The bithorax complex: the first fifty years

EDWARD B. LEWIS*
Division of Biology, The California Institute of Technology, Pasadena, USA

INTRODUCTION

Genetics is a discipline that has successfully used abstractions to attack many of the most important problems of biology, including the study of evolution and how animals and plants develop. The power of genetics to benefit mankind was first recognized by the award of the Nobel Prize in physiology or medicine in 1933 to T. H. Morgan. In the 23 years that had intervened between the time Morgan introduced Drosophila as a new organism for the study of genetics and the award of the Prize, he and his students, especially, A. H. Sturtevant, C. B. Bridges and H. J. Muller, had vastly extended the laws of Mendel as the result of a host of discoveries, to mention only a few: that the genes (Mendel's factors) are arranged in a linear order and can be placed on genetic maps, that they mutate in forward and reverse directions, that they can exist in many forms, or alleles, and that their functioning can depend upon their position. Purely on the basis of breeding experiments, these early workers were able to deduce the existence of inver-

Abbreviations used in this paper: T, thoracic segment; A, abdominal segment; BX-C, bithorax complex; HOM-C, homeotic complex; HOX, homeobox; ANT-C, antennapedia complex; COL, colinearity; COE, cis-overexpression; CIN, cis-inactivation.

*Address for reprints: Biology Division 156-29, California Institute of Technology, Pasadena, CA 91125, USA. FAX: 626-564-9685. e-mail: lewis@cco.caltech.edu
sions and duplications, for example, before it became possible to demonstrate them cytologically. The list of their achievements is a long one and one that has been put into historical perspective by Sturtevant in *A History of Genetics* (Sturtevant, 1965).

All of these discoveries were made with *Drosophila* by taking advantage of its small size, ease of culturing, high fecundity, short life cycle, small chromosome number, wealth of spontaneous and induced mutations, and, after their discovery in 1935, its giant salivary gland chromosomes. Of immense importance also was the existence of standard or "wild-type" strains.

That Morgan’s contributions satisfied the criterion of being of benefit to mankind was evident by the remarkable extent to which the new discoveries with *Drosophila* had direct application to the understanding of the inheritance of many human traits. For example, the inheritance of colorblindness and hemophilia in human beings could be understood for the first time.

The second Nobel Prize for work in the genetics of *Drosophila* was awarded in 1946 to H. J. Muller for his discovery in 1928 that X-rays produce gene mutations and do so in direct proportion to the dose (Muller, 1927). Muller called attention to the genetic risks to the human race posed by indiscriminate use of ionizing radiations, and, prophetically, he argued that such uses would also increase the risk of cancers, if cancer is the result of somatic mutations. The implications of Muller’s work were not overlooked with the advent of the atomic age. As a result, extensive genetic studies were carried out in *Drosophila* and mice to assess the relative rates of mutation in these organisms as a means of assessing the genetic risks to human beings from the use of atomic energy.

The award of the Prize in 1995 for work with *Drosophila* recognizes the growing importance of a field that has come to be called developmental genetics. The work of my co-winners, Eric Wieschaus and Christiane Nüsslein-Volhard, has identified crucial steps in the early development of the organism. Specifically, they have identified major genes involved in setting up the initial axes of the embryo and its germ layers (Nüsslein-Volhard and Wieschaus, 1980) thereby setting the stage for groups of master control genes to eventually lead to a re-examination of the gene concept in terms of the inheritance of many human traits. For example, the inheritance of colorblindness and hemophilia in human beings could be understood for the first time.

In 1925, Sturtevant made two important discoveries that were eventually to lead to a re-examination of the gene concept in terms of the gene’s function (Sturtevant, 1925). In analyzing the progeny of females homozygous for the unstable eye mutation, *Bar (B)*, he predicted that a rare mutation, *double-Bar (BB)*, was a tandem duplication that arose in the progeny of homozygous *B* females as the result of “unequal crossing over”. He then showed that the eyes of BB/+ females are slightly smaller than those of B/B and deduced that the function of a gene can depend upon its position with respect to its neighbors, the first example of the “position effect”, as he named it.

Eleven years later, using the giant salivary gland chromosomes of the *Drosophila* larva, Bridges (1936) and Muller *et al.* (1936) reported that the B mutant was actually a tandem duplication of 7 bands in the X chromosome and that BB was a triplication for that region. Hence BB was arising from unequally paired duplicated regions accompanied by normal rather than “unequal” crossing over. Interestingly, Wright had predicted that B itself would be a duplication before it was demonstrated cytologically (Wright, 1929).

Bridges had earlier called attention to duplication-like structures in the salivary gland chromosomes of wild-type larvae (Bridges, 1935). In particular, he interpreted numerous double banded structures, or “doublets”, as two duplicated bands fused along their edges. Their structure suggests that they are reverse (ABBA), rather than direct (ABAB), repeats of single bands (Fig. 1). Bridges’ cytological evidence for such repeats combined with Sturtevant’s demonstration of position effect suggested that multiple alleles of a given gene might in some cases be resolvable into two or more repeated genes that acted like one because of a position effect. Evidence that multiple alleles might be resolvable into separable loci began to be obtained in the late 1930s by C.P. Oliver at the University of Minnesota. He found a low frequency of revertants to wild type in the offspring of females heterozygous for two recessive *lozenge (lz)* eye mutations. Although the revertants were invariably associated with crossing over in the region, he was unable to detect a reciprocal crossover having both mutants in the same chromosome. He therefore could only suggest that the revertants could be explained as the result of “unequal crossing over or crossing over between ‘repeats.’” (Oliver, 1940).

**Star and asteroid**

I was an undergraduate at that time and Oliver generously gave me a desk in his laboratory and allowed me to work on a new rough-eyed mutant that had been given to me by E. Novitski, who was then at Purdue University. [Novitski and I had begun our work with *Drosophila* in high school around 1935]. Bridges had suggested that it be called *Star-recessive (S’)*, since it acted as an allele of a weakly dominant rough eye mutant, *Star (S)*. Thus, S+/+ flies have slightly smaller eyes that are slightly roughened; S’/S’ flies have eyes reduced to about half their normal size and with a very roughened surface; while S’/S’ flies are nearly eyeless -figured in (Lewis, 1951). Although in a preliminary test, I found a revertant of S’ or of S in 3,235 offspring of S’/S’ females, when flanking markers were introduced I obtained no more wild-type products among 9,294 offspring (Lewis, 1939).

In spite of these inconclusive results, I continued the study of S and S’ as one of Sturtevant’s graduate students at Caltech, commencing in 1939. In the tradition of Morgan, Sturtevant allowed his students considerable freedom to choose their thesis research projects. Quite a risk was involved in choosing to work on S and its “alleles”. Crossovers between them would be rare if they were to occur at all. Even if the wild-type crossover could be recovered, it was expected that it would be very difficult to detect the reciprocal, or double mutant, crossover.

To increase the resolving power of the analysis, I made use of the interchromosomal effect of rearrangements on crossing over.
Introduction of heterozygosity for inversions in chromosome arms other than the left arm of the second chromosome, in which $S$ is located, resulted in an approximately four-fold increase in the frequency of crossing over in the vicinity of $S$. As in Oliver's work on $lz$, the revertants were invariably associated with crossing over between $S$ and $S'$. I renamed the latter «allele,» asteroid (ast).

A tandem duplication for the $S$ region which I had found as an x-ray induced revertant of ast (Lewis, 1941) lent itself to the recovery of the $S$ ast double mutant chromosome (Lewis, 1942a). A striking position effect was in evidence: whereas, $S^{+/+}$ast is nearly eyeless, the complementary genotype, $S$ ast/++, is nearly wild type, except for a slightly smaller and slightly roughened eye indistinguishable from that of $S^{+/+}$ (Lewis, 1942a), figured in (Lewis, 1951).

$S$ and ast proved to be localized to the 21 E 1-2 doublet of the salivary gland chromosomes (Fig. 1), the doublet which Bridges had singled out as being a representative example (Bridges, 1935). These cytogenetic studies of $S$ and ast formed my doctor’s thesis (Lewis, 1942b) published in part in 1945 (Lewis, 1945).

Comparison of the difference in phenotype between cis vs. trans genotypes is usually referred to as the cis-trans test, and the position effect, if present, as the cis-trans effect. For a history of this terminology see Hayes (1968).

Early studies of the bithorax mutants

In 1945, the time seemed ripe to look for more examples of the Star-asteroid type in the genome. An intriguing region of the third chromosome included three loci within less than one centiMorgan: namely, the bristle mutations, Stubble (Sb) and spineless (ss) and a homeotic mutation, bithorax (bx) (Fig. 1). Certain useful combinations of these mutants had already been synthesized by Bridges and maintained in the Caltech stock collection. The recessive alleles of Sb proved to be at a separate locus, that I named stubbloid (sbd), less than 0.1 centiMorgan to the left of the Sb locus. An especially striking position effect occurs: $sbd^{+}/Sb$ flies have extremely short blunt bristles, while $sbd^{+}/++$ flies are wild-type with no trace of the dominant short-bristle phenotype of $Sb/+$ flies.

It soon became evident that the diverse array of existing mutations of the bithorax type held considerable promise of being a cluster of genes rather than a multiple allelic series. It was for this reason that they were chosen for study rather than with any belief that they would tell us something about how genes control development.

The original bx mutant had been found by Bridges in 1915 as a transformation of the third thoracic segment (T3) toward the second (T2), notably causing the halteres to become partially wing-like. Body segments and structures of the wild-type adult are correlated with those of the late embryo in Figure 2. bx was the first example of a mutant that exhibited homeosis, a term Bateson had first coined for conversion of one structure into an homologous one (discussed in Lewis, 1994). In 1919, Bridges found a somewhat similar mutant that fully complemented bx, so he named it bithoraxoid (bxd); i.e., bx/bxd is wild type in phenotype. However, he later showed that bx$^{c5}$, which W. F. Hollander had found, failed to complement either bx or bx(b Bridges, 1944).

Although the original bx mutant has 100% penetrance, it is a highly variable transformation of, as it turns out, only the anterior portion of T3 toward anterior T2. Fortunately, two other bx-like mutants, bx$^{34e}$ (J. Schultz) and bx$^{3}$ (of C. Stern) had also been saved by Bridges (Bridges, 1944). These have 100% penetration and non-variable weak and strong transformations, respectively, of anterior T3 toward anterior T2. The wing-like halter of the bx$^{3}$ homozygote is shown in Figure 3.
Flies homozygous for bxd show 100% penetrance for a partial transformation of only the posterior portion of T3 toward posterior T2. The wing-like halter of a homozygote for an extreme bxd mutation, bxd100, is shown in Figure 3. In addition, bxd flies also have the first abdominal (A1) segment transformed toward T3, occasionally producing tiny rudimentary T3-like legs. A bxd hemizygote has a well-developed T3-like leg on the transformed A1 (Fig. 4D).

A crossing-over analysis showed that bx0 occupies a separate locus between the bx and bxd loci, and therefore it was first renamed Bithorax-like (Bxl) (Fig. 1), and later, Ultrabithorax (Ubx) (Lewis, 1951). This analysis provided a number of cis and trans genotypes that exhibited position effects. Examples are shown in Figure 5.

**Gene evolution by tandem duplication**

These early studies were viewed as supporting a simple hypothesis about how new genes arise from pre-existing genes. Based on the work of Sturtevant and Bridges, already cited above, the hypothesis proposed that new genes evolve from old genes by a two-step process: tandem gene duplication followed by one of the resulting duplicates mutating to a new function (Lewis, 1951). This "new" gene would generally not be easily established in the population unless the other, or "old" gene, was retained to carry out the old function. As a result the genome would be expected to contain clusters of closely linked and functionally related genes that superficially act like a single gene. At the Cold Spring Harbor in 1950, I reported (Lewis, 1951) on the evidence in support of this hypothesis from three studies: of other organisms; of the above mentioned S, Sb and bx regions; and of Iz mutants by Green and Green (1949).

**An early model of the cis-trans effect**

A model (Fig. 6) was also presented at that Symposium to account for the cis-trans effect (Lewis, 1951). It was based on the then generally accepted biochemical dogma that genes were proteins, and that they could catalyze enzymatic reactions. The wild-type alleles of a and b were assumed to control sequential steps in a biochemical pathway in which a substrate, S, is converted into two products, A and B, that are produced at the site of the genes in the chromosome. The a and b mutants are assumed to lower production of A and B, symbolized as <A and <B, respectively (Fig. 6I). As a result, a b/++ (Fig. 6I) is expected to produce enough B to be wild type, or nearly so. By contrast, a+/+b (Fig. 6II) would produce insufficient B, and therefore be mutant in phenotype (Lewis, 1951).

The model could therefore also account for polarized cis-trans effects. For example, when bx3 is opposite an extreme x-ray induced bxd allele, such as bxd100, bx3+/+bxd100 flies have a very slight wing-like transformation of the posterior portion of the halteres. On the other hand, they have no trace of the bx phenotype, even though the latter phenotype is a more sensitive one for the detection of slight effects than is the bxd phenotype. Hence bx3 appears to weakly inactivate bxd, but even extreme bxd mutants do not inactivate bx.

In retrospect the model is no longer compatible with our present knowledge of the structure and function of the gene. However, since no assumptions were made about the nature of the products S, A and B, the model might still be tenable if S, A and B correspond to non-coding RNA transcripts. The real value of this hypothesis was that it led to an experiment that revealed a new phenomenon of "transvection", to be discussed below.

**Contrabithorax - a gain-of-function mutation**

In 1954 (Lewis, 1954a), an x-ray induced mutation was found that had T2 transformed toward T3. This "gain-of-function" (Lewis, 1984) phenotype was therefore the inverse of the T3 to T2 transformation characteristic of the bx and bxd mutations. Surprisingly, mapping showed it to be a double mutation made up of a gain-
of-function mutation, *Contras bithorax* (*Cbx*), the locus of which lies between the *bx* and *Ubx* loci, and a recessive loss-of-function mutation, *postbithorax* (*pxb*), that occupies a new locus distal to that of the original *bxd* mutation. Thus the map expanded to five loci, at which there were mutations with effects on one or more of the segments, T2, T3 and A1 (Lewis, 1955). This cluster of mutant loci came to be called the *Ubx* domain of a much larger cluster, the bithorax complex (BX-C) (Fig. 7). The latter name is derived from “gene complex”, a term invented by Brink for a closely linked cluster of genes that he predicted would be closely related in function (Brink, 1932). Kaufman and his co-workers defined the Antennapedia-complex (ANT-C) that controls the identity of segments anterior to *Ubx* (Fig. 7). The latter name is derived from “gene complex”, a term invented by Brink for a closely linked cluster of genes that he predicted would be closely related in function (Brink, 1932). Kaufman and his co-workers defined the Antennapedia-complex (ANT-C) that controls the identity of segments anterior to those controlled by the BX-C (Kaufman, 1980).

Unlike the *bxd* mutant, *pxb* has only a transformation of the posterior portion of T3 toward posterior T2. The *trans* heterozygote, *bxd*/*pxb*, shows a *pxb* phenotype but no trace of the transformation of A1 toward T3 that is typical of the *bxd* homozygote. Furthermore, *bxd*/*pxb* also shows, albeit weakly, a *pxb* phenotype, but no trace of a *bx* phenotype. In both of these examples the *cis*-heterozygotes are wild type. Thus, polarized inactivation of *pxb* function can be effected in *cis* by either *bxd* or *bx3*.

### The transvection phenomenon

One of the predictions of the early model of the *cis-trans* effect (Fig. 6) was that disruption of somatic pairing might intensify the difference between *cis* and *trans* types. Specifically, heterozygosity for a chromosomal rearrangement that would disrupt pairing in an *a+/b* individual would be expected to cause a more extreme *b* phenotype. The prediction was borne out, and a powerful new method was discovered for detecting chromosomal rearrangements in the first generation after their induction. The method was first used to measure the frequency of induction of such rearrangements in the progeny of males exposed to neutrons from an atomic bomb test (Lewis, 1954b).

The method detects only the majority of rearrangements having one breakage point in a «critical» region of some 500 bands of the salivary gland chromosomes; namely, the region between the centromere of the third chromosome and the locus of the BX-C. Similar findings were later obtained for the *decapentaplegic* (*dpp*) region in 2L (Gelbart, 1982) and for the *eyes-absent* (*eya*) region in that arm (Leiserson et al., 1994).

Although at first only *trans* genotypes showed the phenomenon, it was soon found that *Cbx Ubx*+/+ was also subject to transvection (Lewis, 1955). *Cbx* in this genotype was found to exert a slight gain of function of *Ubx*, chiefly expressed by spread wings and a reduced alula, when the chromosomes are paired. That effect is abolished (wings normal) when pairing is disrupted by transvection-suppressing rearrangements. As a result, it became possible to mutate wild type and to select rearrangements that abolished the weak *Cbx* effect of the *Cbx Ubx*+/+ genotype. Among the resultant rearrangements, some, as expected, had breaks within the BX-C. These breaks were unselected for any effect on function in the BX-C other than suppression of transvection. Such rearrangements, when subsequently tested over deletions of the BX-C, provided the basis for discovering additional *infra-abdominal* (*iab*) regions and ordering all of the known regions from *iab-2* to *iab-8*, inclusive. The *iab-9* region has been identified by means of breakpoints associated with gain-of-function mutants in that region, namely *Uab* (Lewis, 1978) and *Tab* (Celniker and Lewis, 1987) (Fig. 7).

In the process of isolating transvection-suppressing rearrangements, a sex-linked mutant was recovered in two independent cases, whose effect was to enhance the *bithorax* phenotype of *bx3+/Ubx*. This mutant, originally named, *enhancer-bithorax* (*e-bx*), proved to be an allele of the *zeste* (*z*) gene (Judd et al., 1972) and to be like the *z*, or null, alleles of Gans (1953). It was soon found that *z*+/+*, as it is now symbolized, suppresses transvection not only in the case of the BX-C but also *dpp*. The *z* protein has been shown to be a DNA-binding protein that binds in *vitro* to the *Ubx* gene as well as to other genes (Benson and Pirrotta, 1987). Benson and Pirrotta suggest that “transvection effects are a by-product of normal intragenic *z* action” (Benson and Pirrotta, 1988).
Remarkably, tandem duplications for the BX-C region act as powerful suppressors of transvection, when placed opposite the Cbx Ubx chromosome (Lewis, unpublished). Evidently, the duplicate regions pair intrachromosomally with one another and prevent the Cbx mutant from gaining access to the Ubx regions. In organisms which lack somatic pairing between homologous chromosomes, such as the vertebrates, intrachromosomal pairing of tandem repeats may still occur. In that event, transvection may prove to be a general phenomenon applicable to tandemly repeated regions in all organisms.

Mobile elements in the bithorax complex

In 1932 Bridges reported (Bridges, 1932) the discovery of one of the first suppressor mutants in Drosophila. He named it suppressor-of-Hairy wing [now symbolized su(Hw)] and found that it acted as a recessive suppressor of certain alleles of a number of other genes. Although the bx\textsuperscript{3} mutation had been saved as a balanced stock, when I used it in 1946 the homozygote appeared wild type in phenotype, as if the mutant had reverted. In fact, the stock had acquired a suppressor that mapped to the same locus as that of Bridges’ su(Hw). His mutant had been lost, but the new occurrence, named su\textsuperscript{2}-Hw, suppressed the same group of specific alleles as was reported for su(Hw). In addition, we found that it not only suppressed bx\textsuperscript{3}, bx\textsuperscript{3}e and bxd, but also specific alleles of many other genes (Lewis, 1986).

The mechanism by which su\textsuperscript{2}-Hw suppresses specific alleles proved elusive until many years later, when it was shown that such alleles are the result of an insertion of the mobile element, gypsy, almost invariably in the non-coding portion of the gene (Modolell et al., 1983). The wild-type su\textsuperscript{2}-Hw gene codes for a DNA-binding protein (Parkhurst et al., 1988) that is assumed to bind to specific sequences in the gypsy element, thereby lowering the rate of transcription of the gene containing that element (Parkhurst and Corces, 1986). Hence, in the su\textsuperscript{2}-Hw homozygote, failure of the mutant protein to block transcription of that gene would restore the wild type phenotype.

In retrospect, it now seems extremely fortunate that the early mapping of mutants in the Ubx domain was carried out using mutations that were insertions or deletions. Thus, bx\textsuperscript{3} and bxd are gypsy insertions (7 kb in length), Ubx is a «Doc» mobile element (Bender et al., 1983), and pbx and Cbx are a deletion and insertion, respectively, of a 17 kb segment of DNA (Bender et al., 1983). Had they been true point mutations, they might then have been subject to gene conversion, a phenomenon first discovered in fungi and characterized by high negative interference over short map regions and aberrant segregation of alleles in a meiotic tetrad (reviewed by Holliday, 1964). As a result, unambiguous ordering of mutants in the Ubx domain would probably not have been possible.

Half-tetrad mapping of the Ultrabithorax domain

The great diversity of phenotypes represented by mutants at the five known loci of the Ubx domain made it relatively easy to derive double mutants and, in turn, higher multiples, including the quintuple mutant, bx\textsuperscript{3} Cbx Ubx bxd pbx. Although flanking marker recombination provided unambiguous ordering of these loci, the possibility of gene conversion was of sufficient concern that I undertook a half-tetrad analysis of that domain.

Attached autosomal arms had been synthesized, partly to be able to perform such an analysis, by I. Rasmussen and E. Orias, working in my laboratory (Lindsley and Grell, 1968). Females were constructed with the quintuple mutant combination in one of the attached arms and the corresponding five wild-type alleles in the other arm, along with appropriate flanking markers (Fig. 8); their phenotype was indistinguishable from that of Ubx/+ (Lewis, 1967). Among approximately 221,000 female offspring, 19 were the result of exchanges in the regions between the loci of bx\textsuperscript{3} and pbx. Reciprocal crossovers were recovered simultaneously from four out of five of the regions and were easily detected by their having strong cis-trans effects when compared with the maternal Ubx/+ phenotype. None of the half-tetrads showed evidence of gene conversion. As one possible explanation it was suggested that

![Fig. 5. Cis-trans effects involving the bithorax (bx) and Ultrabithorax (here designated Bxl) mutants, illustrated by camera lucida drawings of the dorsal and lateral region of T3 of the adult fly. The pair of genotypes in each vertical column are identical except for the way in which the alleles are distributed between homologous chromosomes. Reprinted from Lewis, 1951.](image)

![Fig. 6. An early model to explain cis-trans effects. Paired homologous chromosomes are diagrammed by the long horizontal lines. Two adjacent loci are shown with either wild-type (+) or mutant (a or b) alleles. The genes at these loci are assumed to catalyze the reaction of the substrate S into product A, and product A to product B. The A product is assumed to remain in the vicinity of the locus where it is produced. The cis configuration (I) produces sufficient B to give a nearly wild-type phenotype. The trans configuration (II) produces insufficient B resulting in a mutant phenotype. Reprinted from (Lewis, 1951).](image)
review of work on the homeotic clusters in a number of organisms (Lewis, 1992). The following sections will be concerned chiefly with the organization and function of the BX-C.

By generating somatic mosaics for the bx phenotype, I was able to show that the effects of the bx mutants are highly autonomous (Lewis, 1963). Thus, when cells mutant for the bx+ function arise from induced somatic crossing over in bx+/+ animals, the cells express the expected mutant phenotype, namely T2-type bristles on T3, which normally lacks any bristles. Morata and García-Bellido provided additional examples and showed that the mutant tissue could arise from exchange events induced as late as the last larval instar (Morata and García-Bellido, 1976). Thus, the wild-type products of at least the Ubx domain are not diffusible to any appreciable extent, and such products continually regulate the development of cuticular structures of T3 into late larval life.

In 1964, borrowing from the then-prevailing biochemical dogma based on the operon model, I interpreted the function of the genes of the BX-C to be to "repress certain systems of cellular differentiation and thereby allow other systems to come into play" (Lewis, 1964). Clearly, that function could also be to activate other systems, as García-Bellido later pointed out (García-Bellido, 1975).

Early studies of the BX-C had reached an impasse until homozygotes for deletions of parts, or of all, of the complex were found to have striking effects on cuticular structures of the late embryo. Simple preparations of late embryos cleared in a drop of lactic acid permitted the study of many embryonic lethal phenotypes.

It became evident that the BX-C included genetic material that programmed the development of not only T3 and A1, but also all of the remaining abdominal segments from A2 through A9, inclusive (Lewis, 1978). Thus, animals lacking the entire BX-C, as the result of being homozygous for deletions that removed all of the 89E1-4 bands, were found to die at the end of embryonic development and to have a striking transformation of the first seven abdominal segments toward the T2 segment. The cuticular structures involved include anterior spiracles, ventral pits, Keilin organs and other sense organs. The A8 and A9 segments transform even more anteriorly toward a head segment, based on their developing tiny rudiments of the mandibular hooks (Fig. 2).

It is always dangerous to deduce the wild type function of a gene from a loss-of-function mutations, especially for genes which affect morphology. The wild-type function of major regions of the BX-C could be inferred by adding them, to a homozygous gene from a loss-of-function mutations, especially for genes which affect morphology. The wild-type function of major regions of the BX-C could be inferred by adding them, to a homozygous

The bithorax complex and its organization

Duncan has provided a comprehensive and thorough review of the complex (Duncan, 1987). I have recently given a brief historical view of the BX-C, in which I discussed the organization and function of the BX-C. The BX-C includes a number of genes that regulate the development of cuticular structures of the late embryo. Simple preparations of late embryos cleared in a drop of lactic acid permitted the study of many embryonic lethal phenotypes.

It became evident that the BX-C included genetic material that programmed the development of not only T3 and A1, but also all of the remaining abdominal segments from A2 through A9, inclusive (Lewis, 1978). Thus, animals lacking the entire BX-C, as the result of being homozygous for deletions that removed all of the 89E1-4 bands, were found to die at the end of embryonic development and to have a striking transformation of the first seven abdominal segments toward the T2 segment. The cuticular structures involved include anterior spiracles, ventral pits, Keilin organs and other sense organs. The A8 and A9 segments transform even more anteriorly toward a head segment, based on their developing tiny rudiments of the mandibular hooks (Fig. 2).

It is always dangerous to deduce the wild type function of a gene from a loss-of-function mutations, especially for genes which affect morphology. The wild-type function of major regions of the BX-C could be inferred by adding them, to a homozygous deletion of the BX-C (Dp-P9) (Lewis, 1978). For example, a duplication, Dp(3)bxd00, that includes a wild-type copy of the Ubx domain proximal to the bxd region, restores the longitudinal tracheal trunks in all segments from T2 to A8, inclusive. The genes of the BX-C control the development of specific structures and organs of the segments rather than segmentation per se. The
particular segments in which a given BX-C gene is expressed is determined by the combined action of trans-regulatory genes.

The analysis of the functions of cis-regulatory regions located distal to the Ubx domain, made use of chromosomal rearrangements having breakpoints in those regions. Loss-of-function mutations in those regions were designated as "infra-abdominal (iab)" mutations; for example, a loss-of-function in the iab-2 region, associated with the transposition, Tp(2;3)P10, causes A2 to transform toward A1. By 1978, three iab regions had been identified, iab-2, iab-3, and iab-4, and a fourth, iab-5, was inferred from an analysis of revertants of a dominant gain-of-function mutation, Miscadastral pigmentation (Mcp) by M. Crosby (Lewis, 1978). Subsequently, the regions of the BX-C controlling abdominal development were divided into two domains, abdominal-A (abd-A) and Abdominal-B (Abd-B), based on lethal complementation studies (Sanchez-Herrero et al., 1985; Tiong et al., 1985).

Rules governing cis-regulation of the BX-C

The BX-C is regulated in cis and in trans. Rules governing its cis-regulation are considered first and were deduced from genetic analysis. Many of the rules are highly unusual and possibly unique. It seems likely that their molecular analysis will reveal hitherto unsuspected regulatory mechanisms.

Colinearity

The rule of colinearity (COL) states that the order of the BX-C loci in the chromosome parallels the order in which the units at those loci are expressed along the antero-posterior axis of the body. Two types of gradients had been invoked to explain this rule: "an antero-posterior gradient in repressor concentration along the embryo and a proximo-distal gradient along the chromosome in the affinities for repressor of each gene’s cis-regulatory element" (Lewis, 1978).

Molecular studies confirmed the rule and extended it to all of the abdominal cis-regulatory regions from iab-2 to iab-8, inclusive (Karch et al., 1985). Associated with the COL rule is the strong tendency for the proteins of the BX-C, once expressed to continue to be expressed more posteriorly in the body except for the terminalia. This is elegantly shown in Figure 9, for the Ubx, abd-A and Abd-B proteins visualized by the use of immunostaining with antibodies specific to each.

Cis-inactivation

The second rule of cis-inactivation (CIN) states that loss-of-function mutations in a given cis-regulatory region tend to inactivate the next more distal region of the complex. Examples have already been cited for the polar inactivation of the bx^3 function by bx^3 and by bxd. Other examples were later found in analyzing rearrangement breakpoints in the iab regions of the BX-C (Lewis, 1986). It has not been possible to establish whether there are CIN effects between major domains of the BX-C.

Cis-overexpression

The third rule of cis-overexpression (COE) is a quite surprising one. The rule states that the loss of function associated with a given cis-regulatory region tends to be accompanied by an overexpression of the function associated with the cis-regulatory region that lies immediately proximal to it. In the abdominal domains, rearrangements with breaks in the iab-3 region, for example, not only have a loss-of-function iab-3 phenotype (A3 transformed toward A2) but a gain-of-function of the iab-2 region that is manifested as a transformation of the A1 segment toward A2. Other examples have been described (Lewis, 1986).

COE effects are known not only for breakpoints of chromosomal rearrangements but for gypsy insertions. An important one is a COE effect of bx^3. Flies homozygous for bx^3 have a reduction in the extreme anterior region of T2. This effect is dominant since it is not suppressed by duplications that totally suppress the recessive bx^3 transformation of T3 toward T2. An x-ray induced mutant, anteroebithorax (abx), was discovered that had a weak bithorax-like phenotype. It is located just proximal to bx, and abx bx^2 double mutants lack the COE effect on T2 seen in the bx^3 single mutant genotype. Until abx had been found, it was not possible to achieve a full transformation of T3 toward T2; i.e., the bx^2 pbx double mutant homozygote fails to transform the most anterior portion of T3. Flies homozygous for the triple mutant, abx bx^2 pbx, were constructed and proved to have virtual complete transformation of the wing and cuticular structures of T3 transformed toward those of T2, resulting in a four-winged fly (Fig. 4).

Negative trans-regulation of the bithorax complex

In 1947, a remarkable x-ray induced dominant mutant, Polycomb (Pc), was found by P. H. Lewis (Lewis, 1949). It had sex combs on not only on the first, but the 2nd and 3rd pair of legs, and rudimentary antennal to leg transformations resembling those of Antennapedia (Antp) mutants. It also had effects that were only later realized to be gain of function of genes in the Ubx domain; namely, reduction in the extreme anterior region of T2 and reduction in the wing similar to that of weak Cbx phenotypes, such as in Cbx Ubx^+/+. It was nearly 30 years before it was realized that Pc is a mutation in a gene that acts as a negative regulator of the BX-C, and of the ANT-C complex as well. Thus, the homozygous Pc embryo has the three thoracic and the first seven abdominal segments all transformed toward A8, presumably as the result of
derepression of the Abd-B-domain (Lewis, 1978) (figured in Duncan, 1982).

Duncan found a second mutant of the Polycombtype, Polycomb-like (Pcl-I) (Duncan, 1982). Pc and Pcl have proved to be but two of a family of genes that act as negative regulators (Jurgens, 1985). That the Pc protein is involved in binding to the BX-C and the ANT-C regions (as well as to other regions) has been elegantly shown by immunostaining of salivary gland chromosome with an antibody to that protein (Zink, 1989). Since the Pc protein is a non-histone chromosomal protein, rather than a DNA-binding protein (Paro and Hogness, 1991) its binding specificity may reside in its complexing with proteins of other genes of the Pc family, some of which first bind specifically to BX-C and ANT-C.

**Positive trans-regulation of the bithorax complex**

Positive trans-regulators were also found, such as Regulator of bithorax (Rg-bx). An analysis of this mutant, and of deficiencies which include the locus, indicate that the wild-type gene is a positive regulator of the BX-C (Lewis, 1981). A partial loss-of-function allele, trithorax (trx), was then found by Ingham (Ingham and Whittle, 1980). The trx gene has been cloned and is a DNA-binding protein of the zinc finger category (Kuzin et al., 1994; Stassen et al., 1995). More recently, Kennison and Tamkun have identified a family of genes like trx that act when mutated as enhancers of bx phenotypes (Kennison and Tamkun, 1988).

Additional classes of trans-regulators of the BX-C have come from the studies of Nüsslein-Volhard and Wieschaus (Nüsslein-Volhard and Wieschaus, 1980). For example, the gap gene, hunchback (hb), is involved in establishing major subdivisions of the body regions. It encodes a zinc finger protein and acts as a negative regulator of the BX-C, keeping the complex turned off in the anterior regions of the body, presumably by the binding of the hb protein to at least one specific motif in the Ubx gene (Qian and Pirrotta, 1988). A dominant mutant, Regulator of postbithorax (Rg-pbx), is now known to be a gain-of-function mutation in the hb gene (Bender et al., 1987). It produces variable pbx-like transformations of the halter (Lewis, 1968).

Another example is the Krupple (Kr) gene of Gloor (1950). It is also a gap gene and encodes a DNA-binding protein (Rosenberg et al., 1986; Schuh et al., 1986). One motif to which it binds is in the iab-2 region and, on two independent occasions, a mutation in a single specific base pair of that motif has resulted in a dominant Hyperabdominal (Hab) phenotype (Schmel et al., 1994). These gain-of-function mutants have poor penetrance, but in some crosses Hab/+ flies occasionally have only four legs and no halteres owing to T3 being transformed toward A2 (Lewis, 1978).

**Molecular analysis of the bithorax complex**

Molecular analysis of the Ubx domain of the BX-C was initiated by D. Hogness and co-workers in 1978 and they soon identified the major features of that region. The bx mutants, Ubx, and several bxd mutants all proved to be insertions of transposable elements (Bender et al., 1983). Molecular studies revealed a single transcription unit coding for proteins in the Ubx domain (O’Connor et al., 1988; Kornfeld et al., 1989). The embryonic distribution of the UBX protein products was determined by White and Wilcox (1984) and by Beachy et al. (1985); see also Figure 9. The transcription unit and protein product of the second domain, abdominal-A, were characterized by Karch et al., (1990). The third domain, Abdominal-B, produces at least four transcripts (DeLorenzi et al., 1988; Kuziora and McGinnis, 1988; Sanchez-Herrero and Crosby, 1988; Celniker et al., 1989; Zvortink and Sakonju, 1989) and two Abd-B proteins (Celniker et al., 1989, 1990; Boulet et al., 1991).

Surprisingly, the cis-regulatory regions are transcriptionally active, as first shown for the bxd region of the BX-C (Lispitz et al., 1987). This region produces a large (26.5 kb) primary transcript, that is then spliced to yield a family of non-protein coding RNAs (i.e., containing multiple stop codons). Similar non-coding transcription units are known for the iab-4 region (Cumberledge et al., 1990).

**The Transabdominal mutation**

King and Wilson (1975) called attention to the possible importance in evolution of creating novel phenotypes solely by rearrangements involving cis-regulatory sequences. A striking example was our discovery of an X-ray induced dominant mutation, Transabdominal (Tab). Tab/+ flies have a sexually dimorphic pattern of pigmented bands in the dorsal thorax of T2 (Fig. 10).
transcription factor (Lee et al., 1993) of the dorsal thoracic disc of T2, which gives rise to the dorsal thorax, show cells in the tergites of segments A5 and A6. Thus, the pigmented bands in the Tab/+ male dorsal thorax are broad as in the A5 and A6 male tergites; whereas, in the Tab/+ female they are narrow as in the corresponding female tergites. The Tab mutant is associated with an inversion having one breakpoint in the iab-9 cis-regulatory region (Fig. 7) and the other near the stripe (sr) locus in 90D which codes for an early growth-response transcription factor (Lee et al., 1995). In situ studies (Celniker and Lewis, 1993) of the dorsal thoracic disc of T2, which gives rise to the dorsal thorax, show cells in Tab/+ animals that express the ABD-BII protein and its RNA. These cells correspond to the sites of the bands in the Tab/+ adult thorax and appear to be the sites of attachment for certain thoracic muscles. Our studies of the RNA and protein distributions in embryos and imaginal discs indicate that the Tab mutation represents a case in which cis-regulatory regions of a gene involved in defining the development of muscle attachment sites is now driving ABD-B protein expression (Lee et al., 1995). Other minor disturbances in the abdominal tergites of Tab/+ flies are believed to involve ectopic expression of the Abd-B protein in such attachment sites for abdominal tergite muscles.

Control of somatic gonad development in Drosophila and Bombyx

As early as 1943, Itikawa reported (Ilikawa, 1943) on a mutant designated E^N whose phenotype when homozygous parallels closely that of the homozygous deficiency for the BX-C in Drosophila (Di-P9). Itikawa's discovery that certain mutants of the E series lacked gonads led me to examine a dominant mutant, Ultra-abdominal^f (Uab^f), which is associated with a recessive iab-3 phenotype. Internally, the Uab^f hemizygote was found to lack gonads (Lewis, 1978). Subsequently, I found that rearrangements with breakpoints in the iab-4 region of the BX-C, when viable as homozygotes appear virtually wild type, but internally they lack gonads (Lewis, 1986). Loss of gonads in iab-2 and iab-3 mutant animals results from cis-inactivation of the iab-4 region (Lewis, unpublished). Since the gonad is of mesodermal origin, its loss was one of the first indications that the BX-C phenotypes were not limited to ectodermal tissues.

A comparative molecular analysis of the iab-4 cis-regulatory with regions controlling gonad formation in Bombyx and other animals may show how the homeotic genes control the development of a specific structure. Thus, since some of the more primitive non-segmented animals, such as the nematode, have somatic gonads, it is likely that control of the initiation of their development will have common features. Of great interest will be the target genes in Drosophila that accomplish such initiation. A promising approach to understanding the process in human beings can be expected to come from analyzing molecularly the basis of inherited defects in the human gonad.

The homeobox and tandem gene duplication

Molecular support for the assumption that tandem gene duplication was responsible for at least the coding portions of the BX-C and the ANT-C complex finally came with the discovery of the homeobox in 1984, by McGinnis et al. (1984) and Scott and Weiner (1984) who independently showed that the proteins encoded by the Ubx and Antp genes contain a remarkably conserved group of amino acids, known as the homeodomain. The DNA sequence encoding the homeodomain was named the homeobox (McGinnis et al., 1984). The homeobox sequence is conserved to a remarkably high degree throughout the animal kingdom and it was used to probe for homologs of the BX-C and ANT-C in many other organisms, including vertebrates as well as invertebrates (Gehring and Hiromi, 1986). Most of these organisms have the homologs of both the BX-C and the ANT-C in a single complex known as the homeotic complex (HOM-C) (Beeman, 1987).

In unsegmented organisms like Caenorhabditis (Kenyon and Want, 1991) there are apparently only a few HOM-C genes. Insects such as the silkworm, Bombyx (Tazima, 1964) and the flour beetle, Tribolium (Beeman et al., 1993) have larger clusters as in Drosophila. The most primitive vertebrates represented by the lancelet, Amphioxus (García-Fernandez and Holland, 1994; Holland et al., 1994) also have a single large HOM-C. However, higher vertebrates have four semi-redundant copies of the HOM-C. In the mouse and human, each copy is on a different chromosome. This redundancy makes it difficult to dissect the function of a given gene in any one of the sets. Remarkable progress is being made by using gene knock-out techniques in mice, to study the role of the HOM-C genes in development. HOM-C gene expression in the mouse,
as in *Drosophila*, obeys the rule of colinearity (reviewed by Lewis, 1992). Their segmental expression limits are also regulated in *trans* by genes that are remarkably parallel to those of the *Pc* Group and *trx* Group (reviewed by Simon, 1995). HOM-C genes are now regarded as master control genes whose proteins bind to the *cis*-regulatory regions of target genes. The latter then activate or repress systems of cellular processes that accomplish the final development of the organism. Even minor mutant lesions in HOM-C genes may be expected to have global effects on such systems. An example is a targeted gene-disruption of the mouse *HOX A3* gene (formerly *HOX 1.5*) that leads to defects in the thyroid glands and surrounding tissues (Chisaka and Capecki, 1991). The resultant group of defects resembles those seen in the congenital DiGeorge syndrome of human beings.

**Complete sequence of the bithorax complex**

The DNA sequence of the BX-C has now been completely determined (Martin et al., 1995) and a preliminary analysis made of it (Lewis et al., 1995). The protein coding regions comprise only 2% of the entire sequence. The other 98% is expected to contain a diverse group of motifs to which *trans*-regulatory proteins bind, thereby conferring the specific spatial and temporal expression of the protein products of each domain. There may also be a regulatory role for non-coding RNA’s of the type identified in the *bxd* and *lab-4* regions.

**The next fifty years**

Only three of the many future challenges will be outlined: (1) molecular and genetic approaches are needed to determine the immediate target genes that are turned on or off by the genes of the HOM-C; (2) since the genes of the HOM-C have tended to remain tightly linked and colinear with their expression patterns along the body axis, it will be exciting to discover underlying mechanisms that have always been there and; (3) comparative DNA sequence analysis of the HOM-C among many different organisms may provide evidence that the *cis*-regulatory regions have evolved by tandem duplication. Ultimately, comparisons of the HOM-C throughout the animal kingdom should provide a picture of how the organisms, as well as the genes of the HOM-C, have evolved.

**Conclusions**

Basic research concerned with testing a simple hypothesis about how new genes arise from old genes led after many circuitous routes to the discovery of the homeotic complex (HOM-C). This cluster of master control genes programs much of the development of all higher animal organisms. Each of the genes contain a homeobox, a remarkably conserved DNA sequence that provides molecular support for the hypothesis that the complex itself arose by a process of tandem gene duplication. The high degree of conservation of the HOM-C, itself, between vertebrates and invertebrates indicates that it arose from an ancestral complex over 500 million years ago, the estimated time of separation of these two great groups of organisms.

It is likely that mutations within the HOM-C’s of human beings are the cause of certain genetically based abnormalities that arise at various stages of human development. Somatic mutations in genes of the HOM-C may conceivably be involved in the generation of tumors. Meanwhile, future genetic and molecular studies of the HOM-C in lower creatures that have but one set of the complex promise to advance our understanding of its role as a master regulator of development.

Much has been learned about the role of the HOM-C in development, and about its molecular products. Nevertheless, we are still unable to make sense of much of the DNA sequence of the bithorax complex (BX-C) or to explain how the complex is itself regulated. Progress will still need to be driven by the logic of genetics, and by further increases in abstraction.

**Acknowledgments**

Recent work on the BX-C has been was supported by research grants from the National Institutes of Health, the ACS and the March of Dimes.

I thank Welcome Bender, Howard Lipshitz, Susan Celinker and Joanne Topol, for a critical reading of the manuscript, and John Knafels, Brian Kearns and Beth Turner for assistance in the preparation of it. Antibodies were kindly provided by R. White, against Ubx protein and by I. Duncan against abd-A. I am indebted to W. Bender for providing Figure 9. While at Caltech, a number of colleagues have directly contributed to our research, namely, Welcome Bender, Marie-Paz Capdevila, Susan Celinker, Loring Craymer, Madeline Crosby, Ian Duncan, Antonio Garcia-Bellido, William Gelbart, Alain Ghysen, Hans Gloor, E. H. Grell, Rhoda Grell, Lily Jan, Burke Judd, Howard Lipshitz, Margit Lohs-Schardin, Rolf Nothiger, Eduardo Orias, Inge Rasmussen and Shige Sakonju. Finally, I want to stress the close cooperation that we have had over the years with David Hogness and colleagues at Stanford University, Welcome Bender at Harvard Medical School and Ian Duncan at Washington University. It was David Hogness’ foresight to launch the molecular analysis of the bithorax complex in 1978 in his laboratory.

**References**


