ON THE STRUCTURE OF GENE CONTROL REGIONS

BRUCE WALLACE[1] AND THOMAS L. KASS

Cornell University, Ithaca, N. Y. 14850

Manuscript received October 19, 1973

ABSTRACT

A model emphasizing the possible genetic role of tandem duplications of reverse repeats has been developed as an extension of CRICK's (1971) general model for the chromosomes of higher organisms. Although developed initially (1) to explain why random differences in the control regions of individual gene loci might confer a selective advantage on heterozygous individuals as well as (2) to offer the species a means by which such differences might be effected without mutational harm, it seems that control regions built on a foundation of tandemly duplicated reverse repeats would exhibit many properties previously observed in studies on mutable loci in various organisms.

APPROXIMATELY fifteen years ago, experimental evidence suggesting that under certain conditions X-ray-induced mutations might have an enhancing effect on viability was obtained for *Drosophila melanogaster* (WALLACE 1957, 1958 and 1959). Control flies that were homozygous for their second chromosomes proved to be less viable than comparable individuals, also homozygous for their second chromosomes, which were *heterozygous* for genetic changes induced by the exposure of one of the two homologues to 500r of X-radiation. At the time, these experimental results were vigorously questioned (MULLER and FALK 1961; FALK 1961) on both theoretical and technical grounds. The initial observations, however, have been confirmed and extended by WALLACE (1963b), MUKAI and YOSHIKAWA (1964) and MARUYAMA and CROW (1974). As the latter say, ". . . it is hard to ignore three independent sets of experiments, in three different laboratories, each significant and pointing in the same direction." In addition to direct measures of the average effect of irradiated chromosomes on viability, a number of studies have been made of the effect of radiation on exposed populations, studies that deliberately interweave the roles of mutation and natural selection in determining observed fitnesses. AYALA (1966) and, more recently, BLAYLOCK and SHUGART (1972) have shown for various Drosophila species that the fitnesses of populations which are exposed to low levels of radiation generally surpass those of unirradiated controls.

In an extremely thorough and systematic analysis of heterosis in *D. melanogaster*, VANN (1966) showed that genetic differences between homologous chromosomal segments—whatever the position of these segments within the two major autosomes—are responsible for hybrid vigor. He found that amounts of

[1] This paper was prepared while the senior author's research was supported under Contract No. AT-(11-1)-3129, U. S. Atomic Energy Commission. This work was also supported by PHS Training Grant No. T 1 GM 1035 from the National Institute of General Medical Sciences.

genetic differences measured in terms of either lengths of chromosomal segments or degrees of relatedness of homologous segments determine heterosis according to the same rules. Thus, however much they differ genetically, chromosomal segments extrapolated to zero length were found to have identical effects on fitness, an effect identical, as well, to that shown by segments of *any* length when the latter involved chromosomes obtained from the same, highly inbred strain of flies.

VANN's experiments and those involving mutagenic radiation suggest that, at a multitude of loci scattered throughout the genome, a dissimilarity of the two alleles (even a dissimilarity that is random in origin) is selectively superior to identity of these alleles. VANN's analysis shows that chromosomal segments at whatever location fit a common pattern. What sort of genetic model might account for beneficial effects of ubiqitous, if not random, differences? What can be the advantage of heterozygosity *per se* to use LERNER's (1954) phrase?

WALLACE (1963a) attempted to answer these questions by focusing attention on the control of gene function rather than on the proteins coded for by structural genes. This early attempt was not notably successful. The models of gene control which were understood at that time were too simple for the task at hand, while suspected control systems of higher organisms were too complex to be understood.

In recent years two excellent papers on gene control have appeared. BRITTEN and DAVIDSON (1969) have postulated a hierarchal system of gene control involving elements each of which can call into play batteries of structural genes (or of subsidiary control elements). CRICK (1971) has described a model in which the recognition sites of gene control are encoded in the DNA thread of each chromosome. These sites ("sensors" in the following discussion; presumably these are adjacent to transcription initiation sites) are deployed as single strands of DNA situated atop hairpin loops of double stranded DNA ("pedestals"); the torsional stress at the apical bend is, according to CRICK, the means by which the two strands of the sensor region are forced apart.

There is no need here to give the details of CRICK's model. Suffice it to say that the image of control regions consisting of linear arrays of sensors, each capable of initiating the transcription of its associated structural gene, suggested an explanation for the advantage of random change in otherwise homozygous individuals (where homozygosis includes identity of the sequential order of corresponding sensors within allelic control regions). Furthermore, in seeking a genetic mechanism that would permit the spontaneous shifting of sensors within control regions without concomitant harm, it appeared that tandem duplications of reverse repeats could serve the purpose. Then, upon examining the more general recombinational properties of such duplications, it appeared that a model capable of accounting for many well known genetic phenomena had been developed.

The sections that follow are intended to elaborate on the above outline. Lest tandem duplications of reverse repeats strike readers as one of the less interesting aspects of genetics, let us regain their attention with the following claim: Tandem

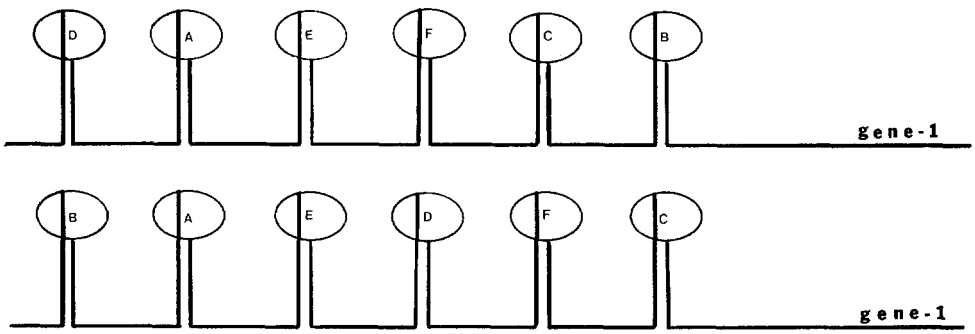


FIGURE 1.—A diagrammatic representation of CRICK's (1971) model of gene control. Each pedestal consists of a hairpin-like loop of double stranded DNA (heavy line) that is topped by a short apical region within which torsional stress causes the two complementary strands (light lines) to separate. The positions of corresponding sensors have been randomized within the control regions of the two alleles of *gene-1* shown here.

duplications of reverse repeats, in our opinion, are to the control of gene action as base pair triplets are to protein synthesis; they are the building blocks of genetic control systems.

Possible advantages of dissimilarly arranged sequences of sensors in the control regions of allelic genes

Although CRICK assigned the pedestals and their recognition sites to globular DNA, the linearity of the DNA strand in these regions is in no way altered. Therefore, the pedestals of double-stranded DNA and their single-stranded sensor regions can be represented in single file as in Figure 1.

Should the number of sensors for a given gene be considerable, the farthest one may not be a trivial distance from the gene whose transcription it controls. Distance represents time and material. Using chromosomal puffs of Dipteran chromosomes as a model, it would appear that tens of minutes are required for puff formation (DUPRAW 1970, page 245); furthermore if transcription follows the DNA molecule from the sensor to the gene, considerable amounts of DNA must be transcribed into RNA before the gene acts.

The supposed interrelations of sensors, structural genes, and systems that call for gene action should be clarified here. Structural genes are located at different loci. An unspecified number of systems can call for gene products; any one system may call for the activation of few or many genes. No one gene need respond to all systems; different genes need not respond to the same number of systems. System *A*, for example, may call for the gene products of genes 1, 2, 4, 6 and 7; system *B* may call for 1, 3, 4, 5, and 6; system *C* may call for 2, 6, 7, 8, and 9. Looking at this example from a different viewpoint, gene 1 must possess sensors for *A* and *B*, 2 for *A* and *C*, 3 for *B*, 4 for *A* and *B*, 5 for *B*, 6 for *A*, *B*, and *C*, 7 for *A* and *C*, 8 for *C*, and 9 for *C*.

Because different genes possess different combinations of sensors, no system can dictate the position of its sensor within the sequence of sensors associated

with a given structural gene. At one locus sensor A may be adjacent to the structural gene but at another locus it may be the most remote of a long series of sensors. To a large extent, the position of a sensor within a series of sensors must be a matter of chance; idealized sequences of all sensors of all systems at all loci may be impossible.

Should the sensors of allelic control regions be distributed in identical sequences, the mean distance between a particular sensor and the gene it controls would be one half the total length of the control region. On the other hand, should sensors be free to move about within the control region of each allele so that allelic sequences are not identical (as shown in Figure 1), the mean distance between the closer of two corresponding sensors and the structural gene it controls would be only one third the total length of the control region (see FEDERER, STEELE and WALLACE 1967, for applicable calculations).

Were sensors arranged identically in allelic control regions, the response of one gene when called upon to act might be nearly instantaneous, that of others considerably delayed. The variation in the length of time between the initial call (say of system A in the earlier example) and the appearance of gene products of the different loci (genes 1, 2, 4, 6, and 7) would be large. Dissimilar arrangements of sensors within allelic control regions at each locus would make the responses more uniform. Consider a system, D, that calls for action on the part of 10 genes each of which has a control region consisting of 10 different sensors. If the sensors in allelic control regions were ordered in identical sequences, the probability that all 10 would be found no farther than 6th place from any structural gene is less than 0.01 ($=0.6^{10}$). On the other hand, if the positions of sensors within allelic control regions differ, then the probability that at least one D-sensor will be found no farther than 6th place from its structural gene at each of 10 loci is about 0.20 [$=(1-0.4^2)^{10}$].

The above example assumes that a system relying on a number of gene products will not function until all such products have been produced. Should it be necessary that these gene products be produced very nearly simultaneously, the dissimilar arrangement of sensors within allelic control regions also offers an advantage. Identical sequences of sensors in each of the 10 pairs of allelic control regions offer but one array of 10 distances from sensor to gene; this would result in a certain variance in the times required for the different genes to respond to the same call (that of system D). Dissimilar sequences in the control regions of allelic genes result, for the 10 genes under discussion, in 1024 ($=2^{10}$) arrays one of which will have a smaller variance than the remaining 1023. Thus, if it is important for genes, when called upon, to respond more or less simultaneously, dissimilarity in the positioning of sensors in allelic control regions greatly increases the probability that they will.

Thus, in summary, the dissimilar positioning of sensors within the control regions of allelic genes tends to reduce the time between the call for a gene product and its synthesis, would better synchronize the actions of sets of genes that are called upon for simultaneous action, and would probably reduce the absolute pool size of nucleotides needed for transcription. Right or wrong, these are the

reasons why CRICK's model, with the additional embellishments, appealed to us.

The dissimilar distribution of sensors within corresponding control regions raises problems with respect to crossing over. If their distributions are not identical, crossing over can lead to duplications and deficiencies; for example, a cross over near *E* (Figure 1) would lead to a loss of *B* or *D* from one recombinant chromatid and its duplication in the other. For the moment, this difficulty will be set aside in order to concentrate on intra-strand recombination. The resolution of the difficulty posed by inter-chromatid crossing over hinges on the elimination of recombinant chromatids carrying duplications or deficiencies of sensors. We shall argue later that paracentric inversions are one means by which this selective elimination of cross-over strands is achieved.

How pedestals might exchange position spontaneously

If, as we have argued above, dissimilar arrangements of pedestals and sensors within allelic control regions are advantageous to the individual, then a means by which these structures can exchange position should be found in the DNA fiber. Because of their unique properties with respect to chromosomal mechanics, tandem duplications and reverse repeats were examined with this role in mind. Neither one alone serves the purpose. Together, however, they do provide a mechanism by which sensors, without accompanying harm, become free to move about within the control region in which they are located.

The structure that meets the requirements mentioned above is shown below:

. . . a b c d 0^1 e f g h h′g′f′e′0^2 d′c′b′a′a b c d 0^3 e f g h h′g′f′e′0^4 d′c′b′a′. . .*

In this structure, 0's represent sensor regions while the material (a b c d, e f g h) on either side of each one forms its pedestal. Except for the sensors which must have specific conformations in order to serve their functions, the pedestals (following CRICK's suggestion) are essentially repetitive. Any pair of pedestals forms a reverse repeat; successive pairs, consequently, represent tandem duplications.

The possible role of reverse repeats in the movement of pedestals within control regions is illustrated in Figure 2. In this figure, a length of double-stranded DNA containing three pedestals and their single-stranded sensors is diagrammed in some detail. Six possible intra-strand recombinations (identified by number) involving pairs of pedestals have been illustrated. The first recombination leaves the sensors unaffected. In contrast, the second causes sensors *A* and *B* to exchange places; following this recombination the order of sensors would be *B A C*. Recombination #3, like #1, leaves the sensors unaffected but #4 exchanges the positions of *B* and *C*, thus leading to the sequence *A C B*. Any mechanism that permits adjacent sensors to exchange places without the loss of either one in effect permits sensors to take up any position within the control region. Through time, control regions that are constructed as shown in Figures 2 and 3 are fluid rather than static areas.

Two additional recombinations are shown in Figure 2. Unlike the first ones

* Our reverse repeat is equivalent to the molecular geneticist's "palindrome"; that is, DNA each of whose strands is internally complementary about a point of symmetry. Primes (′) indicating this complementarity have inadvertently been omitted from the figures.

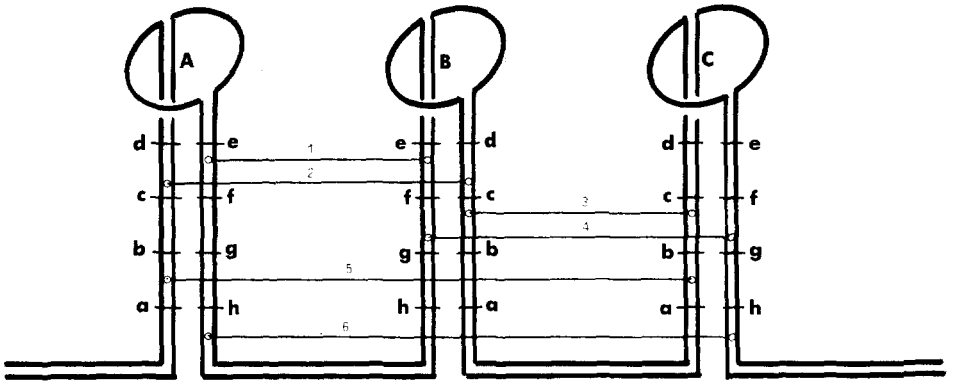


FIGURE 2.—A diagrammatic representation of three sensors and pedestals formed by the tandem duplication of reverse repeats. Double-stranded DNA is represented by double lines; the unpaired single strands of the sensor areas are shown as single lines. (See text for an account of the outcome of the six indicated intra-strand recombinations; see footnote, page 545).

discussed, numbers 5 and 6 involve non-adjacent pedestals and, unlike the others, lead to the loss of pedestals and sensors. Recombination #5 excises sensors *A* and *B* (together with their pedestals) from the control region; #6 excises *B* and *C*.

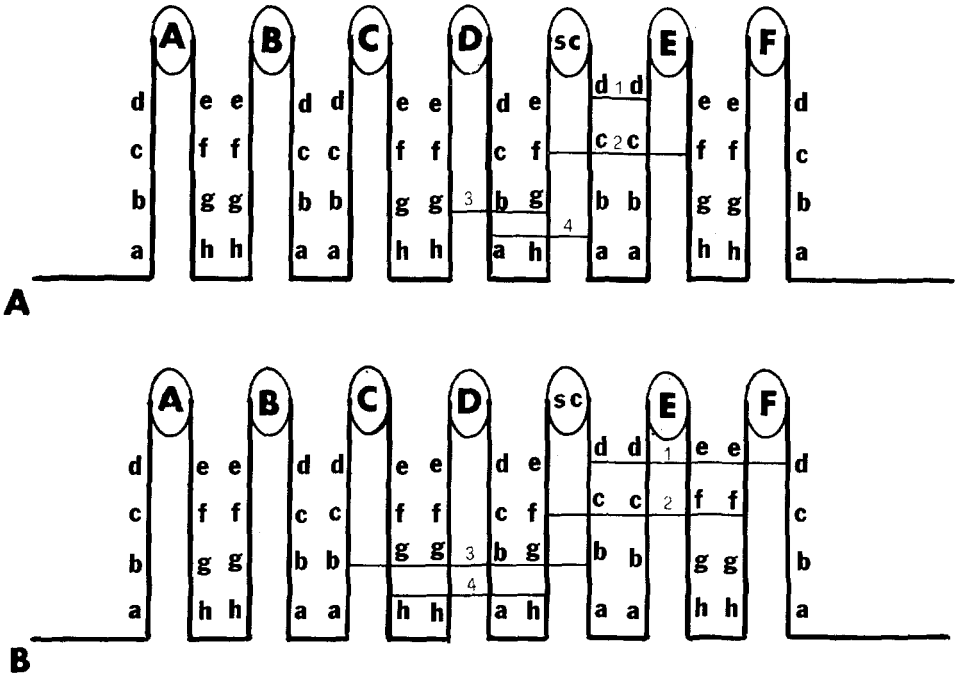


FIGURE 3.—Diagrams suggesting the consequences of intra-strand recombination in control regions within which the sequential pattern of reverse repeats has been disrupted; the outcomes are described in the text. A. recombination between adjacent sensors ("legitimate" recombination); B. recombination between non-adjacent sensors ("illegitimate" recombination).

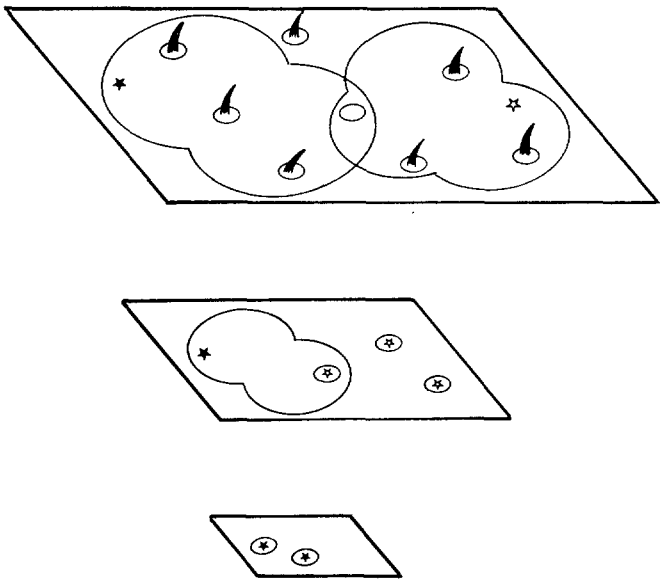


FIGURE 4.—A formal explanation of the mosaic dominance exhibited by two different *scute* alleles in Drosophila. Early to late developmental stages pass from bottom upwards. The solid stars represent the excision of the *sc* sensor of one allele in two cells (small ellipses) at one stage of development and its subsequent absence in descendants of those cells. The open stars represent the later excision of the *sc* sensor of the other allele in three cells, one of which is a descendant of one of the two earlier ones. Adult epidermal bristle-forming cells will form bristles only if they carry at least one *sc* sensor. Bristles that would be missing in adult flies homozygous for the two different *scute* alleles are the four falling within each of the two bounded areas.

Presumably the excision of sensors would represent serious mutational events and therefore recombinations #5 and #6 might be labeled "illegitimate" in contrast to the earlier "legitimate" recombinations. Although these designations are made following the revelation of the contrasting effects of the two types of recombination, the designations are not completely arbitrary. Presumably, histones and other materials must be stripped from DNA before it can replicate. Consequently, the DNA of adjacent pedestals would normally be exposed simultaneously and thus be able to recombine more often than DNA of non-adjacent pedestals.

(We speak here of the accidental excision of sensors as an undesirable, harmful act. Should the stripping of protein from DNA at various places be accurately controlled, however, the excision of unwanted sensors from control regions during development could itself be a valuable method of gene control. GALLY and EDELMAN (1972) have proposed a seemingly analogous mechanism for the genetic control of antibody production.)

Recombination within deficient control regions

Legitimate intra-strand recombination (that is, recombination between adjacent pedestals) within a properly constructed control region merely alters the

relative positions of adjacent sensors. If, however, one sensor is missing because of a prior mutational event, legitimate recombination may lead to the loss of still another sensor. In Figure 3 the sensor labeled *sc* does not fit the otherwise consistent pattern of reverse repeats. Four legitimate recombinations involving *sc* are shown in Figure 3A. It can be readily verified that recombination #1 leaves the sensors unaffected, #2 reverses the positions of *sc* and *E*, #3 excises sensor *D*, and #4 excises sensor *sc*. The presumed rarer illegitimate recombinations are shown in Figure 3B. In this case, recombinations #1 and #2 excise pairs of sensors, #3 exchanges the positions of *sc* and *C*, and #4 leaves the sequence unchanged.

The *sc* sensor of Figure 3, the one borne by the "odd" pedestal, is so labeled because of the similarity between recombinational events described here and the behavior of the scute mutations in *D. melanogaster*. The use of the label *sc* for the one sensor is not strictly accurate because the entire control region plus the structural gene would constitute the scute locus. The interruption of the reverse repeat pattern of the scute control region would produce the scute phenotype provided that the original anomaly would in turn cause the subsequent loss of the *sc* sensor during the development of the individual fly. Thus, the *sc* sensor of Figure 3 is the one responsible for the activation of this gene in bristle-forming cells. Cells that are expected to develop bristles will not do so if they lack sensor *sc*.

Figure 4 offers an explanation for the *mosaic dominance* that is characteristic of scute alleles. Individuals that are heterozygous for two scute alleles (sc^i/sc^j) lack only those bristles that are generally missing in both homozygotes (sc^i/sc^i and sc^j/sc^j). We assume that the time during development at which an *sc* sensor is excised from the aberrant control region depends upon its position within that region. The excision of *sc* from one of the two alleles in two different cells at an early developmental stage is represented by the solid stars in the lower diagram of Figure 4. The additional excision of *sc* of the other allele in three cells at a later developmental stage is represented by the open stars of the center diagram. Note that one of these three cells is a direct descendant of one of those which earlier had lost its other *sc* sensor. The outcome for the adult fly is a collection of epidermal cells (upper diagram) some of which have lost one *sc* sensor, some the other, a few both, and many neither. Only those cells in which both *sc* sensors are missing are unable to form bristles; hence the term *mosaic dominance*.

Because sensors are distributed in linear order along the DNA fiber, careful genetic mapping leads to a fine-structure map of the scute locus. Now, the claimed relationship between the position of *sc* within the control region, the stage at which it is excised, and the distribution of bristleless cells might easily suggest that subgenes at the scute locus are responsible for particular bristles in the adult fly (DUBININ 1932). Furthermore, the loss of the sensor that is adjacent to *sc*, an event about as probable as the loss of *sc* itself, could lead to mutant phenotypes whose responsible genes (alleles of scute) would appear to be within the collection of scute subgenes.

The excised pedestal and sensor

The preceding section dealt with events that might occur within a control

region consisting of tandemly duplicated reverse repeats that has lost one or more pedestals and sensors. The reverse repeat pattern, as we have seen, would be disrupted so that legitimate recombination, instead of being free of harm, would lead to the loss of still another sensor. The emphasis now shifts: What happens to the excised sensor?

Figure 5 follows the loss of a sensor (A) from an incomplete control region.

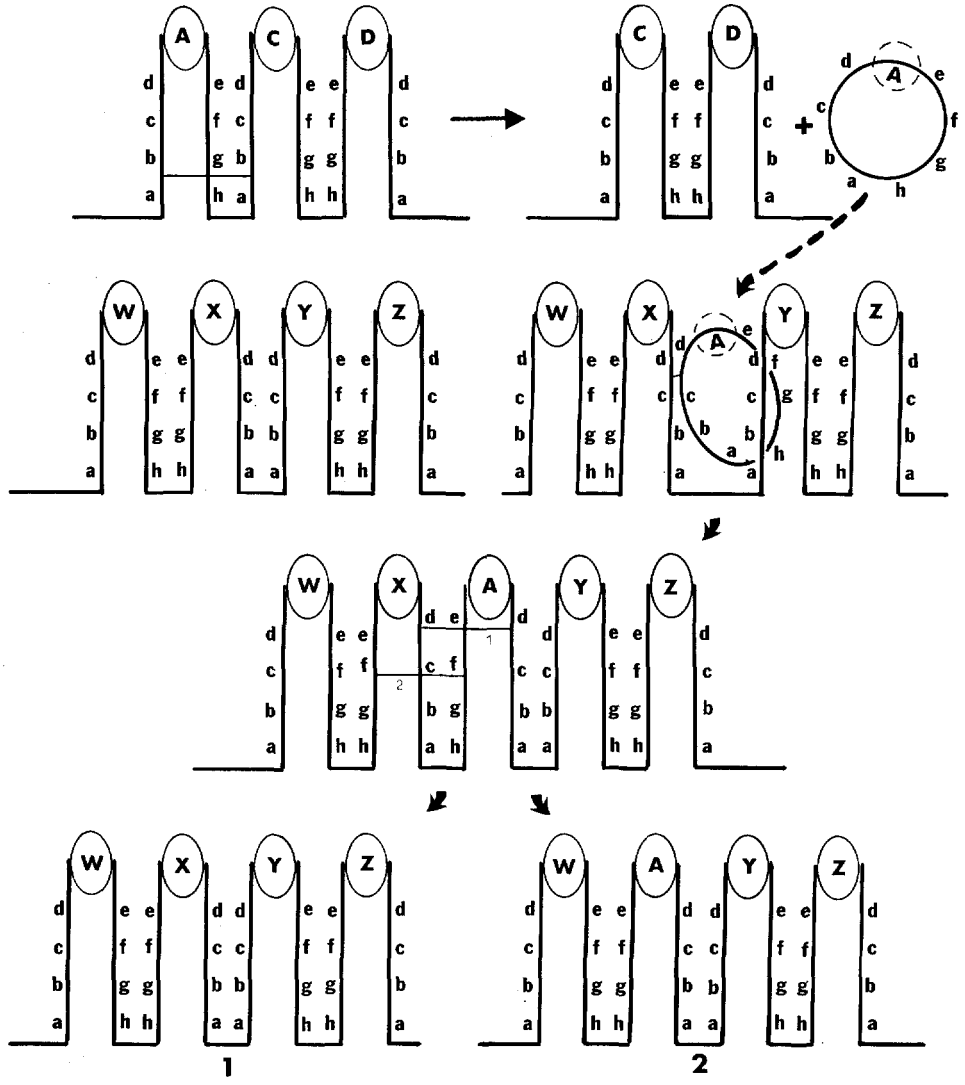


FIGURE 5.—A diagram that illustrates the stabilization of an unstable control region through the excision of sensor A. Released of torsional stress, A and its pedestal form an episome-like circle of double-stranded DNA. The free circle is shown entering a new and heretofore stable control region, an event that makes this region unstable. Stability of the W–Z region can be restored by the excision of (1) A or (2) X. (A and its pedestal need not be truly free-floating, of course, if the two recombinational events—A–C and A–X—occurred during a mispairing of the two control regions.)

A legitimate recombination excises *A* and simultaneously restores the stability of the two remaining pedestals. Because of the loss of *A*, the mutant allele may exhibit still another visible trait.

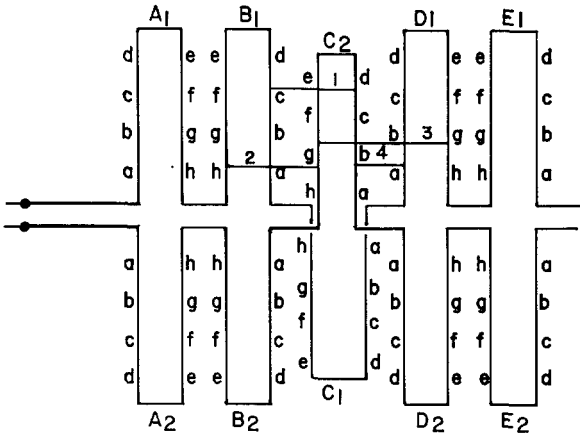
Pedestal and sensor *A*, a small circular fragment of DNA following its excision, is shown being incorporated into a second, previously stable control region (*W–X–Y–Z*). The incorporation is effected in the diagram by means of recombination in the region designated *c–d*. The result is the addition of a new sensor, *A*, to the *W–Z* control region. The gene that is normally controlled by systems whose calls come by way of sensors *W*, *X*, *Y*, and *Z* will now function under the direction of *A* as well and, consequently, may acquire a mutant phenotype. The new mutation will be unstable, however, because the reverse repeat pattern of the previously stable *W–Z* region has been disrupted. Legitimate recombination may excise either *A* (thus causing the gene to revert to its normal state) or *X* (thus stabilizing the new mutant phenotype).

The behavior of the excised pedestal described above is reminiscent of the transposition of control elements in maize described by McCLINTOCK in her classic studies (McCLINTOCK 1950, 1956, 1965). Through somatic transposition, maize plants carrying one *Ac* (*Activator*, one of the control elements studied by McCLINTOCK) on each of two chromosomes occasionally give rise to cells in which either both *Ac*'s are closely linked on one chromosome or both are missing (McCLINTOCK 1949). This transposition and many other anomalies described by McCLINTOCK can be accounted for by recombination in regions of imperfect sequences of reverse repeats. The outcomes of such recombinations are shown in Figure 6. As a result of recombination #1, sensor *C* is deleted from one chromosome and is inserted immediately beside its homologous counterpart, a move that mimics that described above for *Ac*. McCLINTOCK (1965) has summarized her conclusions concerning control elements. Many of the items she lists involve the loss, transposition, or insertion of control elements; these items, in our opinion, can be accounted for on the basis of reverse repeats (and the disruption of sequential reverse repeat patterns) as described here.

Predictions and correlations

The value of any model is measured not only by what it explains but also by what it predicts. The incorporation of CRICK's recognition sites into a matrix of tandemly duplicated reverse repeats has made sense of and has extended our interpretation of past observations. We have not attempted to take up these observations one by one; nevertheless, we might point out the following: A reversal in the relative positions of dissimilar sensors controlling the same structural gene would appear under standard mapping techniques as a reversal in the position of two genes; thus, our model offers a basis for re-examining findings such as those reported by LAUGHNAN (1955) and EMMERLING (1958).

The similarity between the postulated transpositions and other cytological aspects of the tandemly duplicated reverse repeats discussed here and the actual behavior of the control elements studied by McCLINTOCK (*op. cit.*) led to a search into the origins of controlling elements. Their origins lie primarily in material



1a ●–A_1 B_1 ✳ D_2 E_2

b ●–A_2 B_2 C_2 C_1 D_1 E_1

2a ●–A_1 ✳ C_2 D_2 E_2

b ●–A_2 B_2 B_1 C_1 D_1 E_1

3a ●–A_1 B_1 C_1 D_1 B_2 A_2–●

b E_1 C_2 D_2 E_2

4a ●–A_1 B_1 C_1 C_2 B_2 A_2–●

b E_1 D_1 D_2 E_2

FIGURE 6.—Inter-chromatid recombination between corresponding control regions each of which contains an "odd" member (C_1 and C_2) which disrupts the reverse repeat sequence. The outcomes of the four indicated recombinations are shown below the diagram. One outcome, commented on in the text, excises C_1 from its position in one strand and inserts it into the other. (The *a* recombinant strand in each instance begins with the upper centromere (–●–); the *b* strand either begins at the lower centromere or, if *a* is dicentric, depicts the acentric fragment.)

produced during a study of breakage-fusion-bridge cycles. The chromatid breakage-fusion-bridge cycle is illustrated in Figure 7; it is clear from the diagrams that through the repeated formation of dicentric chromatids, the breakage-fusion-bridge cycle generates tandemly duplicated reverse repeats. While it is true that the chromosome breakage-fusion-bridge cycle (see McCLINTOCK 1951, Figure 1) generates tandem duplications, these are often tandem duplications of material in which reverse repeats have already been formed; therefore, this cycle also generates tandem duplications of reverse repeats.

Transcription of DNA that consists of reverse repeats gives rise to reverse complementarity within the newly formed RNA. Assuming that heterogeneous nuclear RNA (HnRNA) is RNA transcribed from control regions of gene loci,

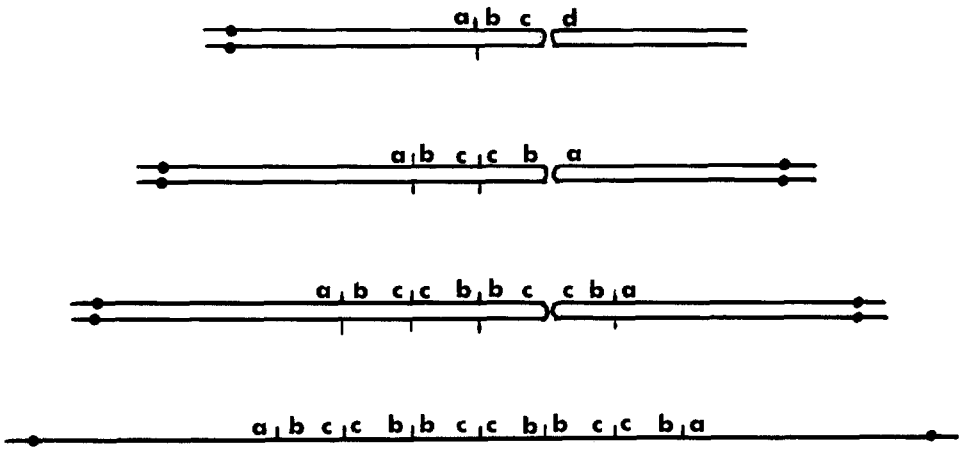


FIGURE 7.—An illustration of the generation of tandemly duplicated reverse repeats by the chromatid breakage-fusion-bridge cycle in maize. The dicentric that is formed by the topmost fusion is symmetric about *c*. The next dicentric is symmetrical about its fusion point and, therefore, is symmetrical for previously existing reverse repeats. The extra chromatin gained by the larger of the dicentrics consists for the most part of sequences of reverse repeats although not necessarily as regularly spaced as those illustrated.

one would predict that newly formed mRNA would be attached to additional RNA consisting of double-stranded hairpin loops. In agreement with this expectation, it appears (JELINEK and DARNELL 1972) that HnRNA does indeed contain double-stranded regions. Furthermore, it seems (DUNN and STUDIER 1973) that the double-stranded RNAase, ribonuclease III of *E. coli*, processes mRNA by destroying attached double-stranded RNA (ROBERTSON *et al.* 1968).

At some future time it should be possible to isolate the HnRNA that is transcribed with a particular mRNA. By restricting the study to times at which a specified locus is subject to call by a single system (as by a hormone or other inducing substance), our model predicts that inbred strains of genetically uniform individuals should produce (for this single locus and single call) HnRNA's of uniform length, but that this length should vary from strain to strain. Furthermore, the corresponding HnRNA of hybrid individuals should show components corresponding to those of the two parental inbred strains, with a possible excess of the shorter one.

Throughout this discussion, sensors have been depicted in the pedestal-receptor-site form suggested by CRICK. If these sensors were fixed in their positions within control regions, one might expect that each sensor would be asymmetrical in base composition, thus assuring that transcription would proceed along the correct strand in the direction of the structural gene. An asymmetrical sensor that initiated transcription on one DNA strand only, however, would not function properly if it were rotated; unfortunately, each shift of a sensor within a control region under our model rotates it as well. Consequently, we are forced to postulate that sensor regions must also be constructed as reverse repeats; rotation in this case would not affect their function.

Sensors that consist of reverse repeats pose problems of their own with respect to transcription: How can transcription be directed toward the structural gene? An inelegant possibility is that transcription proceeds in both directions from the sensor. Transcription of one strand leads eventually to the structural gene and to the transcription of mRNA; that of the other leads to a termination signal and ends with the production of HnRNA only. Although wasteful, this method would work. A large amount (about one half) of newly transcribed HnRNA would not be attached to mRNA.

A more elegant and less wasteful possibility would rely on signals encoded in DNA on either side of the reverse repeat area of the control region. As the polymerase molecule began DNA replication distal to the control region, it would encounter the first signal saying, in effect, "Coat the newly synthesized strand so that it cannot be transcribed." Having completed the replication of the first strand through the structural gene, the polymerase molecule would return on the complementary DNA strand. Between the structural gene and its control region (that is, proximal to the control region), it would encounter a second signal: "Forget signal one; coat me so that I cannot be transcribed." This hypothetical procedure would produce sensors whose exposed single strand of DNA would direct transcription toward the gene.

Although the second scheme is fanciful in the sense that no direct evidence in its support exists, it may explain the striking gene control exerted by reverse repeats generated by the breakage-fusion-bridge cycle. The two signals, remember, were said to lie on the proximal and distal sides of the control region; should the positions of the two signals be reversed, transcription of the gene would be impossible. A minute fragment of chromatin composed of several artificially generated tandemly duplicated reverse repeats might carry both signals in one of the repeated segments. Inserted before a structural gene in the reverse order, this segment would prevent transcription of the gene. If the segment were rotated as a result of recombination within the series of reverse repeats, the gene would be "turned on". The next rotation would turn the gene off once more. Abrupt alterations of this sort ("changes in state") are a frequent characteristic of the controlling elements in maize.

Still other predictions are possible. For example, the model claims that repetitive DNA should contain short, largely unique internal segments which, upon close examination, will themselves prove to be reverse repeats. Justification of the model entirely in molecular terms does not seem necessary, however. The model of gene control proposed by BRITTON and DAVIDSON (1969) is supported by a considerable mass of observational data; the embellishment provided by CRICK (1971) once more focuses attention on properties of the DNA fiber itself in a manner largely neglected since the classical work of WATSON and CRICK (1953); our own stress on the peculiarities in the inheritance of tandemly duplicated reverse repeats has placed control regions in a new light.

Inter-chromosomal recombination

Duplications and deficiencies of sensors may follow inter-chromosomal

recombination within control regions whose sensors are arranged in dissimilar sequences. Five inter-pedestal regions are illustrated in Figure 1; crossing over within any one of these would give rise to alleles with inadequate collections of sensors. The following paragraphs are addressed to this difficulty.

Homozygosis for an entire chromosome in Drosophila is almost always deleterious with respect to individual viability or fitness; the viability of such homozygotes is consistently lower than that of flies heterozygous for two different chromosomes (WALLACE and MADDEN 1953; DOBZHANSKY and SPASSKY 1953). The introduction of a small amount of heterozygosity into an otherwise homozygous background is beneficial not only if the variation is "natural" (VANN 1966) but also if it is artificially induced by radiation (WALLACE 1958; MARUYAMA and CROW 1974).

To cope with the difficulty posed by inter-chromosomal recombination, we begin by postulating that the chromosomes carried by individuals of a hypothetical population are all identical with respect to both the structural genes they carry and the disposition of sensors within allelic control regions. Proceeding from this base, each change in the sequence of sensors of this or that control region should be advantageous. If the probability of a crossover between the displaced sensors of a control region is small compared to the advantage the displacement confers on heterozygous individuals, discrepancies in sensor sequences will grow in number. At some point, however, the harm done by the gain and (especially) loss of sensors through recombination will equal the benefit that differences in sensor sequences confer; the result will be a stable equilibrium.

Should the equilibrium described in the preceding paragraph not exhaust the selective advantage that might possibly accrue to individuals through the dissimilar distribution of sensors in the control regions of still other gene loci, Drosophila has two procedures by which the extent of this type of heterozygosity can be increased. The first of these involves the inversion of gene sequences. Crossing over within the control region of a locus can lead to the loss of sensors in recombinant chromatids if the sensors are arranged in unlike sequences in the two alleles. Crossing over within a parcentric inversion, however, leads to the elimination of recombinant chromatids. Therefore, given (1) the existence of an inversion polymorphism and (2) the ubiquity of loci at which dissimilar arrays of sensors are selectively advantageous, one can imagine that control regions with dissimilar arrays of sensors might preferentially accumulate within the limits of paracentric inversions. Crossing over in the control region of any locus within these limits would be rendered harmless because the crossover strands would be relegated to non-functional polar bodies. Such an accumulation of dissimilar control regions within inverted regions would explain the observed coadaptation of gene arrangements that coexist within local populations (DOBZHANSKY 1948, 1950) and the tremendous hybrid vigor (three-fold differences in the fitnesses of structural heterozygotes and homozygotes are not uncommon) that frequently accompanies heterozygosity for naturally occurring inversions (WRIGHT and DOBZHANSKY 1946).

The notion that dissimilar allelic control regions accumulate within the limits

of paracentric inversions places the role of inversions in Drosophila populations in a new light: At present, inversion polymorphisms in Drosophila are said to be possible because crossing over within the inverted sequences of structurally heterozygous females does not lead to an appreciable egg mortality; the account presented here says that inversion polymorphisms exist because they lead to the harmless elimination of chromosomes that would otherwise be deficient for sensors in gene control regions at various loci.

Drosophila, and presumably other organisms, have a second procedure by which the proportion of loci containing dissimilar control regions might be maximized with minimum concomitant harm. A comparison of the genetic and cytological maps of Drosophila reveals that crossing over does not occur with equal frequency in all segments of the chromosome. Large map distances may represent relatively short chromosomal regions; long chromosomal regions may show little crossing over and hence occupy short regions on the genetic map. The concentration of dissimilar gene control regions in the latter chromosomal regions would greatly reduce the danger posed by recombination and, hence, increase the number of loci which could possess "heterozygous" control regions. The ideas presented here are readily testable.

On the evolution of control systems

Control regions comprising long series of sensors and pedestals could serve as an adequate, if somewhat inefficient, system leading to the proper expression of genes during development and later. A more efficient system would involve a hierarchy of controls such that genes in the first level can each evoke the operation of several in the second, each of which in turn can evoke the action of several loci of the third level, and so on.

The number of different sorts of sensors (and, hence, of sensor-activating substances) that are needed to operate hierarchal and non-hierarchal systems may serve as a measure of efficiency. In a hierarchal system the required number of unique sensors is the *sum* of the loci involved at all levels of control. Were a system as complex as that of the hierarchal system to operate without hierarchal levels, one sensor would be needed for each pathway of control in the hierarchal system; consequently, the necessary number would equal the *product* of the loci postulated for the different levels of the hierarchal system. For complex systems, the product would always exceed the sum.

The evolution of genic control from non-hierarchal to hierarchal systems would not require the development of a new basic mechanism; the reverse repeats that we have discussed would serve equally well in the construction of primitive non-hierarchal control systems and in that of sophisticated hierarchal ones. Thus, the evolution of the one type of system into the other does not call for the origin of new techniques and procedures of gene control. A slight modification of the primitive system adapts it for the more advanced system of control. It is interesting to speculate that organisms (amphibia, for example) possessing what appears to be an inordinant amount of DNA, DNA that cannot readily be explained by polyploidy, are those which have failed to evolve hierarchal control systems.

Neutralism versus *selectionism*

We now turn once more to matters of population genetics, the doorway through which we entered into this extended discussion of tandemly duplicated reverse repeats. At this moment we do not know whether or not selection favors dissimilar homologous control regions as we have argued. If these regions do in fact tend to be dissimilar, however, the effective heterozygosity in natural populations would be exceedingly high. WALLACE (1963a) pointed out that dissimilar control regions C_1, C_2, C_3, . . . at the A locus, for example, effectively convert this locus into a series of alleles A_1, A_2, A_3, . . . because the gene A can function only through its control mechanism. Thus, the dissimilarity in structural alleles revealed as gene-enzyme variation (LEWONTIN and HUBBY 1966) should be combined with the dissimilarity of control regions in order to estimate the total extent of genetic heterogeneity per locus.

With the exception of instances in which sensors have been accidentally lost, it seems that selection operating on their disposition within control regions must for the most part be density-dependent or, in WALLACE's (1968, 1970) terminology, "soft" selection. The importance of gene control concerns the efficiency with which metabolic pools are used, the speed with which genes react when called upon, and the degree to which products of different loci are synchronized for most efficient interaction. A multitude of homeostatic mechanisms exist which normalize development whenever possible; for example, in *D. melanogaster* both bobbed homozygotes and Minute heterozygotes survive although the larval development of the latter may require an additional 48 hours. Consequently, there is no need to look upon inefficient control mechanisms (these are not to be confused with control regions that lack specific sensors) as lethal mutations nor is there any need to assign to them a fixed probability of lethality. Instead, certain well-functioning individuals will—if they exist in a population—displace less efficient ones who, in the absence of this rigorous competition, might otherwise appear thoroughly normal.

Randomness of genetic differences is at times misinterpreted as evidence for the neutrality of such differences. Chromosomal inversions in the genus Drosophila should have served as a warning against this line of reasoning because, although chromosomes differing in gene sequence often exhibit heterosis, the overall distribution of naturally occurring inversions within the genus Drosophila with respect to either size or position on the chromosome is essentially random (FEDERER, STEELE and WALLACE 1967; LEVINS and VAN VALEN 1968; and BAUMANN, personal communication, 1963). Both with respect to the position of sensors within control regions and in the origin of differences between sensors (that is, with respect to recognition sites), a random difference might easily be a selectively favored difference. At times, a genetic difference (*any* difference, not a particular difference) is an appropriate response to a selective challenge. On the other hand CRICK's pedestal serves an essentially structural function within the DNA fiber; consequently, as long as the integrity of the reverse repeat pattern is largely maintained the particular base pairs constituting the pedestals may be unimportant.

In addition to colleagues at Cornell University who have both listened and explained with great patience, the senior author wishes to acknowledge conversations with JAMES C. KING in which some of these ideas were sown—ideas which found no resting place within the genetic knowledge of the time (1950's).

Note added in proof: HUGH ROBERTSON and ELIZABETH DICKSON, at the 1974 Brookhaven Symposium in Biology, presented a paper entitled "RNA processing and the control of gene expression" that contained molecular predictions corresponding to, complementing, and extending those which we have described here.

LITERATURE CITED

AYALA, F. J., 1966 Evolution of fitness. I. Improvement in the productivity and size of irradiated populations of *Drosophila serrata* and *Drosophila birchii*. Genetics **53**: 883–895.

BLAYLOCK, R. J. and H. H. SHUGART, JR., 1972 The effect of radiation-induced mutations on the fitness of Drosophila populations. Genetics **72**: 469–474.

BRITTON, R. J. and E. H. DAVIDSON, 1969 Gene regulation for higher cells: A theory. Science **165**: 349–357.

CRICK, FRANCIS, 1971 General model for the chromosomes of higher organisms. Nature **234**: 25–27.

DOBZHANSKY, TH., 1948 Genetics of natural populations. XVIII. Experiments on chromosomes of *Drosophila pseudoobscura* from different geographic regions. Genetics **33**: 588–602.
——, 1950 Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. Genetics **35**: 288–302.

DOBZHANSKY, TH. and B. SPASSKY, 1953 Genetics of natural populations. XXI. Concealed variability in two sympatric species of Drosophila. Genetics **38**: 471–484.

DUBININ, N. P., 1932 Step alleomorphism and the theory of centres of the gene, *achaete-scute*. J. Genet. **26**: 37–58.

DUNN, J. J. and F. W. STUDIER, 1973 T7 early RNAs and *E. coli* ribosomal RNAs are cut from large precursor RNAs in vivo by ribonuclease III. Proc. Natl. Acad. Sci. U.S. **70**: 3296–3300.

DUPRAW, E. J., 1970 *DNA* and Chromosomes. Holt, Rinehart and Winston, New York.

EMMERLING, M. H., 1958 An analysis of intragenic and extragenic mutations of the plant color component of the R^r gene complex in *Zea mays*. Cold Spring Harbor Symp. Quant. Biol. **23**: 393–407.

FALK, RAPHAEL, 1961 Are induced mutations in Drosophila overdominant? II. Experimental results. Genetics **46**: 737–757.

FEDERER, W. T., R. G. D. STEELE and B. WALLACE, 1967 Mathematical models for chromosomal inversions. Genetics **55**: 783–795.

GALLY, J. A. and G. M. EDELMAN, 1972 The genetic control of immunoglobin synthesis. Ann. Rev. Genet. **6**: 1–46.

JELINEK, W. and J. E. DARNELL, 1972 Double-stranded regions in heterogeneous nuclear RNA from Hela cells. Proc. Natl. Acad. Sci. U.S. **69**: 2537–2541.

LAUGHNAN, JOHN, 1955 Structural and functional bases for the action of the *A* alleles in maize. Am. Naturalist **89**: 91–104.

LERNER, I. M., 1954 *Genetic Homeostasis*. Oliver and Boyd, Edinburgh.

LEVINS, R. and L. VAN VALEN, 1968 The origins of inversion polymorphisms. Am. Naturalist **102**: 5–24.

LEWONTIN, R. C. and J. L. HUBBY, 1966 A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics **54**: 595–609.

MARUYAMA, T. and J. F. CROW, 1974 Heterozygous effects of X-ray induced mutations on viability of *Drosophila melanogaster*. (Mut. Res., in press)

MCCLINTOCK, B., 1949 Mutable loci in maize. Carnegie Inst. Wash. Yearbook **48**: 142–154. ——. 1950 The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. **36**: 344–355. ——, 1951 Chromosome organization and genic expression. Cold Spring Harbor Symp. Quant. Biol. **16**: 13–47. ——, 1956 Controlling elements and the gene. Cold Spring Harbor Symp. Quant. Biol. **21**: 197–216. ——, 1965 The control of gene action in maize. Brookhaven Symp. Biol. **18**: 162–164.

MUKAI, T. and I. YOSHIKAWA, 1964 Heterozygous effects of radiation-induced mutations on viability in homozygous and heterozygous genetic backgrounds in *Drosophila melanogaster*. Japan J. Genet. **38**: 282–287.

MULLER, H. J. and R. FALK, 1961 Are induced mutations in Drosophila overdominant? I. Experimental design. Genetics **46**: 727–735.

ROBERTSON, H. D., R. E. WEBSTER and N. D. ZINDER, 1968 Purification and properties of ribonuclease III from *E. coli*. J. Biol. Chem. **243**: 82–91.

VANN, E. G., 1966 The fate of x-ray induced chromosomal rearrangements introduced into laboratory populations of *Drosophila melanogaster*. Am. Naturalist **100**: 425–449.

WALLACE, B., 1957 The effect of heterozygosity for new mutations on viability in Drosophila: A preliminary report. Proc. Natl. Acad. Sci. U.S. **43**: 404–407. ——, 1958 The average effect of radiation-induced mutations on viability in *Drosophila melanogaster*. Evolution **12**: 532–556. ——, 1959 The role of heterozygosity in *Drosophila* populations. Proc. 10th Intern. Congr. Genet. **1**: 408–419. ——, 1963a Genetic diversity, genetic uniformity, and heterosis. Can. J. Gen. Cytol. **5**: 239–253. ——, 1963b Further data on the overdominance of induced mutations. Genetics **48**: 633–651. ——, 1968 Polymorphism, population size, and genetic load. 87–108. In: *Population Biology and Evolution*. Edited by R. C. LEWONTIN. Syracuse University Press, Syracuse, New York. ——, 1970 *Genetic Load: Its biological and conceptual aspects*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

WALLACE, B. and C. MADDEN, 1953 The frequencies of sub- and supervitals in experimental populations of *Drosophila melanogaster*. Genetics **38**: 456–470.

WATSON, J. D. and F. H. C. CRICK, 1953 The structure of DNA. Cold Spring Harbor Symp. Quant. Biol. **18**: 123–131.

WRIGHT, S. and TH. DOBZHANSKY, 1946 Genetics of natural populations. XII. Experimental reproduction of some of the changes caused by natural selection in certain popluations of *Drosophila pseudoobscura* Genetics **31**: 125–156.

Corresponding editor: B. H. JUDD