Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. — **Source link** ⬏

Roy J. Britten, Eric H. Davidson

Related papers:

- Gene Regulation for Higher Cells: A Theory

- Evolution at two levels in humans and chimpanzees

- Repeated Sequences in DNA

- Selfish DNA: the ultimate parasite

- Selfish genes, the phenotype paradigm and genome evolution.

REPETITIVE AND NON-REPETITIVE DNA SEQUENCES AND A SPECULATION ON THE ORIGINS OF EVOLUTIONARY NOVELTY

By Roy J. Britten * and Eric H. Davidson †

* Department of Terrestrial Magnetism,
Carnegie Institution of Washington,
Washington, D.C. 20015

† Division of Biology, California Institute of Technology,
Pasadena, Calif. 91109

ABSTRACT

Recent experimental information on DNA sequence repetition is reviewed, and the significance of both repetitive and non-repetitive sequence considered. Included are a summary of data on the distribution of genome sizes in animals, new experiments on interspecific DNA homology, the distribution of sequence frequencies, and the interspersion of repetitive sequences within the genome. Aspects of the process of evolution are considered at the level of change in the DNA. The process by which novel structure and function could have arisen during evolution are considered speculatively in terms of the authors' gene regulation theory (Britten and Davidson, 1969).

INTRODUCTION

IN RECENT years experiments have shown that the genome of higher organisms contains a large amount of order which is manifested as repeated nucleotide sequences in the DNA (Britten and Kohne, 1968 a, b). Recently we have considered the possible implications of this order for the mechanism of control of genetic activity and have constructed a model for gene regulation (Britten and Davidson, 1969; also see below). Here we consider some of the implications for the processes of evolution and examine in more detail some of the experimental evidence which

has led to the development of the regulation model.

In the first four sections we take up non-repetitive DNA and the term "complexity"; the growth of the genome during evolution; the origin and divergence of repeated sequences; and their arrangement in the genome. In the final sections we shall make a proposal for the origin of evolutionary novelty and discuss the evolution of regulatory processes. The purpose, as in our previous papers (Britten and Davidson, 1969; Davidson and Britten, 1971) is to construct a conceptual scheme which can be tested. We hope, naturally, that an

111

experimental confrontation will then lead to new factual knowledge as well as theoretical understanding.

In this paper we follow the view that major events in evolution require significant changes in patterns of gene regulation. These changes most likely consist of additions of novel patterns of regulation or the reorganization of pre-existing patterns. We feel that, generally, the appearance of new structural (producer) genes represents a minor part of the changes involved. A sort of syllogism highlights this view: metazoan organization arises through cellular ontogeny; ontogeny results from the operation of genetic regulatory programs; major events in metazoan evolution consist of changes in organization; thus the understanding of major events in evolution requires the examination of the origin of novel programs of gene regulation.

Some support for the view that patterns of gene regulation are central to evolutionary change can be derived from the similarity in a sample of structural gene functions between simple and complex organisms. The only sample of structural gene functions about which information exists are the enzymes. For example, 93 per cent of the enzyme activities listed in the I.E.C. tables are known to be present in both prokaryotes and mammals (see Appendix I for a tabulation). Similarly, biosynthetic pathways such as those for sterol and heme synthesis, among others, are already present in protozoa (Keilin and Ryley, 1953; Mallory, Gordon, and Conner, 1963). Similar sets of chemical tools (enzyme activities), however, are used in systems of greatly increased complexity in the higher forms. The implication is that the major differences are in the programs of gene regulation and thus in the resulting patterns of cellular differentiation. If the bacteria represent the level of complexity from which higher forms evolved, the course of the evolution of the higher forms must have required the oft-repeated addition of new regulatory mechanisms.

To illustrate the link between ontogenic gene regulation and the physiological and morphological reality, consider briefly the ontogeny of the pancreas. The chordate pancreatic rudiment develops as an outpocketing of a certain region of the gut, close to that forming the liver rudiment. Experimental analysis (Rutter, Wessells, and Grobstein, 1964; Wessells and Wilt, 1965) has shown clearly that inductive interaction between the pancreatic rudiment and certain mesenchymal cells is required in order for pancreatic differentiation to occur. If gene activity in the pancreatic rudiment during the early stages of the induction process is blocked with actinomycin, however, the rudiment fails to differentiate. Pancreatic differentiation involves not only the synthesis of a variety of special proteins not found in other cells, but the construction of an elaborate cytoplasmic membrane system which serves as the site for the polyribosomes synthesizing these proteins. Special mechanisms exist as well for the aggregation of the enzymic protein products in granules, and for the release of these granules upon appropriate environmental stimulus. Evidently many genes must be activated in concert in order to give rise to the differentiated pancreatic cell type. We postulate batteries of genes for pancreas properties which are under the control of genomic elements capable of responding to the mesenchymal inducing substance(s). The presence of the pancreas in chordates depends on the presence of the necessary ontogenic programs of gene regulation. Such programs are clearly built into the genome, since they are heritable.

Our theory of gene regulation provides a model of the way in which gene activities may be coordinated in higher cell genomes, and provides a framework for considerations of the nature of change in the regulatory programs. It appears that the model, in addition, contains some testable implications about the kinds of change that may occur. Since the model plays a central role in the argument of this paper, we include here a brief recapitulation of its elements and a definition of the terms used to describe them. The elements of the model taken together appear to have the potentiality of establishing a pattern of gene regulation which determines a particular cell state, and probably are sufficient to establish an orderly process of development (Davidson and Britten, 1971) leading to the full set of cell types and states of an organism.

*Recapitulation of the Model of
Gene Regulation*

Five elements are used in the model. Genes which specify cellular products such as enzymes (i.e., are regulated rather than regulatory in nature) are termed *producer* genes. A set of producer genes whose products carry out a closely related set of functions is termed a *battery*. An example of a battery would be the producer genes coding for the group of liver enzymes required for purine synthesis, which might well be activated simultaneously. The genome is visualized to be in a state of repression except for those genes which are specifically activated. The activity of the producer genes of a battery is controlled through the interaction of particular diffusable regulatory molecules with a DNA sequence termed the *receptor* sequence (or gene) contiguous to each producer gene of the battery. Therefore, the genes of a battery may be located at a distance from each other, and even occur on different chromosomes. The diffusible regulatory molecules of the model are derived from *integrator genes*. These regulatory molecules may be the RNA transcribed from the integrator sequences or they may be proteins derived by translation of such RNA molecules. In either case we refer to the RNA transcribed from the integrator genes as *activator RNA*, since it bears information to be used for gene activation.

There are integrator genes which are transcribed in response to substances arising elsewhere, such as those hormones which affect gene activity. These substances are postulated to interact with a *sensor* structure adjacent to the integrator genes and thereby initiate transcription. Since such substances do not in general have an affinity for particular DNA sequences, the sensor structures must include macromolecules serving as specific binding sites. Chromatin proteins which bind estradiol are good candidates. The reader is referred to the previous description (Britten and Davidson, 1969) of the regulation model for detailed discussion. The following definitions are excerpted from that paper.

Gene. A region of the genome with a narrowly definable or elementary function. It need not contain information for specifying the primary structure of a protein.

Producer gene. A region of the genome transcribed to yield a template RNA molecule or other species of RNA molecules, except those engaged directly in genomic regulation. We are using this term in a manner analogous to that in which the term "structural gene" has been used in the context of certain bacterial regulation systems. Products of the producer gene include all RNA's other than those exclusively performing genomic regulation by recognition of a specific sequence. Among producer genes, for example, are the genes on which the messenger RNA template for a hemoglobin subunit is synthesized, and also the genes on which transfer RNA molecules are synthesized.

Receptor gene. A DNA sequence linked to a producer gene which causes transcription of the producer gene to occur when a sequence-specific complex is formed between the receptor sequence and an RNA molecule called an activator RNA. In this model we do not wish to specify a mode of action for the receptor gene — that is, the nature of the molecular events occurring between the DNA, histones, polymerases, and so forth, present in the receptor complex. This model is concerned primarily with interrelations among the DNA sequences present in the genome. Repetitive DNA sequences involved in regulation have also been invoked by Georgiev (1969), who has suggested that messenger RNA is transcribed together with adjacent repetitive sequences.

Activator RNA. The RNA molecules which form a sequence-specific complex with the receptor genes linked to producer genes. The complex suggested here is between native (double-stranded) DNA and a single-stranded RNA molecule.

The role proposed for activator RNA could well be carried out by protein molecules coded by these RNA's without changing the formal structure of the model. Decisive evidence is lacking for higher cells, and we have chosen the simpler alternative. As the discussion of the evolutionary implications of this model will indicate, however, the probability of formation of new batteries of genes in evolution appears to differ greatly between these two alternatives. *Integrator gene.* A gene whose function is the

synthesis of an activator-RNA. The term integrator is intended to emphasize the role of these genes in leading, by way of their activator RNA's, to the coordinated activity of a number of producer genes. A set of linked integrator genes is activated together in response to a specific initiating event, resulting in the concerted activity of a number of producer genes not sharing a given receptor gene sequence.

Sensor gene. A sequence serving as a binding site for agents which induce the occurrence of specific patterns of activity in the genome. Binding of these inducing agents is a sequence-specific phenomenon dependent on the sensor gene sequence, and it results in the activation of the integrator gene or genes linked to the sensor gene. Such agents include, for example, hormones and other molecules active in intercellular relations as well as in intracellular control. Most of them will not bind to sensor gene DNA, and an intermediary structure such as a specific protein molecule will be required. This structure must complex with the inducing agent and must bind to the sensor gene DNA in a sequence-specific way.

Battery of genes. The set of producer genes which is activated when a particular sensor gene activates its set of integrator genes. A particular cell state will usually require the operation of many batteries.

While the model clearly suggests a function for repeated DNA sequences, a more general point arises. It appears that the existence of gene control functions in the cell (which can hardly be denied) logically requires the existence of repeated DNA sequences in the genome. The only way in which this implication could be totally avoided would be to propose that all gene regulation operates through a kind of "falling domino" process in which each gene activates or represses a single other gene. If, on the other hand, one regulatory gene is assumed to affect more than one other regulated gene, regions of sequence similarity which provide the molecular basis for the recognition process are implied among the regulated genes. This argument in itself suggests that some part of the transcription from unrepeated sequences which is experimentally observed is likely to be involved in cellular control processes. Since evolution of animal forms must have depended

heavily on the evolution of such regulatory systems, an approach to understanding the mechanisms of evolution would be through study of the organization of the genome. Thus we begin with a summary of the current state of knowledge regarding the occurrence of repetitive and non-repetitive sequences in animal DNA (Bolton et al., 1966; Britten, 1969a, b; Britten and Kohne, 1967, 1968a, b, 1969; Britten and Smith, 1970; Corneo, Ginelli, and Polli, 1970a, b; Davidson et al., 1971; Flamm, Walker, and McCallum, 1969, 1971; Kohne, 1970; Laird and McCarthy, 1969; Laird, McConaughy, and McCarthy, 1969; McConaughy and McCarthy, 1970a, b; Ullman, 1970; Wetmur and Davidson, 1968).

NON-REPETITIVE DNA AND GENETIC COMPLEXITY

The term "complexity" is useful to express the amount of diverse DNA sequence in a given preparation. The complexity is an easily defined concept if no sequences are present in more copies than others. "Complexity" is then defined simply as the number of nucleotide pairs present in a single set of all the diverse sequences. In other words, the complexity is equal to the genome size as long as repeated sequences are absent. If repeated sequences are present an obvious extension is to count all of the nucleotide pairs in the single-copy sequences plus all the nucleotide pairs in a single copy of each of the repeated sequences. This is an apparently simple definition. However, the repeated sequences are rarely, if ever, identical to each other. Thus, counting the nucleotide pairs in a single copy only of a set of sequences does not give a measure of the total sequence diversity. The biological significance of the divergence among repeated sequences is not known. It is likely that some of the repeated DNA now exhibits relationships among its sequences only because of past events of replication whereas other sequence homologies are probably functional. No direct evidence is available that shows that the relationships among the members of the set of sequences are actually used by the organism, although there are theoretical reasons (Hood et al., 1971; Britten and Davidson, 1969) for believing that they may be.

A very wide range of degrees of similarity

(or difference) occurs among the members of most sets of related sequences. Thus, the division between single copy and repetitive DNA, for any given genome, depends on the precision of relationship defined by the experimental criterion applied in the measurement of the reassociation of the DNA. As the criterion is lowered, more and more DNA appears to be repetitive. In the limiting case, random associations between unrelated DNA sequences occur. At the other extreme, when a stringent criterion of acceptable relationship is established, little repeated DNA is observed. In the case of the calf genome, which has been examined in some detail, when the criterion temperature is set about 8° C below the native T_m, precise pairs are not rejected to a significant extent. However, most of the known relationships among repeated sequences are rejected at this high criterion, and less than 10 per cent of the genome appears repetitive. On the other hand, if the criterion temperature is 36° C below the native T_m, then 55 per cent of the genome appears to be repetitive. Therefore, not only does the apparent amount of repetitive DNA depend on the conditions of observation, but it is possible that most of the sequences in the DNA of higher organisms exhibit some degree of relationship to many other sequences. In the cases of salmon (Britten and Kohne, 1968a), wheat (Bendich and McCarthy, 1970), and the urodele *Necturus* (Strauss, 1971), more than 80 per cent of the gneome appears to be repetitive at the customary experimental criteria.

Most of the available experimental data, including those cited in this paper, were obtained at a criterion temperature of 60° C (at 0.18 M Na+), about 25° below the T_m for precisely paired sequences (or most native higher-organism DNA). This criterion was originally chosen because it yields an optimal rate for the reassociation of precisely matching sequences. A fair precision of base pairing is required at this criterion, and the DNA of very distantly related organisms exhibits little sequence relationship in cross-species measurements (Martin and Hoyer, 1966). The relationship between the thermal stability of duplexes and the fraction of unmatched base pairs is not yet fully established. Current measurements (Ullman,

1970) suggest, however, that a 25° C reduction in melting temperature could result from as little as 15 per cent of non-conplementary bases. The 60° C, 0.18 M Na+ conditions represent a practical compromise for the exploration of repeated DNA. However, other conditions, particularly less stringent criteria, may supply unexpected insights into the history and patterns of occurrence of repeated DNA.

The most direct approach towards an assessment of the function of repeated or non-repeated DNA is through measurements of transcription of RNA from these sequence fractions. It is known that repeated sequences are transcribed into RNA, although it is not known how many of the individual member sequences of a set are transcribed. The complexity of the DNA which is observed to be transcribed yields a lower limit for the functional complexity of the genes expressed in any given circumstance. For newborn mice (Gelderman et al., 1969, 1971) it appears that more than 12 per cent of the non-repetitive sequences in the genome are transcribed, and thus the functional complexity of the mouse genome at this stage is at least 4×10^8 nucleotide pairs. This very large number is not yet interpretable, since if all of this RNA were messenger it would code for nearly a million different hemoglobin-sized proteins. Other measurements (Davidson and Hough, 1971) indicate that the RNA stored in the oocytes of *Xenopus* represents at least 1.2 per cent of the non-repetitive DNA. The complexity of this RNA is such that it could code for 40,000 different hemoglobin-sized proteins. Thus, it appears that a significant portion of the potential information content of these genomes is actually expressed. It might be noted that in the *Xenopus* oocyte many more copies are present of repetitive sequence transcripts than of non-repetitive sequence transcripts. This observation suggests a functional distinction between the repetitive and non-repetitive sequences and of course for the RNA transcribed from them.

It should be noted that physical separation of repeated from non-repeated DNA is limited in principle by the fact that unless very special techniques are used, one member of each set of related sequences will remain in the non-repeated fraction. This does not usually repre-
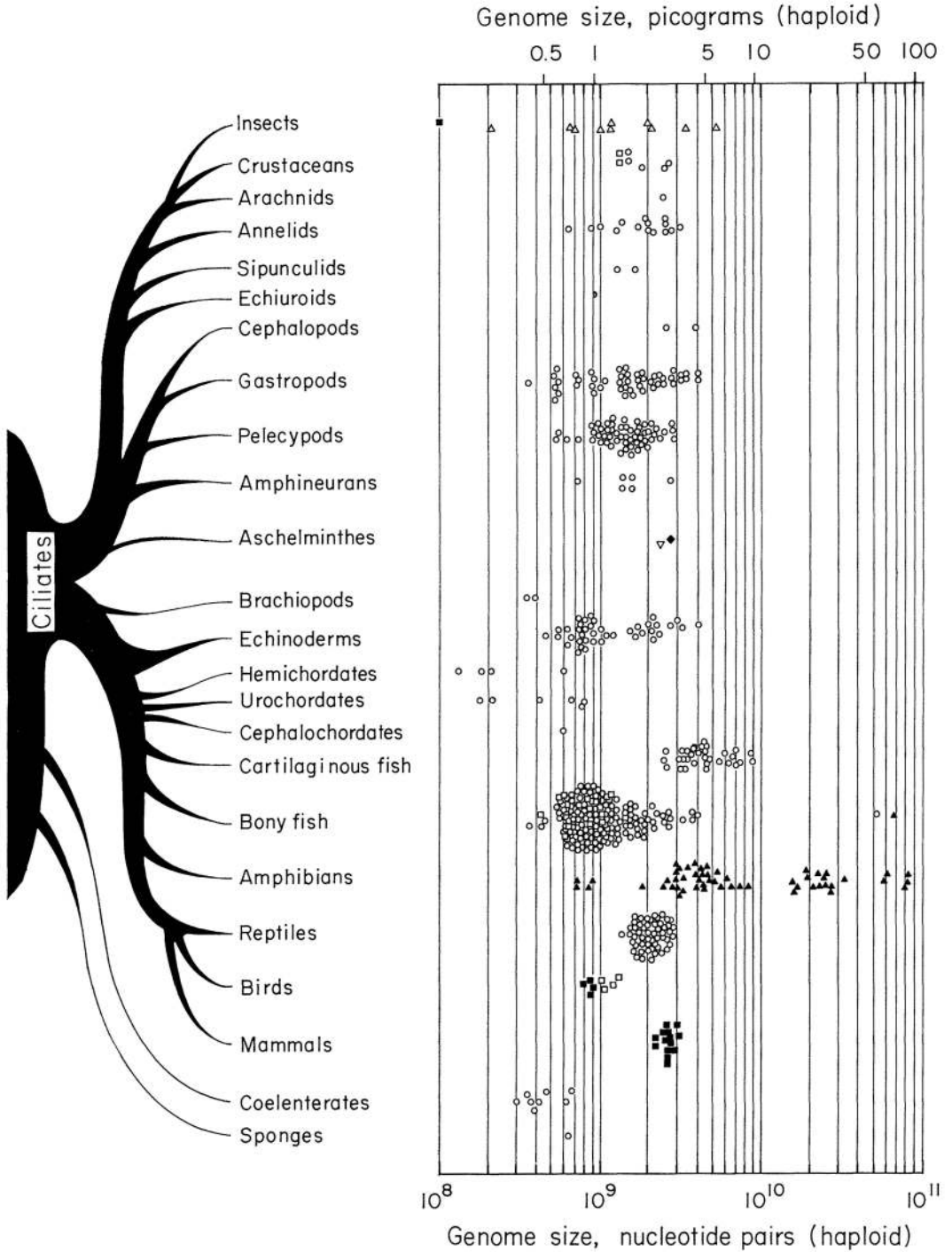
Fig. 1. Distribution of Genome Size in Animals.

Arranged to bring out the striking similarities in the range of genome sizes present in distantly related groups. The data have been assembled from many sources, as indicated in the list of symbols. Ralph Hinegardner has made the majority of these measurements and has very kindly given us permission to

sent a significant fraction of the non-repeated DNA (generally 1 per cent or less), nor would it contribute appreciably to the total complexity.

EVOLUTIONARY GROWTH OF THE GENOME

Fig. 1 summarizes the available data on genome size. It is clear that large changes in genome size have occurred during evolution. The figure shows that a range of genome sizes exists at many levels of biological organization. The pattern of occurrence of non-repeated sequence has been studied most extensively in amphibians (Davidson and Hough, 1969, 1971; Straus, 1971). Table 1 presents data drawn mainly from these studies, which indicate a higher proportion of repetitive sequence present in amphibians of larger genome size. Nonetheless, as measured at 60°C (0.18M Na+) the complexity of these larger genomes is the same as or greater than that of more advanced chordates, such as mammals. This high degree of complexity can be regarded as the consequence of divergence from the enormous repetitive sequence families observable in these genomes. Such a view suggests that not much of the apparent complexity of the *Ambystoma* genome, for example, is actually utilized. The similarity of biological organization between amphibians with large genomes and amphibians of small genome size also leads to this conclusion. The same argument applies to the utilization of repetitive sequences.

The type of replicative event responsible for the presence of repetitive DNA is of course unknown. However, several possibilities can be suggested. The evidence indicates that large-scale multiplication of particular sequences has occurred many times during evolution. These events are responsible for much of the sequence repetition observed and, to an unknown extent have contributed to much of the change in genome size that has occurred. Limited or piecewise sequence duplication is also known to have occurred during the evolution of particular structural genes, e.g., the hemoglobin genes. This phenomenon probably contributes little to the large-scale changes in genome size shown in Fig. 1, however, since the evidence indicates that not many copies of such structural genes are likely to be present and these genes may amount to only a small part of the genome. Large-scale multiplication of genomes by polyploidization has also been authenticated for several groups, among which the amphibians again provide some of the clearest examples (e.g., Beçak, 1969). Polyploidization and related events at the chromosome level, such as chromosome duplication, might figure prominently in the general phenomenon of genome growth. Ancient events of polyploidization, followed by sequence divergence, could give rise to at least some of the non-repetitive sequence in genomes such as those described in Table 1. Large-scale multiplication of particular short sequences, however, is probably the source of most (or all) of the repetitive sequence shown in Table 1, since the number of copies of these sequences is too large to have arisen by polyploidization. At present, it is clearly impossible to evaluate the relative importance to evolutionary genome growth of these replicative mechanisms. There is also evidence that reduction in genome size has sometimes occurred in the course of evolution (Hinegardner, 1968; pers. commun.), and the relative significance of deletion and growth remains to be determined.

ORIGIN AND DIVERGENCE OF REPEATED SEQUENCES

Large populations of repeated DNA sequences appear to be universally present in the genomes of eukaryotes above the fungi. Sets of related but divergent sequences occur in families with 20 to 2,000,000 members. Evidence (Rice, 1971) indicates that new families of repeated DNA sequences have been incorporated

include many unpublished observations. The accuracy of all the measurements plotted is not comparable. For example, some experts believe that there is little or no range of genome sizes among the mammals and that the spread shown is due to experimental error. It is obvious that there are many such questions to be resolved and many areas to be filled in, but the major aspects of genome size distribution are probably revealed by the available data. Phyla for which we were unable to locate genome size data, for example the flatworms, are omitted from the evolutionary tree.

Key: (◯) Hinegardner, R. (1968 pers. commun.); (▢) Mirsky and Ris (1951); (◆) Fairbairn (pers. commun.); (▽) Nigon and Bouvet (1955); (▲) Sexsmith (1968); (■) Sober (1968); (△) Bier, Kunz, and Ribbert (1969); Ribbert and Bier (1969); (●) Dawid and Brown (1970).

TABLE 1

Some amphibian genomes *

ORGANISMS	GENOME SIZE		PERCENTAGE OF GENOME WHICH IS NON-REPETITIVE	COMPLEXITY (NTP)
	PICOGRAMS (PG)/ HAPLOID GENOME	NTP		
Scaphiopus	1.2	1.1×10^9	60	0.7×10^9
Engystomops	2.7	2.5×10^9	50	1.2×10^9
Xenopus	3.0	2.7×10^9	55	1.3×10^9
Bufo	5.0	4.6×10^9	20	0.9×10^9
Rana	8.0	7.3×10^9	22	1.6×10^9
Ambystoma	42.0	3.8×10^{10}	20	7.6×10^9
Bos (cow)	3.5	3.2×10^9	55	1.6×10^9

* Data obtained at a criterion temperature of 60° (0.18M Na⁺). NTP, nucleotide pairs. Genome size can be converted to nucleotide pairs with the constant 9.13×10^8 NTP/pg. Except for *Bos* (Britten and Kohne, 1968a), *Xenopus* (Davidson and Hough, 1969, 1971), and *Engystomops* (Davidson, Chamberlin, and Hough, unpub.), the measurements cited are those of Strauss (1971).

throughout evolution. As a result each evolutionary line appears to possess sequence families, which are shared with other evolutionary lines, as well as some which are confined to it. Sequence relationships have been detected between organisms whose ancestors diverged hundreds of millions of years ago. For example, 5 per cent sequence homology exists between the repetitive DNA's of fish and primates, and 10 per cent between the repetitive DNA's of birds and primates (Hoyer, McCarthy, and Bolton, 1964). Not all of the families of repeated DNA sequences of a given species are present in the DNA of even a closely related species. The more distantly related species share fewer repeated sequences (Hoyer, McCarthy, and Bolton, 1963). In time, divergence of related sequences results in loss of relationship as a result of many base substitutions. However, since each species exhibits sets of repeated DNA sequences which reassociate relatively precisely, it is evident that new families appear regularly. Fig. 2 (Rice, 1971) shows that new repetitive DNA has appeared in the course of recent rodent evolution. The figure shows that mouse DNA contains various repetitive sequence components which are absent from the rat genome.

The incorporation of repeated DNA into the genome does not seem likely to be a continuous process. For example, the "mouse satellite" (described in detail below) makes up 10 per cent of *Mus musculus* DNA and consists of a million copies of an approximately 300-nucleotide-long sequence. This particular component is among those apparently absent or only rarely present (less than 50 copies per genome) in rat DNA (Flamm, Walker, and McCallum, 1969). Strikingly, this component is also essentially absent in two other species of the genus, *Mus caroli* and *Mus cervicolor* (Britten, unpub.). This highly repetitive sequence family must have been produced in a relatively sudden event which has occurred since the lines leading to these species diverged in the last few million years.

Other lines of evidence also indicate that the incorporation of new families of repetitive sequences can occur suddenly (on an evolutionary time scale). Such sudden replicative events are termed *saltatory replications* (Britten and Kohne, 1968a). The incorporation of single extra copies of random segments of DNA cannot explain the observed patterns of repetition, irrespective of the frequency of such events. Instead, particular sequences appear to have been copied many times over. Thus two alternatives present themselves. Either these particular sequences are multiplied in relatively sudden events, or they are multiplied more or less continuously in the course of evolution. The latter alternative is unlikely for a number of reasons. It would imply that each family of repeated DNA would contain ancient members, which differ widely in sequence. Measurements of the repeated sequences held in common
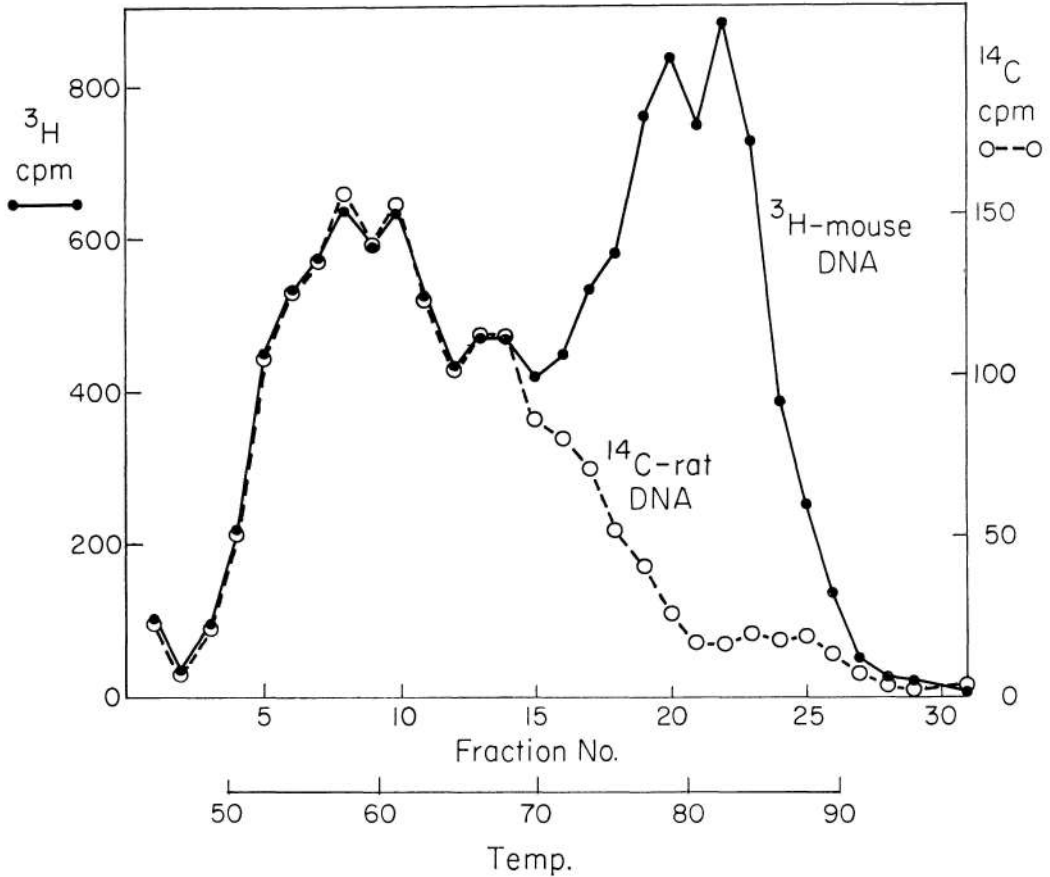
FIG. 2. EVIDENCE FOR THE APPEARANCE OF NEW REPETITIVE DNA SEQUENCES SINCE THE DIVERGENCE OF MOUSE AND RAT (Rice, 1971).

A mixture of ³H-mouse DNA (290 μg/ml) and ¹⁴C-rat DNA (0.86 μg/ml) in 0.14M phosphate buffer (PB) was heated to 100°C for four minutes, annealed at 51°C for 18.3 hours, bound to hydroxyapatite, and eluted with 0.14M PB while the temperature was continuously increased. 58.3 per cent of the total mouse DNA and 51.6 per cent of the rat DNA bound subsequently to hydroxyapatite and were eluted during the linear temperature gradient. ¹⁴C-rat DNA which elutes above about 80°C (3.2 per cent of the total) results from self-reaction.

To emphasize the constant ratio of mouse DNA to rat DNA which is observed during the first half of the gradient, the cpm (counts per minute) scales have been selected so as to produce coincident elution profiles in this region. Only duplexes containing repetitive DNA sequences will have formed under these annealing conditions, and as the temperature at which they are no longer stable is attained they are eluted from the column. Since rat DNA was present at sufficiently low concentration to minimize the formation of rat-rat duplexes, the ¹⁴C elution profile represents the thermal dissociation of mouse-rat duplexes. Those mouse DNA sequences which reassociate to give products of high thermal stability appear not to be present in rat DNA.

between species fail to reveal this pattern (see Fig. 2). Reassociated mouse satellite DNA, for example, has relatively high thermal stability, indicating only a modest amount of sequence divergence. Furthermore, a continuous multiplication process would imply that old and widely divergent families would have more members than recently initiated, less divergent

families. Observations so far indicate that, in fact, relatively closely related, presumably recent, families have the largest number of members. We conclude, therefore, that families of repeated DNA are produced in the relatively sudden evolutionary events which have been termed saltatory replications (Britten and Kohne, 1968a), that is, in events which are ini-

tiated and terminated rather than continuing throughout the course of evolution.

We know nothing of the mechanism of such saltatory replications, but the following processes seem necessary, though they may not have occurred in the order listed. (1) Many copies are made of a DNA sequence and appear in the germ cells of certain individuals. (2) The copies are somehow integrated into the genome so that they are duplicated. (3) Over a sufficient period of time they are disseminated throughout the population and its evolutionary descendants. Dissemination could result from their association with a favorable genetic element or simply because of their multiplicity. (4) Individual sequences become scattered among many chromosomes and interspersed intimately along the length of the DNA of the genome. (5) The growth of the family of repeated DNA is eventually terminated or controlled. Subsequently individual members of the sequence families would diverge from each other through base substitutions. Also, the length of the recognizably related regions is presumably reduced by the events of rearrangement which led to their interspersion throughout the genome.

While this may be an adequate summary of the broad outlines of the history of repeated DNA families, there is little or no information on function and the resulting selection pressures, which are probably important. Further measurements will be required in order to reach to these issues. At this time the view which seems most consistent with classical views of natural selection is that stochastic events initiate the chains of steps leading to each family of repeated DNA. In other words, repetitive sequence families originate accidentally. We assume that, later on, some or all of their member sequences become useful to the species, and only at this stage does selective pressure become important. There is no evidence to decide between this view and the alternative one that certain classes of function in the cell cause the excessive duplication of certain useful sequences in the germ line DNA.

We cannot leave this subject without some comment on possible mechanisms of saltatory replication, although there is again no useful evidence. Four classes of events present themselves as possibilities: (1) erratic behavior of a DNA polymerase, perhaps caught in a closed short loop without adequate termination controls; (2) geometric growth as a result of a series of duplications of an initially short duplicated sequence due possibly to unequal crossing over; (3) the excessive replication of some nuclear element analogous to the episomes of bacteria (not yet observed in higher cells); (4) the integration into the genome of many copies of a viral genome or viral-borne sequence.

REPETITION FREQUENCY COMPONENTS

It is becoming clear that a significant part of the repeated sequences that have so far been observed can be classified into distinct, large, components on the basis of number of copies per sequence. In many cases discreteness is also indicated by other characteristics, such as density in a cesium chloride gradient or thermal stability. The first component definitely shown to be repetitive chromosomal DNA (Waring and Britten, 1966), the mouse satellite, has the following properties (Flamm, Walker, and McCallum, 1969): (1) it amounts to about 10 per cent of the genome; (2) it is discrete in frequency, about a million copies of a sequence approximately 300 nucleotides long; (3) it is discrete in a cesium chloride equilibrium density gradient, due to its particular nucleotide composition and its occurrence in clusters; (4) when reassociated (as well as native), it exhibits a narrow range of melting temperatures; (5) it has a striking asymmetry in the composition of its strands, which distinguishes it from most of the DNA of the mouse; (6) it is mainly localized at chromosome centromeres (Pardue and Gall, 1969; Jones, 1970), and at least in some circumstances it is not transcribed (Flamm, Walker, and McCallum, 1969). A moderate number of such components, generally sedimenting as satellites in a cesium chloride solution, have been coming to light recently. Table 2, which presents a sample of our current knowledge of the frequencies of repetition of discrete sequences, includes several of these. Our present knowledge of discrete repetitive sequence families is in fact biased in favor of satellite-like components, due to their prominence and the ease of separating these components from the rest of the genome. The satellite components of Table 2 contain between 10^5 and over 10^6

TABLE 2

Some discrete components in the spectra of repeated DNA sequences †

SPECIES	PERCENTAGE OF NUCLEAR DNA	NUMBER OF COPIES	COMPLEXITY (NTP PAIRS)	MELTING RANGE	REF.
Calf	38	66,000	17,000	broad	Britten, 1969b;
	5	1,000,000	(150)		Britten and Kohne, 1967
Sea urchin	20	(50)	3,000,000	broad	Britten and Davidson,
	10	1,200	60,000	broad	unpub.
	3	14,000	1,600	broad	
Human	3	(300)	400,000		Britten, 1969a and un-
	15	40,000	10,000	broad	pub.
	10	300,000	1,000	broad	
Mouse	25	(1,000–10,000)		broad	Waring and Britten, 1966; Bolton et al., 1966
	10	1,000,000 *	300	narrow	
Green monkey	20	1,500,000 *	450	narrow	Maio, 1971
Guinea pig	5.5	2,200	80,000	broad	Corneo, Ginelli and
	2.5	160,000 *	500	narrow	Polli, 1970a
	2.5	500,000 *	150	broad	
Ilyanassa	12	20	17,000,000	broad	Davidson et al., 1971
	15	1,000	450,000	broad	

† Data referred to are obtained mainly at 60°C (0.18M Na⁺). Parentheses denote uncertainty.
* Fraction isolated from CsCl density gradients as a satellite.

copies. They are atypical in several respects, besides their unusually high sequence frequency. Their complexity is extremely low (Table 2) and, at least in the cases of mouse and green monkey, the satellite sequences are tightly clustered in the genome. This is in sharp contrast to the other repetitive sequence families, which appear, in calf and sea-urchin, to be abundantly interspersed with non-repetitive sequences. These other families often are not species-specific as the satellites appear to be. Fig. 2 clearly shows both the components that are held in common and those which are not shared between mouse and rat. An additional unusual feature of the satellite components is their small degree of sequence divergence, again compared to repetitive sequence classes of lower frequencies. A growing number of the latter are becoming known, as Table 2 indicates. For example, the principal part of the repeated DNA of the calf consists of one component making up about 38 per cent of the total DNA (Table 2). The evidence for the existence of

this component, and for the other repetitive sequence components included in Table 2, derives from measurements of the rate of DNA reassociation. The discreteness of these sequence frequency classes is their most unexpected and striking feature. This discreteness is established by detailed observations carried out on isolated fractions of the DNA containing primarily the components in question. Examples are displayed in Fig. 3. Due to the limits of the method it is not yet possible to say whether all of the sequences present in such discrete repetitive frequency components are present in exactly the same number of copies. However, the majority do not differ by more than a factor of two from the average frequency of repetition, if they differ at all.

The DNA of organisms below the mammals appears in general to have lower repetition frequencies than that of the mammals. As Table 2 shows, the sea urchin and snail display frequency classes in the range 10 to 10⁴.

What is the significance of the discontinuous
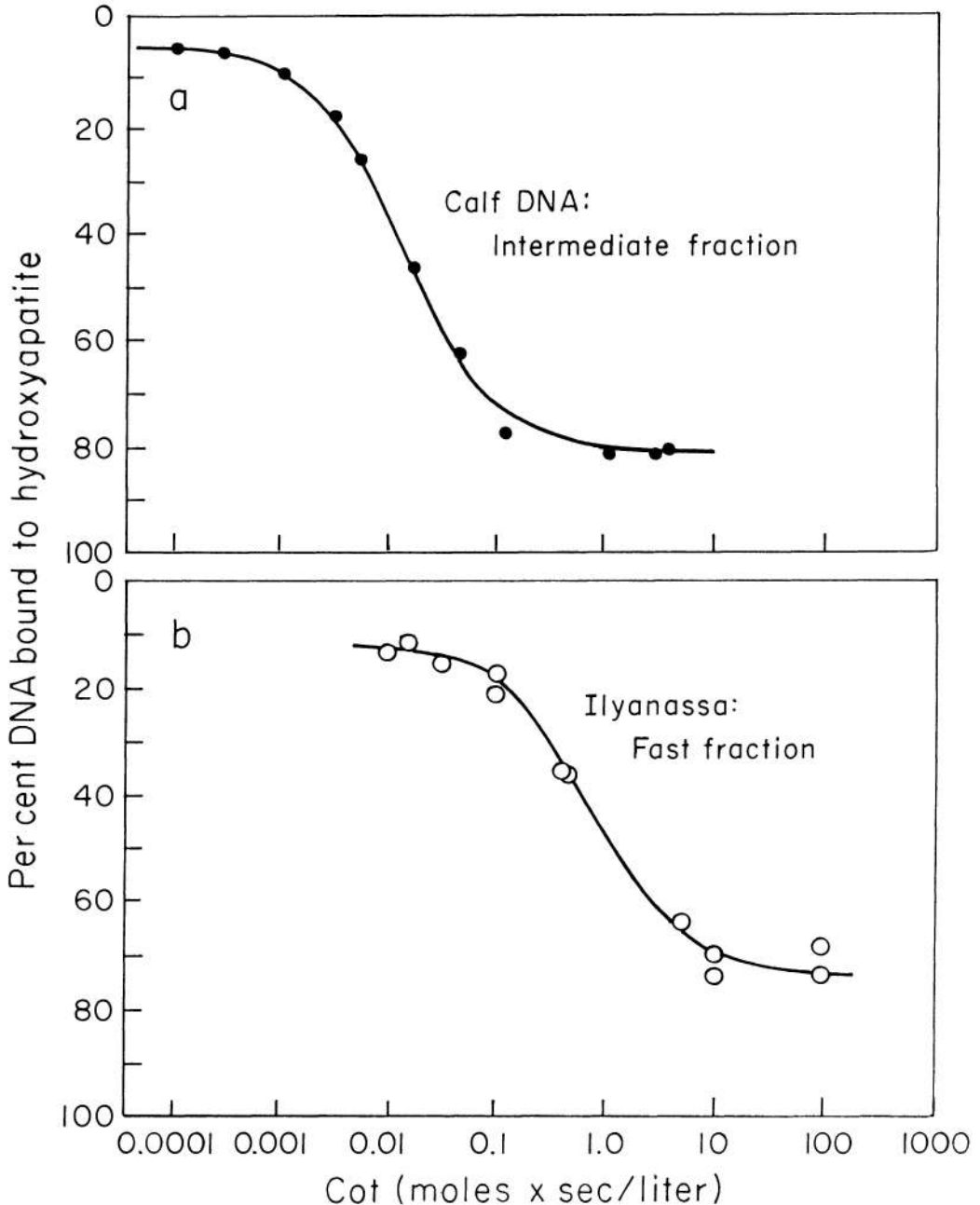
FIG. 3. REASSOCIATION OF REPETITIVE DNA FREQUENCY COMPONENTS

In each case a fraction of DNA containing mainly the specific frequency component shown was isolated from whole DNA by the appropriate annealing and hydroxyapatite separation procedures. The isolated fraction was then denatured, incubated for various times at 60°C in 0.12M phosphate buffer, and passed over a hydroxyapatite column operated at 60°C. The percentage of the DNA bound to hydroxyapatite (HAP) is shown on the ordinates, and the incubation "Cots" are plotted on a log scale on the abscissae. (Cot = moles nucleotide per liter times time of annealing, in seconds, for solutions annealed at 60°C in 0.12M phosphate buffer; see Britten and Kohne, 1968a.) The curves represent ideal second-order reactions fitted to the data by a least-squares method using an IBM 1130 computer.

distribution of sequence frequencies which emerges from these investigations? This tight clustering could simply be the result of the nature of the replication events. Examples such as the mouse satellite, presumably recent, would reflect the event itself with relatively little subsequent change. On the other hand, the frequency pattern we observe may be selectively profitable, in that the repetitive sequences are utilized in a way that requires the preservation of sequence homologies in certain frequency classes. If repetitive sequences are utilized for the programming of gene regulation in the manner which we have proposed, for example, certain frequency size classes might indeed occur. For instance, the number of copies of a given repetitive sequence might approximate the number of cell states, in all the cells at all the stages of an organism, if that sequence is involved in the activation of a producer gene common to all of them. Several meaningful frequencies can be imagined, such as the number of cell types, the number of ontogenic programs, but we do not yet know enough about any metazoan organism to say what these numbers really are. This line of thought presupposes, of course, that the repetitive sequence families are actually utilized. While we can be certain from RNA-DNA hybridization studies that some repetitive sequences are actually transcribed in most, if not all, cell types, we know as yet little more than this. Nor can we at the present state of knowledge determine whether the discrete frequency classes which we observe are the product of selection, rather than being functionless, recent saltatory additions to old genomes.

THE ARRANGEMENTS OF REPEATED SEQUENCES IN
DNA

Very early in the study of repeated DNA sequences an interesting observation was made (Bolton et al., 1966) which indicated that repeated and single copy sequences are intimately intermixed. When moderately fragmented DNA (molecular weight 5 to 10 million) is reassociated so that only repetitive DNA can react, large structures are formed, which we have termed networks. The rapidity of formation and the size of these networks imply that each of the strands contains several different repetitive sequences. The several sequences may each reassociate with a different strand and thus a rapid branching and growth of network particles can occur. The particles are quite large, sometimes being visible to the naked eye and causing a visible haze at low concentrations. They can be efficiently collected by low speed centrifugation or filtration. They usually contain most of the DNA and therefore are made up of both single copy and repetitive DNA sequences. This evidence implies that non-repeated and repeated sequences of various classes are present together on individual DNA strands of the typical size used in these tests, about 20,000 nucleotides long. An intimate intermixing or interspersion of different sequences is thus clearly indicated on this size scale.

Recently the scattering or *interspersion* of repeated and single copy DNA has been explored on an even finer scale (Britten and Smith, 1970; Britten, 1969b). In these measurements a small quantity of relatively long fragments of known size (about 4000 nucleotides long) was incubated with a much larger quantity of short (400 or 500 nucleotides long) fragments. After incubation the reassociated fragments were separated from the remaining single-stranded DNA by hydroxyapatite fractionation, a method which permits separation of duplex structures. Fig. 4b shows the result of such a set of measurements for calf DNA. The ratio of the quantity of short DNA to the long was

(a) The "Intermediate Component" of calf DNA, representing at least 30 per cent of the DNA. From the rate constant of the fitted curve, it can be calculated that the complexity of this repetitive sequence fraction is about 17,000 nucleotide pairs. There are about 70,000 related copies of each sequence in the calf genome (Britten and Smith, 1970).

(b) Reassociation of the "Fast Component" of snail (*Ilyanassa*) DNA, representing at least 15 per cent of the genome. Note that the reaction rate of this component is about an order of magnitude slower than that of the calf's "Intermediate Component." From the fitted second-order curve, the complexity of this fraction can be calculated to be about 450,000 nucleotide pairs, each sequence being present about 1000 times (Davidson et al., 1971).

The detailed steps of the fractionations are described in compact symbols, as follows ("Cot" means incubation of the fraction to the equivalent Cot shown by the following number; "HAPbd" or "HAPunbd" implies that the fraction bound or unbound to HAP was used for the succeeding incubation): (a) Calf: Cot 48 HAPbd; Cot .0004 HAPunbd; Cot 3 HAPbd. (b) *Ilyanassa*: Cot 0.1 HAPunbd; Cot 10 HAPbd.

made so great that practically the only reaction possible was the reassociation of the long fragments with short ones. Thus the binding of the long fragments to the hydroxyapatite is due almost entirely to their reassociation with short fragments. This interpretation is supported by control experiments in which little binding was observed when the long fragments were given an equivalent incubation in the absence of short fragments. Fig. 4b shows that a greater proportion of the long-fragment DNA is bound than of the short fragments. This result indicates that most of the 4000-nucleotide-long fragments contain non-repeated as well as repeated regions of DNA. It appears that at least two-thirds of the non-repeated DNA of the calf occurs adjacent to repeated DNA sequences. In other words, most stretches of non-repeating DNA are interrupted at least every 4000 nucleotides by segments of repeated DNA.

Clearly a great deal of interspersion of the different sequences is present in the calf genome. This set of measurements, of course, does not go so far as to demonstrate that all of the repeated DNA sequences are interspersed with the non-repeated DNA. A minimum of 10 or 20 per cent of repeated DNA scattered rather evenly throughout the genome would suffice to produce these results. This evidence relates almost entirely to the principal component of the repeated sequence spectrum of calf DNA, the sequences of which are present in approximately 66,000 copies (Table 2).

Similar results have been obtained in a series of more extensive measurements which are currently being made on the DNA of the sea urchin *Strongylocentrotus purpuratus* (Britten and Smith, 1971). As mentioned above (Table 2), there are three major components in the repeated sequence spectrum of this sea urchin. Interspersion measurements have been carried out on each of the components, as well as on the whole DNA. The evidence indicates that the intermediate rate component of the sea urchin genome shows the greatest degree of interspersion. This component alone, though it constitutes only about 10 per cent of the DNA, will reassociate with over 80 per cent of the long DNA fragments and cause their binding to hydroxyapatite. This experiment is shown in Figure 4a. The two other components also

show interspersion, but to a somewhat lesser extent. The data suggest that many of the long fragments contain segments of more than one component of repeated DNA as well as single copy sequences.

For each fragment size that has been examined, the individual fragments frequently contain sequences of different kinds. This fact implies that a rather thorough scattering of segments of DNA has occurred in the course of evolution. Of course the basic question immediately follows: does scattering or interspersion of DNA sequences exist (that is, is it positively selected) for functional reasons, or is it simply a measure of the random events of rearrangement which have occurred historically? It should be obvious that the authors believe that this degree of interspersion indicates function.

In this paper we propose to follow the view that control of genetic function is almost assuredly related to the arrangement of sequences in the genome. The original events of rearrangement may well be accidental occurrences, but natural selection is probably important on two levels. The more obvious is simply the preservation of lines carrying rearrangements that lead to some positive benefit and the loss of lines that carry those rearrangements which are deleterious. However, natural selection may also act to preserve lines subject to this sort of rearrangement.

PROPOSALS FOR THE ORIGIN OF
EVOLUTIONARY NOVELTY

According to the gene regulation model, alterations in the organization of the genome by rearrangement would be expected to have profound effects. Rearrangements could lead to the construction of new regulatory networks in the course of evolution. Suppose that a chromosomal rearrangement results in the insertion of a foreign integrator gene into an existing set of integrator genes. Transcription of this new sequence will be under the same control as that of the remainder of the integrator gene set; i.e., transcriptional control by the continguous sensor gene. The transcription products of the integrator set would then include activator RNA from the foreign sequence. This RNA would control the activation of the relevant set of producer genes either directly or after trans-
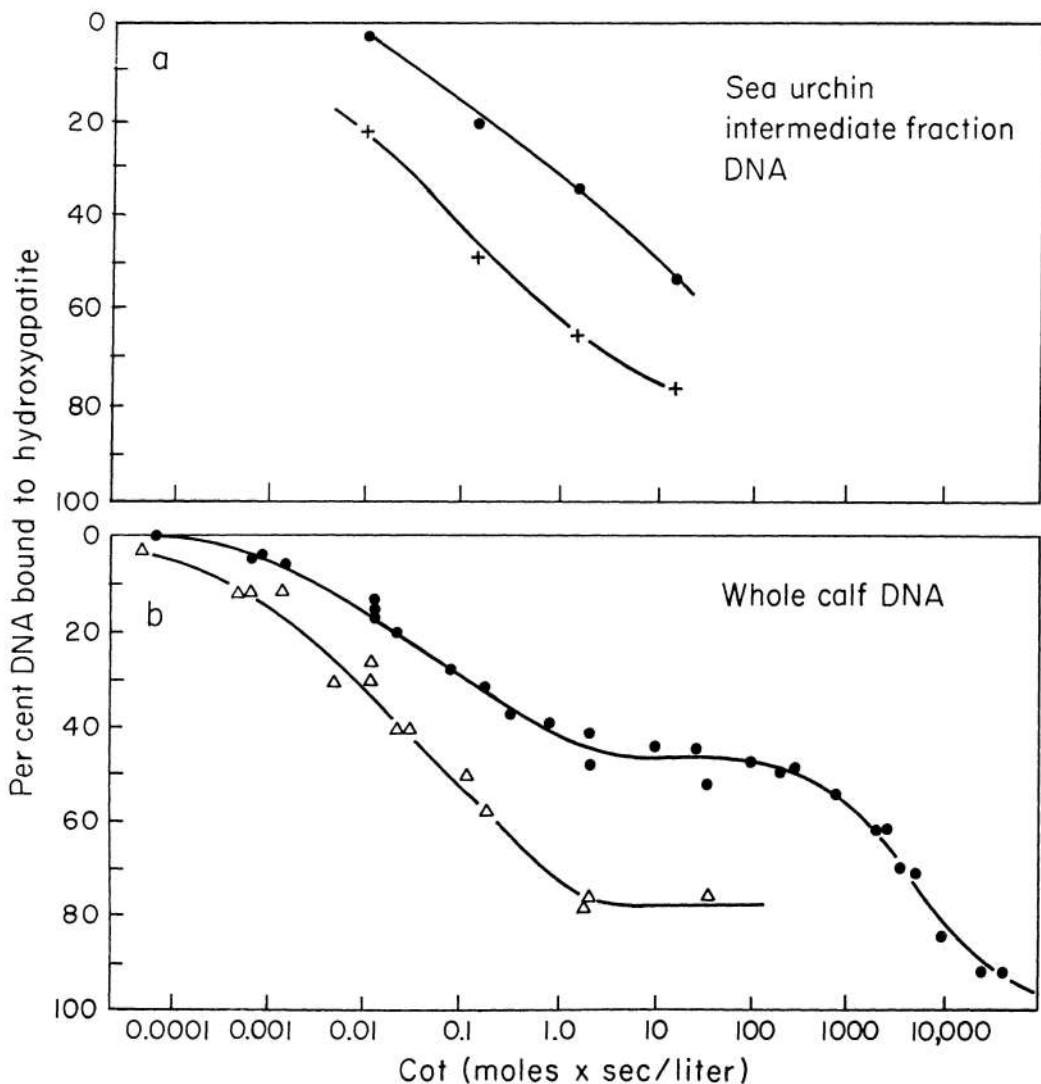
FIG. 4. DEMONSTRATIONS OF FINE-SCALE INTERSPERSION OF SEQUENCES OF DIFFERENT FREQUENCY CLASSES
(Britten and Smith, 1970; Britten, 1969b).

(a) Interspersion of repetitive sequence components of the sea urchin (*Strongylocentrotus*) genome. An intermediate repetitive fraction of unlabelled sea urchin DNA sheared to about 400 nucleotides was incubated with a very small quantity of ^{14}C-labelled, unfractionated long fragments of sea-urchin DNA, lightly sheared to about 4,000 nucleotides. The solid circles represent the binding to hydroxyapatite of the 400-nucleotide intermediate repetitive DNA fragments, as measured by optical density. The crosses represent the binding of the 4,000-nucleotide-long unfractionated fragments, as measured by their radioactivity. Only negligible reassociation of the long fragments with each other did occur at the low concentration in which they were present. Thus binding to the hydroxyapatite of the labelled long fragments indicates their reassociation with the selected class of small fragments. The experiment shows that 80 per cent or more of these long fragments contain sequences belonging to a component or components present in the intermediate repetitive fraction. The intermediate repetitive fraction was prepared by hydroxyapatite fractionation of sheared urchin DNA which reassociated after incubation to Cot 3 but did not reassociate after a subsequent incubation to Cot 0.07. All reassociations in this experiment, as in (b), were carried out at 60°C in 0.12M phosphate buffer. (b) Interspersion of repeated and non-repeated sequences in the whole calf genome. The upper curve shows the reassociation of unfractionated calf DNA fragments sheared to about 400 nucleotides. The lower curve shows the reassociation of a small quantity of labelled 4,000-nucleotide-long fragments with the 400-nucleotide-long fragments which were present in great excess. For the upper curve, data have been included from a number of other measurements (all sheared to 400 nucleotides) in order to give a more complete picture of calf DNA reassociation kinetics.

lation, to yield activator protein. In either case the result would be the same. The set of producer genes controlled by the foreign integrator gene would be activated along with the original set of producer genes under control of the contiguous sensor gene. A new regulative coordination would now be programmed in the genome. That is, stimulation of the sensor gene brings about activity of a gene battery different from that which was previously functional under these circumstances.

The important consequences of such an event can be illustrated by a crude hypothetical example. Suppose that at a remote prechordate level an integrator gene set (A) controls the ontogenic appearance of circulating blood cells, while a battery controlled by a distant integrator element (B) is involved in the ontogeny of cuticle. Suppose further that among the cuticular cell functions is the synthesis of a heme-containing globulin used for trapping oxygen, and that the battery of producer genes required for its synthesis is activated by this integrator element, B. If B is translocated to the A integrator gene set, the result, in this illustration, would be the appearance of the heme compound in the circulating blood cells.

Genomic rearrangement events could have a variety of significant results. Translocation of receptor sequences to sites adjacent to foreign producer genes would have the effect of adding them to pre-existing gene batteries. Translocation of a sensor gene to a new integrator sequence would result in the activation of a whole new set of genes in response to a given stimulus. These examples show that if the genomic regulatory system is organized as we have proposed, then rearrangement events may indeed have the result of constructing novel regulative systems.

Consider the production in the genome of entirely new sets of repeated sequences which have the characteristics required for function as integrator or receptor sequences. Arguments concerning the origin of repetitive DNA (see above) suggest that repetitive sequences probably occur initially in tight clusters. The evidence indicates that rearrangements have occurred so that these clusters are ultimately broken up. In this way the individual sequences of a repetitive sequence family would, in a sense, "diffuse" through the genome.

We propose that such a set of related sequences would be a source of new integrative relationships, as suggested in Fig. 5. For the moment we restrict our considerations to the case where the RNA transcribed from integrator sequences is utilized directly, as activator RNA, rather than serving as template for activator proteins. If activator RNA is the regulatory element, sequence homology between integrator and receptor genes could be the only essential requirement for potential regulatory function. It seems likely that, on occasion, a sequence belonging to a repetitive sequence family would be translocated to a position adjacent to a pre-existing integrator gene set and be transcribed under the control of the associated sensor. Then, if another member of the same repetitive sequence family is translocated into the receptor location adjacent to a producer gene, that producer gene would be transcribed under the control of the particular sensor. This might or might not be an advantage to the organism, and natural selection would operate to preserve or eliminate individuals carrying this new regulatory feature.

The large number of repeats (up to a million copies) observed in some sets of repeated sequences suggests that the rate of occurrence of such tests of new regulatory relationships may be high. Natural selection might promote the survival of populations capable of testing new regulatory relationships, thus maintaining the potentiality of producing new repetitive DNA sequences and translocating them.

One of the most difficult issues for evolutionary theory is the appearance of new organs or of complex systems which carry out novel functions. Here we emphasize that at all times an internally consistent program of development for the organ must be present in the genome. The gene regulation systems which function during ontogeny are crucial. While we cannot, at this stage, describe the actual evolution of an organ, it seems plausible that a source of new integrative relationships would be essential. A million copies of a given set of sequences scattered throughout the genome would represent an exceptionally large number of possible new regulatory relationships. A

1) A portion of the genome containing a new saltatory replication (▨▨ ▨) :

or:
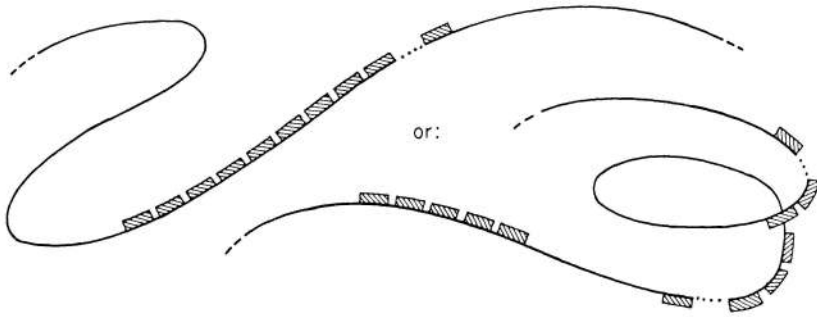
2) "Diffusion" of sequences throughout the genome by subsequent chromosomal rearrangments:
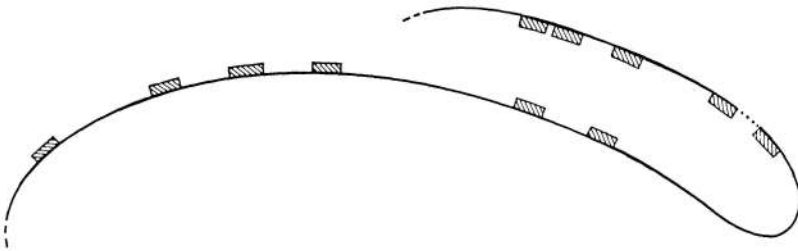
3) Among some local arrangements which might thus arise could be these:

4) In this way new regulative pathways could arise, for example:

aRNA

aRNA

aRNA

aRNA

aRNA

FIG. 5. DIAGRAM PORTRAYING THE EVOLUTIONARY APPEARANCE OF A REPETITIVE DNA SEQUENCE FAMILY, AND THE EVENTUAL INCORPORATION OF SOME OF THE REPEATED SEQUENCES INTO THE GENOMIC REGULATORY SYSTEM (see text).
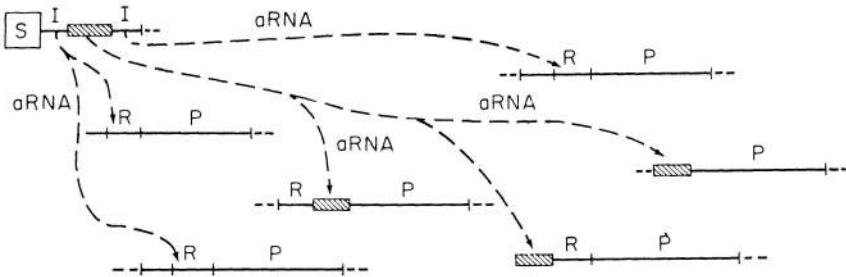
I, integrator gene; R, receptor gene; S, sensor gene; P, producer gene; aRNA, activator gene.

wide variety of experimentation with new regulatory systems could ensue. Many of these regulatory networks would certainly be ineffective or lethal. Nevertheless, gene regulatory programs leading to the development of novel structures of great usefulness could occasionally be produced, and if of lasting significance to the organism would be selectively preserved. The formation of complex systems, such as organs, would not necessarily depend on the recent addition of a large new set of repeated sequences. A variety of repetitive sequences has evidently been present in eukaryotic genomes for hundreds of millions of years and has been available for the formation of new regulatory relationships. In addition, as suggested above, fruitful rearrangements of pre-existing regulatory networks probably have continued to occur.

Existing systems and organs will thus be continually modified and recombined and the rare events with selective advantage will be preserved. It seems possible that really new functions and structures could evolve in this way.

THEORETICAL REQUIREMENT FOR POSITIVE CONTROL

Essential to the process envisioned here is a means for protecting and preserving new genomic sequences which do not yet have a function. Until the time when these sequences can be tested and their usefulness assayed by selection, they must be retained for the many generations required for them to diffuse throughout the genome. This requirement has some significant consequences, for it means that newly arising sequences, whether repeats or divergence products of pre-existing sequences, cannot, in general, be immediately transcribed and tested for selective value. Otherwise, as the following argument indicates, such new sequences would almost certainly be eliminated from the population by natural selection. Consider a replication event which results in the addition of thousands or tens of thousands of new copies of some DNA sequence. Whether the sequences which are replicated in this event served as producer genes or as regulative elements in the ancestral genome, the sudden presence of a large number of transcribed copies would almost certainly be deleterious. In gen-

eral, any novel sequence is unlikely to be anything but deleterious if tested for use as a producer gene.

It is likely, however, that regulation of genes in higher cells is accomplished by a positive control system, that is, through specific activation of a genome which is repressed non-specifically by histones (Britten and Davidson, 1969). Automatic repression of transcription from new sequences is thus exactly what would be expected. Throughout the period following a sequence replication event, while the newly formed repetitive sequences are diffusing through the genome, these sequences would be inactivated by histone, like any other sequences not specified for transcription. Only after they are translocated into a gene regulation system and have become a part of it could they be transcribed and subjected to selective test. In other words, histone coverage can be thought of as conveying protection against premature elimination by natural selection. The range of genome sizes displayed in Fig. 1 and Table 1 provides additional indirect evidence that unused genetic material can be stored and replicated in the genomes of higher organisms. We propose that positive control, with non-specific histone repression, is the vital characteristic which has allowed genome growth, evolutionary experiment, and invention to take place at a high rate in eukaryotes.

GENERAL SUMMARY: THE EVOLUTION OF
REGULATORY PROCESSES

In metazoans, the biological structures which are acted on by natural selection are the product of cellular developmental processes. Cellular development is governed, in turn, by the individual patterns of gene activity in the nuclei of the differentiating cells. As a useful simplification, we can consider that coordination of the cells in development is established in the beginning by spatial instructions present in the egg (localization) and later by interactions among diverse cells (induction). Both of these processes for the control of the direction and timing of cellular differentiation must ultimately operate through the control of genetic activity. In other words, whatever other levels of regulatory activity may occur, what the cell does is finally the product of its gene activity.

The mechanisms that determine its function do so by affecting the spectrum of active genes.

The result of localization and induction processes, then, is to specify patterns of gene activity in cells at given times. The relevant acts of gene regulation consist of the activation of large sets of genes, that is, they involve *patterns* of genetic activity. Since the developmental process is inherited, the genome must contain the *regulative programs* with which to control the *pattern* of genetic activity as well as the genes individually called upon to function in these patterns. Evolutionary changes in the developmental process could certainly come about by alternations of individual genes expressed at given stages of development. It is clear, however, that alterations in the genes which determine the regulative programs could cause enormous changes in the developmental process and that this would be a much more potent source of evolutionary change. We feel that to explain the magnitude of the functional and structural change which has occurred in evolution it is necessary to postulate changes in the regulatory apparatus. Indeed, an explanation of the scale on which significant evolutionary invention has taken place may require in addition whole new classes of mechanism, such as selection in favor of genomes with favorable types of sequence organization (e.g., suitable sources of regulatory genes) or with optimal rates of reorganization. Evolutionists, such as Stebbins (1969) and Mayr (1963), have also considered the role of inherited developmental patterns in the process of evolution.

The gene regulation model proposed earlier (Britten and Davidson, 1969) provides a possible molecular level interpretation of the programming of gene activity in development (Davidson and Britten, 1971). It indicates how specification of particular cell lines in development could occur in the processes of localization and induction. A basic property is the potentiality for evolutionary modification of the regulatory systems that were envisioned. In this model important parts of the regulative relationships depend on the location of sequences in the genome, and a variety of changes can be caused by sequence translocation. As a result, a great many individual cellular properties could be affected by genomic alterations of a type which may occur relatively frequently, namely, trans-

locations and chromosomal rearrangements. It is quite plausible that selectively favorable changes of the former type are much more frequent. Furthermore, the model regulatory system is receptive to the incorporation of new sequences and sets of sequences, and through these additions can achieve new functional interrelationships. It can grow as well as become rearranged. Whether or not the particular gene regulation system that we have proposed turns out to occur in living systems, the properties of growth and change by genomic rearrangement are likely to be necessary attributes of evolving regulatory systems.

In this paper we have summarized some evidence on the structure of today's metazoan genomes. This evidence must contain significant historical clues as to the nature of the events which led to the evolution of modern genomes. We have been guided by the idea that many of the important events that have occurred in the evolution of genomes are those which have induced change and growth in the regulatory systems. A prime example, in our view, is the saltatory origin of repetitive sequence families, followed by their dispersion throughout the genome, with the result that many sequences of different families are highly interspersed with each other. Consequences of such a process for changes in the regulatory programming in the model system have already been discussed (see section, Proposals for the Origin of Evolutionary Novelty), and are illustrated in Fig. 5.

In considering the process by which a new structure or organ might have arisen in evolution, we imagine that an early step would be the rearrangement of existing gene regulatory relationships so that certain cells of an antecedent structure gain additional new characteristics. Obviously there will be severe restrictions on the changes possible, since the embryo must remain viable. The steps that follow from this initial event to the formation of a new organ or structure of course cannot be mapped at present. We do not know if it is possible for internal forces in the genome (e.g., sequence patterns) to lead to continued growth of a newly initiated set of regulatory relationships. Such controversial possibilities, lying outside the processes envisioned in classical evolutionary theory are not necessary in the application of
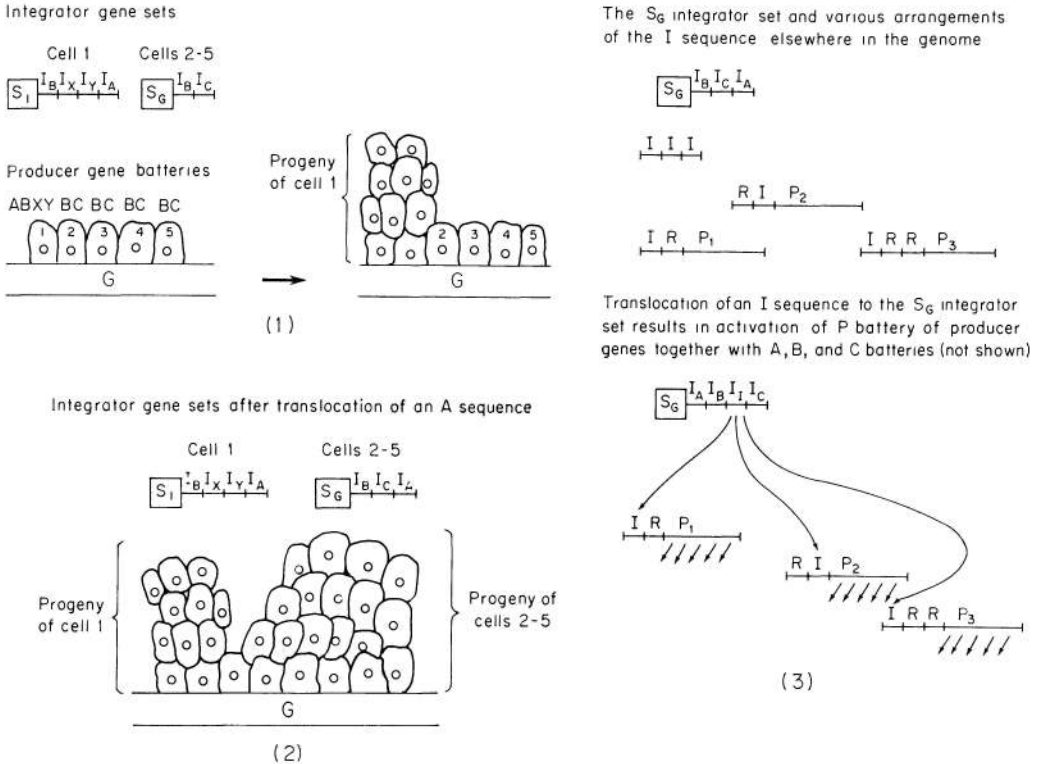
FIG. 6. A HYPOTHETICAL EXAMPLE OF EVOLUTIONARY CHANGE IN GENE REGULATORY PATHWAYS LEADING TO PRODUCTION OF A NEW ORGAN-LIKE STRUCTURE

1. Mesodermal cells (1–5) growing along the gut tube (G) at an early embryonic stage of a primitive organism. The letters above each cell denote the batteries of *producer genes* active in that cell, i.e., the *developmental* program of gene activity responsible for that cell's embryological behavior. The program BC causes cells 2-5 to synthesize a set of mesoderm-specific proteins (battery B) and also to grow along the gut wall (battery C). Battery C could be visualized as directing the synthesis of adherent cell surface proteins, for example. Both batteries are activated in response to an inducer diffusing into the cells from the gut. As diagrammed at the top of the figure, this response would be mediated by the integrator genes (I_B and I_C) which control batteries B and C, and which are activated by the *sensor gene* S_G, the site where the gut inducer binds. Other batteries, A, X, Y, are active in cell 1, which responds to a different sensor gene, S_1. Of particular interest is battery A, the function of which is to induce rapid cell division, with the result shown on the right of (1). The clone descendant from cell 1 might serve as an anterior structural support, for example. The difference in the developmental program expressed at this stage in cell 1 and that in cells 2–5 could be due to differences in the localization of regulatory elements in those regions of egg cytoplasm from which these cells originally descended.

2. Occurrence of a translocation now endows the integrator gene set active in cells 2-5 with a sequence homologous to the integrator sequence A. The new integrator gene sets are shown on the top of (2); thus when the gut inducer activates S_G, the A *producer* gene battery is activated as well as the B and C batteries. The result, as shown in the figure, is the developmental growth of a large cell mass in place of the original flat sheet of mesoderm cells.

3. A multiple new synthetic function appears in the new structure as the result of the addition of a new sequence, I_1, to the integrator gene set (I_B-I_C-I_A) operative in these cells. As shown on the top of (3) this sequence, I, has homologues elsewhere in the genome, some of which are in position to function as receptor elements controlling the contiguous producer genes. Prior to the translocation, the I sequence is silent, i.e., is not transcribed, since its activation is not specified by any existing regulatory processes. P_1, P_2, and P_3 represent various producer genes already engaged in various other gene batteries, and responding to receptor sequences R. Thus, when the gut inducer binds to the sensor element in cells 2–5, transcription of these producer genes is activated by I_1, along with transcription of batteries A, B, and C. The result is that the cell mass descended from cells 2–5 becomes endowed with additional synthetic functions by accumulation of synthetic processes (P_1, P_2, P_3) previously scattered among other cell types. This example has the following properties:

(1) Only pre-existing producer genes are used (including those called for by programs A, B, and C).

this model to evolutionary phenomena. However, the model does suggest such intriguing possibilities. In any case, the production of a complex structure must require a succession of events that amount to the modification of a whole pattern of regulatory relationships. In Fig. 6, a hypothetical example of such a sequence of events is diagrammed, not as a description of the history of any particular structure, but as a crude illustration of the type of large-scale change that could be brought about.

The processes of development appear to be sequential, not only in the obvious time course of the events, but in the basic molecular and cellular mechanisms. In other words, the later stages are built on a foundation consisting of the events occuring earlier in development. As a result, changes in the parts of the developmental program operative at a given stage might result in drastic alterations of later developmental events. We expect, therefore, that the regulatory programs active earlier in development would also have been elaborated at early stages of evolution. Clearly, there would be greater freedom for modification and improvement by natural selection of what are now early

developmental stages before the more complex and dependent later stages of development were superimposed on them. As a corollary, we would expect that, once the later stages evolved, the earlier stages of the developmental regulatory program would be more or less fixed. One can imagine modest alterations or additions to the early parts of the developmental program, but it would be very unlikely that such programs could be supplanted. Therefore the basic developmental patterns would be expected to have been elaborated earlier in evolution and be more widespread, phylogenetically. There is an obvious resemblance here to the proposition of ontogenetic recapitulation of phylogeny. However, what is being suggested here is not recapitulation. We do not suggest that the forms and structures of actual precursor organisms recur during development, but that ancient regulatory programs that have been more or less modified continue to be utilized. As development progresses from stage to stage, progessively less ancient and phylogenetically more restricted genomic regulatory patterns would come into play.

LIST OF LITERATURE

BEÇAK, W. 1969. Genic action and polymorphism in polyploid species of amphibians. *Genetics* (Suppl.), 61: 183–190.

BENDICH, A. J., and B. J. McCARTHY. 1970. DNA comparisons among barley, oats, rye, and wheat. *Genetics*, 65: 545–565.

BIER, K., W. KUNZ, and D. RIBBERT. 1969. Insect oogenesis with and without lampbrush chromosomes. In C. D. Darlington and K. R. Lewis (eds.), *Chromosomes Today*, Vol. 2, p. 107–115. Plenum Press, N.Y.

BOLTON, E. T., R. J. BRITTEN, D. B. COWIE, R. B. ROBERTS, P. SZAFRANSKI, and M. J. WARING. 1966. Biophysics. *Carnegie Inst. of Wash. Yearbook*, 64: 313–348.

BRITTEN, R. J. 1969a. Repeated sequences in human DNA. *Carnegie Inst. of Wash. Yearbook*, 67: 327–330.

——. 1969b. Repeated DNA and transcription. In E. W. Hanly (ed.), *Problems in Biology: RNA in Development*, p. 187–216. University of Utah Press, Salt Lake City.

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

BRITTEN, R. J., and D. E. KOHNE. 1967. Nucleotide sequence repetition in DNA. *Carnegie Inst. of Wash. Yearbook*, 65: 78–106.

——, and ——. 1968a. Repeated sequences in DNA. *Science*, 161: 529–540.

——, and ——. 1968b. Repeated nucleotide sequences. *Carnegie Inst. of Wash. Yearbook*, 66: 73–88.

——, and ——. 1969a. Repetition of nucleotide sequences in chromosomal DNA. In A. Lima-de-Faria (ed.), *Handbook of Molecular Cy-*

Translocation of both pre-existing and new regulative sequences is required, in the successive evolutionary stages diagrammed in parts (2) and (3) of the figure.

(2) A new structure is formed by alteration of the developmental program. This structure bears certain homologies to the pre-existing structure, formed initially by the clone descended from cell 1.

(3) Synthesis of the products of producer genes P_1, P_2, P_3 may occur in far greater quantity than previously possible due to the large number of cells in the new structure, as determined by battery A. The selective advantage could be great. This same factor might tend to favor the evolutionary accretion of further "glandular" activities in this structure.

tology, p. 21–36. North-Holland Publishing Co., Amsterdam.

——, and ——. 1969b. Implications of repeated nucleotide sequences. In A. Lima-de-Faria (ed.), *Handbook of Molecular Cytology*, p. 37–51. North-Holland Publishing Co., Amsterdam.

BRITTEN, R. J., and J. SMITH. 1970. A bovine genome. *Carnegie Inst. of Wash. Yearbook*, 68: 378–386.

——, and ——. 1971. Cattle, sheep, and satellites. *Carnegie Inst. of Wash. Yearbook*, 69: 506–507.

CORNEO, G., E. GINELLI, and E. POLLI. 1970a. Different satellite deoxyribonucleic acids of guinea pig and ox. *Biochemistry*, 9: 1565–1571.

——, ——, and ——. 1970b. Repeated sequences in human DNA. *J. Mol. Biol.*, 48: 319–327.

DAVIDSON, E. H., and R. J. BRITTEN. 1971. Note on the control of gene expression during development. *J. Theoret. Biol.*, in press.

DAVIDSON, E. H., and B. R. HOUGH. 1969. High sequence diversity in the RNA synthesized at the lampbrush stage of oogenesis. *Proc. Nat. Acad. Sci.*, 63: 342–349.

——, and ——. 1971. Genetic information in oocyte RNA. *J. Mol. Biol.*, 56: 491–506.

DAVIDSON, E. H., B. R. HOUGH, M. CHAMBERLIN, and R. J. BRITTEN. 1971. Sequence repetition in the DNA of *Nassaria* (*Ilyanassa*) *obsoleta*. *Develop. Biol.*, in press.

DAWID, I. B., and D. D. BROWN. 1970. The mitochondrial and ribosomal DNA components of oocytes of *Urechis caupo*. *Develop. Biol.*, 22: 1–14.

FLAMM, W. G., P. M. B. WALKER, and M. McCALLUM. 1969. Some properties of the single strands isolated from the DNA of the nuclear satellite of the mouse (*mus musculus*). *J. Mol. Biol.*, 40: 423–443.

GELDERMAN, A., A. RAKE, and R. J. BRITTEN. 1969. Genetic expression of non-repeated DNA sequences in the mouse. *Carnegie Inst. of Wash. Yearbook*, 67: 320–325.

——, ——, and ——. 1971. Transcription of non-repeated DNA in neonatal and fetal mice. *Proc. Nat. Acad. Sci.*, 68 172–176.

GEORGIEV, G. P. 1969. On the structural organization of operon and the regulation of RNA synthesis in animal cells. *J. Theoret. Biol.*, 25: 473–490.

HINEGARDNER, R. 1968. Evolution of cellular DNA content in teleost fishes. *Amer. Natur.*, 102: 517–523.

HOOD, L., K. EICHMAN, H. LACKLAND, R. M. KRAUSE, and J. J. OHMS. 1970. Rabbit antibody light

chains and gene evolution. *Nature*, 228: 1040–1044.

HOYER, B. H., B. J. McCARTHY, and E. T. BOLTON. 1963. Complementary RNA in nucleus and cytoplasm of mouse liver cells. *Science*, 140: 1408–1412.

——, ——, and ——. 1964. A molecular approach in the systematics of higher organisms. *Science*, 144: 959–967.

JONES, K. W. 1970. Chromosomal and nuclear locations of mouse satellite DNA in individual cells. *Nature*, 225: 912–915.

KEILIN, D., and J. F. RYLEY. 1953. Hemoglobin in protozoa. *Nature*, 172: 451.

KOHNE, D. 1970. Evolution of higher organism DNA. *Quart. Rev. Biophysics*, 3: 327–375.

LAIRD, C., and B. J. McCARTHY. 1969. Molecular characterization of the *Drosophila* genome. *Genetics*, 63: 865–882.

LAIRD, C., B. L. McCONAUGHY, and B. J. McCARTHY. 1969. Rate of fixation of nucleotide substitutions in evolution. *Nature*, 224: 149–154.

MAIO, J. 1971. DNA strand reassociation and polyribonucleotide bindings in the African green monkey, *Cercopithecus aethiops*. *J. Mol. Biol.*, 56: 579–595.

MALLORY, F. B., J. T. GORDON, and R. L. CONNER. 1963. The isolation of a pentacyclic triterpenoid alcohol from a protozoan. *J. Amer. Chem. Soc.*, 85: 1326–1363.

MARTIN, M., and B. HOYER. 1966. Thermal stabilities and species specificities of reannealed animal deoxyribonucleic acids. *Biochemistry*, 5: 2706–2713.

MAYR, E. 1963. *Animal Species and Evolution*. Belknap Press, Cambridge, Mass.

McCONAUGHY, B. L., and B. J. McCARTHY. 1970a. Related base sequences in the DNA of simple and complex organisms. V. The specificity of interactions between oligonucleotides and denatured DNA. *Biochem. Genetics*, 4: 409–424.

——, and ——. 1970b. Related base sequences in the DNA of simple and complex organisms. VI. The extent of base sequence divergence among the DNA's of various rodents. *Biochem. Genetics*, 4: 425–446.

MIRSKY, A. E., and H. RIS. 1951. The deoxyribonucleic acid content of animal cells and its evolutionary significance. *J. Gen. Physiol.*, 34: 451–462.

NIGON, V., and P. BOUVET. 1955. La teneur des gamètes en acide désoxyribonucléique chez *Parascaris equorum* Goeze. *Compt. rend. Soc. Biol.*, 149: 129–130.

PARDUE, M. L., and J. G. GALL. 1969. Molecular hybridization of radioactive DNA to the DNA

of cytological preparations. *Proc. Nat. Acad. Sci.*, 64: 600–604.

RICE, N. 1971. Thermal stability of reassociated repeated DNA from rodents. *Carnegie Inst. of Wash. Yearbook*, 69: 472–479.

RIBBERT, D., and K. BIER. 1969. Multiple nucleoli and enhanced nucleolar activity in the nurse cells of the insect ovary. *Chromosoma*, 27: 178–197.

RUTTER, W. J., N. K. WESSELS, and C. GROBSTEIN. 1964. Control of specific synthesis in the developing pancreas. *Nat. Cancer Inst. Monograph*, 13: 51–65.

SEXSMITH, E. 1968. DNA values in karyotypes of Amphibia. Thesis, University of Toronto.

SOBER, H. A. 1968. *Handbook of Biochemistry*. Chemical Rubber Co., Cleveland.

STEBBINS, G. L. 1969. *The Basis of Progressive Evolution*. Univ. of North Carolina Press, Chapel Hill.

STRAUSS, N. 1971. Comparative DNA renaturation kinetics in amphibians. *Proc. Nat. Acad. Sci.*, 68: 799–802.

ULLMAN, J. S. 1970. The effect of chemically altered base pairs on DNA-DNA duplex thermal stability. Thesis, University of Washington, Seattle.

WARING, M., and R. J. BRITTEN. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science*, 154: 791–794.

WESSELS, N. K., and F. H. WILT. 1965. Action of actinomycin D on exocrine pancreas cell differentiation. *J. Mol. Biol.*, 13: 767–779.

WETMUR, J. G., and N. DAVIDSON. 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.*, 31: 349–370.

Appendix I: Overlap of Enzyme Activity Classes in Eukaryotes and Prokaryotes

Compiled by Barbara Hough, from data listed in *The Enzymes,* Dixon and Webb, Academic Press, N.Y. (2nd ed.), 1964.

[The total of known enzymes of a functional class confined to eukaryotes (animals and/or plants) is 55, compared to 771 in prokaryotes. Thus relatively few enzyme activity classes have arisen since the divergence of the eukaryotes. The history of the individual enzymes is of course not known.]

Classes of enzyme activities found in bacteria and in animals or plants
M. Dixon and E. C. Webb, *Enzymes* (1964), p. 670

I.E.C. CLASSI-FICATION	NUMBER OF ENZYMES	ACTIVITY	ACTING ON	ACCEPTOR	EXAMPLE
1.1.1	66	oxidoreductase	—CHOH	NAD or NADP	alcohol dehydrogenases
1.1.2	4	"	"	cytochrome	glycerolphosphate dehydrogenase
1.1.3	8	"	"	O_2	lactate oxidase
1.1.99	5	"	"	other than the above	choline dehydrogenase
1.2.1	20	"	aldehyde or keto	NAD or NADP	formaldehyde dehydrogenase
1.2.3	4	"	" " "	O_2	xanthine oxidase
1.2.4	2	"	" " "	lipoate	pyruvate dehydrogenase
1.3.1	5	"	—CH—CH—	NAD or NADP	dihydro-uracil dehydrogenase
1.3.99	3	"	"	other than NAD, cytochrome, or O_2	succinate dehydrogenase
1.4.1	6	"	CH—NH$_2$	NAD or NADP	glutamate dehydrogenase
1.4.3	6	"	"	O_2	L-aminoacid oxidase
1.5.1	5	"	C—NH	NAD or NADP	tetrahydrofolate dehydrogenase
1.5.3	3	"	"	O_2	spermine oxidase
1.6.1	1	"	NADH$_2$ or NADPH$_2$	NAD or NADP	NAD(P) transhydrogenase
1.6.2	3	"	" " "	cytochrome	NADH$_2$-cytochrome reductase
1.6.4	3	"	" " "	disulfide	cystine reductase
1.6.5	4	"	" " "	quinone or related compound	quinone reductase
1.6.6	8	"	" " "	nitrogenous group	nitrate reductase
1.7.99	3	" "	nitrogenous compounds not —NH or —NH$_2$	other than O_2	hydroxylamine reductase
1.8.3	2	"	sulfur groups	O_2	sulfite oxidase

1.11.1	9	"	(various)	H_2O_2	catalase
1.99.1	15	oxidoreduc-tase	other enzymes using O_2 as oxidant.		hydroxylases
1.99.2	9	"	"	Not specified	oxygenases
2.1.1	10	transferase	one-carbon groups	"	methyltransferases
2.1.2	6	"	"	"	form$_y$ltransferases
2.1.3	3	"	"	"	carboxyltransferases
2.2.1	2	"	aldehydic or ketonic residues	"	transketolase
2.3.1	20	"	acyl groups	"	transacetylases
2.4.1	28	"	glycosal groups	"	hexosyltranferases
2.4.2	15	"	"	"	pentosyltransferases
2.5.1	5	"	alkyl or related groups	"	thiaminase
2.6.1	19	"	nitrogenous groups	"	aminotransferases
2.7.1	50	"	P-containing groups	alcohol group	hexokinase
2.7.2	5	"	"	carboxyl group	aspartate kinase
2.7.4	9	"	"	phospho-group	adenylate kinase
2.7.5	4	"	intramolecular with regeneration of donors	Not specified	phosphoglucomutase
2.7.7	24	"	nucleotidyl groups	"	nucleotidyl transferases
2.8.3	5	"	sulfur-containing groups	"	CoA transferases
			ACTING ON		
3.1.1	20	hydrolase	ester bonds—carboxylic		lipase
3.1.2	8	"	ester bonds—thiolesters		CoA hydrolase
3.1.3	19	"	ester bonds—phosphoric monoesters		alkaline phosphatase
3.1.4	9	"	ester bonds—phosphoric diesters		phosphodiesterase
3.1.6	8	"	ester bonds—sulfuric esters		sterol sulfatase
3.2.1	37	"	glycosyl compounds—glycosides		amylase
3.2.2	6	"	N-glycosyl compounds—glycosides		nucleosidase
3.4.1	4	"	peptide bonds—α-aminopeptide aminoacidhydrolases		aminopeptidase
3.4.4	23	"	peptide-peptide bonds		pepsin
3.5.1	18	"	C—N bonds other than peptides—in linear amides		usease
3.5.2	7	"	C—N bonds other than peptides—in cyclic amides		dihydro-orotase
3.5.3	6	"	C—N bonds other than peptides—in linear amidines		arginase

I.E.C. CLASSI-FICATION	NUMBER OF ENZYMES	ACTIVITY	ACTING ON		EXAMPLE
3.5.4	12	"	C—N bonds other than peptides—in cyclic amidines		cytosine deaminase
3.6.1	11	" ·	acid anhydride bonds, in phosphoryl-containing anhydrides		ATP-ase
3.8.2	1	"	halide bonds in P-halide compounds		DFP-ase
			ACTING ON	TYPES OF ENZYMES	
4.1.1	39	lyase (removes groups from their substrate leaving double bonds)	C—C	carboxyl-lyases	decarboxylase
4.1.2	16	,,	"	aldehyde-lyases	aldolase
4.1.3	8	"	"	ketoacid-lyases	citrate lyase
4.2.1	27	"	C—O	hydro-lyases	carbonic anhydrase
4.2.99	2	"	"	other than hydro-lyases	hyaluronidase
4.3.1	6	"	C—N	ammonia lyases	aspartase
4.3.2	2	"	"	amididine lyases	argininosuccinate lyase
4.4.1	5	"	C—S		cysteine desulfhydrase
5.1.2	3	isomerase	hydroxyacids	racemases and epimerases	lactate racemase
5.1.3	7	"	carbohydrates and derivatives	" "	UDP glucose epimerase
5.1.99	1	"	other substrates than those above	" "	methylmalonyl-CoA racemase
5.2.1	4	"	—	cis-trans isomerases	maleate cis-trans isomerase
5.3.1	14	"	aldose-ketoses	intramolecular oxidoreductases	triose phosphate isomerase
5.3.3	3	"	$C = C$, C—C	" "	steroid Δ-isomerase
5.4.99	2	"	— (not acyl or phosphoryl)	intramolecular transferases	methylmalonyl-CoA mutase
6.1.1	11	ligase	forming C—O bonds	aminoacid-RNA ligases	leucyl-sRNA synthetase
6.2.1	7	"	forming C—S bonds	acid-thiol ligases	acetyl-CoA synthetase
6.3.1	3	"	forming C—N bonds	acid-ammonia	asparagine synthetase
6.3.2	11	"	" " "	acid-aminoacid	glutathione synthetase
6.3.4	5	"	" " "	compounds other than NH_3, aminoacid; or cycloligases	CTP synthetase
6.3.5	3	"	" " "	with glutamine as amino-N donor	NAD synthetase
6.4.1	4	"	forming C—C bonds	—	pyruvate carboxylase

Classes of enzymes found in animals or plants, but not in bacteria

I.E.C. CLASSI- FICATION	NUMBER OF ENZYMES	ACTIVITY	ACTING ON	ACCEPTOR	EXAMPLE	FOUND IN
1.3.2	3	oxidoreductase	—CH—CH	Cytochrome	butyryl-CoA dehydrogenase	animals and plants
1.6.99	1	"	$NADH_2$ or $NADPH_2$	NADP	$NADPH_2$ diaphorase "old yellow enzyme"	yeast
1.7.3	3	"	nitrogenous compounds, other than—NH or—NH_2	O_2	urate oxidase	animals, plants, molds, yeast
1.8.1	2	"	sulfur groups	NAD or NADP	cysteamine dehydrogenase	liver, yeast
1.8.4	1	"	" "	disulfide	glutathione-homo-cystine transhydrogenase	liver
1.8.5	1	"	" "	quinone or related compound	glutathione dehydrogenase	yeast, higher plants
1.8.6	1	"	" "	nitrogenous group	nitrate ester reductase	liver
1.10.3	3	"	diphenols, related sub-stances	O_2	ascorbate oxidase	plants, fungi, some animal tissues
			ACTING ON	TYPES OF ENZYMES		
2.6.2	1	transferase	nitrogenous groups	amidinotransferase	glycine transamidinase	kidney, plants
2.6.3	1	"	" "	oximinotransferase	transoximinase	silkworm tissues, algae
2.7.3	5	"	phosphor groups	nitrogenous acceptor group	guanidino-acetate kinase	animals (esp. worms)
2.7.6	2	"	" "	pyrophosphotransferases	thiamine pyrophosphokinase	liver, yeast
2.7.8	3	"	" "	substituted phospho groups, not nucleotides	choline phosphotransferase	liver, yeast
2.8.1	2	"	sulfur-containing groups	sulfurtransferases	thiosulfate sulfurtransferase	animal tissues
2.8.2	5	"	" "	sulfotransferases	aryl sulfotransferase	liver, animal tissues

I.E.C. CLASSI-FICATION	NUMBER OF ENZYMES	ACTIVITY	ACTING ON	TYPES OF ENZYMES	EXAMPLE	FOUND IN
3.2.3	1	hydrolase	glycosyl compounds	hydrolyzing 5-glycosyls	thioglucosidase	animal tissues, plants
3.3.1	1	hydrolase	ether bonds	thioether hydrolase	adenosylhomo-cysteinase	liver
3.4.2	3	,,	peptide bonds	α-carboxypeptide aminoacidhydrolases	carboxypeptidase	yeast, pancreatic juice
3.4.3	7	,,	,, ,,	dipeptide hydrolases	glycyl-glycine dipeptidase	animal tissues
3.7.1	2	,,	C—C bonds in ketonic substances	——	oxaloacetase	liver
3.8.1	1	,,	halide bonds in C-halide compounds	——	alkylhalidase	liver
3.9.1	1	,,	P—N bonds	——	phosphoamidase	snake venom, plants
4.5.1	1	lyase	C-halide	carbon-halide lyases	DDT-dehydrochlorinase	insects
5.3.2	1	isomerase	keto- and enol groups	intramolecular oxi-doreductases inter-converting keto-enol	phenylpyruvate tautomerase	animal tissues
5.4.1	1	,,	acyl groups	intramolecular transferases	lysolecithin acyl mutase	pancreas mold
5.4.2	1	,,	phosphoryl groups	,,	phosphoglycerate phosphomutase	cereal germ
6.3.3	1	ligase	forming C—N bonds	cyclo-ligase	phosphoribosyl amino-imidazole synthetase	liver