C, but the complexes occur in four sets, each in a different chromosome.[66-69] Moreover, colinearity of spatial order of expression and molecular map order extends over the entire complex in the murine (Fig 5) and in the human cases. Other less extensively characterized clustering of cognates of the HOM-C type occur in sea urchins,[70] nematodes,[71,72] bees,[73] frogs,[74] and chickens.[75] Omitted are a number of other organisms in which cognates have been found for only one or two genes of the HOM-C type.[76,77]

In *Drosophila*, it is not yet known whether temporal order of expression is also colinear with map order. However, there is evidence that colinearity is temporal as well as spatial from experiments involving (1) chicken[75] and murine[78] limb buds; (2) expression of transcripts from HOM-C in the central nervous system and prevertebrae of the mouse embryo[79]; and (3) induction of the HOM-C by retinoic acid in human cell cultures.[80]

WHY HAVE THE HOM-C REMAINED INTACT?

A fundamental question is why genes of the HOM-C have remained linked for an estimated 500 million years or more, assuming they arose before the separation of the vertebrates and invertebrates. Even though they presumably evolved by tandem gene repetition, it is highly unlikely that the linkage would have persisted in so many organisms unless it confers a selective advantage.

Clearly, in *Drosophila*, the existence of two separate complexes, which still remain tightly linked, suggests that at least the *Antp* and *Ubx* genes do not have to be adjacent in this organism. Whether other genes, such as *Ubx* and *abd-A*, can be separated and still function normally is uncertain. Several rearrangements with break points near the boundary of the *Ubx* and *abd-A* domains have been detected by their suppression of transvection[81,82] in a screen that recovers break points that are unselected, with respect to having any mutant effect.[19] These rearrangements show a striking phenotype; namely, development of tiny sense organs on the larval abdominal segments (organs normally found only on the larval thoracic segments), an effect that corresponds to a weak *bxd* phenotype. Bender et al[83] have shown that such break points map near the boundary of the *Ubx* and *abd-A* domains; however, from their *bxd* phenotype, it is clear that they still lie in the *Ubx* domain. Thus, if a break point can fall at or near that boundary and lack a detectable phenotype, our screen has yet to recover one.

A possible clue to understanding why gene clusters of the BX-C and ANT-C type have remained intact comes from studies of gene regulation in another cluster; namely, that coding for the family of β-globin peptides of hemoglobin. Regulation of the genes of this cluster has been shown to depend on enhancer elements residing in the *cis*-regulatory regions surrounding the protein-coding regions of the genes. Nickol and Felsenfeld[84] and Choi and Engel[85] find that the adult and embryonic β-globin genes in chickens share an enhancer.

In the BX-C, Celniker et al[33] have evidence, based on morphology, that the *iab-5* region, which normally regulates the expression of the *Abd-B* gene, may be regulating expression of both *abd-A* and *Abd-B*. They suggest, therefore, that in wild-type, the *abd-A* and *Abd-B* genes share one or more enhancers located in the *iab-5* region. Duncan[18] has discussed this and other cases in the complex where *cis*-regulatory elements seem to be "bifunctional." If enhancer sharing does occur in the BX-C and if it confers a selective advantage, then the genes involved will of course tend to remain linked.

WHY FOUR SETS OF COMPLEXES IN MICE AND HUMANS?

As just noted, homeobox (HOX) clusters have been duplicated in the genome of mice and humans four times, each in a different chromosome in both organisms. In humans, four clusters are designated HOX1, HOX2, HOX3, and HOX4 and are located in chromosomes 7, 17, 12, 2, respectively. Boncinelli et al[67] have speculated that having four sets might provide "a finer-grained spatial control" than would be needed in organisms such as *Drosophila*, which have only one set and perhaps fewer requirements for setting up the body segments than have the higher vertebrates.

MEDICAL IMPLICATIONS

We can look forward to an exciting time when the functions of the genes of the four human HOX clusters are determined. Human genetic abnormalities at virtually any stage of development, but especially those traceable to abnormalities of early embryonic or fetal development, would be expected from mutations in genes of the HOX clusters. The results of genetic experiments on other vertebrates strongly support such a conjecture.

Recently, Chisaka and Capecchi[86] used the new technique of gene targeting to produce a disrupted *Hox1.5* gene of the mouse. Homozygotes for this recessive loss-of-function mutation are found to produce a constellation of defects that they point out is remarkably similar to the human congenital disorder DiGeorge syndrome. The obvious implications are that such a syndrome might result from a mutation in the human cognate of the mouse gene and that other congenital disease syndromes might result from mutations in other HOX genes.

Balling et al[87] overexpressed the murine *Hox1.1* (*Hox2.3* of the HOX2 cluster) gene in transgenic mice and observed craniofacial abnormalities. Kessel et al,[88] studying more severely affected mice from the same experiment,

observed dramatic homeotic effects on the cervical vertebrae. It follows that a gain-of-function mutation in the human cognate of *Hox1.1* is expected to show broadly similar types of defects, depending on the penetrance of the allele.

In *Drosophila*, it is well known that such gain-of-function mutants are not uncommon (Fig 2). The first example in the BX-C was the *Contrabithorax* (*Cbx*) mutant, which causes T2 to transform toward T3, notably turning the wings into halteres.[15,17] All of the *Antennapedia* mutants are also now known to be an overexpression of the wild-type gene in the antenna (Fig 1, D). We would expect that when mutants of HOX genes in humans are found they will often be of the gain-of-function type.

The HOX genes may also be implicated in the genesis of malignancies. Blatt et al[89] have reported a possible example involving a murine myeloid leukemia cell line in which a *Hox2.4* gene has been disrupted by a retroviral insertion. Their finding is confirmed by Kongsuwan et al[90] who have studied the same cell line and who suggest that *Hox2.4* may interfere with terminal differentiation; then, in combination with activation by the same provirus of another gene in the same cell line, conditions would be met for development of the transformed cell. Blatt and Sachs[91] also report that deletion of the *Hox4.1* gene occurs in another murine myeloid leukemia cell line; this gene is the cognate of *Hox2.7* (Fig 5). They have discussed a number of mechanisms that might involve such a deletion in the generation of the malignancy.

To summarize the medical relevance of the discovery of HOX clusters in humans, we can expect that germinal mutations occurring within these clusters are likely to result in abnormalities at many stages of human development. On the basis of preliminary findings in murine leukemia cell lines, it is conceivable that somatic mutations in the HOX clusters may potentially play a significant role in the generation of tumors, both benign and malignant. However, the search for DNA lesions in potential cases of germinal or somatic mutations presents formidable technical problems, not only because of the large size of such clusters but because of their fourfold redundancy.

This study was supported in part by research grant HD06331 from the National Institutes of Health, Bethesda, Md, and by grants from the American Cancer Society, Atlanta, Ga, the March of Dimes, White Plains, NY, and the Lucille P. Markey Charitable Trust, Miami, Fla.

I thank Susan Celniker, PhD, and Welcome Bender, PhD, for a critical reading of the manuscript and Peter Chang, James Hardwick, Helen Park, and Seema Sharma, for assistance in the preparation of it. While at Caltech, a number of colleagues have directly contributed to our research on the bithorax complex, namely, Welcome Bender, PhD, Marie-Paz Capdevila, PhD, Susan Celniker, PhD, Loring Craymer, PhD, Madeline Crosby, PhD, Ian Duncan, PhD, Antonio Garcia-Bellido, PhD, William Gelbart, PhD, Alain Ghysen, PhD, Lily Jan, PhD, Margit Lohs-Schardin, PhD, and Shige Sakonju, PhD. Finally, I want to stress the close cooperation that we have had over the years with Prof David Hogness, PhD, and his colleagues at Stanford University and Prof Welcome Bender, PhD, at Harvard Medical School. It was Professor Hogness' foresight to launch the molecular analysis of the bithorax complex in 1978 in his laboratory.

References

1. Sturtevant AH. The effects of unequal crossing over at the *Bar* locus in *Drosophila*. *Genetics*. 1925;10:117-147.
2. Maeda N, Smithies O. The evolution of multigene families: human haptoglobin genes. *Annu Rev Genet*. 1986;20:81-108.
3. Bridges CB. The *Bar* 'gene': a duplication. *Science*. 1936;83:210-211.
4. Muller HJ, Prokofyeva-Belgovskaya AA, Kossikov KV. Unequal crossing-over in the *Bar* mutant as a result of duplication of a minute chromosome section. *Comptes Rendus (Doklady) el l'Académie des Sciences del l'URSS*. 1936;I:87-88.
5. Rapoport JA. Quadruple-*Bar* in *Drosophila melanogaster*. *Bull Biol Med Exp*. 1936;II:242-244.
6. Lindsley DL, Grell EH. *Genetic Variations of* Drosophila melanogaster. Washington, DC: Carnegie Institution of Washington; 1968:1-521. Publication 627.
7. Bridges CB. Salivary chromosome maps with a key to the banding of the chromosomes of *Drosophila melanogaster*. *J Hered*. 1935;26:60-64.
8. Oliver CP. A reversion to wild-type associated with crossing over in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 1940;26:452-454.
9. Lewis EB. *Star*-recessive, a spontaneous mutation in *Drosophila melanogaster*. *Proc Minn Acad Sci*. 1939;7:23-26.
10. Lewis EB. The *Star* and *asteroid* loci in *Drosophila melanogaster*. *Genetics*. 1942;27:153-154.
11. Lewis EB. Pseudoallelism and gene evolution. *Cold Spring Harbor Symp Quant Biol*. 1951;16:159-174.
12. Green MM, Green KC. Crossing-over between alleles at the *lozenge* locus in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 1949;35:586-591.
13. Hayes W. *The Genetics of Bacteria and Their Viruses: Studies in Basic Genetics and Molecular Biology*. 2nd ed. New York, NY: John Wiley & Sons Inc; 1968.
14. Lewis EB. The relation repeats to position effect in *Drosophila melanogaster*. *Genetics*. 1945;30:137-166.
15. Lewis EB. A gene complex controlling segmentation in *Drosophila*. *Nature*. 1978;276:565-570.
16. Jacob F, Monod J. On the regulation of gene activity. *Cold Spring Harbor Symp*. 1961;26:193-211.
17. Lewis EB. Genetic control and regulation of developmental pathways. In: Locke M, ed. *Role of Chromosomes in Development*. New York, NY: Academic Press Inc; 1964:231-252.
18. Duncan I. The bithorax complex. *Annu Rev Genet*. 1987;21:285-319.
19. Lewis EB. Regulation of the genes of the bithorax complex in *Drosophila*. *Cold Spring Harbor Symp Quant Biol*. 1986;50:155-164.
20. Bender W, Akam M, Karch F, et al. Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science*. 1983;221:23-29.
21. Karch F, Weiffenbach B, Peifer M, et al. The abdominal region of the bithorax complex. *Cell*. 1985;43:81-96.
22. Celniker SE, Lewis EB. *Transabdominal*: a dominant mutant of the bithorax complex produces a sexually dimorphic segmental transformation in *Drosophila*. *Gene Dev*. 1987;1:111-123.
23. 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.
24. White RAH, Wilcox M. Protein products of the bithorax complex in *Drosophila*. *Cell*. 1984;39:163-171.
25. Beachy PA, Helfand SL, Hogness DS. Segmental distribution of bithorax complex proteins during *Drosophila* development. *Nature*. 1985;313:545-551.
26. Lawrence PA, Martinez-Arias A. The cell lineage of segments and parasegments in *Drosophila*. *Philos Trans R Soc Lond Biol*. 1985;312:83-90.
27. Hogness DS, Lipshitz H, Beachy DA, et al. Regulation and products of the *Ubx* domain of the bithorax complex. *Cold Spring Harbor Symp Quant Biol*. 1985;50:181-194.
28. Karch F, Bender W, Weiffenbach B. *abdA* expression in *Drosophila* embryos. *Genes Dev*. 1990;4:1573-1587.
29. Macias A, Casanova J, Morata G. Expression and regulation of the *abd-A* gene of *Drosophila*. *Development*. 1990;110:1197-1207.
30. Celniker SE, Keelan DJ, Lewis EB. The molecular genetics of the bithorax complex of *Drosophila*: characterization of the products of the *Abdominal-B* domain. *Genes Dev*. 1989;3:1424-1436.
31. Delorenzi M, Bienz M. Expression of *Abdominal-B* homeoproteins in *Drosophila* embryos. *Development*. 1990;108:323-329.
32. Boulet AM, Lloyd A, Sakonju S. Molecular definition of the morphogenetic and regulatory functions and the *cis*-regulatory elements of the *Drosophila Abd-B* homeotic gene. *Development*. 1991;111:393-405.
33. Celniker SE, Sharma S, Keelan DJ, Lewis EB. The molecular genetics of the bithorax complex of *Drosophila*: *cis*-regulation in the *Abdominal-B* domain. *EMBO J*. 1990;9:4277-4286.
34. Sanchez-Herrero E, Akam M. Spatially ordered transcription of regulatory DNA in the bithorax complex of *Drosophila*. *Development*. 1989;107:321-329.
35. Lipshitz H, Peattie D, Hogness D. Novel transcripts from the *Ultrabithorax* domain of the bithorax complex. *Genes Dev*. 1987;1:307-322.
36. Kaufman TC, Seeger MA, Olsen G. Molecular and genetic organization of the Antennapedia gene complex of *Drosophila melanogaster*. *Adv Genet*. 1990;27:309-362.
37. Sturtevant AH. Genetic data on *Drosophila affinis*, with a discussion of the relationships in the subgenus *Sophophora*. *Genetics*. 1940;25:337-353.
38. Kaufman TC, Lewis R, Wakimoto B. Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic gene complex in polytene chromosomal interval 84A,B. *Genetics*. 1980;94:115-133.
39. Garber RL, Kuroiwa A, Gehring WJ. Genomic and cDNA clones of the homeotic locus Antennapedia in *Drosophila*. *EMBO J*. 1983;2:2027-2036.
40. Scott MP, Weiner AJ, Hazelrigg TI, et al. The molecular organization of the *Antennapedia* locus of *Drosophila*. *Cell*. 1983;35:763-776.
41. Nüsslein-Volhard C, Wieschaus E. Mutations affecting segment number and polarity in *Drosophila*. *Nature*. 1980;287:795-801.
42. Nüsslein-Volhard C, Frohnhofer HG, Lehmann R. Determination of anteroposterior polarity in *Drosophila*. *Science*. 1987;230:1675-1681.
43. Ingham PW. The molecular genetics of embryonic pattern formation in *Drosophila*. *Nature*. 1988;335:25-34.
44. Lewis EB. Genetic control of developmental pathways in *Drosophila melanogaster*. Presented at the XII International Congress on Genetics; August 26, 1968:96-97; Tokyo, Japan.
45. Bender M, Turner FR, Kaufman TC. A developmental genetic analysis of the gene *regulator of postbithorax* in *Drosophila melanogaster*. *Dev Biol*. 1987;119:418-432.
46. Garcia-Bellido A, Capdevila MP. The initiation

and maintenance of gene activity in a developmental pathway of *Drosophila*. In: Subtelny S, Sussex IM, eds. *The Clonal Basis of Development: 36th Symposium of the Society for Developmental Biology*. Orlando, Fla: Academic Press Inc; 1978:3-21.
47. Qian S, Capovilla M, Pirrotta V. The *bx* region enhancer, a distant *cis*-control element of the *Drosophila Ubx* gene and its regulation by *hunchback* and other segmentation genes. *EMBO J*. 1991;10:1415-1425.
48. Ingham PW, Martinez-Arias A. The correct activation of Antennapedia and bithorax complex genes requires the *fushi-tarazu* gene. *Nature*. 1986;324:592-597.
49. Ingham PW, Ish-Horowicz D, Howard KR. Correlative changes in homeotic and segmentation gene expression in Krüppel mutant embryos in *Drosophila*. *EMBO J*. 1986;5:1659-1665.
50. Martinez-Arias A, White RAH. *Ultrabithorax* and *engrailed* expression in *Drosophila* embryos mutant for segmentation genes of the pair-rule class. *Development*. 1988;102:325-338.
51. Irish VF, Martinez-Arias A, Akam M. Spatial regulation of the Antennapedia-homeotic and *Ultrabithorax*-homeotic genes during *Drosophila* early development. *EMBO J*. 1989;8:1527-1537.
52. Zink B, Paro R. In vivo binding pattern of a *trans*-regulator of homeotic genes in *Drosophila melanogaster*. *Nature*. 1989;337:468-471.
53. Zink B, Engstrom Y, Gehring WJ, Paro R. Direct interaction of the polycomb protein with Antennapedia regulatory sequences in polytene chromosomes of *Drosophila melanogaster*. *EMBO J*. 1991;10:153-162.
54. Paro R, Hogness DS. The *Polycomb* protein shares a homologous domain with a heterochromatin-associated protein of *Drosophila*. *Proc Natl Acad Sci U S A*. 1991;88:263-267.
55. Struhl G. Genes controlling segmental specification in the *Drosophila* thorax. *Proc Natl Acad Sci U S A*. 1982;79:7380-7384.
56. Struhl G, Akam M. Altered distributions of *Ultrabithorax* transcripts in extra *sex combs* mutant embryos of *Drosophila*. *EMBO J*. 1985;4:3259-3264.
57. Ingham PW, Whittle JRS. *Trithorax*: a new homeotic mutation of *Drosophila melanogaster* causing transformations of abdominal and thoracic imaginal segments. *J Mol Gen Genet*. 1980;179:607.
58. Ingham PW. A new homeotic mutation of *Drosophila melanogaster*, II: the role of *trx*$^+$ after embryogenesis. *Wilhelm Rouxs Arch Dev Biol*. 1981;190:365.
59. McGinnis W, Levine M, Hafen E, Kuroiwa A, Gehring WJ. A conserved DNA sequence found in homeotic genes of the *Drosophila* Antennapediaand bithorax complexes. *Nature*. 1984;308:428-433.
60. Scott MP, Weiner AJ. Structural relationships among genes that control development: sequence homology between the *Antennapedia, Ultrabithorax*, and *fushi tarazu* loci of *Drosophila*. *Proc Natl Acad Sci U S A*. 1984;81:4115-4119.
61. Beeman RW. A homeotic gene cluster in the red flour beetle. *Nature*. 1987;327:247-249.
62. Beeman RW, Stuart JJ, Haas MS, Dennell RE. Genetic analysis of the homeotic gene complex (HOM-C) in the beetle *Tribolium castaneum*. *Dev Biol*. 1989;133:196-209.
63. Tsujita M. On the crossing-over between E^B and E^K of the E allelic series in the silkworm. *Jpn J Genet*. 1955;30:227-235.
64. Itikawa N. Genetical and embryological studies on the E-multiple alleles in the silkworm, *Bombyx mori*. *L Bull Sericultural Exptl Station Tokyo*. 1952;14:23-91.
65. Tazima Y. *The Genetics of the Silkworm*. London, England: Logos; 1964.
66. Gaunt SJ, Sharpe PT, Duboule D. Spatially restricted domains of homeo-gene transcripts in mouse embryos: relation to a segmented body plan. *Development*. 1988;104(suppl):169-179.
67. Boncinelli E, Somma R, Acampora D, et al. Organization of human homeobox genes. *Hum Reprod*. 1988;3:880-886.
68. Krumlauf R, Holland PWH, McVey JH, Hogan BLM. Developmental and spatial patterns of expression of the mouse homeobox gene, *Hox-2.1*. *Development*. 1987;99:603-617.
69. Schugart K, Kappen C, Ruddle F. Duplication of large genomic regions during the evolution of vertebrate homeobox genes. *Proc Natl Acad Sci U S A*. 1989;86:7067-7071.
70. Humphreys T. Molecular cloning and developmental expression of homeobox-containing genes in sea urchin embryos. *Zool Sci*. 1987;4:1129.
71. Burglin TR, Ruvkun G, Coulson A, et al. Nematode homeobox cluster. *Nature*. 1991;351:703.
72. Kenyon C, Wang B. A cluster of *Antennapedia*-class homeobox genes in a nonsegmented animal. *Science*. 1991;253:516-517.
73. Walldorf U, Fleig R, Gehring WJ. Comparison of homeobox-containing genes of the honeybee and *Drosophila*. *Proc Natl Acad Sci U S A*. 1989;86:9971-9975.
74. Wright CVE, Cho KWY, Hardwicke J, Collins RH, De Robertis EM. Interference with function of a homeobox gene in *Xenopus* embryos produces malformations of the anterior spinal cord. *Cell*. 1989;59:81-93.
75. Nohno T, Sumihare N, Koyama E, et al. Involvement of the *Chox-4* chicken homeobox genes in determination of anteroposterior axial polarity during limb development. *Cell*. 1991;64:1197-1205.
76. Gehring WJ, Hiromi Y. Homeotic genes and the homeobox. *Annu Rev Genet*. 1986;20:147-173.
77. Scott MP, Tamkun JW, Hartzell GW. The structure and function of the homeodomain. *Biochim Biophys Acta*. 1989;989:25-48.
78. Graham A, Papalopolu N, Krumlauf R. The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell*. 1989;57:367-378.
79. Wilkinson DG, Bhatt S, Cook M, Boncinelli E, Krumlauf R. Segmental expression of Hox-2 homeobox-containing genes in the developing mouse hindbrain. *Nature*. 1989;341:405-409.
80. Simeone A, Acampora D, Arcioni L, et al. Sequential activation of HOX2 homeobox genes by retinoic acid in human embryonal carcinoma cells. *Nature*. 1990;346:763-766.
81. Lewis EB. The theory and application of a new method of detecting chromosomal rearrangements in *Drosophila melanogaster*. *Am Naturalist*. 1954;88:225-239.
82. Lewis EB. Some aspects of pseudoallelism. *Am Naturalist*. 1955;89:73-89.
83. Bender W, Weiffenbach B, Karch F, Peifer M. Domains of *cis*-interaction in the bithorax complex. *Cold Spring Harbor Symp Quant Biol*. 1985;I:173-180.
84. Nickol JM, Felsenfeld G. Bidirectional control of the chicken beta-globin and epsilon-globin genes by a shared enhancer. *Proc Natl Acad Sci U S A*. 1988;85:2548-2552.
85. Choi ORB, Engel JD. Developmental regulation of beta-globin gene switching. *Cell*. 1988;55:17-26.
86. Chisaka O, Capecchi MR. Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *Hox-1.5*. *Nature*. 1991;350:473-479.
87. Balling R, Mutter G, Gruss P, Kessel M. Craniofacial abnormalities induced by ectopic expression of the homeobox gene HOX-1.1 in transgenic mice. *Cell*. 1989;58:337-347.
88. Kessel M, Balling R, Gruss P. Variations of cervical-vertebrae after expression of a HOX-1.1 transgene in mice. *Cell*. 1990;61:301-308.
89. Blatt C, Aberdam D, Schwartz R, Sachs L. DNA rearrangement of a homeobox gene in myeloid leukaemic cells. *EMBO J*. 1988;7:4283-4290.
90. Kongsuwan K, Allen J, Adams JM. Expression of Hox-2.4 homeobox gene directed by proviral insertion in a myeloid leukemia. *Nucleic Acids Res*. 1989;17:1881-1892.
91. Blatt C, Sachs L. Deletion of a homeobox gene in myeloid leukemias with a deletion in chromosome 2. *Biochem Biophys Res Commun*. 1988;156:1265-1270.